Resource Consultant Training Program
Research Report No. 12

RCTP

Analysis of Grades 3-5 Math Curricula for Production of Survey Diagnostic Tests

Richard Parker

Ed Kameenui
Math Project Director

University of Oregon, Division of Teacher Education, Special Education Area,
Eugene, Oregon, 97403-1215

Published by
Resource Consultant Training Program
Division of Teacher Education
College of Education
University of Oregon

Staff
Gerald Tindal, Program Director
Jerry Marr, Editor

Parker, Richard
Analysis of Grades 3-5 Math Curricula for Production of Survey Diagnostic Tests
Research Report No. 12

Preparation of this document was supported in part by the U.S. Department of Education, grant number H-180C80009-89. Opinions expressed herein do not necessarily reflect the position or policy of the U.S. Department of Education, and no official endorsement by the Department should be inferred.

Cover design: George Beltran

Analysis of Grades 3-5 Math Curricula for Production of Survey Diagnostic Tests

Richard Parker
University of Oregon

Ed Kameenui
Math Project Director
University of Oregon

This project produced and pilot-tested survey diagnostic tests for Grades 3, 4, and 5, based on the Open Court math curriculum. Our goal was to bring into better alignment curriculum, goals/objectives, and assessment. A taxonomy was constructed to be used both for curriculum analysis and for item construction. Reasonable inter-scorer reliability was obtained in coding learning activities from each lesson. From the curriculum codes and scanning teacher and student pages, representative items were created. Procedural rules were established for including items in pilot tests which were administered to approximately 250 students in each of Grades 3, 4, and 5. From the pilot-testing, test- subtest- and item-level, information on difficulty and reliability was obtained. Criteria for item deletion reduced the number of items by half. The remaining items were used to prepare combined criterion-referenced/norm-referenced feedback for the teachers. A second test, strictly equivalent to the first, was created for assessment later in the year. Results from the two assessment periods were then compared, and recommendations for improved test development procedures were made.

Introduction

A recent review of research on effective teaching with low-achieving students (Christenson, Ysseldyke, & Thurlow, 1989) concluded that there is no "effective instruction" unique to this population. Drawing from earlier reviews (e.g., Walberg, 1984; Good and Brophy, 1986) and their own research, the authors identify ten critical factors for achievement of handicapped and non-handicapped learners in both regular and special education environments. Directly relevant to the present project are four factors which create a supportive framework for effective instruction. The supportive framework is present when there is alignment or congruence of (a) the curriculum, (b) learning goals/objectives, (c) assessment, and (d) instruction.

Christenson et al. (1989) state that learning goals/objectives should be short-term, clearly articulated, and *closely related to the curriculum. Assessment of student learning* should provide frequent feedback, guide instruction, and be *closely linked to both mastery goals/objectives, and the curriculum. Instruction* of students should occur at those locations in the curriculum where students can show regular improvement toward curriculum goals and objectives. Figure 1 depicts the interrelationships among Curriculum, Instruction, Assessment, and Goals/objectives. The figure helps illustrate and expand Nitko's (1989) discussion of "tripartite congruence" between instruction, objectives, and test items (p. 458). In the figure, major influences among the four elements are represented as dark arrows; the light arrows describe minor or less frequent influence. Excluded from this model are other influences on instruction such as teacher variables, class composition, and the program structure and resources (Smylie, 1988, 1989).

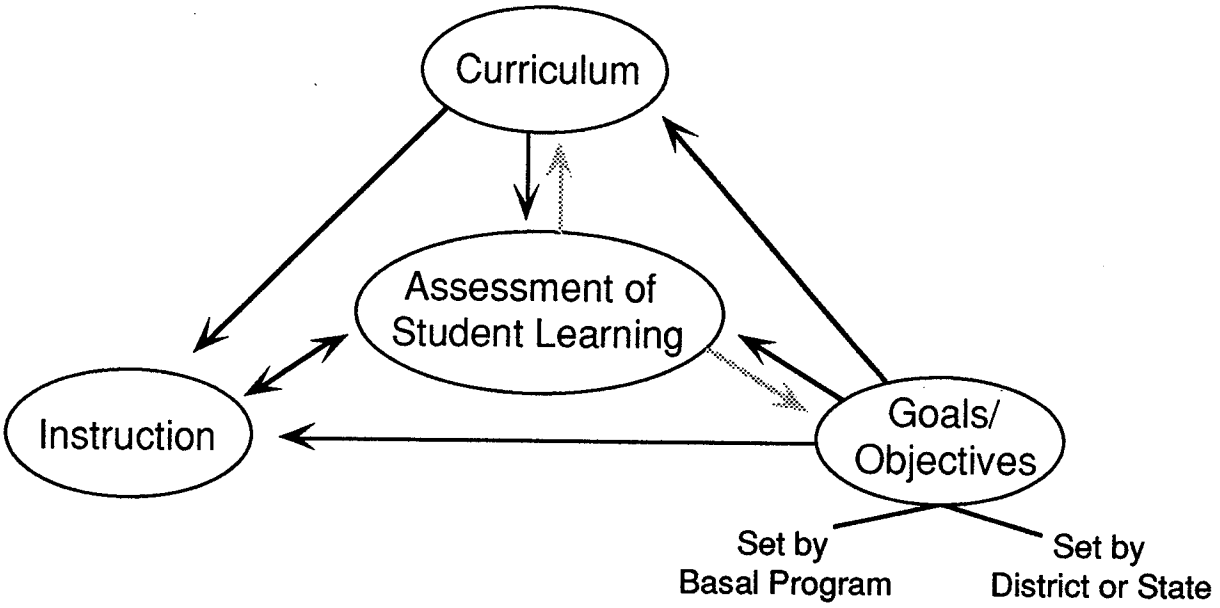


Figure 1. Influences Among Curriculum, Instruction, Assessment, and Goals/Objectives

Unfortunately, *Curriculum, Instruction, Assessment,* and *Goals/Objectives* often are poorly aligned in elementary education (Nitko, 1989). Mismatches are common between *Goals/objectives* and *Instruction,* between *Goals/objectives* and *Assessment,* and between *Assessment* and *Curriculum.* The aim of this project was to develop procedures for improving the alignment of the *Curriculum,* learning *Goals/Objectives,* and *Assessment* in Grade 3-5 mathematics within regular classrooms. Research indicates that better congruence among all of these factors could improve the regular class performance of students "at-risk" and with handicaps.

Project Rationale

The arrows in Figure one depict three main influences on instruction: the curriculum, assessment results, and goals and objectives (from within or outside the curriculum). Of these influences, assessment occupies a pivotal position, as it influences and is influenced by each of the other three variables. Test content and format should be strongly influenced by the curriculum used, the goals and objectives pursued, and what actually is taught. The light arrows show lesser influence in the opposite direction; assessment results may influence selection/adaptation of curriculum materials, or modification of instructional goals.

This project focused on only three of the critical variables—curriculum, assessment, and goals/objectives. It did not attempt to account for the fourth variable—instruction. It was expected that by improving alignment of the other three variables, instruction also would be indirectly improved. The general approach for aligning curriculum, assessment, and goals/objectives was to develop and use a single classification scheme for performance objectives, curriculum activities, and test items. That goal is not new. It is embedded in Instructional Quality Inventory (IQI) procedures, which have been applied to both military training (Merrill, Reigeluth & Faust, 1979) and public school curricula and assessment (Roid & Haladyna, 1982). Three maturing methodologies also exist to help reach the goal of congruence: curriculum analysis, item construction, and test construction.

After briefly reviewing the relevant literature, this report describes (a) the development of curriculum analysis, item analysis, and test construction procedures in elementary mathematics, (b) the initial application of these procedures to the *Open Court Math* basal program at Grades 3, 4, and 5, and (c) a summary of student performance data from administering these tests to 600 students in 6 schools. The goal of the report is to provide guidance for other university or district-based efforts to produce useful math tests that are closely aligned with the curriculum.

Mismatch Problems

Assessment Mismatch

No cohesive national math curriculum can be identified by popular tests or basal programs. Content analyses of five nationally standardized math tests and four major basal math programs demonstrated wide differences at the level of specific objectives (Freeman, Kuhs, Porter, Floden, Schmidt, & Schwille, 1983). It is also well known that "different curricula are associated with different patterns of achievement" (Walker & Schaffarzick, 1974). These

differences prevent standardized test scores from being interpreted in a straightforward manner, as they represent "opportunity to learn" as well as actual student learning (Romberg & Carpenter, 1986).

Poor alignment of math test content can be seen in inappropriate item presentation and response formats. Most standardized math tests rely on the multiple choice response format, although it seldom occurs in instruction and curriculum-based assessment (Murnane & Raizen, 1988; Alexander & James, 1987). Problem-solving applications especially require free response formats, as multiple choice selections reflect different abilities. Another disadvantage of the multiple-choice test format is its influence on how teachers present math content, and how students study for tests (Frederiksen, 1984; Kirkland, 1971).

The mismatch between curriculum and assessment has the following unwanted results: (a) standardized tests are relatively insensitive to achievement within a particular curriculum, causing underestimation of student improvement (Porter, Schmidt, Floden, & Freeman, 1978), and (b) standardized tests are differentially sensitive to achievement in different curricula. Because standardized achievement tests are "content-biased" (Schmidt, 1983), they occasion unfair evaluations of instructional programs. First, successes obtained by an instructional program may be overlooked, simply not measured. Second, in evaluating competing instructional programs, an unfair edge will be obtained by the program with content that overlaps most of the test content (Airasian & Madaus, 1983).

Curriculum Mismatch

Although teachers often are permitted to depart from the content and sequence of basal texts, they seldom do so (Stake & Easley, 1978; Stephens, 1982). The math textbook is perceived by teachers as *the* authority on knowledge and the guide to learning (Romberg & Carpenter, 1986; Good, Grouws, & Ebmeier, 1983). However, basal texts often suffer from problems of internal misalignment. Their listed "scope and sequence" objectives may be too broad, too ambiguous, or simply too inaccurate to reflect prescribed lesson activities (Popham, 1984; Roid & Haladyna, 1982). A second major problem is that learning activities often provide only exposure, not measurable skill growth toward mastery: "A very large percentage of the topics taught receive only brief, perhaps cursory, coverage" (Porter, 1989, p. 12). When teaching for "exposure" and "review" replace teaching for skill development and mastery, the alignment of goals/objectives, curriculum, and instruction becomes tenuous. Furthermore, assessment lacks a satisfactory foundation: Should tests be based on goals/objectives or on actual activities? Difficulties in assessment also cause problems for school and teacher accountability for student learning (Nitko, 1989; Porter, 1989).

Another type of curriculum mismatch is that between basal learning activities and learning objectives mandated by outside authorities with social/political bases (Nitko, 1989; Jaeger, 1989). The extent to which state and district level core curriculum goals or competencies are being achieved will vary according to the particular basal program in use (Freeman, et al., 1983). Tests developed to assess attainment of state or district objectives also may be biased toward a particular curriculum unless common content areas are first identified across curricula. Item presentation and response format, and use of symbols and cues all may bias a test. Although logical curriculum analysis can help create a fair test with maximum overlap of goals/objectives and learning activities, empirical analysis also is required. Using test results from students instructed through different basal programs, individual item-types may be examined for curriculum bias in the same way that racial and sexual bias is assessed (Cole & Moss, 1989).

Curriculum Analysis

The increasing popularity of curriculum analysis can be ascribed to three recent trends. First, the basal text has been identified as a strong influence on classroom instruction (Durkin, 1978-79; Komoski, 1985). Second, for test construction purposes, recent advances in criterion referenced testing (CRT) allow that technology to better support curriculum analysis (Hsu & Yu, 1989; Nitko, 1989). Finally, cognitive processing views of classroom learning have recently suggested curriculum analyses with a cognitive orientation. Instead of categorizing only content, curriculum analysts now are interested in identifying what mental operations are occurring (Kameenui & Griffin, 1989; Snow & Lohman, 1989).

Description and tabulation of at least three curriculum features appear necessary for valid test construction: (a) the subject content presented, (b) the activities and required student performance (cognitive and behavioral), and (c) the location of these activities in the curriculum. The first two features assist in *item construction*, and the third in systematic item sampling for *test construction* and efficient use of test results (Roid & Haladyna, 1982). The second feature specifies important characteristics of learning activities: stimulus presentation, learner response type, and inferred "cognitive operations" or "reasoning" (Hively, Patterson, & Page, 1968; Osburn, 1968). By describing and quantifying these three features, curriculum analysis can define a curriculum-refer-

enced domain of behaviors for CRT test construction and a strategy for item sampling (Berk, 1980; Nitko, 1980).

In mathematics education there is considerable agreement on 12 major skill areas (Denmark & Kepner, 1980). By crossing content areas with levels of cognitive process (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) a taxonomy or item classification matrix has been created which has served as the basis for the National Assessment of Educational Progress (NAEP) math tests (1983).

An alternative math content description has been offered by Glennon and Wilson (1972), consisting of seven hierarchically organized domains. These domains have been crossed with Williams and Haladyna's (1982) LOGIQ matrix (including level of abstraction, intellectual operation, and response mode) to create a complete test item typology (Tindal,1989).

A third alternative taxonomy for math content analysis has been developed at the Institute for Research on Teaching at Michigan State University (Kuhs, et al., 1979). The three dimensional taxonomy for test items includes "general intent" (e.g. conceptual understanding or application), "nature of content" (e.g. fractions or decimals), and the "operation the student must perform (e.g. estimate or multiply)." The taxonomy proved reliable in application to both basal texts and standardized tests.

Item Creation and Classification

The information provided by curriculum analysis yields data on curriculum content, curriculum location & focus, and activity characteristics, all of which serve the next two procedures, *item creation* and *test construction*. With few modifications, the same taxonomy used in curriculum analysis is suitable for item creation. Specifications for test items include subject content, presentation and response formats, and reasoning or cognitive operations involved in task completion. These features, often combined with various others, constitute "amplified" behavioral objectives (Popham, 1978), item construction rules (Bormuth, 1970; Millman, 1980), or item templates to guide item creation (Nitko, 1980).

Once items are created, they can be classified according to these same characteristics to assist in constructing a variety of tests with different purposes. For example, items can be selected by content (such as word problems involving subtraction of 2- to 4-digit numbers), with regrouping, or by cognitive operation (such as problems requiring recall of facts and rules). Items also can be selected by a combination of presentation and response formats, as , for example, math problems presented verbally, requiring a written response within a time limit, and without scratch pad or calculator. Indeed, items can be selected by any sensible combination of characteristics.

A curriculum analysis alone does not guarantee good items; item writing also requires both familiarity with the curriculum and content-area expertise. Item-writing skills have been defined by Roid & Haladyna (1982), Hambleton & Eignor (1978), Ebel (1971), and Haladyna & Downing (1989). Roid (1989) recommends either hiring expert item writers or conducting a summer workshop for a group effort, and requiring item critique and interchange between experts in content and those in psychometrics.

Test Construction

Ideally, curriculum-based tests should be used by teachers with students who may be working at various points in the curriculum, who have demonstrated varying degrees of content mastery, and for whom only selected curriculum goals/objectives may be appropriate. Therefore, any test items created also should be categorized by placement in the curriculum: by lesson, unit, grade level, and type of material (basal text, supplementary practice book, etc.). This requirement is important also because, for most teachers, a basal's lesson and unit sequence dictates what to teach next. Information on instructional focus can also be useful to identify where in the curriculum sequence a skill is first introduced, practiced, and reviewed.

The information on curriculum location and focus complements the item information described earlier on subject content, task format, response mode, and cognitive level (Hambleton & Swaminathan, 1985; Stone, 1989). Information on curriculum location and focus allows certain items to be selected and grouped together to form a test covering a selected portion of the curriculum.

The Project

This project involved integrated curriculum analysis, item creation and indexing, and test production in elementary mathematics for the purpose of improving the congruence among teaching goals, assessment, and instructional activities. Levels 3, 4, and 5 of the *Open Court Math* curriculum were analyzed. The curriculum analysis and subsequent item creation was intended to allow production of a variety of tests. However, this project involved production and piloting of only survey diagnostic tests at each of the three grade levels. Tests were desired which would: (a) accurately reflect curriculum content for an academic year, (b) provide both norm-referenced and criterion-referenced data on major subskill performance, (c) provide for adminis-

Table 1. Curriculum Analysis Taxonomy (Abbreviated Form)

Curriculum Content	Curriculum Location & Focus	Activity Characteristics
a. Numeration b. Addition c. Subtraction d. Multiplication e. Division f. Multiple/Basic Facts g. Multiple/Multi-digit h. Fractions i. Decimals j. Word Problems k. Measurement l. Geometry m. Percent n. Applications o. Algebra p. Relationships q. Reasoning r. Statistics s. Calculators (The above19 main categories included 115 separate subcategories and content codes)	•Grade Level •Lesson Number •Basal Source a. Student Pages b. Mental Math c. Thinking Story d. Demonstration •Task Intent a. Introduced b. Practiced c. Reviewed •Assessment a. Unit Test b. Review Test	Cognitive Level a. Memory/Rote Learning b. Skill/Procedural c. Conceptual understanding d. General understanding e. Problem-solving applications Task Format a. Mental Math b. Paper/Pencil c. Manipulative d. Discussion Response Mode a. Oral b. Written c. Show/Demonstrate Response Type a. Selection b. Production Cues/Hints a. Examples from problem b. Pictures c. Objects/Manipulables

tration of equivalent forms two or three times per year, and (d) possess test and subtest reliability.

Administration of test prototypes and final versions involved approximately 250 students at Grades 3, 4, and 5 from 31 classrooms and 6 elementary schools in a Pacific Northwest school district. Eight steps were taken to produce sensitive and valid curriculum-based tests: 1. Curriculum analysis, 2. Item creation and indexing, 3. Test construction, 4. Administration of a test prototype, 5. Item review & revision, 6. Test Construction, 7. Test administration, and 8. Summarization of test results for teachers.

Curriculum Analysis

All Grade 3, 4, and 5 *Open Court* math program lessons (approx 140 per grade level) were analyzed, relying mainly on the teacher's manuals, which included reprints of student workbook pages. The analysis entailed affixing multiple codes to each lesson activity. The codes were based on a three-dimensional taxonomy (presented in Table 1): (a) *curriculum content,* (b) *curriculum location and focus,* and (c) *activity characteristics.* The full taxonomy of the 115 content subcategories is presented in Table 2.

Besides the curriculum-based taxonomy, a second was also considered—that for the State of Oregon's Comprehensive Curriculum Goals in Mathematics for Grades 3-5 (Oregon Dept of Ed., 1987) (see Appendix A). Oregon's 59 math goals for Grades 3-5 were cross-referenced with the 115 Content Categories developed for the *Open Court* curriculum. The cross-referenced codes for the two taxonomies are presented in Appendix B. All State goals were reflected in the content codes. Only six content codes were not covered by the State goals: (a) percent word problems, (b) percent estimation, (c) percents and relation signs, (d) applications of averages, (e) parentheses in algebra, and (f) relationship symbols. Classification of the State of Oregon goals was done only indirectly, through cross-referencing the state goals with curriculum content codes.

Coding of *Open Court* lesson activities was undertaken by four education graduate students from the University of Oregon, who, after practice required approximately 4 minutes per lesson, or 9 hours per grade level. Codes were entered on a form (Appendix C-2) with guidelines (Appendix C-1). Coded data were then transferred to Excel® (Microsoft Corporation, 1990) spreadsheet software on the Macintosh® computer (Apple Computers) for

Table 2. Full Math Content Taxonomy (115 Subcategories)

Open Court Math Program Content Codes

A	**NUMERATION**	F6	estimating	**L**	**GEOMETRY**
A1	counting (forward/backward, skip count, estimate)	F7	using relation signs	L1	shapes (including concave, convex)
A2	reading & writing (standard to written/written to standard)	**G**	**OPERATIONS - MULTIPLE/MULTI - DIGIT** (Includes chains & inverse)	L2	perimeter
A3	place value (understand, operationalized)	G1	addition and/or subtraction	L3	area
A4	rounding	G2	addition/subtraction/multiplication	L4	angles (including congruent triangles)
A5	negative numbers	G3	multiplication/division	L5	lines (parallel, perpendicular, symmetry)
A6	special cases (roman numerals, prime numbers	G4	addition/subtraction/multiplication/division	L6	word problems (all of the above)
B	**OPERATIONS - ADDITION**	G5	estimating	L7	estimating
B1	estimating	G6	using relation signs	**M**	**PERCENT**
B2	basic facts	**H**	**FRACTIONS**	M1	relation to decimals/fractions (concept & conversion)
B3	two/three digit	H1	conceptual understanding (comparing, recognizing, relation to time)	M2	computation (% of, % increase/decrease, sales tax, discounts)
B4	multi-digit (4+)	H2	equivalent fractions	M3	word problems
B5	column addition	H3	proper/improper/mixed numbers	M4	estimating
B6	regrouping	H4	addition/subtraction w/like denominators	M5	using relation signs
B7	using relation signs	H5	addition/subtraction w/unlike denominators	**N**	**APPLICATIONS**
C	**OPERATIONS - SUBTRACTION**	H6	multiplication	N1	averages
C1	estimating	H7	3,4,5, & 6	N2	ratios
C2	basic facts	H8	estimating	N3	graphing & charting
C3	two/three digit	H9	using relation signs	N4	money
C4	multi-digit (4+)	H10	fractions of whole numbers	N5	maps/scale drawings
C5	regrouping	**I**	**DECIMALS**	N6	word problems (involving all of above)
C6	using relation signs	I1	conceptual understanding (e.g., tenths/ hundredths, relation to money & metric, comparing size)	N7	making problems to fit information
D	**OPERATIONS - MULTIPLICATON**	I2	relation to fractions	**O**	**ALGEBRA**
D1	estimating	I3	addition/subtraction	O1	missing term problems
D2	basic facts	I4	multiplication (decimal x whole, decimal x decimal)	O2	parentheses
D3	single multiplier (1x22, 1x333, 1x444)	I5	division	O3	functions (solve for *n*)
D4	double multiplier (11x22, 11x333)	I6	mixed operations	**P**	**RELATIONSHIPS**
D5	multiple multipliers (3+)	I7	estimating	P1	relation signs (< > =)
D6	multipliers with 0's	I8	using relation signs	P2	conversions - metric
D7	powers & multipliers of 10	**J**	**WORD PROBLEMS**	P3	conversions - traditional
D8	relation signs	J1	addition/subtraction	P4	conversion - time
E	**OPERATIONS - DIVISION**	J2	multiplication	P5	conversions - money
E1	estimating	J3	division	P6	symbols
E2	basic facts	J4	mixed operations	**Q**	**REASONING**
E3	single digit divisor	J5	fractions	Q1	thinking stories
E4	single digit divisor with remainders	J6	decimals	Q2	"Do I have enough?"
E5	two-digit divisor	J7	estimating	**R**	**STATISTICS**
E6	two-digit divisor with remainders	**K**	**MEASUREMENT**	R1	probability
E7	relation signs	K1	length.weight.distance (metric, traditional)	R2	interpreting
E8	powers/multiples of 10	K2	volume	**S**	**CALCULATORS**
F	**OPERATIONS - MULTIPLE/BASIC FACTS** (Includes chains & inverse)	K3	time	S1	use of/ practice
F1	addition only	K4	temperature	S2	use with other operations
F2	addition/subtraction	K5	word problems (all of the above)		
F3	addition/subtraction/multiplication	K6	choosing appropriate unit		
F4	multiplication/division	K7	estimating		
F5	addition/subtraction/multiplication/division	K8	using relation signs		

summarization; a sample spreadsheet page is included in Appendix D.

Only those lesson activities with clearly stipulated student performance were coded. Lectures or discussions which did not contain or result in required student responses were excluded; many such activities were found. On the average, each lesson yielded about five scorable activities. Multiple content codes (up to three) were permitted for each activity. Counts of activities and related content codes were as follows: Grade 3 - 874 and 988; Grade 4 - 549 and 678; Grade 5 - 662 and 775.

Interrater reliability estimates for coding lesson activities were calculated for a random sample of 20 Grade 3 - 5 activities. The index for categorical agreement was Cramer's *V*, which rescales Pearson's Phi from 0 (no association) to +1 (perfect association) (Hays, 1981). Cramer's *V* reliability coefficients for the taxonomy dimensions were: *Grade Level, Lesson #,* and *Basal Source,* 1.00; *Task Intent,* .59; *Cognitive Level,* .55; *Task Format,* .73; *Response Mode,* .69; *Response Type,* .70; *Cues/Hints,* .73. After establishing clearer definitions and examples, changing some categories, and further training, agreement indices were increased: *Grade level, Lesson #,* and *Basal Source,* 1.00; *Task Intent,* .84; *Cognitive Level,* .80; *Task Format,* .92; *Response Mode,* .94; *Response Type,* .98; *Cues/Hints,* .85. The improved reliability indices were obtained later in the study, however, and are not reflected in the data in this report.

Interrater reliability for coding *Curriculum Content* required a different method because multiple codings were permitted. Thirty activities were selected representing as many content categories. Four trained scorers independently coded each activity with one, two, or three codes. Multiple codes were permitted in order to help provide feedback for the further development of the content taxonomy. Approximately 80% of the ratings were single codes only, 15% consisted of double codes, and 5% were triple codes. The ratio of obtained to possible agreement pairs (including null responses) was used to produce an agreement index for each lesson activity. Based on the Rand statistic of agreement for unequal numbers of categories (Rand, 1971) these indices ranged from 0 to 1. The computation method and examples are presented in Appendix E. Approximately 40% of the resulting indices showed perfect agreement (1.00). The remaining indices were distributed as follows: 1.00 >.75 (30%), .75 > .50 (20%), .50 > .00 (10%). The average agreement index was .76.

Once the lesson activity codes were entered onto the Excel® spreadsheet, several useful summaries were prepared, including (a) the distribution of curriculum content by lesson (including *task intent—introduction, practice,* or *review*), (b) changes in content focus across the grades, and (c) the relative frequency of occurence of content categories summarized by year. These three summaries are presented in Table 3, Figure 2, and Tables 4 through 6, respectively. The information in Table 3 is most useful for constructing tests covering one or more units or groups of lessons. However, the goal of this project was to produce grade level survey tests which covered the full year's content. Table 3 also provides an index to the curriculum which is useful during item creation. Through the index, concrete examples of item types can be located. The bar graph in Figure 2 presents information for the 19 main content categories across the three grade levels. These data were useful in determining the number of each item type to include in the test. Tables 4 through 6 show in more detailed tabular form information similar to that contained in the bar graph. From 80 to 90 Content sub-codes are represented in lesson activities for each of the three grades.

Item Creation & Categorizing

After the curriculum analysis, the development team began to create test items for pilot testing. The items were then strategically selected to construct the grade level survey achievement tests. Only written responses were permitted. In addition, the test was structured to allow students to progress through items independently, rather than through teacher-pacing. Because of the low reading skills of some students, text was minimized within Applications or "word problems." To be diagnostically useful, the survey test required that performance on individual items or small item clusters be interpretable. Multiple choice or true-false response types therefore were not included, because of the effects of guessing; production responses only were permitted.

It was first predicted that the *activity characteristics* and *curriculum content* information would be sufficient to describe "item forms" (Hively, 1974) or "item types" from which a number of representative individual items could be created. However, some categories of the curriculum taxonomy proved essential for the purpose of item creation, while others were not useful.

Cognitive Level

Cognitive level was valuable as a screener for identifying certain general understanding and memory/Rote learning activities which could not be assessed in a group paper and pencil test. Skills, problem-solving applications, and conceptual understanding proved to be very useful descriptive categories, the latter two of which were later integrated into the Curriculum Content codes.

Table 3. Curriculum Codes on Excel Spreadsheet, Sorted by Content

Grade	Content Code	Lesson	Content	Basal	Task Intent
5	a1	1	concepts of order	w	p
5	a1	2	number sequence	s	p
5	a1	2	count & num. seq	m	p
5	a2	3	place value-stand.	s	p
5	a3	3	place val & regroup	w	p
5	a3	6	place value	m	p
5	a3	16	place values	s	r
5	a3	25	rounding	s	i
5	a3	25	rounding to nearest 1000	s	p
5	a3	25	rounding to nearest 10	s	p
5	a3	25	rounding to nearest 100	s	p
5	a3	25	how much and when to round	s	p
5	a3	25	rounding t nearest whole #	s	p
5	a4	38	+ & - with neg #'s	s	i, p
5	a4	38	negative #'s	s	i, p
5	a4	38	+ & - with neg #'s	s	p
5	b2	60	squaring #'s /addition	w	p
5	b3	6	add 3 digit #'s	w	p
5	b3	6	multi-dig add/group	m	p
5	b3	6	multi-dig add/col	s	p
5	b3	6	multi-dig add/col.	s	r
5	b3	24	multi-digit addition	m	p
5	b3	25	multi-digit addition	m	p
5	b3	40	multi-digit addition	m	p
5	b3	112	addition	s	p
5	b3	134	mental addition	w	p
5	c3	7	multi-digit subtr.	s	p
5	c3	7	multi-dig -/regroup	s	r
5	c3	37	multi-digit subtr	m	p
5	c3	38	multi-digit subtr	m	p
5	d1	11	mult 2-dig <> 1002	m	p
5	d1	23	approx wrong ans. thumb up/d	m	p
5	d1	67	approx multidigit *	w	p
5	d2	8	mult w/fact 0-10	w	p
5	d3	9	mult 3 dig by 1	s	p
5	d3	9	mult 3 dig by 1	s	r
5	d4	10	multidigit mult	s	p
5	d4	10	multiplications	w	p
5	d4	10	* 2 whole #/algorit	s	r
5	d4	11	multi-dig. mult	w	p
5	d7	9	mult 10 & 100	m	p
5	d7	9	mult by powers 10	s	p, r
5	d7	9	mult by mults 10	s	p, r
5	d7	10	mult by mult 10 fact	m	p
5	d7	10	mult multiples 10	s	p, r
5	d7	61	mult & div by pow 10	s	p
5	d7	62	mult & div by pow 10	m	p

Task format

Task Format proved to be of little use in item construction. The discussion and mental math categories provided information available in other categories, and manipulative activities had to be omitted or converted to paper/pencil exercises for the group test.

Response Mode & Type

Response Mode was also of little use because it provided redundant information, and the Oral and Show/Demonstrate descriptors applied to activities that could not easily be group-tested. Both Task Format and Response Mode would be useful in conducting curriculum analyses for other purposes,

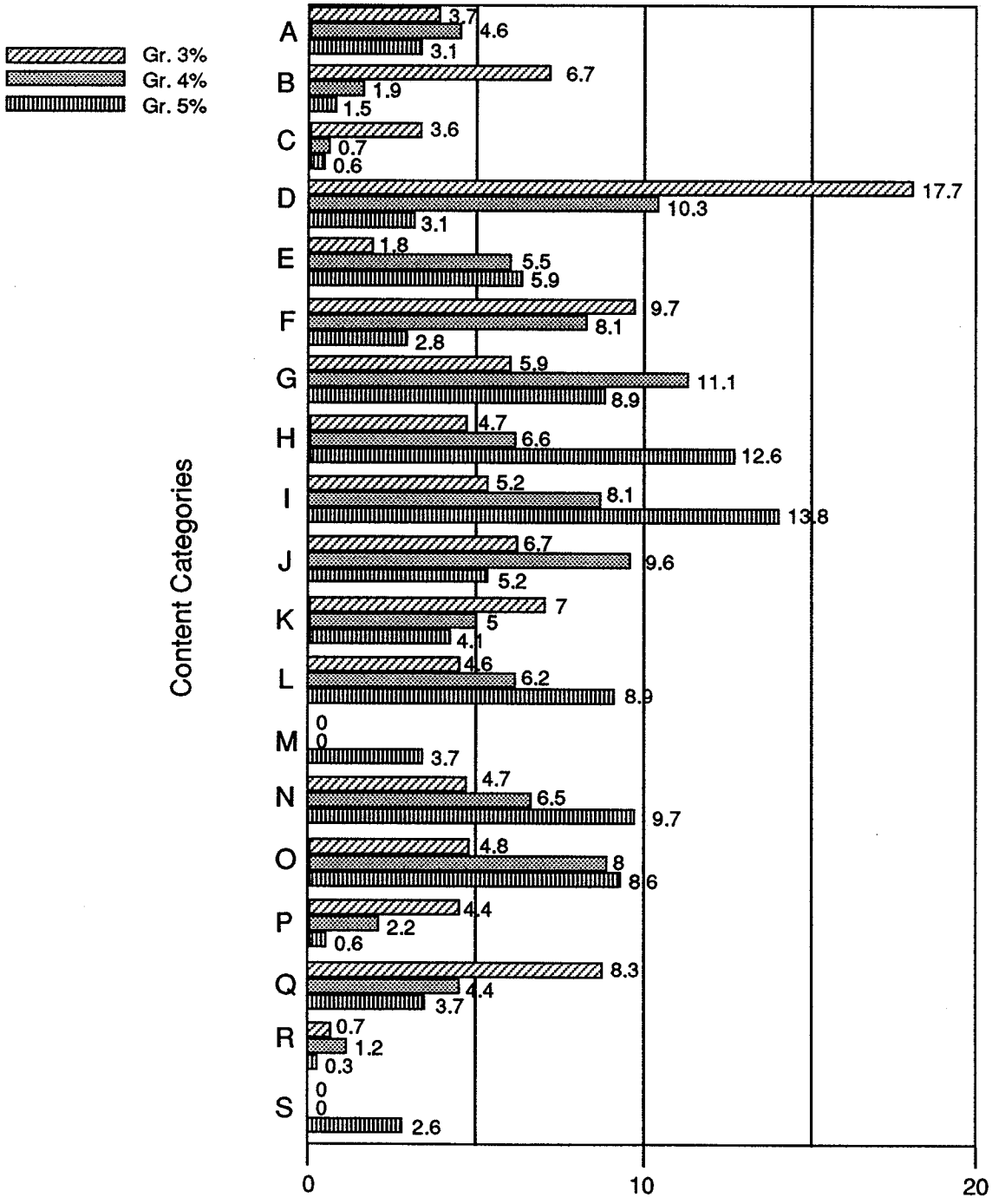


Figure 2. Graphic Display of 19 Main Content Category Frequencies Across Three Grades

Table 4. Grade 3 Open Court Curriculum Content Codes (ordered), Including Number of Activities Representing, Percent of Activities Representing, and Cumulative Percent of Activities Representing.

Order	Content	# Activ.	% Activ.	Cum %	Order	Content	# Activ.	% Activ.	Cum %
1	Q1	82	8.3	8.3	46	I8	6	0.6	89.8
2	D2	75	7.6	15.9	47	L5	6	0.6	90.4
3	G1	50	5.1	21.0	48	B7	5	0.5	90.9
4	H1	38	3.8	24.8	49	B1	5	0.5	91.4
5	D7	36	3.6	28.4	50	B6	5	0.5	91.9
6	F5	32	3.2	31.7	51	N5	5	0.5	92.4
7	F2	31	3.1	34.8	52	L2	4	0.4	92.8
8	I3	27	2.7	37.6	53	J3	4	0.4	93.2
9	O1	26	2.6	40.2	54	L6	4	0.4	93.6
10	K1	26	2.6	42.8	55	F3	4	0.4	94.0
11	D6	25	2.5	45.3	56	D1	4	0.4	94.4
12	J1	25	2.5	47.9	57	F7	4	0.4	94.8
13	N3	23	2.3	50.2	58	G5	4	0.4	95.2
14	A1	22	2.2	52.4	59	C2	4	0.4	95.6
15	F4	20	2.0	54.5	60	F1	3	0.3	96.0
16	O3	20	2.0	56.5	61	E4	3	0.3	96.3
17	D3	20	2.0	58.5	62	D8	3	0.3	96.6
18	B5	20	2.0	60.5	63	J5	3	0.3	96.9
19	P5	19	1.9	62.4	64	A4	3	0.3	97.2
20	L3	16	1.6	64.1	65	N4	3	0.3	97.5
21	J4	14	1.4	65.5	66	P3	3	0.3	97.8
22	P2	14	1.4	66.9	67	F6	2	0.2	98.0
23	I1	13	1.3	68.2	68	I4	2	0.2	98.2
24	E2	13	1.3	69.5	69	L4	2	0.2	98.4
25	L1	13	1.3	70.9	70	E3	2	0.2	98.6
26	B3	13	1.3	72.2	71	I6	2	0.2	98.8
27	J2	13	1.3	73.5	72	I2	1	0.1	98.9
28	C3	12	1.2	74.7	73	C6	1	0.1	99.0
29	C5	11	1.1	75.8	74	A3	1	0.1	99.1
30	D4	11	1.1	76.9	75	O2	1	0.1	99.2
31	K7	11	1.1	78.0	76	G2	1	0.1	99.3
32	A2	11	1.1	79.1	77	G4	1	0.1	99.4
33	B4	11	1.1	80.3	78	G6	1	0.1	99.5
34	N7	9	0.9	81.2	79	K4	1	0.1	99.6
35	K2	9	0.9	82.1	80	J7	1	0.1	99.7
36	K3	9	0.9	83.0	81	D5	1	0.1	99.8
37	H2	8	0.8	83.8	82	G3	1	0.1	99.9
38	C4	8	0.8	84.6	83	P6	1	0.1	100.0
39	K5	7	0.7	85.3					
40	R1	7	0.7	86.0					
41	B2	7	0.7	86.7					
42	N6	6	0.6	87.3					
43	K6	6	0.6	88.0					
44	P1	6	0.6	88.6					
45	J6	6	0.6	89.2					

but not for item construction. Response Type coding produced little variation; nearly all activities in the basal required Production responses, and only Production responses were permitted in the group test.

Cues/Hints

Cues/Hints were of little use in preparing a survey test with the scope of a full year. Cues/Hints information would be useful in constructing items for narrower diagnostic skills tests or Unit-level tests. Most lesson activities which included cues or hints

Table 5. Grade 4 Open Court Curriculum Content Codes (ordered), Including Number of Activities Representing, Percent of Activities Representing, and Cumulative Percent of Activities Representing.

Order	Content	# Activ.	% Activ.	Cum %	Order	Content	# Activ.	% Activ.	Cum %
1	D2	32	4.7	4.7	46	A4	5	0.7	86.9
2	G1	30	4.4	9.1	47	K2	5	0.7	87.6
3	Q1	30	4.4	13.5	48	L4	5	0.7	88.4
4	O3	27	4.0	17.5	49	P1	5	0.7	89.1
5	O1	25	3.7	21.2	50	H4	5	0.7	89.8
6	F5	21	3.1	24.3	51	I5	5	0.7	90.6
7	N3	20	2.9	27.2	52	G3	4	0.6	91.2
8	G5	19	2.8	30.0	53	H2	4	0.6	91.8
9	E3	18	2.7	32.7	54	D6	4	0.6	92.3
10	I4	18	2.7	35.3	55	D1	4	0.6	92.9
11	N4	17	2.5	37.8	56	D5	3	0.4	93.4
12	G4	15	2.2	40.1	57	H5	3	0.4	93.8
13	A1	15	2.2	42.3	58	I2	3	0.4	94.3
14	E4	15	2.2	44.5	59	I7	3	0.4	94.7
15	K7	14	2.1	46.5	60	J7	3	0.4	95.1
16	J4	13	1.9	48.5	61	C3	3	0.4	95.6
17	K1	12	1.8	50.2	62	E5	2	0.3	95.9
18	F2	12	1.8	52.0	63	F6	2	0.3	96.2
19	H1	12	1.8	53.8	64	K5	2	0.3	96.5
20	J5	12	1.8	55.5	65	O2	2	0.3	96.8
21	D4	11	1.6	57.1	66	B5	2	0.3	97.1
22	J1	11	1.6	58.8	67	J	2	0.3	97.3
23	A3	11	1.6	60.4	68	B2	2	0.3	97.6
24	L3	11	1.6	62.0	69	L1	2	0.3	97.9
25	H8	10	1.5	63.5	70	H3	2	0.3	98.2
26	I8	10	1.5	64.9	71	P5	1	0.1	98.4
27	I1	9	1.3	66.3	72	H9	1	0.1	98.5
28	J2	9	1.3	67.6	73	G6	1	0.1	98.7
29	P2	9	1.3	68.9	74	E1	1	0.1	98.8
30	L5	8	1.2	70.1	75	H6	1	0.1	99.0
31	D7	8	1.2	71.3	76	C2	1	0.1	99.1
32	F1	8	1.2	72.5	77	C6	1	0.1	99.3
33	L2	8	1.2	73.6	78	E6	1	0.1	99.4
34	D3	8	1.2	74.8	79	L6	1	0.1	99.6
35	J6	8	1.2	76.0	80	K3	1	0.1	99.7
36	R1	8	1.2	77.2	81	B4	1	0.1	99.9
37	N1	7	1.0	78.2	82	B6	1	0.1	100.0
38	F3	7	1.0	79.2					
39	I3	7	1.0	80.3					
40	J3	7	1.0	81.3					
41	B3	7	1.0	82.3					
42	H10	7	1.0	83.4					
43	L7	7	1.0	84.4					
44	G2	6	0.9	85.3					
45	F4	6	0.9	86.2					

for students were followed by more summative activities where these cues/hints were removed. Therefore, no cues/hints were included with items prepared for the survey test.

In summary, item creation relied mainly on Curriculum Content and Cognitive Level data. Much of the Cognitive Level data was integrated into revised Curriculum Content codes. As well as using these two codes, skimming the actual lessons was necessary for item creation.

Practical considerations also dictated how items were created. No lengthy discussion-type items

Table 6. Grade 5 Open Court Curriculum Content Codes (ordered), Including Number of Activities Representing, Percent of Activities Representing, and Cumulative Percent of Activities Representing.

Order	Content	# Activ.	% Activ.	Cum %
1	O3	56	7.2	7.2
2	L4	35	4.5	11.7
3	I5	32	4.1	15.9
4	N3	28	3.6	19.5
5	I4	27	3.5	23.0
6	H3	22	2.8	25.8
7	Q1	22	2.8	28.6
8	G4	20	2.6	31.2
9	H10	19	2.5	33.7
10	F5	18	2.3	36.0
11	G5	17	2.2	38.2
12	E1	16	2.1	40.3
13	H5	16	2.1	42.3
14	I2	16	2.1	44.4
15	M2	14	1.8	46.2
16	D7	13	1.7	47.9
17	J4	13	1.7	49.5
18	K7	13	1.7	51.2
19	G1	12	1.5	52.8
20	L3	12	1.5	54.3
21	N1	12	1.5	55.9
22	A3	11	1.4	57.3
23	L1	11	1.4	58.7
24	N4	11	1.4	60.1
25	N5	11	1.4	61.5
26	B3	10	1.3	62.8
27	H2	10	1.3	64.1
28	M1	10	1.3	65.4
29	Z2	10	1.3	66.7
30	E8	9	1.2	67.9
31	G3	9	1.2	69.0
32	H1	9	1.2	70.2
33	I7	9	1.2	71.4
34	Z1	9	1.2	72.5
35	E5	8	1.0	73.5
36	I6	8	1.0	74.6
37	K2	8	1.0	75.6
38	E4	7	0.9	76.5
39	G	7	0.9	77.4
40	H6	7	0.9	78.3
41	H8	7	0.9	79.2
42	I3	7	0.9	80.1
43	J5	7	0.9	81.0
44	J6	7	0.9	81.9
45	N2	7	0.9	82.8
46	O1	7	0.9	83.7
47	Q2	7	0.9	84.6
48	A4	6	0.8	85.4
49	E3	6	0.8	86.2
50	L7	6	0.8	87.0
51	N6	6	0.8	87.7
52	A1	5	0.6	88.4
53	D4	5	0.6	89.0
54	H4	5	0.6	89.7
55	I1	5	0.6	90.3
56	J3	5	0.6	91.0
57	K3	5	0.6	91.6
58	C3	4	0.5	92.1
59	G2	4	0.5	92.6
60	D1	3	0.4	93.0
61	F6	3	0.4	93.4
62	I8	3	0.4	93.8
63	J2	3	0.4	94.2
64	J7	3	0.4	94.6
65	K5	3	0.4	95.0
66	L5	3	0.4	95.4
67	M3	3	0.4	95.7
68	O2	3	0.4	96.1
69	P2	3	0.4	96.5
70	D3	2	0.3	96.8
71	H9	2	0.3	97.0
72	J1	2	0.3	97.3
73	K1	2	0.3	97.5
74	L2	2	0.3	97.8
75	R1	2	0.3	98.1
76	A2	1	0.1	98.2
77	A10	1	0.1	98.3
78	B2	1	0.1	98.5
79	B6	1	0.1	98.6
80	C6	1	0.1	98.7
81	D2	1	0.1	98.8
82	F2	1	0.1	99.0
83	H7	1	0.1	99.1
84	K4	1	0.1	99.2
85	M	1	0.1	99.4
86	M4	1	0.1	99.5
87	O7	1	0.1	99.6
88	P1	1	0.1	99.7
89	P6	1	0.1	99.9
90	S	1	0.1	100.0

were included because of the amount of reading involved, and the need for orchestration by the teacher, if presented verbally. Assessment of "Reasoning Skills" (Q codes) involving discussion was therefore not attempted, given the prescribed format and length of the diagnostic survey test. However, multi-part application problems were presented as text format.

Items were produced in two steps. They first were sketched in 2x3 inch boxes (see Appendix F) and then were electronically "drawn." Fidelity to the curriculum was possible through the inclusion of diagrams, special symbols, and drawings (Stone, 1989). All items were screened internally by the research/development team for clarity, fairness, and relationship to the curriculum.

Items were drawn as individual "pict" files using Canvas® (Deneba software) on a Macintosh® (Apple Corporation) computer. They were stored and indexed as a simple, flat data base in Microsoft Word® 4.0 tables. These tables served not only as storage, but also to format the final tests for printing.

Test Production

Valid and reliable survey diagnostic tests need to meet at least four important criteria: (a) The form and content of test items must match curriculum learning activities; (b) the relative frequency or predominance of various item types must match the curricular emphasis on those skills; (c) the test format must assure that students can't get items correct by guessing alone; and (d) the test must be reliable and useful, yielding test and subtest scores that are stable and discriminate well among students. The first two criteria relate to content and face validity, and pertain mainly to criterion-referenced tests. The second two criteria are important for norm-referenced tests. A survey diagnostic test must meet both criterion-referenced and norm-referenced test requirements. The second criterion is met by test production—the strategic selection of items previously produced from curriculum analysis data.

Curriculum analysis data summaries from the Excel® spreadsheet (see Figures 4-8) helped ensure that the survey test closely reflected the curriculum in content and emphasis. Items were selected for each grade-level survey test according to the relative frequency of occurence of Curriculum Content and Cognitive Level categories. Other more detailed information on Task Intent (introduced, practiced, reviewed) and Lesson # was not required, as each test spanned the full curriculum year.

Both major (19) and minor (115) curriculum content categories were considered in selecting test items. A severe limitation on item selection was the 60-90 minute length of the test, restricting each test to approximately 150 items. In addition, at least three or four examples of each item type were desired to increase reliability in reporting sub-skill scores. Yet, at all three grade levels, several subskills were represented only infrequently in learning activities. For example, Figure 8 shows that, of the 90 sub-codes represented in the Grade 5 curriculum, the least frequent 15 were each found in only one activity, and the least frequent 31 were each found in three or fewer activities.

This practice of "exposure," or presenting many skills, but with few opportunities for practice, has been noted by others. The practice presents a dilemma for test producers; all of the many different subskills cannot be reliably measured in a test of reasonable length. The second dilemma is knowing whether mastery can be expected and should be tested for a skill that is given exposure in only one or two activities. For these reasons, a number of arbitrary decisions were needed to guide selection and omission of item types.

The survey test produced for each grade level was designed for untimed administration during three sessions of approximately 20-30 minutes. Because of these limitations, a ceiling of 150 items per test was permitted. The number of items per test were: Grade 3 (155), Grade 4 (147), and Grade 5 (147). Curriculum Content codes and frequencies for the

Table 7a. Content of Grade-Level Survey Diagnostic Tests by Frequency of Main Curriculum Content Codes (Multiple coding permitted).

Grade 3 Test		Grade 4 Test		Grade 5 Test	
Code	# Items	Code	# Items	Code	# Items
D	55	N	33	I	32
K	23	D	29	D	32
C	21	O	26	H	30
B	20	I	25	N	21
N	19	A	24	E	20
F	15	G	15	O	19
O	14	E	13	M	17
H	10	H	12	G	16
I	9	B	10	J	14
J	8	L	9	L	10
E	6	F	9	C	4
A	6	K	7	K	2
L	5	J	1	F	2
		C	1	B	2
Total Items:	155		147		147
Total Codes:	211		214		221

Table 7b. Content of Grade-Level Survey Diagnostic Tests by Frequency of Curriculum Content Sub-Codes (Multiple coding permitted).

Grade 3 Test		Grade 4 Test		Grade 5 Test	
Code	# Items	Code	# Items	Code	# Items
D2	24	O3	23	O3	19
F2	15	G4	15	G4	16
O3	14	N4	14	I4	15
N4	12	D7	12	M1	10
D4	12	I4	12	I4	10
K1	10	N3	10	H1	10
I3	9	D2	8	N3	9
C5	9	L5	7	D4	9
D6	9	H10	6	M2	7
K6	8	A4	6	H3	7
H1	7	A3	6	E8	7
C3	6	D4	6	E2	7
N3	6	N6	6	D2	7
E2	6	I5	6	J5	6
D7	6	H4	6	I5	6
B5	6	A1	6	D7	6
K3	5	E4	6	D6	6
L3	5	A2	6	I2	5
B3	5	K1	5	N6	4
J1	4	E2	4	N4	4
D3	4	I3	4	N1	4
B2	3	B5	4	E5	4
C2	3	F2	3	D5	4
B6	3	O2	3	J4	3
C4	3	F3	3	J3	3
J7	3	F5	3	I7	3
A1	3	D3	3	I3	3
A3	3	38	3	H9	3
H10	3	B6	3	H4	3
N6	1	N1	3	H10	3
J2	1	I8	3	J6	2
		L2	2	H5	2
		B4	2	F5	2
		B2	1	C5	2
		C2	1	C3	2
		K5	1	K5	1
		K3	1	K3	1
		J3	1	H6	1
				H2	1
				E6	1
				E4	1
				B6	1
				B4	1
Total Items:	155		147		147
Total Codes:	211		214		221

resulting tests are presented in Table 7a; the sub-codes and related frequencies are presented in Table 7b.

The resulting three grade-level tests, with teacher directions, are presented in Appendix G. A second reliability exercise was conducted on coding the items in final test format. Three raters independently indexed all items by Content sub-codes, while being blind to their curriculum source. Multiple codes were permitted; no limit was set on the number of possible codes. Across the three tests, *perfect* agreement was reached on 98 of the 156 items. Over all items, 71% agreement was obtained, using the modified Rand statistic described in Appendix E. Higher agreement would have been obtained if only one code were permitted per item.

Thus far, this report has focussed on how test items were created and selected to logically match the curriculum. Besides logical justification, newly constructed tests need to be supported by empirical, psychometric evidence of their usefulness in assessing math skills.

Item Analysis

Tests were administered by teachers to approximately 240 students at each grade level, and scored by the research/development team. Prior to item analyses, typical norm-referenced results were promptly provided to participating schools. Approximately 2 weeks after receiving the test results, raw "number correct" scores for each student were returned to participating teachers (see Appendix H). Frequency distributions and decile line graphs were also provided to allow normative interpretation of the raw scores (see Appendix I). Deciles only (rather than percentiles) were provided to acknowledge the lack of precision in this "first run" set of items. Administrators were also provided with box plots, allowing comparisons between schools at each grade level (see Appendix J).

Although "consumer feedback" was not formally solicited from teachers, they informally reported that the test required significantly more time than had been anticipated. Clearly, the test had to be shortened. They also helped identify a few ambiguous

Table 8. Results of Item Analysis with Desired Criteria.

Criteria	Test-level Data	Subtest Data	Item-level Data
1. High overall test reliability (internal consistency): above .90.	Spearman-Brown/ Cronbach's Alpha		
2. Most items not too hard or too easy: 15 to .85 difficulty index range.			Item Difficulty Index
3. Contains only item-types taught in the curriculum—from early to late in the year.	Curriculum analysis data.		Item Content Codes.
4. Overall test length reduced from 6 to 4 pages (approx. 60 minutes).	Count items		
5. Frequency of item-types roughly reflects emphasis (number of activities) within curriculum.	Curriculum analysis		Item Content Codes
6. Some representation desired for all common item-types.	Curriculum analysis		Item Content Codes
7. Test difficulty level not to exceed 30% to avoid excessive frustration by low achievers.	Test Difficulty Index		
8. Exclusion of ambiguous items.		Spearman-Brown/ Cronbach's Alpha	Item Reliability Index
9. Reasonably high subtest reliability (internal consistency) (for 10-15 scores): .80 - .90.		Spearman-Brown/ Cronbach's Alpha	
10. Each subtest to include range of item difficulties.			Item Difficulty Index
11. Each subtest to be homogeneous in item - types.			Item Content Codes

and otherwise inappropriate items, which were deleted.

Alternate forms of the survey diagnostic test were to be administered at two or three points in the school year to provide two kinds of information on student performance and growth: (a) criterion-referenced information on individual items and groupings of items (subtests), and (b) norm-referenced test and subtest scores. To obtain useful norm-referenced scores requires reliability at the subtest and total test levels. Without test and subtest reliability, comparisons of fall, winter, and spring test results cannot be made with reasonable confidence; judgements of student growth are not possible. Test and subtest reliability is also needed to interpret the accuracy of a single obtained norm-referenced score, i.e. within a confidence interval.

Table 9. Test & Subtest Information from Winter Administration.

Grade Level	# Main Problems	# Items	# Subtests	Test Difficulty	Spearman Brown	Cronbach's Alpha
3	54	95	12	38%	.97	.96
4	50	77	13	32%	.96	.94
5	52	74	10	29%	.97	.95

Grade 3 Subtest Data

Subtest:	place value	addition	subtraction	multiplication	combine op.	fractions
Code:	A3	B	C	D	F	H
# Items:	3	2	3	30	5	6
Difficulty:	.24	.77	.78	.45	.74	.14
*Reliability:	too few	too few	too few	.96 / .93	.66 / .56	.93 / .85

Subtest:	decimals	measure.	word probs.	applications	algebra	relations
Code:	I	K1,3,6	K5,J1	N	O	P
# Items:	7	12	9	6	6	6
Difficulty:	.18	.28	.45	.34	.11	.33
*Reliability:	.89 / .83	.81 / .81	.87 / .79	.70 / .62	.85 / .75	.89 / .89

Grade 4 Subtest Data

Subtest:	numeration	rounding	addition	multiplication	division	combine op.
Code:	A	A4	B	D	E	F,G
# Items:	13	2	2	8	2	6
Difficulty:	.48	.02	.76	.34	.04	.36
*Reliability:	.92 / .83	too few	too few	.82 / .74	too few	.73 / .70

fractions	decimals	measure.	word probs.	averaging	chart/graph	algebra
H	I	K	N	N1	N3	O
3	7	1	8	1	6	18
.07	.27	.04	.26	.01	.32	.31
too few	.81 / .72	too few	.75 / .66	too few	.73 / .66	.90 / .85

Grade 5 Subtest Data

Subtest:	multiplication	division	combine op.	fractions	decimals	word probs.
Code:	D	E	G	H	I	J,N
# Items:	3	3	10	8	11	11
Difficulty:	.44	.29	.29	.37	.23	.24
*Reliability:	too few	too few	.79 / .79	.94 / .87	.88 / .80	.77 / .73

Subtest:	geometry	averaging	chart/graph	algebra
Code:	L	N1	N3	O
# Items:	4	3	3	18
Difficulty:	.23	.14	.29	.34
*Reliability:	.92 / .89	too few	too few	.89 / .84

Note. Reliability coefficients: *Spearman-Brown Prophesy / Cronbach's Alpha.*

The examination of item level results entailed several steps. First, classical item analysis was conducted on Testat® software (SYSTAT, Inc.) to help establish and improve, through item elimination, test and subtest reliability. The analysis yielded test, subtest, and item-level information. Test and subtest level information included: (a) reliability (internal consistency, via Spearman-Brown Prophesy Formula, Cronbach's Alpha), (b) difficulty (percent correct), and (c) measurement error for 15 score intervals (SEmeas). Item-level information included: (a) difficulty (percent correct on item), (b) standard deviation (SD), (c) item-total test correlation, (d) item reliability index, and (e) test reliability if item were deleted.

These data were used to help decide which items to eliminate. The desired criteria and relevant data are presented in Table 8 (page 15). Using these criteria, the Grade 3 test was reduced to 61% of its original length, and the Grade 4 and 5 tests to 52% and 50%, respectively, of their original lengths. Summary information on the three revised tests is presented in Table 9.

A minimum of four items was required to compute subtest internal consistency (Spearman-Brown and Alpha). Unfortunately, 13 of the 34 subtests (across the three grades) contained too few items.

Reliability

At Grade 3, Spearman-Brown coefficients varied from .66 (*combined operations*) to .96 (*multiplication*). At Grade 4, the range was .73 (*combined operations*) to .92 (*numeration*). The range at Grade 5 was .77 (*word problems*) to .94 (*fractions*). Generally, *word problems/ applications* and *combined operations* had lowest reliability. There was no overall relationship between subtest reliability and difficulty.

Perhaps the main use for reliability coefficients is in calculation of a confidence interval around obtained scores. Confidence intervals are presented and described later for spring testing data.

Difficulty

Item difficulty levels are summarized in Table 10. This table shows that whole number basic arithmetic operations—both singly and in combination—were relatively easy, while averaging and decimals were among the most difficult skill areas. Word problems/applications were comparatively of medium difficulty at all grade levels. Grade 4 and 5 tests were difficult; only two Grade 4 skills exceeded 50% correct, and none did so at Grade 5. Because extreme test difficulty tends to decrease reliability, "pruning" the least reliable items from difficult tests is essential.

Following item analysis, schools received the second batch of test results. The first batch of results had been based only on raw scores and percentiles for the entire test. This second batch of results was based on those items which survived the screening.

Table 10. Subtest Difficulty Levels of Winter Tests.

Gr. 3 Content	Difficulty	Gr. 4 Content	Difficulty	Gr. 5 Content	Difficulty
algebra	.11	averaging	.01	averaging	.14
fractions	.14	rounding	.02	decimals	.23
decimals	.18	division	.04	geometry	.23
place value	.24	measure.	.04	word probs.	.24
measure.	.28	fractions	.07	division	.29
relations	.33	word probs.	.26	chart/graph	.29
applications	.34	decimals	.27	combined op.	.29
multiplication	.45	algebra	.31	algebra	.34
word probs.	.45	chart/graph	.32	fractions	.37
combined op.	.74	multiplication	.34	multiplication	.44
addition	.77	combined op.	.36		
subtraction	.78	numeration	.48		
		addition	.76		

These latter results also included subtest scores with total test scores. Also criterion-referenced "percent correct" scores were provided, along with grade-level standards to permit normative score interpretations. Appendix K is a sample of results sent to teachers. Teachers also received information on the difficulty level of each item, next to a facscimile of the item (see Appendix L). This latter information can potentially provide more direction to instruction than test or subtest information alone. However, no consumer feedback was obtained on its usefulness.

Test Production: Spring Version

Strictly equivalent test forms were required for subtest score comparisons between fall/winter and spring. An equivalent spring test was produced by duplicating each item on the fall test, and changing only numerals and names. Problem pairs for fall/ winter and spring were identical in such features as the use of decimals and zeros, amount and type of regrouping required, number of digits in a numeral,

Table 11. Test & Subtest Information from Spring Administration.

Grade Level	# Main Problems	# Items	# Subtests	Test Difficulty	Spearman Brown	Cronbach's Alpha
3	37	95	12	73%	.97	.96
4	44	77	13	56%	.96	.95
5	36	76	10	58%	.95	.95

Grade 3 Subtest Data:

Subtest:	place value	addition	subtraction	multiplication	combine op.	fractions
Code:	A3	B	C	D	F	H
# Items:	3	2	3	30	5	6
Difficulty:	.36	.88	.86	.85	.82	.62
*Reliability:	too few	too few	too few	.90 / .88	.66 / .60	.95 / .89

Subtest:	decimals	measure.	word probs.	applications	algebra	relations
Code:	I	K1,3,6	K5,J1	N	O	P
# Items:	7	12	9	6	6	6
Difficulty:	.85	.62	.67	.66	.69	.76
*Reliability:	.92 / .83	.82 / .82	.86 / .78	.66 / .59	.87 / .83	.86 / .86

Grade 4 Subtest Data:

Subtest:	numeration	rounding	addition	multiplication	division	combine op.
Code:	A	A4	B	D	E	F,G
# Items:	13	2	2	8	2	6
Difficulty:	.62	.06	.83	.73	.66	.55
*Reliability:	.82 / .73	too few	too few	.74 / .70	too few	.66 / .67

fractions	decimals	measure.	word probs.	averaging	chart/graph	algebra
H	I	K	N	N1	N3	O
3	7	1	9	1	5	17
.35	.52	.06	.40	.40	.46	.67
too few	.84 / .77	too few	.82 / .75	too few	.79 / .68	.92 / .90

Grade 5 Subtest Data:

Subtest:	multiplication	division	combine op.	fractions	decimals	word probs.
Code:	D	E	G	H	I	J,N
# Items:	3	3	10	9	11	11
Difficulty:	.73	.58	.57	.71	.47	.48
*Reliability:	too few	too few	.84 / .82	.83 / .78	.88 / .81	.79 / .75

Subtest:	geometry	averaging	chart/graph	algebra
Code:	L	N1	N3	O
# Items:	4	3	4	18
Difficulty:	.62	.51	.56	.63
*Reliability:	.86 / .90	too few	.75 / .73	.83 / .83

*Note: Reliability coefficients: Spearman-Brown Prophesy / Cronbach's Alpha

etc. The equivalent spring tests are included as Appendix M. Summary information on these tests is presented in Table 11.

Test Reliability

As expected, the spring test's overall reliability was similar to that for the winter testing—around .96. Reliability figures for subtests were also similar overall, although individual subtest reliability differences between winter and spring of .06 to .10 were common.

Reliability coefficients are needed to calculate confidence intervals around obtained mean scores and individual student scores. Confidence intervals around mean scores are based on the standard error of the mean (SEM), while the standard error of measurement (SEmeas) is the basis for confidence intervals around individual scores (Salvia & Ysseldyke, 1978). We can predict with 64% certainty that a student's true score lies somewhere within a confidence interval of ± 1 SEmeas.

Although a single SEmeas can be calculated for all individual student scores achieved on a given test, this value is not as accurate as conditional SEmeas calculated separately for different raw score levels (Feldt, Steffen, & Gupta, 1985). Table 12 presents conditional SEmeas for raw score levels demarked in .5 SD units from the mean (Automatically calculated by Testat® software).

From Table 12, based on the conditional SEmeas, we can be 64% certain that a Grade 5 student with an obtained raw score of 26 (26/76 = 34% correct) has a true score of 26±3.1, or within the range of 23 to 29. We can also translate from raw scores to percentile ranks. Percentile ranks were calculated using StatView® statistical software (see Appendix N). The raw score confidence interval can be translated to percentiles by using Appendix N. We can be 64% certain that an obtained score at the15th percentile represents a true score somewhere between the 12th and 18th percentiles. This range of uncertainty represents good precision for a survey test.

Table 12. Conditional Standard Error of Measurement for 10 One-Half SD Raw Score Intervals: Spring Tests.

Grade 3

Raw Score Interval			Standard Error
11	to	19	2.5
20	to	28	3.3
29	to	37	3.9
38	to	46	3.8
47	to	55	3.5
56	to	64	3.3
65	to	73	3.1
74	to	82	3.5
83	to	91	2.2
92	to	>92	2.1

Grade 4

Raw Score Interval			Standard Error
1	to	7	1.4
8	to	15	3.0
16	to	23	2.2
24	to	31	3.5
32	to	39	3.8
40	to	46	3.3
47	to	54	2.9
55	to	62	2.5
63	to	70	4.1
71	to	>71	2.0

Grade 5

Raw Score Interval			Standard Error
1	to	8	1.6
9	to	16	3.0
17	to	24	2.8
25	to	32	3.1
33	to	40	4.6
41	to	47	4.1
48	to	55	3.7
56	to	63	3.4
64	to	>64	2.5

Subtest Reliability

Confidence intervals can be constructed for subtests also. The somewhat typical Grade 3 Algebra subtest (# items - 6; difficulty - .69; reliability - .87 / .83) was selected for demonstration.

Based on the conditional SEmeas in Table 13, we can be 64% certain that a Grade 3 student with an obtained algebra subtest score of 2 (17th percentile) has a true score between 1 (14th percentile) and 3 (23rd percentile). Because of the small number of items in the algebra subtest, it necessarily lacks the precision of the total test score. However, considering the short length of the subtest, the degree of precision obtained is quite good.

Difficulty

At all three grade levels, students found the spring test much easier. Winter and spring difficulty levels were, respectively, Grade 3: .38 and .73; Grade 4: .32 and .56; Grade 5: .29 and .58. Therefore, *increase* in "percent-correct" scores was, for Grade 3: 35%, for Grade 4: 23%, and for Grade 5: 29%. As was done for winter test items, spring item difficulty levels were sorted and summarized in Table 14. From difficulty levels of subtests at winter and spring, student improvement was calculated. Subtests were then ranked in from least to most improvement (see Table 15).

Students improved in all subtests from winter to spring, but the amount of change ranged from 8 to 67 percentage points (99 points maximum). Few commonalities were noted across the three grades in amount of subtest improvement. Whereas Grade 3 students improved most in "decimals," Grade 5 students improved least on that same subtest.

For Grade 3 students, the amount of subtest improvement from winter to spring was easily predicted by the winter subtest scores; the lower the winter subtest score, the more change in the spring. This phenomenon did not hold for Grades 4 and 5, however.

Summary and Conclusions

The purpose of this research and development project was to produce efficient, valid and reliable diagnostic survey tests for Grades 3, 4, and 5, based on Open Court math curricula. The impetus for this project was twofold: (a) the poor match documented between the content of published math curricula and most commercial math tests, and (b) the technical

Table 13. Conditional Standard Error of Measurement and Percentile Ranks for Spring Grade 3 Algebra Subtest (6 items)

Raw Score Interval			Standard Error	Raw Score	Percentile Rank
.6	to	1.6	1.0	0	.07
1.7	to	2.6	1.1	1	.14
2.7	to	3.6	1.2	2	.17
3.7	to	4.6	0.8	3	.23
4.7	to	5.6	1.0	4	.34
				5	.56
				6	.85

inadequacy of progress tests/unit tests/probes that accompany basal math programs. Although the procedures followed in this project potentially can be used to produce a variety of types of tests for different decision needs, here only survey diagnostic tests were produced.

The following desirable attributes were sought in a survey diagnostic test:

Efficiency/Utility:

(a) the test can be efficiently group-administered within two 30-minute sessions.

(b) it exists in at least 2 or 3 strictly equivalent forms, allowing fall, winter, and spring testing;

(c) Subtest scores are produced for major skill/content areas;

(d) total test and subtest scores permit both criterion-referenced (percent of items correct) and normative (percentile rank) interpretations;

Validity:

(e) it represents all major skills in the full year's math program;

(f) the number of test items of various problem types represents the curriculum focus;

Reliability:

(g) total test and subtest scores have known standard errors of measurement, reflecting sufficient sensitivity for 2 or 3 administrations per year;

(h) the use of only production responses effectively to eliminate student guessing.

This section will summarize the procedures and difficulties in attempting to produce a test with these attributes.

Curriculum analysis proved to be an essential first step, because the content/skill index to the instructional materials provided by the publisher was both incomplete and inaccurate. The taxonomy first created for the curriculum analysis was too detailed and insufficient for identifying item types. Taxonomy information on Curriculum Location & Focus was not needed, since full-year tests were to be produced. The most useful parts of the taxonomy were Cognitive Level and Curriculum Content. Even with the taxonomy codes, it proved useful to skim the teacher and student materials to more precisely identify item types. The need to review instructional materials depends on how closely the item types need to be linked to instructional content. In general, the curriculum could be coded with at least moderate interrater reliability.

Two unanticipated results arose from the curriculum analysis. First, many of the lesson activities could not be assessed, because there was no identifi-

Table 14. Subtest Difficulty Levels of Spring Tests.

Gr. 3 Content	Difficulty	Gr. 4 Content	Difficulty	Gr. 5 Content	Difficulty
place value	.36	rounding	.06	decimals	.47
fractions	.62	measure.	.06	word probs.	.48
measure.	.62	fractions	.35	averaging	.51
applications	.66	averaging	.40	chart/graph	.56
word probs.	.67	word probs.	.40	combined op.	.57
algebra	.69	chart/graph	.46	division	.58
relations	.76	decimals	.52	geometry	.62
combined op.	.82	combined op.	.55	algebra	.63
decimals	.85	numeration	.62	fractions	.71
multiplication	.85	division	.66	multiplication	.73
subtraction	.86	algebra	.67		
addition	.88	multiplication	.73		
		addition	.83		

able student performance. Second, a number of lesson activities had expected responses which could not be assessed by the group test because they required verbal responses, selection responses, or group responses—which could not be easily translated to individual pencil/paper tasks. The decision to use only written test responses was based on efficiency, and the decision to use only production responses was intended to eliminate guessing. Unfortunately, the structure of some lesson activities encouraged selection responses, and permitted "accurate guessing"—these skills were not included in the tests.

instructional emphasis. The rationale for this weighting was to make the total test score more accurately reflect progress in the curriculum. This rationale holds only for the total test score—not for subtests. Changes are recommended for future efforts: Individual item frequency should not attempt to reflect the curriculum focus. Instead, subtest scores should be weighted to reflect curriculum focus, and then averaged for the total test score. Some item types require more exemplar items than others, because they represent a broader skill/content domain. This requirement is much more defensible than weighting the total test score by the number of items.

Table 15. Change in Subtest Difficulty Levels From Winter to Spring*

Gr. 3 Content	Change in Difficulty	Gr. 4 Content	Change in Difficulty	Gr. 5 Content	Change in Difficulty
combined op.	.08	measure.	.02	decimals	.24
subtraction	.08	rounding	.04	word probs.	.24
addition	.11	addition	.07	chart/graph	.27
place value	.12	chart/graph	.14	combined op.	.28
word probs.	.22	numeration	.14	algebra	.29
applications	.32	word probs.	.14	division	.29
measure.	.34	combined op.	.19	multiplication	.29
multiplication	.40	decimals	.25	fractions	.34
relations	.43	fractions	.28	averaging	.37
fractions	.48	algebra	.36	geometry	.39
algebra	.58	averaging	.39		
decimals	.67	multiplication	.39		
		division	.62		

The electronic spreadsheet summary of curriculum analysis codes helped determine how many of which item types could representatively test the curriculum. The flexibility and power of spreadsheet summaries was barely tapped in this project. This technique permits more sophisticated curriculum analyses for a variety of purposes, such as comparing or evaluating curricula on internal criteria (e.g., continuity and repetition of essential skills).

A major difficulty encountered in creating an efficient, representative test was the large number of subskills coded in the curriculum for one school year—even one unit. For example, at Grade 5, 90 separate subskills were coded , although only 59 of these subskills made up 93% of the lesson activities (see Figure 8). The remaining 31 subskills each were found in only 1, 2, or 3 lessons—could these 31 subskills be mastered?—should they be assessed? The first, overly-long tests contained about 150 items (75-95 items after screening). The limited test length and the large number of identified subskills was a problem, given that 3 or 4 items were desired to reliably test each item type.

Compounding this problem was our desire to weight the test (by number of items) to reflect

Creating items as electronic "pict" files, database storage, and formatting and laser printing proved to be relatively easy. The power of this technique will not be realized, however, until items are repeatedly selected in different combinations for a variety tests and purposes. The dedicated Macintosh® software program, LXR-Test 4.0® (Logic Extension Resources, 1989) acts as a data base for pict items and formats them for test printing. It has received positive reviews (*MacGuide Magazine*, summer, 1988), and should be considered where a variety of tests are desired. The cut-and-pastes between Canvas® and Microsoft Word® could thus be eliminated.

The criteria for item screening (see Figure 10) were numerous and difficult to apply objectively. This was due in part to the desire to proportionally represent curriculum focus by number of test items, a goal which proved unnecessary, and which should be abandoned, in favor of the arithmetic weighting of subtest scores. Given the item analysis software we used, Testat®, another criterion should have been added: a minimum of 4 items per subtest; fewer cannot be analyzed by Testat®.

The rationale for the difficulty criteria (see Figure 10, #2 & #7) was mainly to obtain high classical

reliability for total test scores. Although a worthwhile pursuit from a norm-referenced viewpoint, an opposing criterion-referenced viewpoint would strenuously object to eliminating very hard items which to be introduced only later in the school year. Criteria #5 and #6 safeguard against this unwise elimination. In fact, several of the 11 criteria in Figure 10 are mutually inconsistent, designed to counterbalance one another. This inconsistency or incompatibility of norm-referenced and criterion-referenced assessment approaches is commonly described. However, acknowledging this *theoretical* inconsistency, *in practice*, we pursued the middle road of combined criterion/norm-referenced assessment.

The large initial group of test items permitted considerable flexibility in item deletion, leading to construction of tests with desirable criterion and norm-referenced attributes. Total test reliability (internal consistency) coefficients of around .95 were gratifying, considering the wide range of item types and subskills. Subtest reliability coefficients were not uniformly as high as desired. First, a minimum of 4 (preferably 6) items appeared necessary for stable reliability estimates. Second, reliability appeared to be a function of the total number of items in the subtest relative to the number of different item types. For example, for the Grade 3 Applications subtest, several different item types were represented in only 6 items, yielding low reliability coefficients of .70 / .62. Fewer different item types and/or more items in total would probably have yielded higher reliability. Still, 15 of the 22 subtest coefficients (Spearman-Brown) which could be calculated (had sufficient items) were .80 or greater. A reasonable goal for subtest reliability (internal consistency) is probably .85-.90. With a short subtest (6 items) .85-.90 reliability reflects sufficient test sensitivity for measurement of growth 2 or 3 times per year (see Table 13 and related text).

Two dedicated statistics programs were used for analysis of test data. A general statistics program, Statview® (Brainpower, Inc.), was used for calculation of percentile ranks only. The item analyses required specialized software, Testat®, which provided test, subtest, and item-level difficulty and reliability statistics. Although only classical item analyses were performed, this software also allows Rasch (1960) one-parameter item response theory (IRT) analysis. Rasch analysis adds additional power when subtests or items are to be selectively administered to students of different math abilities. However, for the present purpose of a diagnostic survey test, the IRT analysis appeared to offer little advantage. Since the item-level data are already entered, IRT can be conducted whenever needed. Production of percentile ranks would be a welcome addition to item analysis software, to eliminate the need for Statview®.

Student test scores returned to teachers (% correct for the test and subtests, by student, class, and grade) were easily produced from an Excel® spreadsheet; no statistics program was needed. The perceived usefulness of the information provided to teachers and the accessibility of the presentation format were not evaluated—no feedback was requested or obtained from schools. This shortcoming should be remedied in the future so test feedback "boilerplates" can be efficiently tailored. The use of IRT analyses opens possibilities of a wide range of test data to schools for various purposes. Feedback from schools is needed to help determine which data are worth summarizing and reporting.

A handicap in the use of item analysis for test production and feedback is the time required for (a) manual scoring of test protocols (since all responses were production) and (b) coding and inputting a 0/1 score for each item for each student. These activities contributed to delayed the return of subtest-level data to teachers. However, the turnaround time was reduced considerably from winter (when item screening also had to be carried out) to spring assessment. The delay between schools' receipt of the test protocols and receipt of the test scores was about 4 weeks (nearly 2 weeks of which was expended on test administration). Following are the times required for spring, from testing to delivering results to the schools:

(a) 10 days (May 14-25): administration and return of all tests (from 6 schools, 31 teachers, 750 students).
(b) 10 days (May 22-June 2): scoring and inputting of item-level data.
(c) 2 days (June 3-4): final checking of input data.
(d) 1 day (June 5): analysis and summary of test data.
(e) 2 days (June 6-7): formatting and printing student-level test summary to 31 teachers.
(f) 1 day (June 8): collating and delivering test scores.

The inclusion of State of Oregon learning goals with the curriculum analysis was included as a project sideline in order to meet an immediate district responsibility and to test the ability of the spreadsheet to cross-reference the Oregon and curriculum taxonomy content codes. Both goals were successfully achieved.

In pursuing this line of development further, we recommend the following priorities:

1. Streamlining the curriculum coding scheme, and applying it to other basal curricula.
2. Producing two other types of tests from existing

item information: Unit tests (with equivalent forms), and skill diagnostic tests.
3. Using a minimum of 4 to 6 items per test for any subskill for which scores are reported.
4. Weighting subtest scores prior to total test score calculation to accurately reflect curriculum emphasis.
5. Piloting LXR-Test software for flexible in-school production of tests from an established item bank.
6. Developing scoring masks and procedures for in-school test correction.
7. Obtaining consumer-type feedback from teachers on the utility of various types of test data, in various formats.

References and Selected Bibliography

Airasian, P. W. & Madaus, G. F. (1983). Linking testing and instruction: Policy issues. *Journal of Educational Measurement, 20,* 103-118.

Alexander, L., & James, H.T. (1987). The nation's report card. Cambridge, MA: National Academy of Education.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. Handbook: The cognitive domain.* New York: David McKay Co.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook of formative and summative evaluation of student learning.* NY: McGraw-Hill.

Bormuth, J. R. (1970). On the theory of achievement test items. Chicago: University of Chicago Press.

Christenson, S.L., Ysseldyke, J.E., & Thurlow, M.L. (1989). Critical instructional factors for students with mild handicaps: An integrative review. Remedial and Special Education, 10(5), 21-31.

Cole, N. S., & Moss, P. A. (1989). Bias in test use, In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 201-220), NY: Macmillan, Inc.

Denmark, T., & Kepner, H. S., Jr. (1980). Basic skills in mathematics: A survey. *Journal for Research in Mathematics Education,* (March) 104-123.

Durkin, D. (1978-79). What classroom observation reveals about reading instruction. *Reading Research Quarterly, 14,* 481-533.

Durkin, D. (1981). Reading comprehension instruction in five basal reader series. *Reading Research Quarterly, 16,* 515-544.

Ebel, R.L. (1979). *Essentials of educational measurement* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Feldt, L. S., Steffen, M., & Gupta, N. C. (1985). A comparison of five methods for estimating the standard error of measurement at specific score levels. *Applied Psychological Measurement, 9*(4), 351-361.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39,* 193-202.

Freeman, D.J., Kuhs, T. M., Porter, A.C., Floden, R.F., Schmidt, W.H., & Schwille, J.R. (1983). Do textbooks and tests define a national curriculum in elementary school mathematics? *The Elementary School Journal, 83*(5), 501-513.

Glennon, V. J., & Wilson, J. W. (1972). Diagnostic-prescriptive teaching. In National Council of Teachers of mathematics, *The slow learner in mathematics: The 35th Yearbook.* Reston, VA: NCTM.

Good, T., & Brophy, J. (1986). School Effects. In Merlin C. Wittrock (Ed.), *Handbook of research on Teaching* (Third Edition). (pp. 570-602). New York: Macmillan Publishing Company.

Good, T. L., Grouws, D. A., & Ebmeier, H. (1983). *Active mathematics teaching* (Research on Teaching Monograph Series). New York: Longman.

Haladyna, T.M., & Downing, S.M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied measurement in education, 2,* 37-50.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 147-200, NY: Macmillan.

Hambleton, R. K., & Eignor, D. R. (1978). *competency text development, validation, and standard-setting.* Paper presented a the Minimum Competency Testing Conference of the American Education Research Association. Washington, DC.

Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston: Kluwer-Nijhoff.

Hays, W. L. (1981). *Statistics,* NY: Holt, Rinehart and Winston.

Hiscox, M. D. (1985). Computer-based testing systems: Much ado about nothing? *Educational Measurement: Issues and Practice, 4*(1) 27-28.

Hively, W., Patterson, H. L., & Page, S. A. (1968). A "universe defined" system of arithmetic achievement tests. *Journal of Educational Measurement, 5,* 275-290.

Hsu, T. & Yu, L. (1989). Using computers to analyze item response data. *Educational measurement: Issues and Practice, 8*(3), 21-28.

Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 147-200, NY: Macmillan.

Kameenui, E. J. & Griffin, C. C. (1989). The national crisis in verbal problem solving in mathematics: A proposal for examining the role of basal mathematics programs. *The Elementary School Journal, 89*(5), 575-593.

Kirkland, M. C. (1971). The effects of tests on students and schools. *Review of Educational Research, 41*, 303-350.

Komoski, P. K. (1985). Instructional materials will not improve until we change the system. . *Educational Leadership, 12*, 31-37.

Kuhs, T., Schmidt, W., Porter, A., Floden, R., Freeman, D., & Schwille, J. (1979). *A taxonomy for classifying elementary school mathematics content* (Research Series No. 4). East Lansing: Institute for Research on Teaching, Michigan State University.

Merrill, M. D., Reigeluth, C. M., & Faust, G. W. (1979). The instructional quality profile: A curriculum evaluation and design tool. In H. F. O'Neil, Jr. (Ed.), *Procedures for instructional systems development*. NY: Academic Press.

Millman, J. and Arter, J. A. (1980). Issues in item banking. *Journal of Educational Measurement, 21*, 315-330.

Millman, J. (1984). Individualized test construction and administration by computer. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 78-96). Baltimore, MD: Johns Hopkins University Press.

Mislevy, R. J. & Bock, R. D. (1984). *BILOG: Maximum likelihood item analysis and test scoring with logistic models*. Mooresville, IN: Scientific Software.

Mislevy, R. J. & Stocking, M. L. (1987). *A consumer's guide to LOGIST and BILOG* (Research Rep. No. 87-43). Princeton, NJ: Educational Testing Service.

Murnane, R. J., & Raizen, S. A. (1988). *Improving indicators of the quality of science and mathematics education in grades K-12*. Washington, DC: National Academy Press.

Nitko, A. J. (1980). Distinguishing the many varieties of criterion-referenced tests. *Review of Educational Research, 50*(3), 461-485.

Nitko, A. J. (1989). Designing tests that are integrated with instruction, In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 447-474), NY: Macmillan.

Osburn, H. G. (1968). Item sampling for achievement testing. *Educational and Psychological Measurement, 28*, 95-104.

Popham, J. (1978). As always, provocative. *Journal of Educational Measurement, 15*, 297-300.

Popham, W. J. (1984). Specifying the domain of content or behaviors. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 29-48). Baltimore: Johns Hopkins University Press.

Porter, A. C. (1989). A curriculum out of balance: The case of elementary school mathematics. *Educational Researcher, 18*(5), 9-15.

Porter, A. C. Schmidt, W. H. Floden, R. E. & Freeman, D. J. (1978). Practical significance in program evaluation. *American Educational Research Journal, 15*, 529-539.

Qualls, A. L. & Ansley, T. N. (1985). *A comparison of items and ability parameter estimates derived from LOGIST and BILOG*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

Reckase, M. D. (1989). Adaptive testing: The evolution of a good idea. *Educational Measurement: Issues and practices, 8*(3), 11-16.

Roid, G., & Haladyna, T. (1982). *A technology for test item writing*. NY: Academic Press.

Roid, G. H. (1989). Item writing and item banking by microcomputer: An update. *Educational Measurement: Issues and practices, 8*(3), 17-20.

Romberg, T. A., & Carpenter, T. P. (1986). Research on teaching and learning mathematics: Two disciplines of scientific inquiry (pp. 850-873). In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.), NY: Macmillan.

Salvia, J., & Ysseldyke, J. E. (1978). Assessment in special and remedial education. Boston: Houghton Mifflin.

Schmidt, W. H. (1983). Content bias in achievement tests. *Journal of Educational Measurement, 20*, 165-178.

Smylie, M. (1988). The enhancement function of staff development: Organizational and psychological antecedents to individual teacher change. *American Educational Research Journal, 25*(1), 1-30.

Snow, R. E. & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 447-474), NY: Macmillan.

Stake, R. & Easley, J. (Eds.) (1978). *Case studies in science education*. Urbana: University of Illinois.

Stephens, W. (1982). *Mathematics knowledge and school work: A case study of the teaching of developing mathematical processes* (Project Report). Madison: Wisconsin Center for Education Research.

Stone, C. A. (1989). Testing software review: MicroCAT Version 3.0. *Educational Measurement: Issues and practice, 8*(3), 33-38.

Suen, H. K. & Ary, D. (1989). *Analyzing quantitative behavioral observation data*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tindal, G. (1989). Evaluating the effectiveness of educational programs at the systems level using curriculum-based measurement. In M. Shinn (Ed.), *Curriculum-based measurement*. NY: The Guilford Press-School Practitioner Series.

Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*, 19-30.

Walker, D. & Schaffarzick, J. (1974). Comparing curricula. *Review of Educational Research, 44*, 83-111.

Williams, R. G., & Haladyna, T. M. (1982). Logical operations for generating intended questions (LOGIQ): A typology for higher level test items (pp. 161-185). In G. Roid & T. Haladyna (Eds.), *A technology for test item writing*, NY: Academic Press.

Wingz® (1989). Informix Software, Lenexa, Kansas.

Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika, 50*, 275-291.

Appendices

A. State of Oregon Comprehensive Curriculum Goals

B. Cross-Referenced Codes for State of Oregon Goals and Curriculum Taxonomy (Content Only)

C-1. Guidelines for Coding Math Curriculum

C-2. Form for Entering Math Curriculum Codes

D. Sample Excel® Spreadsheet Containing Curriculum Codes

E. Interrater Agreement Index for Coding Curriculum Content

F. Draft Sketches of Test Items

G. Winter Tests: Grades 3-5

H. Intitial Total Score Feedback for Teachers

I. Decile Line Graph and Frequency Distribution Feedback for Teachers: Grades 3-5

J. Box Plots for Building-Level Feedback: Grades 3-5

K. Fall/Winter Test & Subtest Results for Teachers

L. Fall//Winter Item-Specific Results for Teachers

M. Spring Tests Grades 3-5

N. Percentile Ranks for Raw Scores: Spring Tests

Appendix A
State of Oregon Comprehensive Curriculum Goals

Mathematics
Comprehensive Curriculum Goals
A Model for Local Curriculum Development
October 1987

Sample page 48:
2.0
Appropriate Computational Skills: Students select and use the most appropriate form of computation—manipulative, mental paper/pencil, estimation, or calculator usage to solve problems and check all computations for reasonability.

2.1
Use mental, paper and pencil, estimation, and calculator computations to solve appropriate problems (ELS 1.4 and 1.7).

Grade 3	Grade 4	Grade 5
a. Use mental, manual, or calculator processes to perform grade-level arithmetic operations.	a. Use mental, manual, or calculator processes to perform grade-level arithmetic operations.	a. Use mental, manual, or calculator processes to perform grade-level arithmetic operations.
b. Select the most appropriate method of computation (manipulative, mental, paper/pencil, estimation, calculator) to use in a given situation.	b. Select the most appropriate method of computation (manipulative, mental, paper/pencil, estimation, calculator) to use in a given situation.	b. Select the most appropriate method of computation (manipulative, mental, paper/pencil, estimation, calculator) to use in a given situation.
c. Use estimating skills, such as rounding, to make approximate whole number computations.	c. Use rounding and other techniques useful in mental computation to estimate and make approximate whole number, fraction, and decimal computations	c. Use rounding and other techniques useful in mental computation to estimate and make approximate whole number, fraction, and decimal computations
d. Apply acquired strategies including modelling patterns (such as "counting on," "doubles," "neighbors," etc.) and properties (commutativity and associativity), to aid in quick recall of addition, subtraction, and multiplication facts.	d. Apply acquired strategies to aid in quick recall of all basic facts.	d. Apply acquired strategies to aid in quick recall of all basic facts.
e. Solve mentally, appropriate addition and subtraction problems involving place value understanding, e.g,, add or subtract 10 or 100 to (from) any 3-digit number;	e. Use mental arithmetic skills to solve appropriate problems (multiples of 10 and 100, addition of fractions with like denominators, etc.).	e. Solve mentally, appropriate whole number, fraction, and decimal problems, e.g., 10x64; 60x20; 14,000-7,000; 5,000+261; 3,000x7; $^1/_4 + 3/_4$; $5/_8 - 4/_8$; 3-0.5.

Appendix B
Cross-Referenced Codes for State of Oregon Goals and Curriculum Taxonomy (Content Only)

	O.C.	State	O.C.	State	O.C.	State	O.C.	State
1	A1	1.1B	F3	2.1F	J4	3.2E	N5	6.1F
2	A1	1.1C	F4	2.1D	J5	3.1A	N5	7.1D
3	A2	1.1A	F4	2.1F	J5	3.1B	N6	3.1A
4	A3	1.1C	F5	2.1D	J5	3.1F	N6	3.1B
5	A3	1.2A	F5	2.1F	J5	3.2A	N6	3.1F
6	A3	1.2B	F6	2.1C	J5	3.2B	N6	3.2A
7	A4	2.1C	F7	7.2A	J5	3.2E	N6	3.2B
8	A5	1.1A	G1	2.1F	J6	3.1A	N6	3.2E
9	A6	7.2D	G2	2.1F	J6	3.1B	N6	5.1C
10	B1	2.1C	G3	2.1F	J6	3.1F	N7	3.2C
11	B2	2.1D	G4	2.1F	J6	3.2A	O1	2.1F
12	B2	2.1F	G5	2.1C	J6	3.2B	O2	
13	B2	2.2A	G6	7.2A	J6	3.2E	O3	7.3B
14	B3	2.1F	H1	1.1A	J7	3.1D	P1	7.2A
15	B3	2.2A	H1	1.2A	J7	3.2A	P2	5.2D
16	B4	2.1F	H2	1.1A	J7	3.2B	P3	5.2D
17	B4	2.2A	H3	1.1A	J7	3.2E	P4	5.1B
18	B5	2.1F	H4	2.1F	K1	2.1A	P5	5.1A
19	B5	2.2A	H4	2.2D	K1	5.2A	P6	
20	B6	2.2A	H5	2.1F	K1	5.2B	Q1	7.2C
21	B7	7.2A	H6	2.1F	K1	5.2G	Q1	7.4A
22	C1	2.1C	H7	1.1A	K2	4.3B	Q1	7.4B
23	C2	2.1D	H7	2.2D	K3	2.1G	Q2	7.2C
24	C2	2.1F	H7	2.2F	K3	5.1B	Q2	7.4A
25	C3	2.2A	H8	2.1C	K4	5.2G	Q2	7.4B
26	C3	2.1F	H9	7.2A	K5	3.2A	R1	6.1A
27	C3	2.2A	H10	2.1F	K5	3.2B	R1	6.1B
28	C4	2.1F	I1	1.1A	K5	5.1C	R2	6.1J
29	C4	2.2A	I1	1.2A	K6	5.2A	R2	6.1L
30	C5	2.2A	I1	1.2B	K7	5.2A	S1	2.1H
31	C6	7.2A	I1	2.2C	K8	7.2A	S2	2.1H
32	D1	2.1C	I2	7.2F	L1	4.1A		
33	D2	2.1D	I3	2.1F	L1	4.1B		
34	D2	2.1F	I3	2.2C	L1	4.1F		
35	D2	2.2A	I4	2.1F	L1	7.1A		
36	D3	2.1F	I4	2.2C	L2	4.3A		
37	D3	2.2A	I5	2.1F	L2	4.3B		
38	D4	2.1F	I5	2.2C	L2	4.3C		
39	D4	2.2A	I6	2.1F	L2	5.2C		
40	D5	2.1F	I6	2.2C	L3	4.3A		
41	D5	2.2A	I7	2.1C	L3	4.3B		
42	D6	2.1F	I8	7.2A	L3	4.3C		
43	D6	2.2A	J1	3.1A	L3	5.2C		
44	D7	2.1E	J1	3.1B	L4	4.1C		
45	D8	7.2A	J1	3.1F	L4	5.2B		
46	E1	2.1C	J1	3.2A	L5	4.1C		
47	E2	2.1D	J1	3.2B	L6	3.2B		
48	E2	2.1F	J1	3.2E	L6	3.2E		
49	E2	2.2A	J2	3.1A	L7	4.3B		
50	E3	2.1F	J2	3.1B	M1	7.2F		
51	E3	2.2A	J2	3.1F	M2	7.2G		
52	E4	2.1F	J2	3.2A	M3			
53	E4	2.2A	J2	3.2B	M4			
54	E5	2.1F	J2	3.2E	M5			
55	E5	2.2A	J3	3.1A	N1			
56	E6	2.1F	J3	3.1B	N2	7.2G		
57	E6	2.2A	J3	3.1F	N3	4.1E		
57	E7	7.2A	J3	3.2A	N3	6.1E		
59	E8	2.1E	J3	3.2B	N3	6.1F		
60	E8	2.2A	J3	3.2E	N3	7.1D		
61	F1	2.1D	J4	3.1A	N3	7.2D		
62	F1	2.1F	J4	3.1B	N3	7.3B		
63	F2	2.1D	J4	3.1F	N4	2.2B		
64	F2	2.1F	J4	3.2A	N4	5.1A		
65	F3	2.1D	J4	3.2B	N5	5.3A		

Appendix C-1
Guidelines for Coding Math Curriculum

Guidelines for Reliably Coding Open Court Basal Math Activities (10/1/89)

Basal Source:

Dem=demonstration/seminar/whole group activity.
St.pg=student pages.
Ment=mental math/response exercis (includes student pages).
Thnk=thinking story.
Wkshp=workshop/games.

Assess:

U.Rev.=unit review.
U.Test=unit test.

Task Intent: *(What is its purpose?) (For "introduce and/or review," check both.)*

Intro=introduction. (Includes "Teach," "Show," "Opportunity to Discover") (Must include teacher talk.)
Pract=practice. (Includes students working problems as a review.)
Review=review. (Includes "Evaluate.") (Must include teacher talk.)

Cognitive Level:

Mem=memorize, rote learning.
Skill=skill (procedural) development & practice. Apply an algorithm.
Con.und=conceptual understanding of a specific skill, procedure, or concept/principle. (The focus here is on a specific skill or procedure, but not simply on practicing the skill. The focus is instead on a broader understanding of "what is happening" or "what it means" when the skill is performed.
Gen.und=general understanding. (includes any general logic-related or math-related discussions/presentations which do not focus on a specific math procedure, skill, or concept/principle.)

Task Format:

Ment=mental. (Includes "estimating.")
Pa/Pen=paper/pencil.
Manip=manipulative. (Usually concrete objects used to help students solve a problem. The purpose is to use the concrete object to understand the concept and/or solve the problem.) (The Response Wheel is not considered a manipulative, but instead a way of showing the answer to a mental operation.) (Throwing dice is not a manipulative.)

Response Mode:

Oral=oral.
Write=written.
Show=show/demonstrate.

Response Type:

Select.=selection type.
Produc=production type.

Cues or Hints: *(Assist in doing the problem. They must clearly demonstrate or be an integral part [aide] in solving the problem.)*

Examp=cues/examples from problem.
Pict=pictures.
Object=objects/manipulables.

Other Notes: Thinking stories are always practice. Their content area is reasoning and they are always done in a written/discussion format, with an oral/written response for 4th and 5th grade levels.

Appendix C-2
Form for Entering Math Curriculum Codes

CODE #	GRADE LEVEL	LESSON #	CONTENT CONCEPT/SKILL CATEGORY	BASAL SOURCE	ASSESS		TASK INTENT				COGNITIVE LEVEL(S)					TASK FORMAT				RESPONSE MODE OR # PROBS			RESPONSE TYPE		CUES OR HINTS			COMMENTS	STATE OBECT
				DEM. ST.PG MENT. THNK WKSH	U. REV.	U. TEST	INTRO	PRACT	REVIEW	EXA/EXP	MEM	SKILL	CON. UND.	GEN. UND.	PROB (M)	MENT	PA/PEN	MANIP	DISC	ORAL	WRITE	SHOW	PRODUCT.	SELECT.	EXAMP.	PICT.	OBJECT		

BASAL SOURCE: DEM=Demonstration/Seminar/Whole Group Activity; ST.PG.=Student Pages; MENT=Mental Math Response Exercise; THNK=Thinking Story; WKSHP=Workshop/Games
ASSESS: U. REV.=Unit review; U. TEST=Unit Test
TASK INTENT: INTRO=Introduction; PRACT=Practice; REVIEW=Review; EXA/EXP=Explanation/Exploration
COGNITIVE LEVEL: MEM=Memorize, Rote Learning; SKILL=Skill (Procedural) Development & practice; CON. UND=Specific Conceptual Understanding; GEN. UND=General Understanding
TASK FORMAT: MENT=Mental; PA/PEN=Paper/Pencil; MANIP=Manipulative; DISC.=Discussion
RESPONSE MODE: ORAL=Oral; WRITE=Written; SHOW=Show/Demonstrate
CUES OR HINTS: EXAMP=Cues/]Examples from Problem; PICT=Pictures; OBJECT=Objects/ManipulablesP=Problem-Solving/Applicaton to Novel Situations [(M)=Using multiple processes]

Appendix D
Sample Excel Spreadsheet Containing Curriculum Codes

Code #1	Code #2	Grade	Lesson	Content	Basal	Tch. Intent	Cog. Level	T. Format	Resp. Mode	Cues/Hints
k1	k5	5	1	Estim. & Meas.	s	p	s	pa, d	o, w	
a1		5	1	concepts of order	w	p	m	pa	w	
k7		5	1	estimating measure	m	p	s	me	o	
a1		5	2	count & num. seq.	m	p	m	me	o	
k7		5	2	estimating meas.	m	p	s	me	o	
n3		5	2	organize & int. data	s	p	s	pa, d		
a1		5	2	number sequence	s	p	m	pa	w	
f5		5	2	basic facts *-+/	s	p	m	pa	w	
f5		5	3	basic facts	m	p	m	me	o	
a2	a3	5	3	place value-stand.	s	p	g	pa, d	o, w	e
a3		5	3	place val. & regroup	w	p	s	pa	w	
f5		5	3	basic facts-horiz.	s	p	m	pa	w	
f5		5	3	basic facts-vert.	s	p	m	pa	w	
f5		5	4	chair caal +/-* mix	m	p	m, s	me	o	
j4		5	4	word prob./basic	s	p	p	pa	w	
q1		5	4	reasoning	t	p	p	pa, d		
f5		5	5	basic facts	m	p	m	me	o	
o2		5	5	parenthesis	s	l, r	g	d	o	
o2		5	5	solv parent & prob.	s	p	s	pa	w	e
o2		5	5	put in parenthesis	s	p	s	pa	w	e
f5		5	5	basic facts	w	p	m	me	o	
a3		5	6	place value	m	p	m	me	o	
b3	b6	5	6	multidig. add/group	m	p	m	me	o	
b3		5	6	multidig. add/col.	s	r	g	d	o	
b3		5	6	multidig. add/col.	s	p	s	pa	w	e
e1		5	6	multidig. add/approx.	s	p	s	pa	w	
b3		5	6	add 3 digit #s	w	p	s	pa	w	
g1		5	7	add & subtract	m	p	m	me	o	
c3	c6	5	7	multidig -/regroup	s	r	g	d	o	
c3		5	7	multidig sub.	s	p	s	pa	w	e
g1		5	7	add & sub. multi #s	w	p	s	pa	w	
n5		5	8	map read -++-	s	p	s	pa	w	
		5	8	add & sub	m	p	s	me	o	
g1		5	8	multidig add or sub	s	p	s	pa	w	
j4		5	8	word problems	s	p	s	pa	w	
d2		5	8	mult w/fact 0-10	w	p	s	pa, me	w	
d7		5	9	mult by powers 10	s	p, r	s, g	pa, d	w	
d7		5	9	mult by mults 10	s	p, r	s, g	pa, d	w	
d7		5	9	mult 10 & 100	m	p	s	me	o	
d3		5	9	mult 3 dig by 1	s	r	g	d	o	

Appendix E
Interrater Agreement Index for Coding Curriculum Content

Method for Computing Agreement Among Multiple Raters When
Number of Judgments per Rater May Vary
(based on W. M. Rand [1971], *Journal of American Statistical association, 66* 846-850)

The Ratio: $\frac{\text{Obtained Agree pairs (including nulls [Ø])}}{\text{Possible Agree pairs (including nulls [Ø])}}$

Possible points three raters:

0 = no agreements among three choices
1 = two of three choices agree (1 pair).
3 = all three choices agree (3 pairs).

Example 1:

Rater 1	Rater 2	Rater 3
a b	a b	a c

→

Example 1

Agreements: a-a-a = 3points
b-b = 1 point
Total = 4 points

Possible points (perfect agreement):

3 + 3 + = 6

$\frac{\text{Agree Pts.}}{\text{Possible Pts.}} = \frac{4}{6} = .67$

To help computation in the examples below, the null set symbol, Ø, has been placed as a space holder for "no rating." The number of space holders is determined by the greatest number of ratings by one rater.

Example 2:

Rater 1	Rater 2	Rater 3
a b Ø Ø	a b Ø Ø	a c d e

→

Agreements:
a-a-a = 3 points; b-b = 1 point
Ø Ø = 1 point; Ø Ø = 1 point
Total = 6 points

Possible points (perfect agreement):
3 + 3 + 3 + 3 = 12

$\frac{\text{Agree Pts.}}{\text{Possible Pts.}} = \frac{6}{12} = .50$

Example 3: a b | a Ø | a c — $\frac{3}{6} = .50$

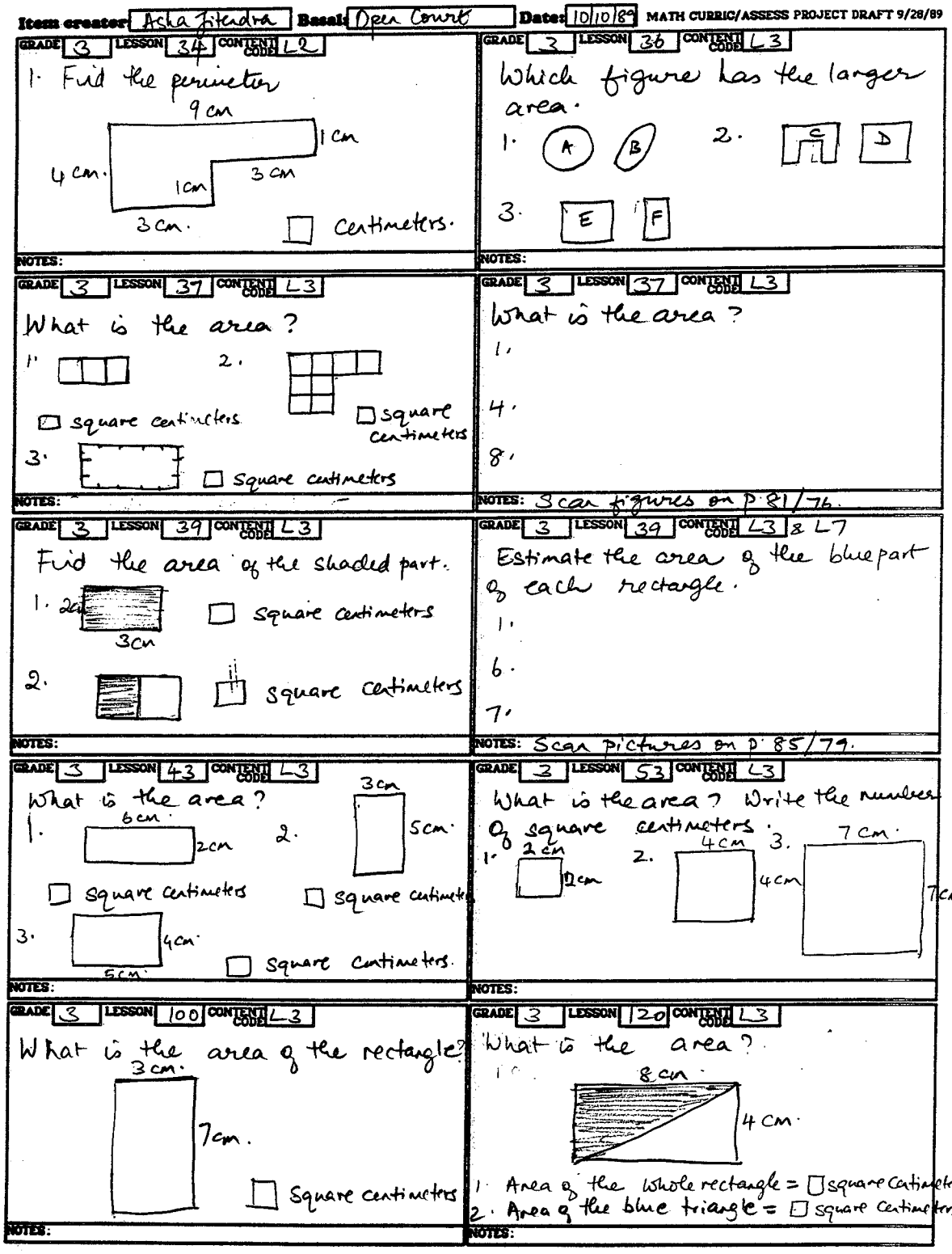
Example 4: a Ø | a Ø | a c — $\frac{4}{6} = .67$

Example 5: a b Ø Ø Ø | b c d e f | a Ø Ø Ø Ø — $\frac{5}{15} = .33$

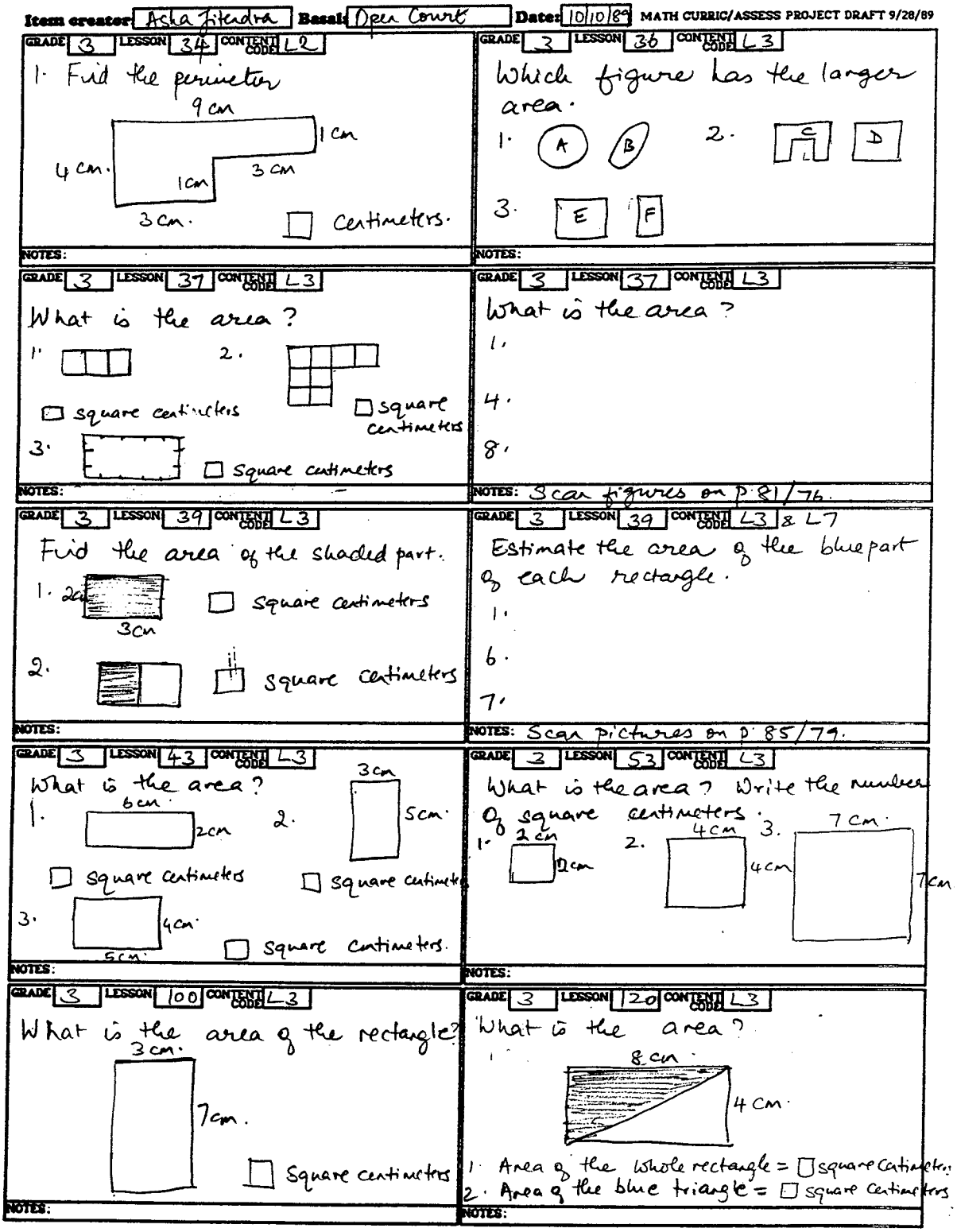
Example 6: a b c Ø Ø | b h e f g | a b d Ø Ø — $\frac{4}{15} = .27$

Example 7: a b c Ø Ø | b h e f g | a b c Ø Ø — $\frac{6}{15} = .47$

Appendix F
Draft Sketches of Test Items



Appendix F
Draft Sketches of Test Items



Item creator: Asha Jitendra Basal: Open Court Date: 10/10/89 MATH CURRIC/ASSESS PROJECT DRAFT 9/28/89

GRADE 3 LESSON 34 CONTENT CODE L2

1. Find the perimeter

9 cm, 1 cm, 4 cm, 1 cm, 3 cm, 3 cm.

☐ centimeters.

NOTES:

GRADE 3 LESSON 36 CONTENT CODE L3

Which figure has the larger area.

1. (A) (B) 2. C D

3. E F

NOTES:

GRADE 3 LESSON 37 CONTENT CODE L3

What is the area?

1. 2.

☐ square centimeters ☐ square centimeters

3. ☐ Square centimeters

NOTES:

GRADE 3 LESSON 37 CONTENT CODE L3

What is the area?

1.

4.

8.

NOTES: Scan figures on p. 81/76.

GRADE 3 LESSON 39 CONTENT CODE L3

Find the area of the shaded part.

1. 2cm 3cm ☐ Square centimeters

2. ☐ Square centimeters

NOTES:

GRADE 3 LESSON 39 CONTENT CODE L3 & L7

Estimate the area of the blue part of each rectangle.

1.

6.

7.

NOTES: Scan pictures on p. 85/79.

GRADE 3 LESSON 43 CONTENT CODE L3

What is the area?

1. 6cm 2cm ☐ square centimeters

2. 3cm 5cm ☐ square centimeters

3. 4cm 5cm ☐ Square centimeters.

NOTES:

GRADE 3 LESSON 53 CONTENT CODE L3

What is the area? Write the number of square centimeters.

1. 2cm 2cm

2. 4cm 4cm

3. 7cm 7cm

NOTES:

GRADE 3 LESSON 100 CONTENT CODE L3

What is the area of the rectangle?

3cm. 7cm. ☐ Square centimeters

NOTES:

GRADE 3 LESSON 120 CONTENT CODE L3

What is the area?

1. 8cm 4 cm

1. Area of the whole rectangle = ☐ square centimeters

2. Area of the blue triangle = ☐ square centimeters

NOTES:

Appendix G
Winter Tests: Grades 3-5

DIRECTIONS FOR ADMINISTRATION
OPEN COURT MATH ASSESSMENT PROJECT: GRADES 3, 4, & 5

GENERAL INFORMATION FOR TEACHERS

Each of your students will complete a six-page test, printed on three separate sheets of paper. The tests will be in three separate packets. These tests should be administered on three different days to avoid test fatigue. On the first day of testing students will complete pages 1 & 2 from the first packet, the second day, pages 3 & 4 from the second packet, and on the third day, pages 4 & 5 from the third packet. For example, this test could be completed on Tuesday, Wednesday, and Thursday. Each day, the two pages of the test should be completed by students in one sitting.

Students may work on their test sheets and will write their answers directly on the test. However, they should have scratch paper available for working longer problems, if necessary. Please have some extra paper available for this purpose.

THIS IS NOT A TIMED TEST. Enough time should be allowed for most students to finish. The original plan called for each sheet to require about 20 minutes for students to complete. That may be an underestimation. Within *reasonable* limits, please allow sufficient time for students who are able to complete all items on the test.

Because these tests cover curriculum content from the entire year, some of the problems will test material that your students have not yet been taught. To reduce frustration, students will be asked to circle those items which they have no idea about, and go on. Please monitor the testing sessions closely, encouraging students to try all items which they have some knowledge about. Guessing should be discouraged.

All completed tests should be returned immediately to Rich Davidson, Special Education Director. Please attempt to administer make-up tests for students who are absent during the week of testing. If absent students cannot have their tests made-up within one extra week, please return blank or partially completed tests to the district office.

We have pledged to get the results form this test back to you promptly. This is the first of three tests that will be administered this year. If you have any questions about this test or the project, please contact Rich Davidson.

Thank you and your students very much for your participation.

Materials Needed for Testing

- One test sheet per student for each of the three days of testing
  (Day 1: Pages 1 & 2; Day 2: Pages 3 & 4; Day 3: Pages 4 & 6)
- A sharp pencil and eraser for each student
- Scratch paper for working problems

Appendix G
(continued)
Grade 3 Winter Test

Student _____ Grade _____ *Page 1 [10/22/89] Grade 3*

School _____ Teacher _____ Date _____

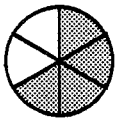
1 Multiply:

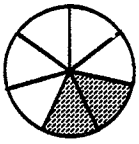
8	0	9
x 2	x 5	x 7

2 Multiply:

8	7	6
x 9	x 3	x 9

3 What part of each circle is shaded?

a.  ☐ b.  ☐

c.  ☐ d.  ☐

4 Add:

26	14	18
19	52	28
+ 41	+ 23	52
		+ 74
☐	☐	☐

5
12 + 13 - 7 = ☐
14 - 5 - 7 = ☐
9 + 8 - 5 = ☐

6
3 + 3 + 6 - 4 - 4 = ☐
5 - 5 + 4 + 4 = ☐
12 - 3 - 6 - 1 = ☐

7 Multiply:

256	234	806
x02	x70	x 64

8 How many <u>new</u> marks must be put on the line segment to show the following parts: 

a. thirds? ☐
b. fourths? ☐
c. fifths? ☐

9 How much is:

a. $\frac{1}{3}$ of 18? ☐
b. $\frac{1}{4}$ of 12? ☐
c. $\frac{1}{5}$ of 10? ☐

10 Multiply:

a. 3 x 10 = ☐
b. 8 x 0 = ☐
c. 10 x 7 = ☐

Appendix G
(continued)
Grade 3 Winter Test

Student _____ Grade _____ *Page 2 [10/22/89] Grade 3*

School _____ Teacher _____ Date _____

11 Rewrite to show no hundreds:

a. 507 = _____ tens and _____ ones.

b. 826 = _____ tens and _____ ones.

c. 351 = _____ tens and _____ ones.

12 Multiply:

a. 4 x 100 = ☐

b. 53 x 1000 = ☐

c. 100 x 78 = ☐

13 Multiply:

9	5	3
x 9	x 5	x 3

14 Solve for n:

a. n ÷ 8 = 8 n = ☐

b. 64 ÷ n = 8 n = ☐

c. 1 ÷ 1 = n n = ☐

15 Solve for n:

42 = n x 6 n = ☐

n = 0 x 2 n = ☐

63 = 7 x n n = ☐

16 Subtract:

a. 4.53 - 2.82 = ☐

b. 7.3 - 5.06 = ☐

c. 14.8 - 5.9 = ☐

17 Multiply:

643	789	367
x 7	x 2	x 5

18

a. 3 l = ☐ ml

b. 2 km = ☐ cm

c. ☐ km and 82 cm = 1082 cm

19 Write the name of the unit that makes the most sense.
Write **K** for Kilometers, **m** for meters, **c** for centimeters, **kg** for kilograms, or **g** for grams.

a. The man was about 2 ☐ tall.

b. The pills weighed about 6 ☐ .

c. The worm was about 8 ☐ long.

d. The road was about 30 ☐ long.

20

a. Ron is 18 years old. His sister Rita is 2 years younger than him. How old is Rita? ☐

b. Jack bought an ice cream cone for 50¢ and some pop for 20¢. How much money did he spend? ☐

Appendix G
(continued)
Grade 3 Winter Test

Student _____ Grade _____ *Page 3 [10/21/89] Grade 3*

School _____ Teacher _____ Date _____

21 a. Jan bought a plant that was 25 cm tall. Now it is 32 cm tall. How much did it grow since Jan bought it? ☐

b. Rob lives 20 km west of Salem. His sister lives 15 km west of Salem. About how far does Rob live from his sister? ☐

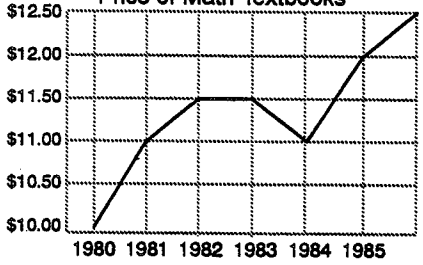
22 Count up or down. Fill in missing numbers.

a. 41 ☐ 43 44 ☐ ☐ 47

b. ☐ ☐ 92 ☐ ☐

c. 437 438 ☐ ☐ 441 ☐

23 Price of Math Textbooks

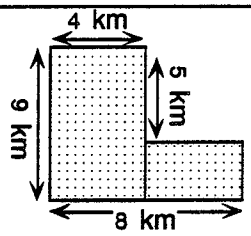


a. What was the price of a math textbook in1982? ☐

b. How much did the price of a textbook increase from 1980 to 1984? ☐

c. Between what two years did the price go down? ☐ and ☐

24



a. What are the areas of the two rectangles? ☐ ☐

b. What is the total area of the figure? ☐

25 Change to all dollars or all cents:

a. 3745 ¢ = $ ☐

b. $ 21.45 = ☐ ¢

c. 703 ¢ = $ ☐

26

a. $3.20 = $ ☐ and ☐ ¢

b. $ ☐ and 30¢ = $4.30

c. $15.30 = $ ☐ and ☐ ¢

27 Divide:

a. 64 ÷ 8 = ☐

b. 27 ÷ 9 = ☐

c. 63 ÷ 7 = ☐

28

a. Ms. Kelly made curtains for 1 window. She used 13 meters of cloth. How many meters will she need to make curtains for 4 more windows? ☐

b. Ms. Kelly is buying 23 meters of cloth. The piece at the store is 15 meters long. What length of cloth will be left? ☐

c. The cloth costs $42.00. Ms. Kelly gave the storekeeper three $20.00 bills. How much change should she get? ☐

Appendix G
(continued)
Grade 3 Winter Test

Student _____ Grade _____ *Page 4 [10/21/89] Grade 3*

School _____ Teacher _____ Date _____

29 Raymond has 2 one-dollar bills and 12 dimes.

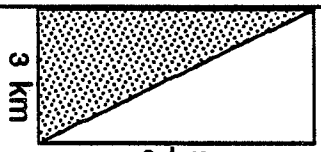
a. Does Raymond have enough money to buy a book costing $3.06 ? ☐

b. Does he have enough to buy two baseballs, if they cost $1.30 each ? ☐

30 Multiply:

7	7	4
x 9	x 7	x 6

31



a. What is the area of the rectangle? ☐

b. What is the area of the shaded triangle? ☐

32

Items	Prices	
	Store A	Store B
Milk (1 gallon)	$1.85	$1.95
Ivory Soap (4 bars)	$1.75	$1.50
Orange Juice (12 oz)	99¢	$1.05

a. Which store charged more for the three items in total? ☐

b. Did every item cost more in the more expensive store? ☐

c. If you buy just milk and soap, which store is more expensive? ☐

33 Change to all dollars or all cents:

a. 3745 ¢ = $ ☐

b. $ 21.45 = ☐ ¢

c. 703 ¢ = $ ☐

34 Add:

6	49	38
5	11	16
+ 9	+ 3	+ 42
☐	☐	☐

Appendix G
(continued)
Grade 3 Winter Test

Student _____ Grade _____ *Page 5 [10/21/89] Grade 3*

School _____ Teacher _____ Date _____

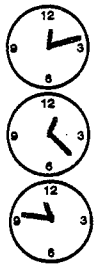
35 Subtract:
a. 76 - 12 = ☐
b. 33 - 13 = ☐
c. 19 - 10 = ☐

36 Raymond has 2 one-dollar bills and 12 dimes.
a. Does Raymond have enough money to buy a book costing $3.06 ? ☐
b. Does he have enough to buy two baseballs, if they cost $1.30 each ? ☐

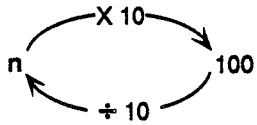
37 Multiply:
a. 7 x 10 = ☐
b. 1000 x 80 = ☐
c. 30 x 100 = ☐

38 Draw lines to match the clocks with the correct time.

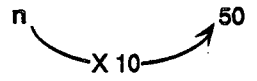
12 : 21 •
12 : 12 •
"twelve minutes 'til 12." •
11 : 48 •
"twelve after 12:00" •



39 Find the value of n.



n = ☐



n = ☐

40 Multiply:
a. 10 x 15 = ☐
b. 32 x 10 = ☐
c. 76 x 10 = ☐

41 Fred caught 3 fish. Each fish weighed between 5 and 6 pounds.
a. Can Fred have 16 pounds of fish altogether? ☐
b. Can Fred have 20 pounds altogether? ☐
c. Can he have 12 pounds of fish altogether? ☐

42 Multiply:

0	6	3
x 2	x 3	x 5

43 Subtract:

13.7	91.20	7.5
-5.26	-2.33	-4.8

44 a. 1 foot = ☐ inches.
b. 1 Yard = ☐ feet.
c. 1 Pound = ☐ ounces.

Appendix G
(continued)
Grade 3 Winter Test

Student _____ Grade _____ *Page 6 [10/21/89] Grade 3*

School _____ Teacher _____ Date _____

45 Add:

$$\begin{array}{rrr} 4.06 & 6.98 & 10.09 \\ +0.74 & +2.59 & +3.33 \end{array}$$

46 Add:

$$\begin{array}{rrr} 340 & 399 & 579 \\ +209 & +500 & +220 \end{array}$$

47 Multiply:

$$\begin{array}{rrr} 490 & 932 & 666 \\ \times\ 5 & \times\ 4 & \times\ 6 \end{array}$$

48 Solve for n:

6 = 7 - n n = ☐

5 = n - 6 n = ☐

n = 12 - 4 n = ☐

49 Solve for n:

42 = n x 6 n = ☐

n = 0 x 2 n = ☐

63 = 7 x n n = ☐

50 Add:

$$\begin{array}{rrr} 4 & 6 & 0 \\ +8 & +3 & +9 \end{array}$$

51

12 + 13 - 7 = ☐

14 - 5 - 7 = ☐

9 + 8 - 5 = ☐

52

3 + 3 + 6 - 4 - 4 = ☐

5 - 5 + 4 + 4 = ☐

12 - 3 - 6 - 1 = ☐

53

11 + 12 - 12 = ☐

0 - 0 + 7 = ☐

10 + 10 - 5 = ☐

54 Complete each sentence by choosing the best measurement unit:

cm, mm, l, in., ml, km, lbs, km

a. The car John drives is about four _____ long.

b. My hat is about .5 _____ around.

c. The new baby weighed about nine _____ .

d. The milk bucket holds about two _____ .

Appendix G
(continued)
Grade 4 Winter Test

Student _____ Grade _____ *Page 1 [10/22/89] Grade 4*

School _____ Teacher _____

1 Draw the right sign in the circle.
Draw one of these: < , > , =

a. 0.59 ◯ 0.72

b. 36.3 ◯ 4.63

c. 19.2 ◯ 19.20

2 Find the average of these numbers:

a. 1324 349 128 1031 ☐

b. 32 32 30 2 ☐

c. 5 27 62 3 3 ☐

3 Write in decimal form:

a. 4 dimes and 3 cents. ☐

b. 8 dimes and 0 cents. ☐

c. 1 dime and 64 cents. ☐

4 Add:

6099	5943	5943
+2009	+ 999	+ 55

5 Five people together win $6,273.00 from the lottery. How much will each person receive? ☐

6 It took 52 minutes for John to drive to the state fair. It took 1 hour, and 5 minutes to drive back home at night. How much slower was he in driving back home? ☐

7 a. 9.234 ÷ 10000 = ☐

b. 2245 ÷ 1000 = ☐

c. 36.3 ÷ 100 = ☐

8 a. 66.67 x 10000 = ☐

b. .4371 x 1000 = ☐

c. 336 x 1000 = ☐

9 a. 33.67 x 10 = ☐

b. 8.371 x 1000 = ☐

c. .136 x 100 = ☐

10 Write in Meters only:

a. 3 m, 4 dm, 7 cm = ☐

b. 6 mm, 2 cm, 9 dm, 3 m = ☐

c. 200 cm, 10 dm = ☐

Appendix G
(continued)
Grade 4 Winter Test

Student _____ Grade _____ *Page 2 [10/22/89] Grade 4*

School _____ Teacher _____

11 Write in standard form:

a. 6 tens, 1 one, 5 tenths. ☐

b. 7 ones, 6 tenths, 9 hundredths, 4 thousandths. ☐

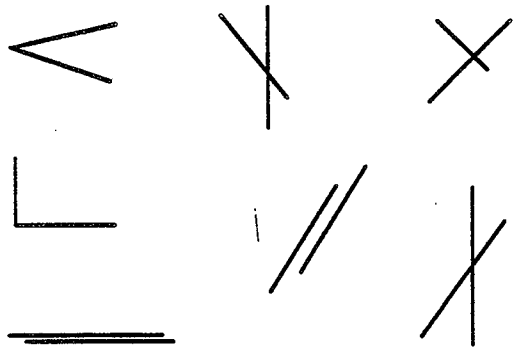
c. 9 hundreds, 4 tens, 0 ones, 8 tenths, 1 hundredth. ☐

12 Complete the chart for long distance phone charges.

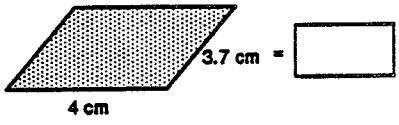
Time	Charge for 3 minutes	Charge for 1 minute
Morning	267¢	
Afternoon	243¢	
Evening	162¢	

13 a. Draw a circle around the perpendicular lines.

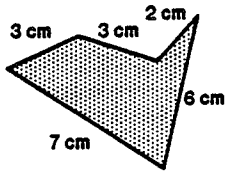
b. Draw a big X on the parallel lines.



14 Find the perimeters:



= ☐



= ☐

15 Divide. Show remainders.

$8\overline{)92}$ $4\overline{)288}$ $5\overline{)8213}$

16 Multiply:

a.	b.	c.
490	932	666
x 5	x 4	x 6

Appendix G
(continued)
Grade 4 Winter Test

Student _____ Grade _____ *Page 3 [10/22/89] Grade 4*

School _____ Teacher _____

17 Write the number that comes before:

☐ 560 ☐ 91 ☐ 438

Write the number that comes after:

7019 ☐ 700 ☐ 810 ☐

18

40 x 1000 = ☐

1,000,000 x 1 = ☐

100 x 10 = ☐

19 Add:

$\frac{7}{9} + \frac{1}{9} =$ ☐

$\frac{5}{7} + \frac{2}{7} =$ ☐

$\frac{12}{33} + \frac{20}{33} =$ ☐

20 Complete the charts:

x → (-7) → y	
x	y
	6
17	
12	

x → (x 6) → y	
x	y
4	
	48
7	

21

3 x 3 + 6 - 3 = ☐

15 ÷ 5 + 4 - 3 = ☐

12 x 3 + 7 - 1 = ☐

22 Find the value of y:

4 → (+7) → y ☐

y → (-9) → 8 ☐

9 → (x 0) → y ☐

23 Add:

26	14	18
19	52	28
+ 41	+ 23	52
☐	☐	+ 74
		☐

24 Divide. Show remainders.

$7\overline{)216}$ $3\overline{)7241}$ $5\overline{)557}$

25

a. Tom gets $3.00 each time he cuts someone's hair. How many haircuts will Tom need to give to earn $24.00? ☐

b. If Jan earns 75¢ for each 'A' on her report card, how many 'A's will it take to earn $24.75? ☐

26

2,000 ÷ 1,00 - 2 - 1 = ☐

10 ÷ 10 - 1 + 2,00 = ☐

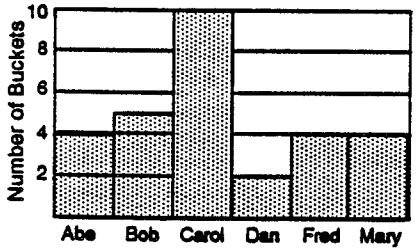
32,000 ÷ 32 - 500 -15 = ☐

Appendix G
(continued)
Grade 4 Winter Test

Student _____ Grade _____ *Page 4 [10/22/89] Grade 4*

School _____ Teacher _____

27



a. How many buckets did Bob pick? ☐

b. Who picked most buckets of cherries? ☐

If each person is paid $2.56 per bucket,

c. How much money did Fred make? ☐

d. How much did Abe make? ☐

28

a. There are 14 weeks to Christmas. Kelly can save $5.00 in allowance each week. How much will she have saved by Christmas? ☐

b. If Kelly also earns $14.50 from babysitting, how much will she have altogether at Christmas? ☐

c. If Kelly decides to spend half of all her money on candy, how much will she have left at Christmas? ☐

29

Multiply:

a. 256 x32

b. 234 x72

c. 806 x64

30

Solve for n:

a. 36 ÷ (10 - 4) = n ☐

b. (45 ÷ 9) + 7 = n ☐

c. (6 x 7) ÷ 7 = n ☐

31

The yard is 46 m wide, and 62 m long. The janitor has 250 m of fencing.

a. Does he have enough to fence around the entire yard? ☐

b. How much extra does he have *or* how much more will he need? ☐

32

Solve for n:

6 = 7 - n n = ☐

5 = n - 6 n = ☐

n = 12 - 4 n = ☐

33

Solve for n:

35.6 + 0.9 = n n = ☐

6.4 = n + 3.02 n = ☐

n + 3.01 = 4.00 n = ☐

34

How many tens in:

a. 100 ☐ d. 320 ☐

b. 3,000 ☐ e. 110 ☐

c. 4,200 ☐ f. 5,020 ☐

Appendix G
(continued)
Grade 4 Winter Test

Student _____ Grade _____ *Page 5 [10/22/89] Grade 4*

School _____ Teacher _____

35 Complete the charts:

x 4	
In	Out
3	
6	
8	
1	

x 9	
In	Out
4	
0	
7	
5	

36 Find the function rule: ____

8 → ? → 2

36 → ? → 9

24 → ? → 6

37

15 x 30 - 16 - 3 = ____

160 ÷ 10 + 220 - 150 = ____

4 x 12 + 9 - 20 = ____

38 a. Round to the tenths place:

45.77 ____ 100.632 ____ 99.50 ____

b. Round to the hundreds place:

1,293.55 ____ 223.00 ____ 652.33 ____

39

CITY	YEAR 1945	1965	1985
Sydney	950,000	1,940,000	3,450,000
Melbourne	760,000	1,840,000	3,150,000
Brisbane	330,000	790,000	1,155,000
Perth	125,000	645,000	1,010,000

a. Name the city with the fewest people in 1965: ____

b. Which city grew the least between 1945 and 1965? ____

c. Which city grew the most between 1945 and 1985? ____

40 Write the following numbers in standard form:

7,000 + 60 + 3 ____

4,000 + 200 ____

3,000 + 900 + 9 ____

Appendix G
(continued)
Grade 4 Winter Test

Student _____ Grade _____ *Page 6 [10/22/89] Grade 4*

School _____ Teacher _____

41 $250 \div 10 - 20 + 5 =$ ☐

$10 \times 8 - 20 + 15 =$ ☐

$240 \div 8 + 9 - 20 =$ ☐

42 Subtract:

$\frac{2}{3} - \frac{1}{3} =$ ☐

$\frac{6}{7} - \frac{2}{7} =$ ☐

$\frac{14}{20} - \frac{11}{20} =$ ☐

43 Find the function rule: ☐

9 —(?)→ 15

5 —(?)→ 11

44 Solve for n:

$n = 367 + 526$ $n =$ ☐

$350 \div 7 = n$ $n =$ ☐

$n = 748 - 267$ $n =$ ☐

45 The clothing store is having a sale. Mary buys a sweater that is normally $64.

a. How much will she pay if the sale is $\frac{1}{4}$ off? ☐

b. How much will she pay if the sale is $\frac{1}{2}$ off? ☐

c. How much will she pay if she pays $\frac{3}{4}$ the regular price? ☐

46 Multiply:

a.	b.	c.
59	87	73
x32	x26	x37

47 Solve:

a. $\frac{1}{3}$ of 27. ☐

b. $\frac{4}{5}$ of 45. ☐

c. $\frac{4}{6}$ of 9. ☐

48 Solve for n:

a. $n \div 8 = 8$ $n =$ ☐

b. $64 \div n = 8$ $n =$ ☐

c. $1 \div 1 = n$ $n =$ ☐

49 a. $.234 \div 1000 =$ ☐

b. $22.455 \div 100 =$ ☐

c. $363 \div 10 =$ ☐

50 $3 + 3 + 6 - 4 - 4 =$ ☐

$5 - 5 + 4 + 4 =$ ☐

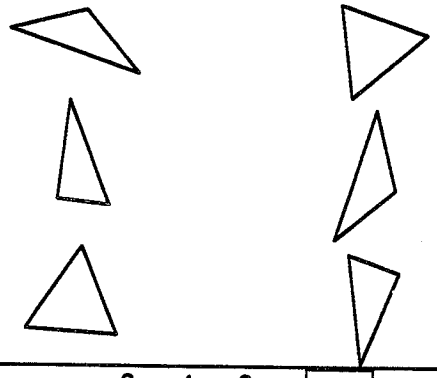
$12 - 3 - 6 - 1 =$ ☐

Appendix G
(continued)
Grade 5 Winter Test

Student _____ Grade _____ *Page 1* *[10/22/89] Grade 5*

School _____ Teacher _____

1 Match the congruent triangles:



2 Circle the correct answer by eliminating the three answers that are clearly wrong:

a. 64,414 - 37,583 =

36,730 5,431.50 101,997 56,831

b. 255 x 2,13 =

5,431.50 54.31 543.15 400.15

c. 5875 + 4024 + 4152 =

14,051 4,162 20,333 13.051

3

a. $\frac{2}{7} + \frac{4}{7} + \frac{2}{7} =$ ☐

b. $\frac{5}{13} + \frac{7}{13} =$ ☐

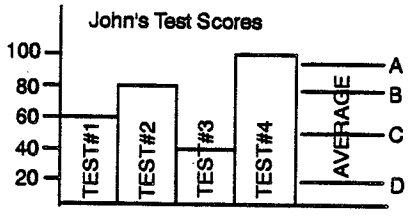
c. $\frac{12}{27} - \frac{4}{27} =$ ☐

4 The hot water tank at Kevin's house holds 45 gallons of hot water. It takes 2 hrs. of continuous use to finish up the hot water.

a. How much hot water is used up in 1 hr? ☐

b. If 5 people live in his house and each of them takes a shower in the morning, what is the most time each should spend in the shower to save enough hot water for everyone? ☐

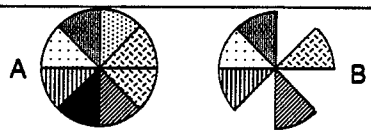
5



a. Which line (A,B,C,D) is closer to the average height of the four Test Bars? ☐

b. What is the ratio of the lowest to the highest test score? ☐

6



a. What fraction of pie B is missing ? ☐

b. What fraction of pie B is still there? ☐

c. What is the ratio of the number of pieces missing to the number still there? ☐

Appendix G
(continued)
Grade 5 Winter Test

Student _____ Grade _____ *Page 2* *[10/22/89] Grade 5*

School _____ Teacher _____

Multiply:

60.00 x2.000 700.8 x99.9 5.943 x .050

Divide.

.08 | 640 .40 | 2.88

1 If Tom, Eric, and Mary each eat $\frac{1}{3}$ of the cake, how much will be left for Phil? ☐

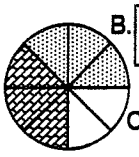
3 Mrs. Vargas has a dozen donuts. She has promised to give them to her three children.
a. If she gives 1/3 to Sally, and 2/12 to RoseMary, what fraction will be left for Julio? ☐
b. How many donuts will Julio get? ☐

5 Change the following mixed numbers into improper fractions:
a. 2 3/4 = ☐
b. 5 7/20 = ☐
c. 6 2/9 = ☐

8 Jody's scores are: 50, 90, 85, 95.
What is her average? ☐

10 2,000 ÷ 1,00 - 2 - 1 = ☐
10 ÷ 10 - 1 + 2,00 = ☐
32,000 ÷ 32 - 500 -15 = ☐

12 If you watch 6 hrs. of television each day, what fraction of the 24 hr. day do you spend in front of the TV? ☐

14 ☐ A.  B. ☐ C. ☐
In each box, write the fraction of the circle that has that pattern.

16 If Sue takes 5 tacks from a box of 75 tacks, what fraction of the box is left? ☐

Appendix G
(continued)
Grade 5 Winter Test

Student _____ Grade _____ *Page 3* *[10/22/89] Grade 5*

School _____ Teacher _____

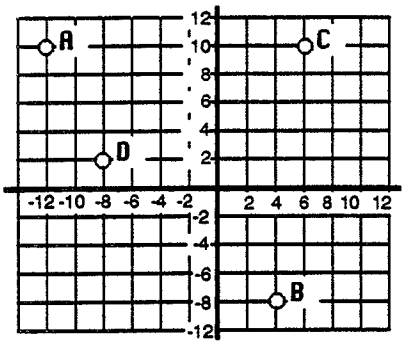
7 Match the triangle to its correct name:

Equilateral

Isosceles

Scalene

18



Find the coordinates of these points on the graph above:

A ( ) C ( )

B ( ) D ( )

9 Convert these improper fractions to mixed numbers:

a. $\frac{13}{6}$ ☐ b. $\frac{28}{3}$ ☐

c. $\frac{47}{9}$ ☐ d. $\frac{18}{8}$ ☐

20 Jane and Al ordered two medium pizzas. Al's pizza was cut in 16 pieces. He ate 4 of them. Jane's pizza was cut in 8 pieces.

a. How many pieces does Jane have to eat to eat the same amount as Al? ☐

b. If Henry also orders a medium pizza, and it is cut into only 4 pieces, how many pieces will it take to eat more than Al? ☐

1 257 ÷ 10 - 20 + 5 = ☐

10 x 100 - 100 + 20 = ☐

182 ÷ 13 + 9 - 20 = ☐

22 2,000 + 1,000 - 2 + 1 = ☐

10 + 10 - 1 + 2,000 = ☐

32,000 + 32 - 500 - 15 = ☐

3 Write the decimal equivalents of the following fractions:

$\frac{4}{100}$ ☐ $\frac{1}{10}$ ☐

$\frac{5}{10}$ ☐ $\frac{25}{100}$ ☐

24 Pam bought a stereo priced at $760.00. The sales tax was 6%.

a. How much was the tax? ☐

b. How much did Pam pay altogether? ☐

Appendix G
(continued)
Grade 5 Winter Test

Student _____ Grade _____ *Page 4* *[10/22/89] Grade 5*

School _____ Teacher _____

25 Solve for n:

35.6 + 0.9 = n n = ☐

6.4 = n + 3.02 n = ☐

n + 3.01 = 4.00 n = ☐

26 Multiply:

490	9.32	.666
x 0.5	x .04	x 6.0

27 Draw one of these signs in each oval: < , > , =

$\frac{7}{8}$ ◯ $\frac{3}{4}$

$\frac{4}{9}$ ◯ $\frac{5}{9}$

$\frac{3}{4}$ ◯ $\frac{2}{3}$

28 Give the equivalent decimal:

a. 55 % ☐ d. $\frac{3}{4}$ ☐

b. 20 % ☐ e. 0.4 % ☐

c. 11.5 % ☐ f. 0.02 % ☐

29 a. 33.67 x 10 = ☐

b. 8.371 x 1000 = ☐

c. .136 x 100 = ☐

30 Complete the table:

Fraction	$\frac{1}{2}$		$\frac{1}{4}$		$\frac{2}{3}$
Decimal		.333			
Percent				75%	

31 Complete the function charts:

y = 7x		y = $\frac{x}{4}$	
x	y	x	y
4	28	12	3
6		20	
7		32	

32 a. 2.34 ÷ 1000 = ☐

b. 22.455 ÷ 100 = ☐

c. 363 ÷ 10 = ☐

33 Find the value of y:

46 → (x 12) → y ☐

33 → (x 10) → y ☐

9 → (x 431) → y ☐

34 Find the value of y:

408 → (+ 34) → y ☐

4296 → (+ 16) → y ☐

800 → (+ 100) → y ☐

Appendix G
(continued)
Grade 5 Winter Test

Student _____ Grade _____ *Page 5* *[10/22/89] Grade 5*

School _____ Teacher _____

.5 Find the average of these numbers:

a. 1324 349 128 1031 ☐

b. 32 32 30 2 ☐

c. 5 27 62 3 3 ☐

.7 Complete the function charts:

$y = \frac{x}{3}$	
x	y
	24
6	
12	

$y = 4x$	
x	y
12	
	20
	32

.9 The 7th grade of Moshone school is planning a picnic. The total cost of the picnic, including renting the park and buying the food, is $126.00.

a. If only one class (of 28 people) goes, how much will it cost each person? ☐

b. If two classes go (one with 28 people, and one with 21) how much will it cost each person? ☐

.1 Multiply:

a. $603.18 x $44.00 = ☐

b. $500.03 x $1.45 = ☐

c. $21.30 x .23 = ☐

36 Complete the chart:

$y = \frac{x}{7} - 4$				
x	28	0	70	56
y				

38 a. Scott is making chili for 12 people. He wants each person to have 0.25 kilograms of chili. How much chili will he need to make to have enough for everyone? ☐

b. The baseball stadium buys caps in cases of 50 caps for $62.50. If the caps are sold for $3.00 each, how much profit does the stadium make on each cap? ☐

c. How much profit does the stadium make altogether on two cases of hats? ☐

40 Solve for n:

$n = \frac{1}{3}$ of 24 ☐

$n = \frac{3}{4}$ of 60 ☐

$n = \frac{4}{10}$ of 90 ☐

42 Solve the following composite functions:

x → (÷3) → n → (+5) → y

If x is 6, what is y? ☐

x → (x4) → n → (-7) → y

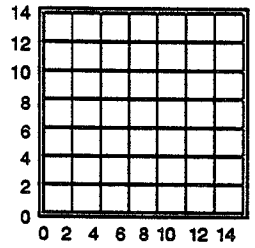
If y is 21, what is x? ☐

Appendix G
(continued)
Grade 5 Winter Test

Student _____ Grade _____ *Page 6* *[10/22/89] Grade 5*

School _____ Teacher _____

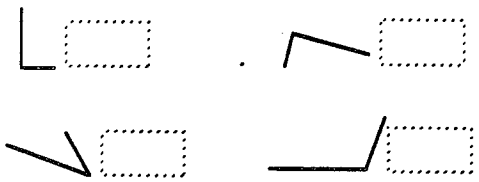
3



Plot and label the following coordinates on the graph:

A = (2,8) B = (9,3) C = (6,12)

5 Label each angle as acute, obtuse, or right:



7 Divide.

.25 ⟌ .7800 0.30 ⟌ 1.4760

9

a. 66.67 x 1,000 = _____

b. .4371 x 100 = _____

c. 336 x 100 = _____

11 It takes 4 glasses of water to fill a jug. If the jug is already 1/8th full, how many more glasses of water will it take to fill the jug? _____

44 Solve for n:

n = 13 x 22 _____

16,000 ÷ 400 = n _____

n = 927 + 3839 _____

8536 - 99 = n _____

523 - 89 = n _____

46

a. 9.234 ÷ 10000 = _____

b. 2245 ÷ 1000 = _____

c. 36.3 ÷ 100 = _____

48

a. Cathy needs hats for the birthday party. The hats come only in packages of 8. Cathy needs 25 hats. How many packages will she need to buy? _____

b. The three Smith brothers earned $12.75 for doing yard work. If they divide the money equally,how much will each brother get? _____

50 Multiply:

a. 0.005 x 30 = _____

b. 83 x 6.3 = _____

c. 0.034 x 223.11 = _____

52

15 x 30 + 16 - 3 = _____

160 + 10 + 224 - 150 = _____

45 x 12 + 9 - 200 = _____

Appendix H
Initial Total Score Feedback for Teachers

Open Court-Based Math Assessment (11/28/89)

Raw scored results for teachers from October teat sampling the full year's work.
Total possible: Grade 3—155, Grade 4—147, Grade 5—147.
(Detailed results by Mastery Objectives will be provided later in the month.)

Directions

Raw score is number correct on entire test. To interpret the student ID number, see accompanying confidential code page.

Student ID		Raw Score
A.L.5.01	☞	27
A.L.5.02	☞	36
A.L.5.03	☞	51
A.L.5.04	☞	74
A.L.5.05	☞	31
A.L.5.06	☞	50
A.L.5.07	☞	22
A.L.5.08	☞	50
A.L.5.09	☞	43
A.L.5.10	☞	16
A.L.5.11	☞	77
A.L.5.12	☞	56
A.L.5.13	☞	47
A.L.5.14	☞	29
A.L.5.15	☞	35
A.L.5.16	☞	27
A.L.5.17	☞	70
A.L.5.18	☞	46
A.L.5.19	☞	24
A.L.5.20	☞	40
A.L.5.21	☞	20
A.L.5.22	☞	17

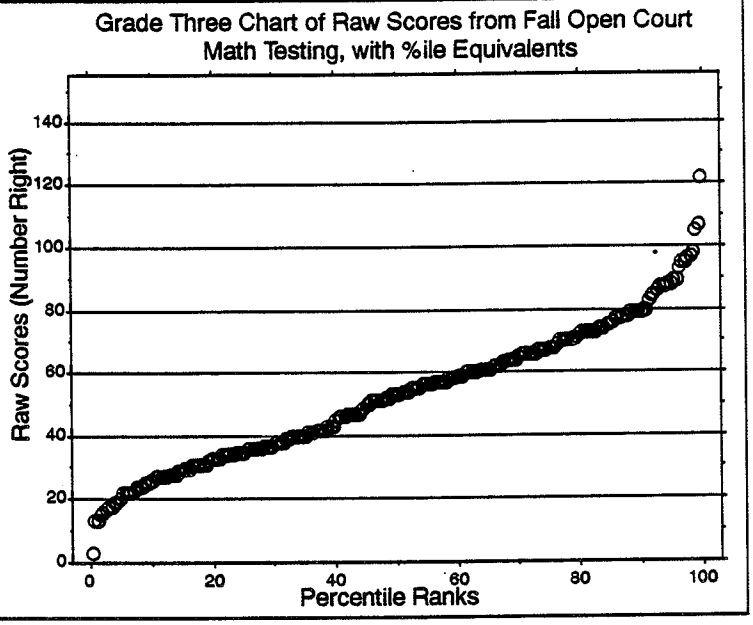
Appendix I
Decile Line Graph and Frequency Distribution Feedback for Teachers: Grade 3

Fall Open Court Math Testing:
Grade 3 Results [11/28/89]

Part 1: Norms to Help calculate Percentiles:

Grade 3

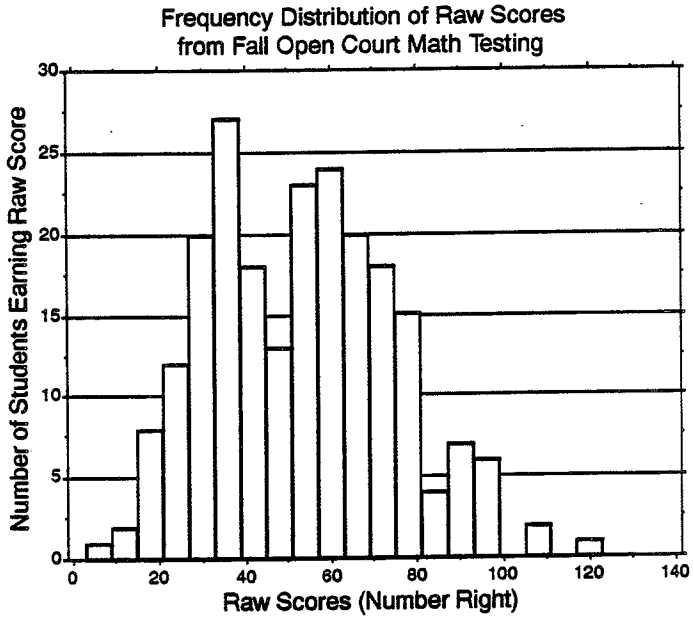
%ile Rank	Raw Score
1	13
5	21
10	26
20	33
30	37
40	44
50	53
60	59
70	65
80	72
90	79
95	88
99	105



Part 2: Other Results:

Statistical Summary for All Gr. 3 Students

Mean:	Std. Dev.:	Count:
52.71	21.61	221
Minimum:	**Maximum:**	**Range:**
3	122	119



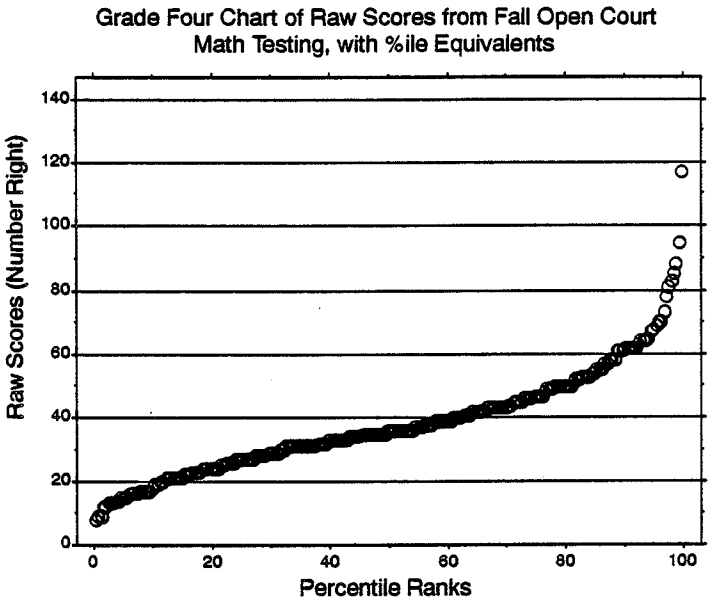
Appendix I
(continued)
Decile Line Graph and Frequency Distribution
Feedback for Teachers: Grade 4

Fall Open Court Math Testing:
Grade 4 Results [11/28/89]

Part 1: Norms to Help calculate Percentiles:

Grade Four

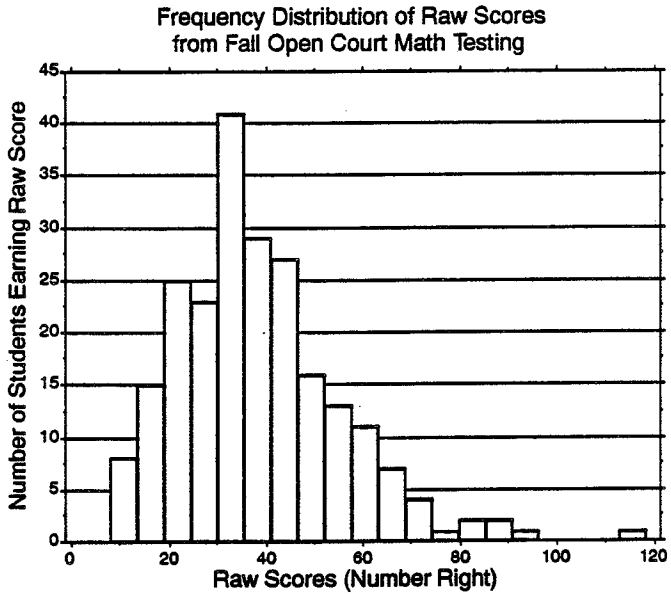
%ile Rank	Raw Score
1	9
5	15
10	18
20	24
30	29
40	33
50	36
60	39
70	43
80	50
90	61
95	68
99	89



Part 2: Other Results:

Statistical Summary for All Gr. 4 Students

Mean:	Std. Dev.:	Count:
38.2	16.92	226
Minimum:	**Maximum:**	**Range:**
8	117	109



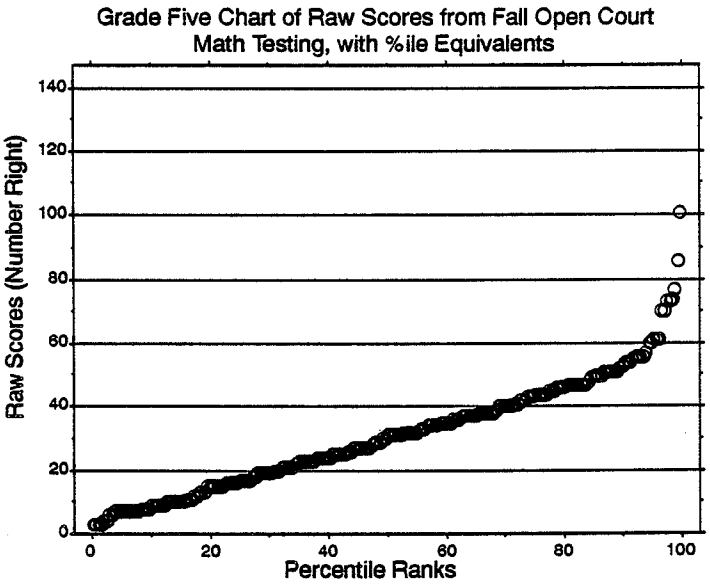
Appendix I
(continued)
Decile Line Graph and Frequency Distribution
Feedback for Teachers: Grade 5

Fall Open Court Math Testing:
Grade 5 Results [11/28/89]

Part 1: Norms to Help calculate Percentiles:

Grade Five

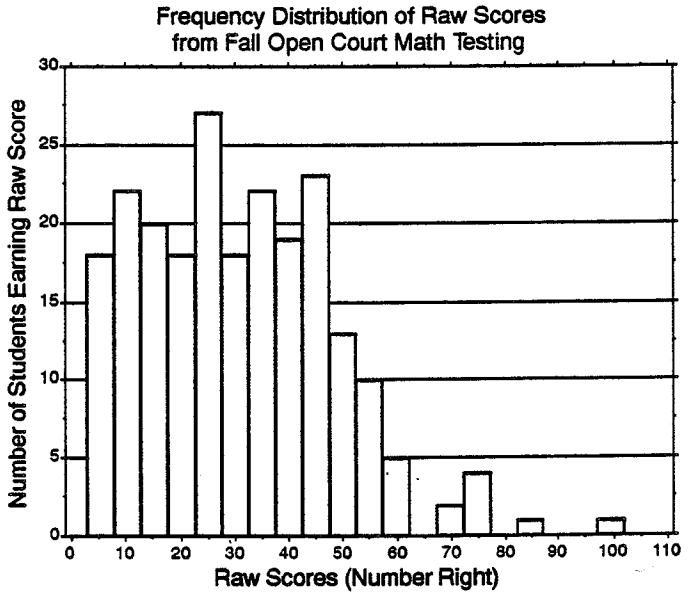
%ile Rank	Raw Score
1	3
5	7
10	8
20	15
30	19
40	24
50	31
60	35
70	40
80	46
90	53
95	60
99	79



Part 2: Other Results:

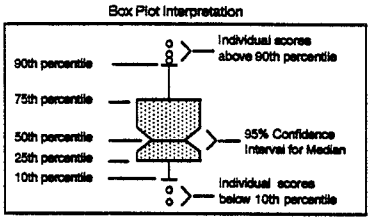
Statistical Summary for All Gr. 5 Students

Mean:	Std. Dev.:	Count:
31	17.77	223
Minimum:	**Maximum:**	**Range:**
3	101	98



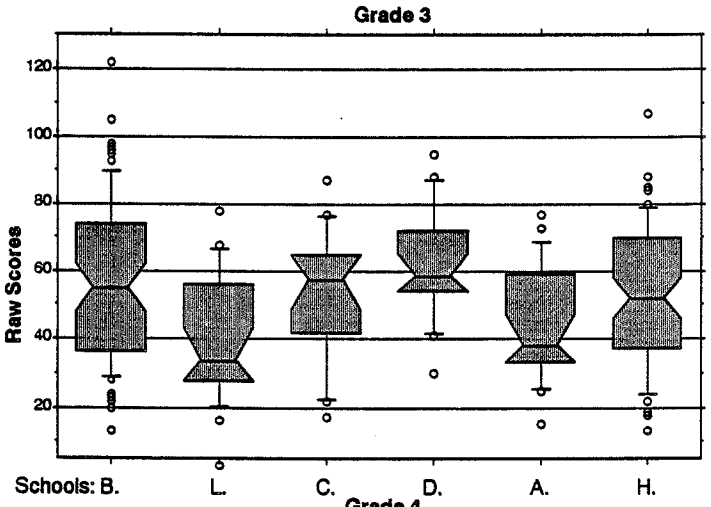
Appendix J
Box Plots for Building-Level Feedback: Grades 3 - 5

Fall Open Court Math Testing:Grade 3, 4, & 5 results.
Comparisons among Schools [11/28/89]



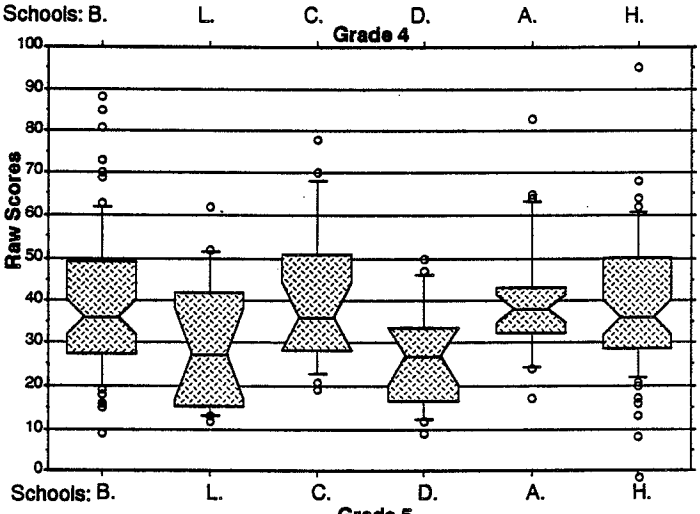
Grade 3 Results

School	# Students:	Mean:	Std. Dev.:
B	73	55.8	24.1
L	22	39.8	19.0
C	16	53.5	19.6
D	20	62.0	16.4
A	21	44.1	17.2
H	69	53.2	20.4



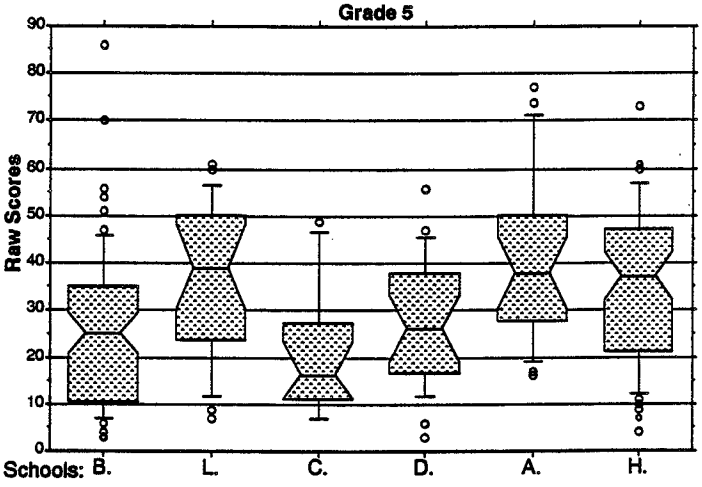
Grade 4 Results

School	# Students:	Mean:	Std. Dev.:
B	73	39.8	17.2
L	16	29.9	15.78
C	21	40.1	17.17
D	16	26.5	12.5
A	26	40.1	14.23
H	74	39.6	17.45



Grade 5 Results

School	# Students:	Mean:	Std. Dev.:
B	78	25.7	16.62
L	22	37	16.31
C	13	20.3	13.71
D	23	27.3	14.08
A	22	40.3	17.94
H	65	35.5	18.44

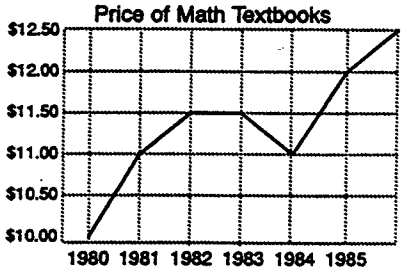


Appendix K
Fall/Winter Test & Subtest Results for Teachers

CODE	D	E	G	H	I	J, N	L	N1	N3	O	Full Test
OBJECTIVES	multiply	divide	combined	fractions	decimals	word probs.	geometry	averaging	graph/chart	algebra	
# PROBLEMS	3	3	10	8	11	11	4	3	3	18	74
GRADE AVG.	44%	29%	29%	37%	23%	24%	23%	14%	29%	34%	29%
Percent Correct for each student:											
Student Code											
"BJ5.01"	33%	67%	30%	50%	9%	18%	0%	0%	33%	50%	31%
"BJ5.02"	100%	0%	30%	13%	55%	36%	50%	0%	0%	61%	41%
"BJ5.03"	0%	0%	20%	0%	0%	27%	0%	0%	67%	44%	20%
"BJ5.04"	33%	0%	40%	63%	0%	73%	0%	0%	0%	39%	34%
"BJ5.05"	100%	0%	40%	0%	64%	45%	0%	0%	0%	22%	31%
"BJ5.06"	100%	0%	90%	63%	45%	45%	0%	0%	100%	67%	57%
"BJ5.07"	0%	0%	0%	63%	9%	45%	100%	0%	100%	44%	35%
"BJ5.08"	100%	0%	60%	0%	36%	18%	50%	0%	0%	56%	36%
"BJ5.09"	33%	0%	20%	63%	18%	36%	0%	0%	0%	50%	31%
"BJ5.10"	33%	0%	40%	63%	18%	36%	0%	0%	0%	39%	31%
"BJ5.11"	100%	0%	30%	0%	36%	36%	0%	33%	0%	44%	31%
"BJ5.12"	0%	0%	20%	25%	0%	18%	0%	0%	33%	39%	19%
"BJ5.13"	100%	33%	90%	75%	73%	73%	100%	100%	0%	83%	77%
"BJ5.12"	100%	67%	60%	50%	64%	73%	0%	0%	0%	44%	51%
"BJ5.15"	100%	67%	30%	63%	45%	27%	0%	0%	67%	39%	41%
"BJ5.16"	100%	100%	30%	75%	55%	45%	0%	0%	33%	50%	49%
"BJ5.17"	100%	67%	30%	75%	45%	73%	50%	33%	33%	50%	54%
"BJ5.18"	67%	0%	20%	38%	55%	36%	0%	0%	67%	44%	36%
"BJ5.19"	67%	67%	70%	38%	18%	9%	75%	0%	67%	56%	43%
"BJ5.20"	33%	0%	70%	50%	64%	36%	0%	0%	0%	50%	43%
"BJ5.21"	67%	67%	40%	25%	45%	45%	0%	0%	0%	39%	36%
"BJ5.22"	100%	100%	80%	75%	91%	55%	100%	100%	67%	89%	82%
"BJ5.23"	67%	0%	80%	13%	45%	45%	0%	0%	67%	72%	49%
"BJ5.24"	100%	0%	80%	63%	18%	36%	0%	33%	0%	44%	42%

Appendix L
Fall/Winter Item-Specific Results for Teachers

Code: N
Objectives applicat.
Problems: 6
Grade Av'ge: 34%



a. What was the price of a math textbook in1982? ☐

b. How much did the price of a textbook increase from 1980 to 1984? ☐

Item: 23a 29% :Difficulty
Item: 23b 12% :Difficulty

Items	Prices	
	Store A	Store B
Milk (1 gallon)	$1.85	$1.95
Ivory Soap (4 bars)	$1.75	$1.50
Orange Juice (12 oz)	99¢	$1.05

b. Did every item cost more in the more expensive store? ☐

c. If you buy just milk and soap, which store is more expensive? ☐

Raymond has 2 one-dollar bills and 12 dimes.

a. Does Raymond have enough money to buy a book costing $3.06 ? ☐

b. Does he have enough to buy two baseballs, if they cost $1.30 each ? ☐

Item: 32b 37% :Difficulty
Item: 32c 32% :Difficulty

Item: 29a 43% :Difficulty
Item: 29b 53% :Difficulty

Code: O
Objectives algebra
Problems: 6
Grade Av'ge: 11%

Solve for n:

b. 64 ÷ n = 8 n = ☐

c. 1 ÷ 1 = n n = ☐

Item: 14b 4% :Difficulty
Item: 14c 7% :Difficulty

Solve for n:

42 = n x 6 n = ☐

n = 0 x 2 n = ☐

Item: 15a 9% :Difficulty
Item: 15b 15% :Difficulty

Solve for n:

6 = 7 - n n = ☐

n = 12 - 4 n = ☐

Item: 48a 15% :Difficulty
Item: 48c 18% :Difficulty

Code: P
Objectives relations
Problems: 6
Grade Av'ge: 33%

Change to all dollars or all cents:

a. 3745 ¢ = $ ☐

b. $ 21.45 = ☐ ¢

c. 703 ¢ = $ ☐

Item: 25a 26% :Difficulty
Item: 25b 26% :Difficulty
Item: 25c 26% :Difficulty

a. $3.20 = $ ☐ and ☐ ¢

b. $ ☐ and 30¢ = $4.30

c. $15.30 = $ ☐ and ☐ ¢

Item: 26a 40% :Difficulty
Item: 26b 41% :Difficulty
Item: 26c 40% :Difficulty

Appendix M
Grade 3 Spring Test

Name_____ School_____ Teacher_____

Page 1
Gr. 3 [8/9/90]

1 Multiply:

$$\begin{array}{ccc} 7 & 0 & 6 \\ \times\,3 & \times\,4 & \times\,8 \end{array}$$

2 Multiply:

$$\begin{array}{ccc} 9 & 6 & 6 \\ \times\,4 & \times\,5 & \times\,9 \end{array}$$

3 What part of each circle is shaded?

a. ☐ b. ☐

c. ☐ d. ☐

4 Add:

$$\begin{array}{r} 56 \\ 17 \\ +\,31 \\ \hline \square \end{array} \qquad \begin{array}{r} 38 \\ 28 \\ 52 \\ +\,62 \\ \hline \square \end{array}$$

5 11 + 21 - 6 = ☐

7 + 9 - 3 = ☐

6 4 + 3 + 5 - 4 - 7 = ☐

7 How much is:

a. $\frac{1}{3}$ of 18? ☐

b. $\frac{1}{4}$ of 12? ☐

8 Multiply:

a. 4 x 10 = ☐

b. 7 x 0 = ☐

c. 10 x 8 = ☐

9 Rewrite to show no hundreds:

a. 604 = _____ tens and _____ ones.

b. 915 = _____ tens and _____ ones.

c. 252 = _____ tens and _____ ones.

10 Multiply:

a. 6 x 100 = ☐

b. 72 x 1000 = ☐

c. 100 x 59 = ☐

11 Multiply:

$$\begin{array}{ccc} 8 & 6 & 4 \\ \times\,8 & \times\,6 & \times\,4 \end{array}$$

12 Solve for n:

a. 72 ÷ n = 8 n = ☐

b. 3 ÷ 3 = n n = ☐

Appendix M
(continued)
Grade 3 Spring Test

Name_____ School_____ Teacher_____

Page 2
Gr. 3 [8/9/90]

13 Solve for n:

a. 54 = n x 6 n = ☐

b. n = 0 x 5 n = ☐

14 Subtract:

a. 5.54 - 2.63 = ☐

B. 12.7 - 3.8 = ☐

15 Write the name of the unit that makes the most sense.
Write **km** for Kilometers, **m** for meters, **c** for centimeters, **kg** for kilograms, or **g** for grams.

a. The horse was about 2 ☐ tall.

b. The paper clips weighed about 9 ☐

c. The bird was about 11 ☐ long.

d. The river was about 40 ☐ long.

16 a. Leah bought a cat that was 15 cm tall. Now it is 21 cm tall. How much did it grow since Leah bought it? ☐

b. Silas lives 35 km east of Salem. His sister lives 12 km west of Salem. About how far does Silas live from his sister? ☐

17 a. Bart is 11 years old. His sister Mary is 2 years older than him. How old is Mary? ☐

b. Lewis bought a Big Gulp for 70¢ and some gum for 30¢. How much money did he spend? ☐

18 Choose the best measurement unit:
km, **m**, **c**, **kg**, **g**

The car John drives is about four _____ long.

19 Change to all dollars or all cents:

a. 4392 ¢ = $ ☐

b. $ 11.50 = ☐ ¢

c. 606 ¢ = $ ☐

20 a. $8.15 = $ ☐ and ☐ ¢

b. $ ☐ and 60¢ = $9.60

c. $22.50 = $ ☐ and ☐ ¢

Appendix M
(continued)
Grade 3 Spring Test

Name_____ School_____ Teacher_____

21 Sandra has 5 one-dollar bills and 14 dimes.

a. Does she have enough money to buy a book costing $6.06 ? ☐

b. Does she have enough to buy two pins if they cost $3.10 each ? ☐

22 a. Ron made a sleeping bag. He used 11 meters of canvas. How many meters will he need to make 5 more sleeping bags? ☐

b. Ron is buying 15 meters of canvas. The piece at the store is 23 meters long. What length of canvas will be left? ☐

23 Subtract:

a. 57 - 15 = ☐

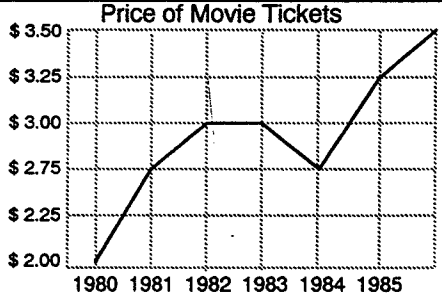
b. 46 - 16 = ☐

c. 18 - 10 = ☐

24 Multiply:

8	7	5
x 9	x 6	x 8

25 Price of Movie Tickets



a. What was the price of a movie ticket in 1982? ☐

b. How much did the price of a ticket increase from 1980 to 1984? ☐

26

Items	Prices	
	Store A	Store B
Milk (1 gallon)	$1.85	$1.95
Mint cookies (1 pkg)	$1.05	$1.50
Orange Juice (12 oz)	$1.26	$1.18

a. Did every item cost more in the more expensive store? ☐

b. If you buy just cookies and juice, which store is more expensive? ☐

Appendix M
(continued)
Grade 3 Spring Test

Name_____ School_____ Teacher_____

Page 4
Gr. 3 [8/9/90]

27 Draw lines to match the correct time with clocks.

12 : 17 •
12 : 32 •
"twelve minutes 'til 12." •
11 : 48 •
"seventeen after 12" •

28 Trish caught 4 fish. Each fish weighed between 2 and 4 pounds.

a. Can Trish have 12 pounds of fish altogether? ☐

b. Can Trish have 20 pounds altogether? ☐

c. Can she have 18 pounds of fish altogether? ☐

29 Multiply:

a. 10 x 85 = ☐

b. 27 x 10 = ☐

c. 81 x 10 = ☐

30 Subtract:

(a)	(b)
61.30	8.7
-4.23	-2.8

31 Multiply:

8	7	3
x 0	x 4	x 6

32 Add:

(a)	(b)	(c)
3.08	4.76	10.06
+0.52	+6.37	+2.13

33

a. 1 foot = ☐ inches.

b. 1 yard = ☐ feet.

34 Multiply:

a. 50 x 100 = ☐

b. 1000 x 70 = ☐

c. 30 x 10 = ☐

35 Multiply:

(a)	(b)	(c)
380	742	555
x 6	x 3	x 5

36 Solve for n:

(a) 5 = 8 - n n = ☐

(b) n = 14 - 6 n = ☐

37

(a) 13 - 12 + 12 = ☐

(b) 11 + 11 - 4 = ☐

Appendix M
(continued)
Grade 4 Spring Test

Name _____ Grade _____ *Page 1 [8/9/90] Grade 4*

School _____ Teacher _____

1 Draw the correct sign in the circle.
Draw one of these: < , > , =

29.2 ◯ 29.20

2 Find the average of these numbers:

14 14 40 4 ☐

3 Write in decimal form:
a. 3 dimes and 5 cents. ☐
b. 7 dimes and 0 cents. ☐
c. 1 dime and 48 cents. ☐

4 Add:

4942
6007
+ 989

5 47.5 ÷ 100 = ☐

6 Write in standard form:
6 tens, 1 one, 5 tenths.
☐

7 41.76 x 10 = ☐

8 552 x 1000 = ☐

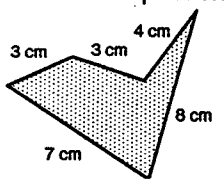
9 It took 86 minutes for Shannon to drive to the park. It took 1 hour and 6 minutes to drive back home at night. How much faster was she driving back home?
☐

10 Write in meters only:
3 m, 4 dm, 7 cm = ☐

11 Divide. Show remainders.

$7\overline{)93}$ $3\overline{)266}$

12 Find the perimeter:



= ☐

13 Write the number that comes after:
6019 ☐ 800 ☐ 710 ☐

14 $\frac{6}{8} + \frac{1}{8} =$ ☐

15 Multiply:

(a)	(b)	(c)
590	832	656
x 6	x 5	x 4

16 (a) 50 x 1000 = ☐
(b) 1,000,000 x 1 = ☐
(c) 10 x 100 = ☐

Appendix M
(continued)
Grade 4 Spring Test

Name _____ Grade _____ *Page 2 [8/9/90] Grade 4*

School _____ Teacher _____

17 Find the value of y:

(a) 5 —(+8)→ y ☐

(b) y —(-7)→ 9 ☐

(c) 4 —(x 0)→ y ☐

18 Complete the charts:

(a) x —(-7)→ y

x	y
	6
17	
12	

x —(x 6)→ y (b)

x	y
4	
	48
7	

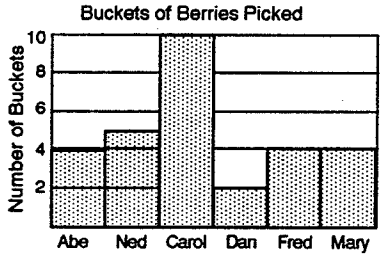
19 18 ÷ 6 + 5 - 3 = ☐

20 Tom gets $4.00 each time he washes a car. How many cars will Tom need to wash to earn $32.00? ☐

21 Add:
27
28
52
+ 63
☐

22 Multiply:
232
x64

23



(a) How many buckets did Ned pick? ☐

If each person is paid $1.49 per bucket,

(b) How much money did Fred make? ☐

(c) How much did Abe make? ☐

24

CITY	YEAR 1945	1965	1985
Sydney	950,000	1,940,000	3,450,000
Melbourne	760,000	1,840,000	3,150,000
Brisbane	330,000	760,000	1,255,000
Perth	325,000	745,900	1,910,000

(a) Name the city with the fewest people in 1965: ☐

(b) Which city grew the least between 1945 and 1965? ☐

Appendix M
(continued)
Grade 4 Spring Test

Name _____ Grade _____ *Page 3 [8/9/90] Grade 4*

School _____ Teacher _____

25 The yard is 38 m wide and 53 m long. The prison guard has 180 m of wire

How much more wire does he need to fence around the entire yard?

26 Solve for **n**:

(a) 45 ÷ (12 - 3) = **n**

(b) (48 ÷ 6) + 7 = **n**

(c) (6 x 7) ÷ 7 = **n**

27 Solve for **n**:

25.7 + 0.9 = **n** **n** =

28 Solve for **n**:

(a) 6 = **n** - 5 **n** =

(b) **n** = 13 - 4 **n** =

29 Complete the chart:

x 8	
In	Out
6	
0	
7	
5	

30 (a) There are 27 weeks until Christmas. Jason can save $3.00 in allowance each week. How much will he have saved by Christmas?

(b) If Jason also earns a total of $12.00 from baby-sitting, how much will he have altogether at Christmas?

(c) Jason spent half of all his allowance and baby-sitting money on Christmas gifts. How much does he have left?

31 (a) 240 ÷ 10 + 200 - 150 =

(b) 3 x 15 + 4 - 22 =

32 How many tens in:

(a) 100 (d) 240

(b) 8,000 (e) 101

(c) 5,500 (f) 6,030

33 Write the following numbers in standard form:

(a) 8,000 + 40 + 1

(b) 6,000 + 500

(c) 2,000 + 800 + 8

34 The clothing store is having a sale. Kristen buys a sweater that normally costs $60.

(a) How much will she pay if the sale is $\frac{1}{4}$ off?

(b) How much will she pay if the sale is $\frac{1}{2}$ off?

Appendix M
(continued)
Grade 4 Spring Test

Name _____ Grade _____ *Page 4 [8/9/90] Grade 4*

School _____ Teacher _____

35 (a) 360 ÷ 10 - 30 + 5 = ☐

(b) 10 x 7 - 20 + 17 = ☐

36 Subtract:

$\frac{7}{8} - \frac{3}{8} =$ ☐

37 Find the function rule: ☐

9 → ? → 15

5 → ? → 11

38 Solve for n:

(a) n = 365 + 426

n = ☐

(b) n = 648 - 256

n = ☐

39 Solve:

$\frac{1}{3}$ of 27 ☐

40 Multiply:

84
x 48

41 3 + 4 + 7 - 5 - 5 = ☐

42 (a) Round to tenths place: 61.66 ☐

(b) Round to hundreds place: 751.59 ☐

43 Find the function rule: ☐

9 → ? → 3

36 → ? → 12

24 → ? → 8

44 Solve for n:

(a) 72 ÷ n = 9 n = ☐

(b) 1 ÷ 1 = n n = ☐

Appendix M
(continued)
Grade 5 Spring Test

Student _____ Grade _____ *Page 1 [6/9/90] Grade 5*

School _____ Teacher _____

1

(a) $\frac{2}{6} + \frac{3}{6} + \frac{2}{6} =$ ☐

(b) $\frac{8}{23} + \frac{2}{23} =$ ☐

(c) $\frac{16}{17} - \frac{7}{17} =$ ☐

2

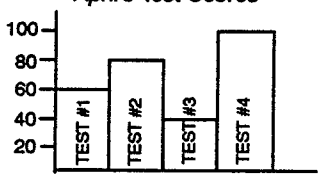
The hot water tank at Kevin's house holds 30 gallons of hot water. Two hrs. of continuous running will use up all the hot water.

(a) How much hot water is used up in 1 hr? ☐

(b) If 4 people each take a daily shower, what is the most time each can spend in the shower to save enough hot water for everyone? ☐

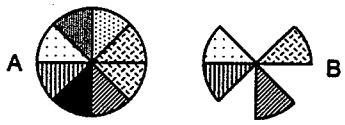
3

April's Test Scores



What is the ratio of the highest to the lowest test score? ☐

4



(a) What fraction of pie B is missing ? ☐

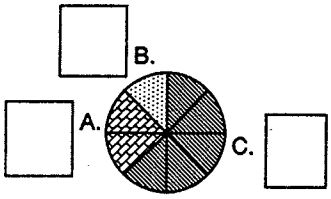
(b) What fraction of pie B remains ? ☐

5

(a) 50.00 × 3.000

(b) 6.953 × .040

6



In each box, write the fraction of the circle with each pattern.

7

Jody's scores are: 65, 20, 80, 55. What is her average? ☐

8

If you watch 4 hrs. of television each day, what fraction of the 24 hr. day do you spend in front of the TV? ☐

Appendix M
(continued)
Grade 5 Spring Test

Student _____ Grade _____ *Page 2 [8/9/90] Grade 5*

School _____ Teacher _____

9 If Sam, Eric, Sue, and Tom each eat $\frac{1}{4}$ of the cake, how much will be left for Phil? ☐

10 (a) 4,000 ÷ 100 - 2 - 3 = ☐

(b) 100 ÷ 100 + 200 - 3 = ☐

11 If Sue takes 8 tacks from a box of 64 tacks, what fraction of the box is left? ☐

12 (100 x 10) - 100 + 20 = ☐

(168 ÷ 12) + 9 - 20 = ☐

13 Fred and Jane each ordered a medium pizza. Fred's pizza was cut in 12 pieces. He ate 4 of them. Jane's pizza was cut in 6 pieces.

(a) Jane ate the same amount as Fred. How many pieces did she eat? ☐

(b) Henry also ordered a medium pizza, and had it cut into 24 pieces. He ate more pizza than Fred. At least how many pieces did Henry eat? ☐

14 Complete the function chart:

$y = \frac{x}{3}$

x	y
12	4
21	
33	

15 Write the decimal equivalents of the following fractions:

(a) $\frac{25}{100}$ ☐ (b) $\frac{5}{10}$ ☐

16 4,000 + 2,000 - 1 + 10 = ☐

10 + 10 - 1 + 3,000 = ☐

45,000 + 45 - 500 - 25 = ☐

17 Draw one of these signs in the oval: < , > , =

$\frac{2}{3}$ ◯ $\frac{3}{4}$

18 (a) 8.331 x 1000 = ☐

(b) .173 x 100 = ☐

19 Solve for n:

(a) 52.4 + 0.9 = n n = ☐

(b) 12.8 = n + 4.01 n = ☐

(c) n + 2.01 = 4.00 n = ☐

20 Multiply:

.666
x 6.0

Appendix M
(continued)
Grade 5 Spring Test

Student _____ Grade _____ *Page 3 [8/9/90] Grade 5*

School _____ Teacher _____

21 Find the value of y:

(a) 34 → (x 21) → y ☐

(b) 9 → (x 322) → y ☐

22

(a) 4.44 ÷ 1000 = ☐

(b) 32.655 ÷ 100 = ☐

(c) 474 ÷ 10 = ☐

23 Find the average of these numbers:

(a) 41 41 40 2 ☐

(b) 6 26 32 3 3 ☐

24 Find the value of y:

(a) 528 → (+ 33) → y ☐

(b) 700 → (+ 100) → y ☐

25 Complete the function chart:

y = 3 x	
x	y
11	
	39
	18

26 A seventh grade class at Sunrise school is planning a picnic. The total cost of the picnic, including park rental and food, is $126.00.
If the costs are shared evenly among the 28 students, how much will it cost each student? ☐

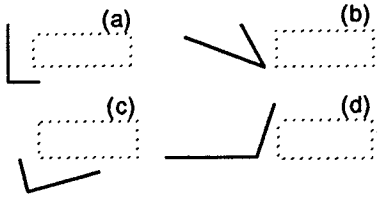
27 Solve for n:

n = $\frac{1}{4}$ of 32 ☐

n = $\frac{3}{4}$ of 80 ☐

28 Lou is making ice cream for 13 people. He wants each person to have 0.25 km of ice cream to eat. How much ice cream will he need for everyone to have enough? ☐

29 Label each angle acute, obtuse, or right:



30 Solve these functions:

X → (÷ 2) → n → (+ 5) → y

If x is 8, what is y? ☐

X → (x 5) → n → (- 6) → y

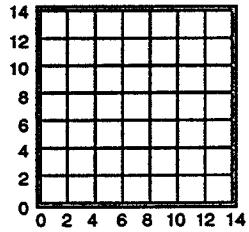
If y is 24, what is x ? ☐

Appendix M
(continued)
Grade 5 Spring Test

Student _____ Grade _____ *Page 4 [8/9/90] Grade 5*

School _____ Teacher _____

31



Plot and label the following coordinates on the graph:

A = (4,10) B = (9,2) C = (11,12)

32

Solve for n:

(a) n = 12 x 23 ☐

(b) 20,000 ÷ 400 = n ☐

(c) n = 627 + 3,837 ☐

(d) 7,535 - 89 = n ☐

(e) 413 - 86 = n ☐

33

(a) Ms. Finch needs math tests for her class of 25. If the tests come only in packages of 7, how many packages will she need for the class? ☐

(b) The three Plimp brothers earned $7.08 for doing yard work. If they divide the money equally, how much will each brother get? ☐

34

(a) 8.222 ÷ 10,000 = ☐

(b) 6,246. ÷ 1000 = ☐

(c) 82.2 ÷ 100 = ☐

35

(a) 22.67 x 100 = ☐

(b) 0.4371 x 1000 = ☐

(c) 3366 x 1000 = ☐

36

(a) 12 x 40 + 16 - 2 = ☐

(b) 240 ÷ 10 + 204 - 120 = ☐

(c) 50 x 22 + 7 - 200 = ☐

37

Multiply:

$403.13 x $54.00 = ☐

Appendix N
Percentile Ranks for Raw Scores: Spring Tests

Grade 3		Grade 4		Grade 5	
Raw Score	Percentile	Raw Score	Percentile	Raw Score	Percentile
10	<1	4	<1	1	<1
12	<1	5	<1	2	.01
15	.01	6	.01	4	.01
24	.01	7	.02	6	.02
25	.02	11	.03	7	.02
28	.03	12	.03	8	.03
30	.03	13	.04	10	.04
31	.04	14	.05	11	.04
32	.05	15	.06	13	.05
33	.06	17	.07	14	.06
34	.07	18	.08	15	.07
36	.07	19	.09	16	.07
39	.08	20	.11	17	.08
40	.09	23	.12	18	.09
41	.10	24	.14	19	.09
42	.10	25	.16	21	.10
43	.11	26	.16	22	.11
44	.12	27	.17	23	.12
45	.12	28	.18	24	.12
46	.13	29	.19	25	.13
48	.13	30	.20	26	.15
49	.14	31	.22	27	.16
50	.15	32	.23	28	.17
51	.16	33	.25	29	.18
52	.18	34	.27	30	.19
53	.19	35	.29	31	.21
54	.20	36	.30	32	.21
55	.20	37	.32	33	.22
56	.21	38	.35	34	.24
57	.23	39	.37	34	.24
58	.24	40	.40	35	.26
59	.25	41	.43	36	.28
60	.26	42	.45	37	.30
61	.28	43	.46	38	.32
63	.28	44	.47	39	.33
64	.29	45	.48	40	.34

Appendix N
(continued)
Percentile Ranks for Raw Scores: Spring Tests

Grade 3		Grade 4		Grade 5	
Raw Score	Percentile	Raw Score	Percentile	Raw Score	Percentile
65	.30	46	.50	41	.36
66	.32	47	.54	42	.38
67	.33	48	.58	43	.41
68	.35	49	.61	44	.43
69	.36	50	.64	45	.45
70	.38	51	.66	46	.47
71	.41	52	.68	47	.49
72	.43	53	.71	48	.52
73	.45	54	.74	49	.55
74	.47	55	.76	50	.57
75	.51	56	.77	51	.60
76	.54	57	.78	52	.64
77	.56	58	.80	53	.68
78	.57	59	.82	54	.71
79	.60	60	.85	55	.73
80	.64	61	.88	56	.75
81	.69	62	.90	57	.78
82	.72	63	.91	58	.80
83	.75	64	.93	59	.82
84	.79	66	.94	59	.82
85	.83	67	.96	60	.84
86	.86	68	.98	61	.86
87	.89	69	.98	62	.88
88	.91	70	.99	63	.90
89	.93	72	.99	64	.93
90	.95			65	.95
91	.96			66	.96
92	.98			68	.98
93	.99			70	.99
95	.99				