DOCUMENT RESUME

ED 342 684 SE 052 917

AUTHOR	Becker, Henry J.
TITLE	Effects of Computer Use on Mathematics Achievement. Findings from a Nationwide Field Experiment in Grades Five to Eight. Classes: Rationale, Study Design, and Aggregate Effect Sizes. Report No. 51.
INSTITUTION	Center for Research on Elementary and Middle Schools, Baltimore, MD.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Nov 90
CONTRACT	OERI-G-90006
NOTE	53p.
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Computer Assisted Instruction; Computers; Computer Software; *Computer Uses in Education; Elementary Education; Mathematical Applications; *Mathematics Achievement; Mathematics Education; *Mathematics Instruction; *Mathematics Skills; *Teaching Methods
IDENTIFIERS	Mathematics Education Research

ABSTRACT

Since the mid-1980's, computers in elementary and middle grade schools have functioned to a large extent as a medium for student practice in the skills and concepts of basic mathematics and logic. This report presents the results of a 2-year nationwide field experiment designed to provide credible evidence about the effects of using computers in mathematics instruction in grades five through eight. Ninety-six classes (48 pairs of "computer" and "traditional" classes) taught by 56 teachers in 31 schools from 25 districts in 16 states participated in the first year of the study; 11 teachers from 9 schools employed the same "teacher-control" design through the second year of the experiment. In 73% of the pairs of classes, the same teacher taught both the "traditional" and "computer" classes. Twenty-three of the 31 schools used Apple II series software, 9 used IBM software, 1 used Texas Instruments cartridge software, and 1 school used networked Radio Shack TRS-80 hardware and software. Most classes used the computers for drill-and-practice programs, and several used problem-solving tasks built into programs. On the average, students spent 36 hours during the school year on computer-based mathematics activities. Results from analysis of pre- and post-test results indicate that the overall effect sizes found at the end of the first year on five measures of mathematics achievement were not substantially above zero for the study populations a whole, except for the estimations subtest. Selected posttests are appended. (4 tables/figures) (MDH)

5F

ED342684

Effects of Computer Use on Mathematics Achievement

Findings from a Nationwide Field Experiment in Grade Five to Eight Classes: Rationale, Study Design, and Aggregate Effect Sizes

Henry J. Becker

Report No. 51

November 1990

Grant No. OERI-G-90006

Published by the Center for Research on Elementary and Middle Schools, supported as a national research and development center by funds from the Office of Educational Research and Improvement, U.S. Department of Education. The opinions expressed in this publication do not necessarily reflect the position or policy of the OERI, and no official endorsement should be inferred.

Center for Research on Elementary and Middle Schools
The Johns Hopkins University
3505 North Charles Street
Baltimore, Maryland 21218

The Center

The mission of the Center for Research on Elementary and Middle Schools is to produce useful knowledge about how elementary and middle schools can foster growth in students' learning and development, to develop and evaluate practical methods for improving the effectiveness of elementary and middle schools based on existing and new research findings, and to develop and evaluate specific strategies to help schools implement effective research-based school and classroom practices.

The Center conducts its research in three program areas: (1) Elementary Schools, (2) Middle Schools, and (3) School Improvement.

The Elementary School Program

This program works from a strong existing research base to develop, evaluate, and disseminate effective elementary school and classroom practices; synthesizes current knowledge; and analyzes survey and descriptive data to expand the knowledge base in effective elementary education.

The Middle School Program

This program's research links current knowledge about early adolescence as a stage of human development to school organization and classroom policies and practices for effective middle schools. The major task is to establish a research base to identify specific problem areas and promising practices in middle schools that will contribute to effective policy decisions and the development of effective school and classroom practices.

School Improvement Program

This program focuses on improving the organizational performance of schools in adopting and adapting innovations and developing school capacity for change.

This report describes a nationwide field experiment conducted on the use of computers in mathematics classrooms in grades five through eight.

Abstract

This report presents the results of a two-year nationwide field experiment designed to provide credible evidence about the effects of using computers in math instruction in grades five through eight. Ninety-six classes (48 pairs of "computer" and "traditional" classes) taught by 56 teachers in 31 schools from 25 districts in 16 states participated in the first year of the study; eleven teachers from nine schools employed the same "teacher-control" design through the second year of the experiment. The overall effect sizes found at the end of the first year on five measures of math achievement, although generally above zero for the methodologically superior implementations, were not substantially above zero for the study population as a whole, except for the estimations subtest.

Since the mid-1980's, computers in elementary and middle grade schools have functioned to a large extent as a medium for student practice in the skills and concepts of basic mathematics and logic. Nearly one-half of American elementary and middle-grades students use computers in their school mathematics activities (Martinez and Mead, 1988), and every week perhaps three to four *million* youngsters are engaged in answering math problems posed to them on computer screens. Yet, in spite of this widespread teaching practice, we really do not know (a) whether and (b) under what conditions students have learned more by practicing math and logic skills with the computer programs that their teachers provided than if their math lessons had been totally without computer use or had they been exposed to using computers in mathematics in very different ways.

Although it is true that there is a growing research literature about the use of computers in mathematics, I would maintain that that literature is inadequate to questions about the effectiveness of typical practice. In "computers in basic mathematics," as in other areas of computer-assisted-instruction, three types of research dominate:

* **Informal, formative assessments of learning, usually in conjunction with software development.** Studies accompanying development projects often incorporate very sensitive measures of learning and performance and are valuable for informing and improving the software and curriculum development that they accompany. However, the questions asked by this type of research tend to be of the nature, "Can computer-based activities enable students to gain certain competencies or improve their capacities?" rather than the question, "Do such activities typically result in those improvements?" These development projects often involve innovative ways of using computers and incorporate theoretically rich ideas about what students need to do to truly understand concepts and relationships. But they typically proceed under the assumption that students will <u>not</u> attain these competencies by traditional instruction, and therefore they do

not collect data comparing achievement by students experiencing the computer-based approach with learning by other students experiencing other teaching approaches, or by the same students during other time intervals. Moreover, the careful site selection and extentive monitoring that often occurs during development phases of instructional programs makes it unclear whether the program might be effective under more typical conditions.

* **Large-scale (sometimes "system-wide") evaluations of a recently implemented computer-based instructional program, usually involving basic skills drill-and-practice.** Most evaluations of system-wide computer-based interventions are planned <u>after</u> the implementation of the program. As a result, these studies are typically left in a weak position with respect to the inferences that they are trying to make. Some of these evaluations are made based only on changes in standardized test score percentiles "pre-" and "post-" treatment for the same children. Particularly useless are non-control-group studies that use fall and spring test score comparisons because there is substantial evidence of inadequate norming by test publishers resulting in typical fall-to-spring gains of 8 NCE points regardless of the students' instructional experiences (Gabriel, et al, 1985). Other system-wide evaluations examine differences in achievement between successive grade-cohorts (the "pre-implementation" cohort and the "post-implementation" cohort) or between schools participating in the program and "comparable" schools not participating. These designs rule out some threats to inappropriate conclusions but most of these studies **inadequately measure other factors** that might also account for differences in student outcomes between the schools or school-years in which the computer approach is used versus those where it is not used--differences in teacher quality, clientele, socio-emotional climate, test preparation activities, and curricular program. Selective publicity to the most favorable statistics, grade levels, and implementations also results from our reliance on vendor-reported data as well as differential reporting of the more favorable vs. the less favorable district-produced evaluations.

* The third type of research data common in the computer-based-instruction literature consists of **small-scale comparison studies, often involving only a handful of teachers in one school, often encompassing just several days or weeks of "treatment," and more often than not lacking in strict experimental control.** Like the formative studies and the system-wide post-hoc evaluations, many small-scale comparison studies do not employ random assignment to treatments so that teachers using the computer-based approach may not only be different individuals than the teachers of comparison classes but possibly different in their teaching effectiveness. Students enrolled in experimental and control classes may be assigned non-randomly and be different in relevant ways (e.g., starting competence). And sometimes (and this is often difficult to know from the sketchy information provided in published reports) experimental and control classes spend unequal amounts of time on the subject-matter tested (i.e., the experiment is an "add-on" activity). In spite of their frequent design deficiencies, these studies purport to draw conclusions about differences in student learning outcomes that <u>result from</u> the computer-based approach studied.

The value of these small-scale studies depends largely on the strength of their internal validity, but even the best designed studies have provided evidence that computer-based interventions can work, but for two major reasons they have not told us **what** works **when** and whether these approaches **generally** do work.

* **Generalizability.** Classroom environments in which computer software is used differ sharply from one to another. Far too often people improperly infer that results of a study done with a few teachers in a narrow range of circumstances are predictive of the kinds of results that will occur under different and possibly less favorable circumstances. Teachers differ in how well they can implement new approaches to their subject-matter. Similarly, classrooms vary in the amount of support and assistance that students require for attending to learning tasks and actually

learning. And the computer experiences themselves vary from one setting to another, even when the same type of computer software is in use and certainly when different software is used. When well-designed but small-scale studies obtain different conclusions about the effectiveness of the computer-based approach that they studied--as they almost always do--these variations are partly due to the random variation that occurs among small samples of students and teachers and partly due to the systematic variation that exists between one circumstance and another. Even the best-designed small studies suffer from both an inability to identify with confidence outcomes that have practical significance (because results that are important for policy purposes may still not be statistically significant) and an inability to identify the range of circumstances within which benefits are likely to occur.

* Germaneness. Many of these studies are reports of short-term interventions. Studies of short duration may indicate the value of using one program for teaching one concept or skill, but do not address the question of whether current school investments in using computers more broadly for instruction in one or more academic subjects are paying off. It may be that, in a particular subject domain, only topic-specific, occasional uses of computers are worthwhile. Although it is important to know which specific uses actually work and are better than existing alternatives, schools also need to know about the value of longer term, more integral uses of computer software which are more likely to justify and support major investments of computer hardware than the occasional, specialized use.

Many mathematics educators have argued that, in elementary and middle-grade mathematics, the germaneness issue is not answered merely by studying long-term (e.g., full year) interventions rather than isolated single-topic units. They argue that the mathematics curriculum itself is hardly germane to the most important numerical and symbol manipulation competencies that adults should have--for example, that instruction is overburdened with teaching manual algorithmic skills for which students now have other tools (e.g., hand calculators). However, regardless of the merits of those curriculum arguments, school administrators who are answerable to many interests and publics that do not accept such arguments need to know whether their use of computers to support instruction in their existing mathematics curriculum is a more effective a way of using their computer resources than other ways that students and teachers might use school computers. Therefore, it is important that researchers address the general question of "what effects are 'typical' computer-based instructional programs in mathematics having on students and under what circumstances," even when answering that question does not directly tell us about the potential value of using computers in other ways that might support other curricular goals.

In any event, to provide valid conclusions about the consequences of using computer approaches for any given curricular goal, it is essential that such research have high internal validity, incorporating comparisons with alternative instructional approaches, control over "teacher effects," and randomized designs. And to understand the limits of applicability of those conclusions, it is important that the research be conducted over a large and representative domain of settings.

Meta-Analyses. One approach to providing data about the mean and range of effects of using computers has been to undertake secondary re-analyses of existing small-scale research. Often these studies are in the form of a meta-analysis, imposing a common statistical metric on diversely reported research and being as inclusive as possible in an effort to narrow the confidence limits of the estimated mean effect. The most recent meta-analysis of computer-based instructional research, the first to really focus on microcomputer-based interventions, was done by Roblyer, Castine, and King (1988). These researchers examined roughly 200 studies published or reported between 1980 and 1988 and included 82 of those studies, about one-half dissertations and one-half published articles, in their meta-analysis. (Studies were excluded that failed to meet certain minimal requirements such as the presence of a comparison group and sufficient statistical data to be able to report effects in a common effect-size metric.)

Many of the 82 studies included in the meta-analysis have limited value for understanding the effects of computer-based approaches to instruction. Studies of a single pair of classes were particularly common, and others were studies of brief durations and limited computer treatment. One-third of the studies in this meta-analysis were based on fewer than 20 students per treatment (if treatments were taught by a single teacher) or fewer than 35 students per treatment (if experimental and control teachers were different). About 20% of the studies involved a treatment period of fewer than 8 weeks and fewer than 10 hours total instructional time.

Moreover, the great diversity of subject-matter, grade levels, and types of software included in this aggregation of studies limits the utility of a concept of "mean effect size." Just as one can study too narrow a range of implementations to be able to make valid generalizations across different school settings, one can mix too much diversity into a single pot. For the purpose of providing policy-relevant data for schools and school districts that are considering alternative ways to use computers, it seems wrong to combine studies of math C.A.I. in elementary school, writing actitivies in high school, and logic and problem-solving puzzles in the middle grades.

In Robylyer, Castine, and King, the number of longer-term studies of at least several classes of students in any one age range, for any one subject-matter, and using any one type of software is quite small. For example, 26 studies covered students mainly in grades 5 through 8. Of those, 17 passed the minimal size and duration criteria discussed in the previous paragraph. But among those 17 studies were 8 studies of mathematics using C.A.I., 4 studies of problem-solving using Logo or various logic practice puzzles, 6 studies of reading C.A.I., and 5 studies of language arts C.A.I.

Requirements for internal validity further reduce the number of studies of any one subject-matter substantially. Of the eight studies of middle grade mathematics, for example, only one (a small study of two classes per treatment) involved random assignment of students to computer and non-computer classes, only that study and one other used a design where the same teacher taught both the "computer" and "non-computer" classes, and only two other studies may have partially compensated for non-random assignment and teacher differences between treatments by being very large (e.g., 300 students per treatment or more).

And each of these studies was conducted under different conditions, collecting different data about student achievement and implementation characteristics, by different researchers.

Field Experiment at a Distance. Another approach to developing an understanding about the effectiveness of typical current practices of using computer-based instruction under a modestly limited but still representative set of conditions is to conduct simultaneous experiments in a great many schools, employing a high-quality research design, and collecting identical data from site to site on many of the conditional variables that might affect the experiment's outcomes. Of course, research designs involving randomized assignment are difficult to implement because many other conflicting considerations affect the willingness and ability of schools to adapt to researchers' needs and preferences. For example, many schools with existing computer-based programs feel obliged to give all students access to their program. However, not every school finds the research requirements for experimental designs impossible to meet. Many teachers and administrators, for one reason or another, are willing to make necessary adjustments in their prior scheduling and assignment practices to allow high-quality research designs to occur. The trick is to identify those research-amenable schools and develop a model for conducting research that permits their involvement, whereever they might be located.

For this research project, our goal was to recruit from the tens of thousands of schools and hundreds of thousands of classrooms around the country about 50 pairs of classes to participate in the first of several planned "National Field Studies of Instructional Uses of School Computers." This study was to focus on mathematics instruction in grades 5 through 8 in both elementary and middle/junior-high school environments. Participants had to meet several requirements: sufficient computer hardware so that a computer-using mathematics class could have regular access to at least one computer for every four students in the class; sufficient computer software for students to receive a substantial computer "treatment" during the year (e.g., 30 hours of computer time per child); computer knowledgeable teachers with two years of computer-use experience; an alternative, "traditional" instructional approach that would have to be applied in one-half of the participating classes and would have to address the same curriculum; and some form of randomized or matched assignment of students to participating pairs of classes to insure equal ability classes--classes that would, in turn, be randomly assigned to the "computer" or "traditional" treatment. Where teachers were responsible for several sections of same grade-same level mathematics, each teacher could act as his or her own control; where teachers taught self-contained classes, two participating (and thus computer-knowledgeable) teachers would be randomly assigned to a treatment. Students in the "traditional" classes were free to use computers for activities outside of mathematics. But in mathematics, the distinction between the computer and traditional approaches was to be maintained from September, 1987 to May, 1988. The selection of computer software and the fit between software and the curriculum was left up to each school. This was to be a study of the effects of computer use as actually practiced across the United States under reasonably rich but not atypical circumstances during the 1987-88 school year.

The researcher's role--accomplished largely by telephone and mail communication--was to obtain the involvement of the best combination of schools and teachers to study the question; mandate the experimental design, assuring the appropriate mixture of forced equivalence and randomization in student and teacher assignments; provide appropriate pre- and posttests; and collect data from teachers and students about their background and about the details of their mathematics class experience during the year.

We recruited schools primarily through announcements in education publications and direct mail requests sent to research directors, other administrators in school systems, a representative national sample of school-level administrators, and 40 publishers of educational software. Sixty-eight school districts or individual schools indicated their willingness to participate and ability to conform to the research design, and 27 of those "applications" were accepted. Altogether 96 classes (48 pairs of "computer" and "traditional" classes), taught by 56 teachers in 31 schools from 25 districts in 16 states participated in the first year of the study. Each school had between 2 and 8 participating classes. The classes were distributed across grade five (26 classes), grade six (30), grade seven (24), and grade eight (16). Two pairs were in parochial schools; the remainder, public schools.

The schools in the study include a nationwide span, but a slight majority were located on the East Coast. One was in New England. Three were in the New York metropolitan area (2 in Brooklyn). Five were in Pennsylvania or Maryland (two of those in Metropolitan areas). Two were in Greenville, N.C., a relatively poor community of under 50,000. Five schools were located in Florida, spread out among the Tampa area, the less wealthy northern part of the state, and Palm Beach. Two schools were in Kentucky (one of those on a military installation), and one was in Texas, near Dallas. Seven schools were in the Midwest, primarily in small communities or suburban areas and in the Iowa cities of Des Moines and Davenport. The Pueblo, Colorado area had two participating schools, one school was near Portland, Oregon, and two schools from California were in the study, one in a small town north of Santa Barbara and one in suburban Los Angeles.

After the conclusion of the first year, we encouraged schools that could employ the "same-teacher-control" design to participate in a second year of the experiment, and eleven teachers from nine schools agreed to do this. The second year of data includes five 5th grade, four 6th grade, and one 7th grade pairs. Data from the other 7th grade pair of classes was not analyzed because of technical problems.

In selecting schools to participate from among those recruited, we had two basic criteria: (1) that the conditions of the study seemed like they could be fulfilled (i.e., that the experimental design could be implemented, that the teachers had computer experience, and that there was sufficient "good-quality" hardware and software available and access time to the equipment); and (2) that the overall study population had a representative balance in terms of geographic location, socio-economic context, grade level, and types of hardware and software in use. In addition, consideration was given to situations that could provide the strongest design (same-teacher control rather than randomly assigning teachers to treatments) and to proximity to the researcher's location.

With only a three-month recruitment period and a goal of having roughly 50 pairs of classes across four grade levels, it was not possible to produce a set of sites that was satisfactory on all accounts. In particular, we were dissatisfied with the small number of large-city school districts we could incorporate (one) and with the relatively few schools and classes with many black students. Blacks numbered under 10% of the enrollment in 22 of the 31 schools. At five of the schools, blacks were at least 25% of the enrollment; but only one had a majority of black students. Other minority groups were better represented. Hispanics were a majority in two schools, and Asian-Americans constituted one-third of the population at another school. Altogether, minority groups were 20% or more of the enrollment at one-third of the schools in the study.

Socio-economic representativeness was reasonably satisfactory. At only five schools was enrollment largely upper-middle class. At those schools, the principal reported that nearly all parents were professionals or white-collar workers with a majority of families earning more than $30,000 per year. But the other 26 schools constituted a more typical blue- and white-collar heterogeneous mixture. In none of the others were white-collar parents clearly in the majority and in all but one the estimated high-income proportion was under 50 percent. Two kinds of communities, though, were poorly represented in the study--only two schools were in very low-income areas with many families receiving financial assistance, and no schools were in rural areas enrolling mainly children from farming families.

The classes participating in the study constituted a broad mixture of student ability levels within each school's population. In year one, fifteen pairs were heterogeneous mixtures of their school's grade level population; 10 more represented the middle-range of students in the school, but not the upper- or lower-thirds; eight represented the school's "upper-third" (one including the middle as well); and fifteen represented mainly "below-average" or "bottom-third" students or a range including middle and bottom-thirds. (See Table 1.) Seventeen pairs were in elementary schools, 27 pairs in middle or junior-high schools; and four in K-8 schools. Sixth-grade classes were split among all three types.

The one other aspect of the study population that fell short of our goals was in teacher experience in using computers <u>for mathematics.</u> Although most of the teachers recruited had previously used computers in some way (generally to do word processing or teach computer literacy), only about half of them had used computers in mathematics teaching and only one-quarter had done so on a regular (more than weekly) basis. Thus, one important characteristic of the study's first year is that it was an examination of student achievement gains during <u>the implementation year</u> of an instructional program. One major purpose of the second year's data collection was to determine whether effects were different in a setting that would be regarded by participants as being more routine.

On the other hand, we exceeded our initial expectations in terms of recruiting schools able to establish the highest quality research designs (student-level randomization, same-teacher control) and having a very favorable computer-student ratio. The ideal design of randomized assignment of individual students to "computer" and "traditional" treatment classes was accomplished in about one-third (11) of the schools participating in the study. At all of the other sites, we came very close to approximating this ideal design. The most common design, for example, (at 14 schools) involved randomized assignment of classes to treatments, with classroom student composition being determined locally but based on a systematically balanced or random assignment procedure, as certified by local school personnel. At two schools, adjacently ranked homogeneously grouped classes were randomly assigned to treatments. At the remaining four schools, because of limitations on the availability of computer facilities, the class assigned to the computer treatment was fixed. At two of those, students were randomly assigned to classes by the researcher; at the other two, local procedures were claimed to produce equal ability classes.

In 73% of the pairs of classes in the study's first year, the same teacher taught both a "traditional" and a "computer" class categorized as "same ability" either by random assignment or by a local randomization or matching procedure. In most of the remaining pairs of classes, two eligible teachers ("computer knowledgeable") taught self-contained classes of same-ability students, one teacher (randomly selected) teaching mathematics using computers, the other teacher using traditional instructional media. In the (4) remaining pairs, teachers taught both "computer" and "traditional" classes but of different ability levels; formally, each teacher was paired twice with another teacher from the same school who taught two classes at the same level but using the opposite "treatment." In year two, all pairs involved the same teacher teaching matched-ability computer and traditional classes.

(The superiority of same-teacher controls over random assignments to treatment between teacher volunteers is suggested by the fact that same-teacher pairs had effect sizes whose standard deviations were much smaller than those for "different teacher" pairs. That is, pairs where teacher effects were added to treatment effects produced much larger differences in mean achievement between "computer" and "traditional" class. The standard deviation of effect sizes for same-teacher pairs was, on average, only 69% as large as for different-teacher pairs across the four achievement variables used.)

Except for one school, computer-using classes in the study had the intended minimum of one computer available for every four students in the class. Many of the classes had substantially more computer resources available. The median number of computers simultaneously accessible to computer-using teachers in the study's first year was 16. The median class-size of computer-using classes was 23. All but five classes had no worse than a 2 to 1 ratio of students to computers. During the second year, all classes had at least a 2 to 1 ratio or better.

All participating schools used microcomputers rather than mini-computer systems. Only 8 schools used computers on a local-area network, and only one school (2 pairs of classes each year) used an individualized, computer-managed, "integrated instructional system." At most sites, teachers or students had to load programs into each computer's disks individually. Twenty-three of the 31 schools used software that ran on Apple II series computers; nine used software for I.B.M. and compatible computers. (Two schools used both types of machines.) One school used Texas Instruments cartridge software along with its Apples. And one school used networked Radio Shack TRS-80 hardware and software.

The mathematics that was taught using computers and the emphasis on computation, concepts, applications, and general problem-solving skills varied from grade to grade, teacher to teacher, and site to site. Across the two years of the study, the most commonly used software products were the two series from Milliken Publishing Co. (Milliken Math and WordMath, together constituting approximately 19% of all software used), MECC software (13%), IBM Math Concepts and Math Practice series (12%), Sunburst problem-solving software (7%), and the managed network-based system from E.S.C. (now Jostens Learning Corp., 6%). Other software that was used by three or more classes for a substantial portion of their computer time included the Mathematics Curriculum Project software from the University of Northern Iowa (U.N.I.), Scholastic Software ("Math Shop"), S.R.A. Math, Softwriters' Skills Bank, and Educational Activities software. The single diskette most widely used across the classes in the study was MECC's "Number Munchers." Although most classes used primarily drill-and-practice programs, programs such as U.N.I.'s Math Curriculum Project and the Jostens E.S.C. program had substantial tutorial (explanation and demonstration) functions. Problem-solving tasks built into programs such as Sunburst's Factory, Survival Math, and King's Rule; Scholastic's Math Shop; and the Math Activities Courseware series from Houghton-Mifflin were each part of the computer experience at several sites.

Although nearly every "computer" class had students use computers at least once each week, variations among sites in the amount of time students used computers over the course of the school year were substantial. Some classes had only a single 30 minute period per week on the computer. One class, at the other extreme, used computers for an hour nearly every day during the school year. On the average, students spent 36 hours during the school year on computer-based mathematics activity. Figure 1 presents a cumulative frequency distribution of computer time for individual students across the 48 computer-using classes in the first year of the study.

Since students spend roughly fifty minutes per day on mathematics, the computer-specific activity on the average amounted to only 25% to 30% of their mathematics experience. Still, the amount of computer time each student in these classes had is substantially more than students in most computer-using middle-grade mathematics classes in the country have. (Unpublished data from a national 1989 survey conducted by the author indicate that a student in a typical computer-using secondary mathematics class uses computers for about 30 minutes per week, or under 20 hours per year.)

Students in paired traditional classes spent the same amount of time learning mathematics as students in the computer classes. In some of the pairs, traditional classes used the time that their paired computer class had for computers on activities and lessons that the computer class did not have, such as group projects or problem-solving tasks. In other pairs, the traditional classes just spent more time on the same assignments (drills, lessons, seatwork) as the computer class. Overall, in exchange for their computer time, the computer classes spent an estimated 40% less time on small group activities (which was, however, a small proportion in either "computer" or "traditional" classes), 33% less time on class drill, and 25% less time on whole class lessons and seatwork than did the traditional classes.

The data collected during the first year included national standardized mathematics achievement test data, both pre- and posttests; daily logs (20% sample of days) from teachers in both treatment groups concerning the class activities; the set of homework and classwork assignments and quizzes and tests given to students throughout the year (a copy of each class' textbook was also provided); questionnaire data from teachers and students at the beginning and end of the year; a brief experimenter-made brief posttest of estimation skills and mental mathematics; and 47 experimenter-made, curriculum-specific posttests, each one based on the instructional content in one particular pair of classes, developed from the assignment and test materials supplied by the teachers throughout the year and the computer programs used by those students. During the second year, the same pre- and posttests and the spring student and teacher questionnaires were fielded as in the first year, but the teachers were asked to give less detailed weekly reports about computer use patterns.

During the first year, five schools (13 pairs of classes) were visited, but because the research sites were spread widely throughout the country, we relied on the periodic self-reports described in the previous paragraph to provide systematic data about how each site accomplished instruction in computer and traditional classes during the year. And because no researcher was on site, it was necessary to rely on teacher and student written feedback to validate that the experiment was actively implemented.

The first year's weekly reports provided information needed to create curriculum-specific posttests for each pair of classes. In addition, the weekly reports and the spring questionnaires of students and teachers enabled us to code curricular and organizational properties of the computer and traditional instruction classes in the absence of on-site observation by a research team. And this data allowed us to examine hypotheses about different aspects of instructional practice that might help account for differences between sites in the relative effectiveness of computer-based approaches. About 90% of the 56 teachers participating during the first year sent back two reports each week for a 22-week period from November through April.

Between the teachers' weekly reports and the spring questionnaires completed by students--providing their own estimates of how often they used computers for different subjects during the school year--we determined that during the first year of the study, a number of sites were not successful in implementing a substantial computer experience and in maintaining a sharp distinction between computer and traditional treatments. Eight teachers' computer class students reported minimal computer experience during the year--15 hours or fewer, by our estimate. One of those classes had fewer computers (6) than any other class in the study; in another case, an 8'th grade math teacher only used a few pieces of software that he felt complemented what he was teaching. And in two of the others, cooperation by a pair of experienced math teachers was quite grudging in that their participation was imposed on them by their principal.

In eight other instances, at least one-quarter of the students in the traditional treatment classes reported having used computers for math on more than five occasions. However, further inquiry indicated that most of those occurrences involved computer use <u>after the posttests</u> when the teachers were free to break down the distinction between treatments. And in all of those cases where there was some contamination, the computer class students reported substantially more computer experience during the year than did the traditional class students. Altogether, the combination of relatively little computer use (15 or under hours during the year) and indication of computer use by the traditional math class by many of those same classes led us to drop eight pairs of first year classes from the data analysis. In the classes studied during the second year, there was substantial computer use in all classes and less evidence of contamination. So all 10 pairs from year two were used in the data analysis, bringing the total number of pairs of classes studied to an even 50 (48 first year class-pairs minus 8 dropped plus 10 second-year class-pairs).

It should also be noted that adding the omitted eight pairs back into the analysis results in absolutely no change in the mean value of effect statistics calculated for the main contrast between experimental and control treatments. All five major "effect size" measurements (see below) are changed by a maximum of 0.01 units (posttest standard deviations) when the eight omitted pairs are included. Moreover, the effect sizes measured on the pairs of classes omitted from the analysis are more clustered around the zero point than were the effect sizes for the other pairs, suggesting that in fact there was less distinction between "experimental" and "control" classes for those pairs. (And the 20 pairs evidencing the most faithful implementation of the design--more than 30 hours of use; few reports of treatment contamination--had effect sizes that least clustered around zero.)

Pretests and Posttests. Standardized tests. Although a common set of posttest measures was used at all sites during both years of the study, in the first year of the study, different pretests were used in different sites. For 54 of the 96 classes (in 19 of the 31 schools), students took the Stanford Achievement Test (math computation and math applications parts) in September, and the tests were scored by the researcher's staff. The remaining schools supplied the project with other pretest data--tests taken during the previous Spring or the Fall of the study year--using a variety of standardized tests (CTBS (3 schools); CAT (2); Iowa (1); SRA (3); Metropolitan (2); and Stanford (2)). Fall pretest scores were obtained from 4 schools, while scores from the previous Spring were used for students at 9 schools. In year two, all sites used the researcher-scored Stanford tests. Posttests each year included the math computation and applications sections of the Stanford. Math concepts were tested through the curriculum-specific test prepared for each pair (see below).

The fact that different schools supplied different pretest data does not affect our ability to assess achievement gains made in computer-treatment classes compared to traditional-treatment classes. Each class pair's effect size is calculated based on their particular differential pretest-controlled posttest scores and therefore is independently measured from effect sizes in other schools. But it does impinge in two ways to limit analysis. First, one cannot do careful studies of absolute achievement gain except in schools using the Stanford pretest. For example, only in those schools can we examine whether the effectiveness of the computer-based program is higher or lower for teachers whose traditional class gained more or less than the average teacher's. (That is, do the computer programs in use help the "better" teachers or the "not so successful" ones.) Also, absence of uniform pretest metrics limits our ability to assess the value of the computer-treatments for categories of students grouped according to previous academic performance--the "high achieving," "average achieving," or "lower achieving" students. Since we have no easily obtained common standard on which to compare the prior academic achievements of students in different schools, these categorizations must be treated as roughly made divisions.

Curriculum-specific test. Besides the Stanford math computation and math applications posttests, three researcher-constructed posttests were used: a curriculum-specific test, a test of fluency in mental mathematics, and a test of estimation skills, the latter two combined in a single orally administered test. Each curriculum-specific test was produced through an informal domain sampling procedure that included conceptual, computational, and problem-solving tasks contained in the teachers' textbook assignments, worksheets, tests, and computer program assignments. The produced test attempted to include a balance of problems given to each class in a pair but not the other, but it de-emphasized those skills already covered on the standardized achievement posttests. Thus the tests focused as much as possible on concepts and on applying math in real and complex situations, consistent with the need to cover only what that teacher actually taught and to balance the experience of the computer class and the traditional class in that pair.

This test was not multiple choice--instead students were asked to supply their own answer, and to use the supplied test paper to do their calculations. Many questions had several parts or involved students supplying several answers (e.g., circling all fractions among a set of 16 that were greater than one-half). Related and multiple-decision questions were combined so that the test was scored as a set of between 15 and 22 "items" per test. Each item was allocated a number of points (most often "2" or "3"; sometimes "1" or "4") based on an assessment of its complexity. Rules for partial credit were established for answers to questions with multiple parts, steps, or decision-points. The researcher scored all tests.

The content of tests varied substantially from pair to pair. There were variations in attention to higher-order concepts and complex applications and in the degree to which computer programs formed the source of test items. (Appendix A contains a sample of several of the curriculum-specific tests.) On average, a test was composed of about 9 items from the computer programs used by the computer class in the pair, about 5 items from the traditional class' special activities (where they did have tasks that the computer class did not) and the remaining 4 items from tests or worksheets used by both classes in a pair. The imbalance between computer and traditional class sources is partly due to the fact that teachers of many pairs did not report any assignments given only to the traditional class in the pair. In other cases, the content of all traditional-class-only assignments was already tested on the Stanford Achievement test. Also, in some cases--where teachers emphasized mathematics computations or simple single-step word problems in both their traditional and computer classes or where the teacher did not consistently provide weekly data about assignments--we composed part of test with problem-solving tasks or concept items from other pairs.

Teachers reported an "opportunity to learn" variable for each test item--that is, they indicated for each test item which of their classes (computer, traditional, both, or neither) had been presented with instruction for which that test item was appropriate. Excluding seven pairs whose teachers did not provide opportunity-to-learn data for both classes, the teachers indicated that the content of 80% of the items had been covered in their classes. On average, the opportunity-to-learn score was 5 percentage points higher for computer classes than for traditional classes, indicating a slight bias in test content favoring the computer treatment. For part of the analysis, differential exposure to the test content (between computer and traditional class of the same pair) and differential opportunity-to-learn were taken into account in analyzing observed effect sizes.

During year two, instead of freshly deriving a test based on the new year's instruction and materials, the same curriculum-specific posttest was given to the teacher's classes as was given in the first year. Opportunity-to-learn measures for the second year were comparable to those during the first year, averaging 83% for the 8 teachers responding, but with the mean for traditional classes still below that for computer classes.

The curriculum-specific tests were expected to be difficult--and they were. Even with partial credit scoring, students in classes studied during the first year averaged only 36 percent on this test. Of course, the tests varied substantially in how well students could answer the questions: students in one-fourth of the pairs scored lower than 25%; students in 5 pairs scored over 50%. Overall, the low scores provide evidence that mathematical problem-solving requiring fluency in dealing with numbers and logical relationships is not successfully taught in most school classrooms. However, a discussion of the substantive mathematics education issues revealed by performance on these tests is reserved for a future paper. In terms of the tests' statistical properties, the item reliability of most tests was satisfactory or better. All but six tests had alpha reliabilities above .60. The mean reliability was .71 and the test with the highest reliability had an alpha value of .86.

Mental Math/Estimation tests. Two distinct rationales lay behind the other administered posttest. First, several of the computer programs used in many of these classes focused on rapid solution to basic math facts. Many of these were presented in an arcade-game format. It seemed appropriate to test students' ability to do rapid mental arithmetic, for example by presenting each problem for a fixed limited number of seconds. Secondly, competence in producing round-number estimates of answers, although not as much a part of the standard mathematics curriculum as mathematics educators recommend, also seemed appropriately measured by an orally administered test. These two goals were combined in a common test given to all pairs. Teachers presented each problem visually on a "flip chart," each one for 10 seconds plus 5 seconds between presentations. The test contained seven mental mathematics items and 13 estimation tasks (and a practice problem for each part). Two versions of the mental math portion of the test were used--one for grades 5 and 6 and the other for grades 7 and 8. The same estimation items were used for all grade levels. Appendix B contains the items in the Grade 7-8 version of this test.

The orally administered test was also a "fill-in," and in the estimation part, "double" credit was given for optimal estimates. Scores on the mental math subtest averaged about 30% and, with the double-credit scoring, the average estimation score was roughly 55%. Mental math and estimation subscales correlated .4 with each other on the individual level and among classes. But because mental computation and estimation are different skills, we treat them as separate dependent variables along with the Stanford computation test, the Stanford applications test, and the curriculum-specific test.

Correlations among the five posttest raw scores are substantial, although the unreliability of the short mental math test produces attenuated statistics. Table 2 shows the mean student-level correlations among the 57 pairs of all posttest-posttest correlations. It also shows, in row one, the mean pretest-to-posttest intercorrelations, using as a pretest variable the simple sum of all pretests for that pair. All correlations among the pretest total, both Stanford posttests, and the curriculum-specific posttest average in the range of .49 to .61. Correlations with the mental math and estimation tests are all in the range .28 to .41.

Pretest Match, and Calculations of Achievement Gains and Effect Sizes. Randomized assignment of between 30 and 60 students to any one pair of classes does not assure that the paired classes are in fact equal in ability, not even when students are initially stratified by test scores, as were students in many of the schools in this study. Furthermore, the tests used to make assignments of students to classes were generally not the same tests as used for the pretest. So variations between "traditional" and "computer" class pretest means even for classes of randomly assigned students would not be unexpected. Moreover, only a minority of sites actually permitted researcher-accomplished randomization. The remainder used local manual or district-driven computer-based assignment procedures to produce "equal ability" classes. In fact, quite a few pairs of classes showed pretest differences. Overall, among the 58 pairs over both years, 24 pairs had pretest mean differences of greater than one-quarter of a standard deviation; 11 of those exceeded one-half of a standard deviation. Moreover, the 24 class pairs randomized by the researcher were somewhat less likely to show large pretest differences (> |.25| s.d.) than were the remaining pairs (38% vs. 45%). In addition, a greater number of large pretest differences favored the traditional class (i.e., indicated higher achievement levels there) than favored the computer class (15 vs. 9). So, for all of these reasons, it seemed particularly important to take pretest differences into account in computing effect sizes.

Consequently, the performance gains accomplished by each student during the school year studied were measured by computing posttest raw scores for that student net of their own pretest-indicated performance level. Separate regression equations were calculated for each of the five posttest measurements used--Stanford computation, Stanford applications, curriculum-specific, mental math, and estimation--and separate regressions were computed for class pairs receiving different pretests or the same pretest but at different grade levels. In year one, for example, the largest pretest group was formed by 10 pairs of 7th grade students pretested in the Fall with the Advanced version of the Stanford pretest. Because there were so many combinations of pretests and grade levels in year one, distinct regression equations were calculated for 20 groups of classes. Most of those were calculated across students in only one pair of classes (e.g., two 5'th grade students taking the CTBS). In year two, since only Stanford pretests were used, groups were formed merely by grade level. In year one, all pretest subscales that existed for that particular pretest (e.g., computation, concepts, and/or applications) were used as separate predictors for each posttest outcome measure. In year two, each Stanford sub-test was regressed only on the corresponding Stanford pretest. Parameters of each equation were then applied to each student in the group's classes yielding a residual posttest score (actual posttest minus posttest score predicted from the regression equation).

Students who did not have at least one pretest sub-scale were not included in the analysis. Students who were added to the class during the first part of the school year (through November) were included if pretest scores were available for them. Students who added later, who changed between "traditional" and "computer" class sections during the year, who left the class prior to the posttest, or who were absent from any one posttest were excluded from calculations of the effect size for that posttest, although they were included in descriptive statistics about the class.

Altogether, pretest data were obtained for 2919 students (combining both years). Of those, at least one posttest was scored for 89%. Attrition for each specific posttest varied between 14% and 18%. (One teacher did not administer the mental math/estimation test and one teacher's curriculum-specific posttest was not scored because of clear evidence of test taking misbehavior.)

For each pair of classes an effect size was calculated for each posttest by computing the difference in mean residuals between the computer and traditional class and dividing that difference by the pooled raw posttest standard deviation for both classes.

Major Results: Effect Sizes for the Study Population. Table 3 gives the mean and standard deviations of the effect sizes observed for each posttest outcome for the full 58 pairs of classes (actually 57 data points because one teacher taught two pairs of classes which were combined for the analysis). The table also provides mean effects for the 50 pairs judged to have implemented the study design satisfactorily, for the 20 pairs judged to have implemented the design <u>most</u> faithfully, for the 9 pairs studied during their second year of using computers as part of the study (one second year teacher was a "traditional" teacher during year one), for the 24 pairs that incorporated researcher-controlled student-level randomization in their design, and for the 29 pairs whose pretest differences between the traditional and computer classes of the same pair were "minor" (under .25 s.d.).

For the study population as a whole, effect sizes for all five outcome variables were negligibly different from zero. For all 57 pairs, they ranged from -.02 to +.07. For the satisfactorily implementing pairs, they ranged from -.07 (mental math) to +.07 (estimation). However, for the 20 most faithful implementations, the effect size means were somewhat more positive (none was less than zero), ranging from +.03 to +.18, although only for the estimation outcome was ES > .10.

Teachers in the second year of the study had more success, in terms of effect sizes, than did the pool of teachers in their first year of the study. However, the teachers continuing for a second year were not a representative sample of first year participants. And when we compare effect sizes in their second year classes with those of their own first year classes, the results for the second year clearly show a decline in effect sizes--not an improvement. (See Table 3.)

When we look at the group of sites where student-level randomization was accomplished under the researcher's control, we see a more consistently positive set of effect sizes. But, except for the estimation posttest, for none of the others was ES > .10. And when we examine only those pairs with a close pretest match between traditional and computer classes, we see modest departures of effect sizes from those for the full study population, but in both positive and negative directions.

The final element in Table 3 reports the results of a multiple regression analysis of three of these methodological factors (all except "second year teachers") on student achievement effect sizes among the 50 pairs of classes that met our minimal criteria. The numbers shown are predicted effect sizes for a class-pair having the best of these implementation/design attributes: a faithful implementation (frequent computer use and no treatment confounding), researcher-controlled student-level randomization, and minor pretest differences. The predicted effect sizes for all achievement outcomes are above zero, and four of the five are near or above +.10. Still, only one predicted ES is above .20, which is a lower bound for what might be called a substantively important effect.

In summary, for the study population as a whole, even when we take into account that many sites had weak implementations or less than ideal study designs, the overall effect sizes, although generally above zero for the methodologically superior implementations, are not substantially above zero, except for the estimations subtest, the outcome variable with the highest standard error. We postpone until later in this paper a discussion of the implications of these results, but one thing is certain--we cannot conclude from these results that "computers are a waste of money." First, the sample, although probably more representative of the range of actual practice than can be found elsewhere, is still not a sophisticated national probability sample of teachers and classrooms. Second, 5th through 8th grade mathematics is only one curricular application of computers. And third, there are other potentially valuable ways to improve students' understanding of mathematics through computers besides the typically diskette-based drill-and-practice programs that constitute the most common approach employed with our study population. Still, on the average, for this population of teachers and students who used computers as they did, it seems as if--for the group considered as a whole--the use of computers did not make much difference for the students' performance on tests of mathematics skill and applications.

Differences in Effect Sizes: Are They Random? Although the mean effect size among these pairs of classes across five outcome variables was fairly close to zero, not all effect sizes were close to zero. One class pair had an effect size of over +1.00 for four of the five outcome variables, and another class pair had an effect size below -.60 for three of the five variables. A third pair's effect sizes were +.62 for Stanford computation, +.43 for the curriculum-specific test, and +.63 for the mental math test, but (negative) -.43 for Stanford applications. The standard deviation of effect sizes ranged from .35 for the Stanford subtests to .65 for the estimation test.

Effect sizes based on any one pair of classes are subject to a variety of situational effects independent of the actual effects of instructional experience. Moreover, even if one were to use randomly produced test scores, effect sizes computed on the basis of a single pair of classes have a non-negligible chance of being greater than one-quarter standard deviation. (Using a monte carlo simulation, I computed effect sizes for this study population based on random test scores, pretest controls, and pre- and posttests correlated between .4 and .6 and found the typical standard deviation of effect sizes to be about .27.) Thus, effect sizes for single class pairs between, say, -.3 and +.3 are hardly meaningful, and even those in the range of |.3| to |.5| are not statistically significant. However, because the standard deviation of effect sizes observed was larger than that likely to be obtained by chance, it is plausible that there are some systematic patterns of effects--that the variations are not merely due to random fluctuations. By combining class pairs that are similar on some characteristic (for example, grade level, type of software used, frequency of computer use, etc.) we can produce empirically based speculations about what factors make a difference in the effectiveness of the range of computer-based approaches to middle grade mathematics instruction employed by the schools in our study population.

Further analyses with this data will examine seven types of variations for clues concerning the differential effects on student achievement of typically employed computer-based approaches to middle grade mathematics. Those seven categories are (1) school and community environments, (2) student characteristics (including grade level), (3) teacher characteristics, (4) the social organization of computer use, (5) curriculum coverage, (6) computer software, and (7) computer hardware and hardware organization. For each of these seven categories, there are at least several variables that are plausibly linked to possible variations in the effectiveness of computer-based instructional programs in mathematics, over the domain of practice that was studied. For example, take student characteristics. Do the computer-based approaches that teachers are now using work better with the younger students in grade 5 or the older ones in grades 7 and 8? What about student ability levels--are the computer-based approaches now in use better suited for students behind grade level in math achievement or for their on-grade or above-grade peers? Or take teacher characteristics. It is plausible that teachers responsible for teaching several subjects to a self-contained class may profit by using computer-assisted instruction more than math specialists. On the other hand, perhaps math specialists make better use of computer-based tools in the subject that they know best. Again, we emphasize that the questions we address concern variations in effectiveness in terms of the range of practices actually studied. We cannot say what effects better-prepared teachers would produce or whether the effects of software more carefully developed to elicit mathematical understanding would be greater than the effects found for the software actually in use by the teachers in these 50 pairs of classes.

References

Gabriel, R.M., Anderson, B.L., Benson, G., Gordon, S., Hill, R., Pfannensteil, J, and Stonehill, R.M. (1985). Studying the sustained achievement of Chapter I students. Washington, D.C.: U.S. Department of Education.

Martinez, Michael E., and Mead, Nancy A. (1988). Computer competence: the first national assessment. Princeton, N.J.: Educational Testing Service (April).

Roblyer, M.D., Castine, W.H., and King, F.J. (1988). Assessing the impact of computer-based instruction. New York: Haworth Press.

TABLE 1

The 48 Pairs of Classes: Grade Levels and Ability Levels

Class' Ability Level (as reported by school)	Grade Level			
	5th	6th	7th	8th
Top 1/3 of their school's grade level	–	3	4	–
Top and Middle Thirds	1	–	–	–
Middle Third	–	5	4	1
Heterogeneous	11	4	–	–
Middle & Lower Thirds	1	1	–	2
Below Average	–	–	1	3
Lower Third	–	2	3	2

TABLE 2

Pre- and Post-Tests: Mean Individual-Level Correlations Among Class Pairs (N=57 class pairs)

	Stanford Computation	Stanford Applications	Curriculum-Specific Posttest	Mental Math	Estimation
Pretest (sum of scores)	.56	.61	.55	.37	.39
Stanford Computation		.54	.49	.38	.34
Stanford Applications			.53	.28	.37
Curriculum Specific				.36	.41
Mental Math					.39
Estimation					

TABLE 3

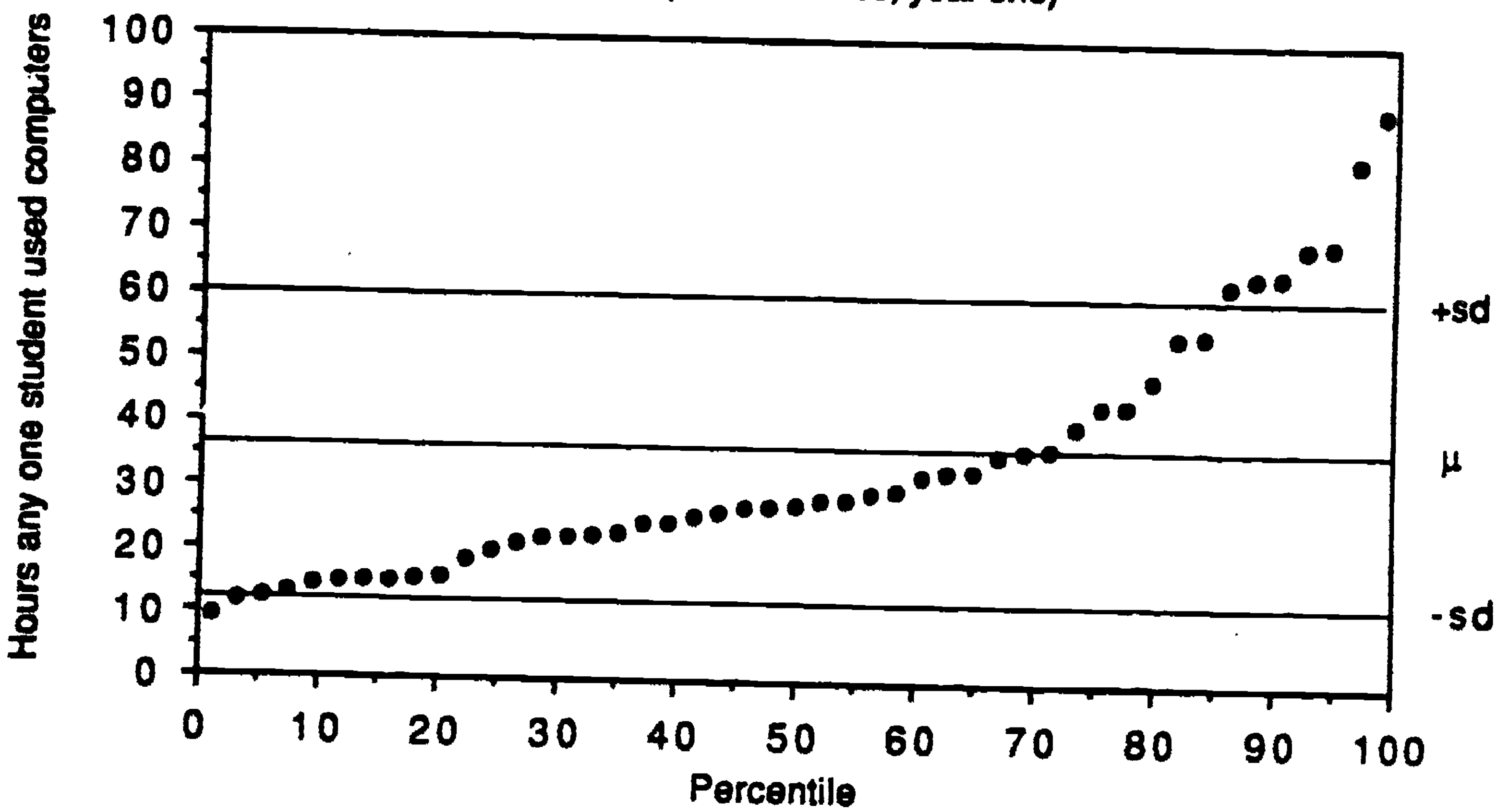
Aggregate Effect Sizes

	N*	Stanford Computation x̄	Stanford Computation s.d.	Stanford Applications x̄	Stanford Applications s.d.	Curriculum-Specific Posttest x̄	Curriculum-Specific Posttest s.d.	Mental Math x̄	Mental Math s.d.	Estimation x̄	Estimation s.d.
All Pairs	57	+.04	.35	+.04	.35	−.00	.43	−.02	.55	+.07	.65
Pairs Kept in Study	50	+.03	.36	+.04	.36	−.01	.46	−.07	.54	+.07	.68
Most Faithful Implementations	20	+.07	.40	+.03	.46	+.04	.56	+.06	.59	+.18	.77
Teachers in Second Year	9	+.11	.32	+.18	.42	−.02	.56	+.01	.32	+.16	.41
Same Teachers in their First Year	9	+.31	.34	+.13	.40	+.23	.36	+.11	.57	+.32	.90
Researcher-Randomized at Student Level	24	+.06	.35	+.08	.44	+.08	.41	+.10	.60	+.26	.75
Only minor pretest differences between traditional & computer classes	29	+.03	.38	+.11	.41	+.02	.46	−.17	.58	+.08	.77
Regression output: linear model predictor for the following condition: most-faithful implementation, researcher randomized, minor pretest difference between traditional and computer classes	df=45	+.09		+.17		+15		+.01		+.35	

*N's for some posttests are 1 fewer because of occasionally missing data.

Pair 40 → •

FIGURE 1. Hours that any one student used computers.
(N=48 classes, year one)



Note: Reported number is mean between (a) estimate from periodic teacher reports of computer use and (b) class mean of student retrospective report of weeks used, days per week, and minutes per turn.

APPENDIX A: Selected Curriculum-Specific Posttests

Johns Hopkins University National School Year Review

P3	Pollock	Class ID:	Name:

1 There are 98 pictures in Pam's photo album. There are 8 pictures on all of the pages except for one special page that has 10 pictures instead of 8. How many pages are in the album, including the special page?

ANSWER:

_____ pages

2 It's baseball season:

a A scorecard costs $1.25. Paula bought one and gave the seller a $5 bill. How much change should she get?

ANSWER:

$ ____ . ____

b The White Sox have played 25 games. The 9 starting players together have made 225 hits so far. What is the average number of hits that a White Sox starting player has had so far?

ANSWER:

_____ hits

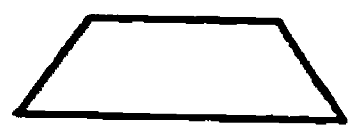
3 Complete the table below by showing the average scores.

Bowling Scores

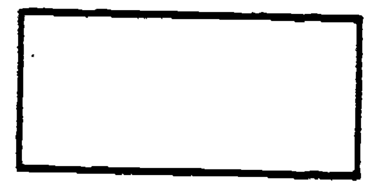
	June	Zach
Game 1	168	165
Game 2	142	145
Game 3	154	160
Average		

4 Draw all lines of symmetry for each figure:

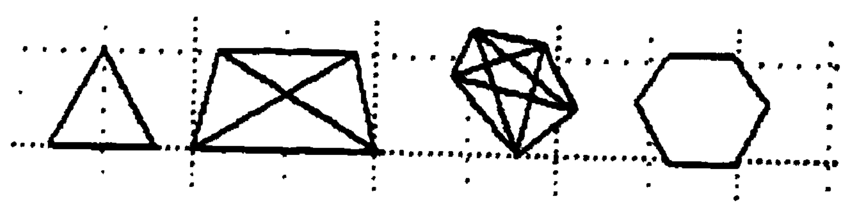
a



b



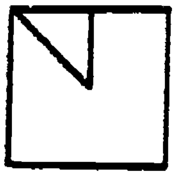
5 Look at these figures.



How many diagonals does a 6-sided figure have?

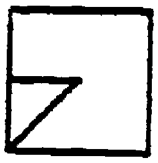
ANSWER:

6 The square below can be rotated around its center.

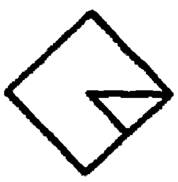


Question: Can it be rotated so that it looks like these squares below.

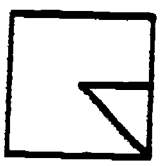
Circle Yes or No for each one.
If Yes, tell how many degrees it should be rotated (clockwise).

a 

Yes — if "Yes" How many
No degrees ____°

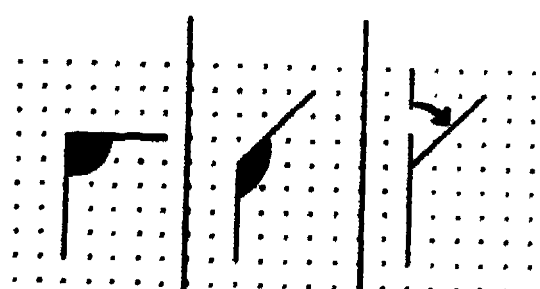
b 

Yes — if "Yes" How many
No degrees ____°

c 

Yes — if "Yes" How many
No degrees ____°

7 a How many degrees?

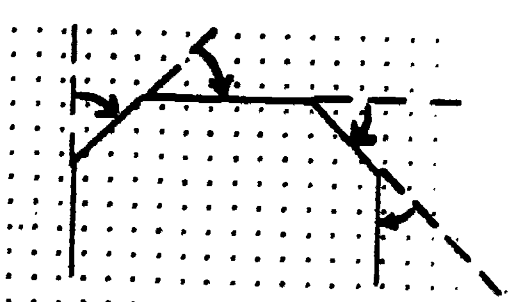


We'll call these inside degrees

and this, turning degrees

____° ____° ____°

b How many turning degrees in all?

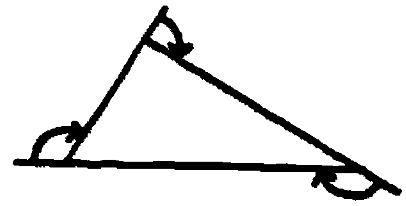


____°

c How many degrees on the inside of a triangle?

____°

d How many turning degrees in a triangle?



____°

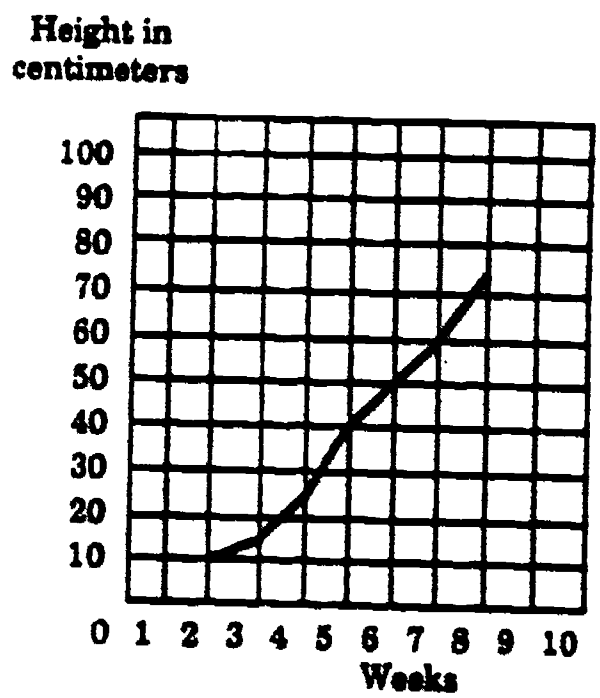
8 Round these decimals as specified

a 5.9610 to the nearest 10th ____

b 4.0019 to the nearest 100th ____

c 5.595 to the nearest 100th ____

9 Corn Plant



a During which weeks was there the most growth?

weeks ____ and ____

b Approximately how tall will the corn plant be after 9 weeks?

____ cm.

10 Complete these fractions so that they are less than, but as near to, the value of 1 as they can be.

a $\dfrac{\square}{6}$ b $\dfrac{9}{\square}$

11 What size fraction is halfway between these fractions

a halfway between $\frac{3}{5}$ and $\frac{4}{5}$: $\square$

b halfway between $\frac{1}{2}$ and $\frac{3}{4}$: $\square$

c halfway between $\frac{5}{8}$ and $\frac{11}{16}$: $\square$

12 Write <, > or = for each pair of fractions to make a correct statement.

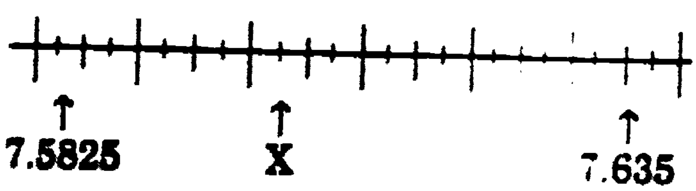
a $\frac{5}{9}$ ◯ $\frac{2}{3}$

c $\frac{5}{6}$ ◯ $\frac{5}{8}$

b $\frac{1}{2}$ ◯ $\frac{2}{5}$

d $\frac{3}{10}$ ◯ $\frac{2}{5}$

13 Two points on this number line have been marked. What is the value of the point labelled "X"



ANSWER

14 Solve these fraction problems:

a $\frac{3}{10} - \frac{4}{20} =$

b $\frac{7}{8} + \frac{3}{5} =$

c $\frac{10}{12} - \frac{5}{6} =$

d $\frac{3}{5} + \frac{1}{3} =$

15 You are having 17 people over for a birthday party. A 2-liter bottle of soda is enough for 3 people. How many bottles of soda do you need to buy?

ANSWER

16 You have been hired to program a chicken-scratch-makingmachine. It can make two kinds of marks.

It makes a | when you command it to CHIRP. It makes a — when you command it to CHEEP.

You can also program it to make a pattern of marks by giving a name to a set of commands. For example,

GOOX = (CHIRP CHEEP CHIRP).

Then when you command it to GOOX it would make | — |.

You can use GOOX in future instructions to the machine. The commands CHEEP GOOX CHEEP would make

— | — | —

a Give a set of 3 commands that would make

|| — ||

We'll call that GLEEK.

b What is the simplest set of commands (CHIRPs, CHEEPs, GOOXs, and GLEEKs) that would make:

| — | — || — ||

Johns Hopkins University National School Year Review

P15	Jordahl	Class ID:	Name:

1 Multiply or divide.

a $3.256 \div 100 =$ ☐

b $\begin{array}{r} .045 \\ \times\ 2.1 \\ \hline \end{array}$

c $.03\ \overline{)\ 2.727}$

2 Jeff loves a ride at the amusement park called the Merry Backbreaker. The ride moves 26 meters per second and thrills 34 people for half a minute. How many meters will the ride travel altogether?

☐ meters

3 Circle all factors of 39 in the grid below.

95	52	114	65	15
27	1	47	139	27
35	26	72	13	1
12	19	139	35	18
1	52	42	63	39

4 Circle the prime numbers.

1	8	5	3	11	12	16
19	21	28	4	2	7	13
31	39	18	14	15	30	36
42	9	17	22	23	25	27

5 Write each set of fractions in order from smallest to largest.

a $\frac{1}{2}, \frac{2}{3}, \frac{3}{5}$ ☐ ___, ___, ___

b $\frac{3}{4}, \frac{4}{5}, \frac{7}{10}$ ☐ ___, ___, ___

6 Add or subtract and write the answer in lowest terms.

a $\frac{4}{5} + \frac{3}{15} =$

b $\frac{7}{9} + \frac{1}{6} =$

c $\frac{8}{11} - \frac{1}{2} =$

d $\frac{1}{16} - \frac{1}{24} =$

7

$\left(\frac{4}{5} + \frac{1}{6}\right) \times \left(\frac{3}{4} + \frac{1}{3}\right) =$ ☐

8 Express as a terminating or repeating decimal.

$\frac{5}{6}$ ☐

$\frac{48}{200}$ ☐

9 Solve these proportions for

$\frac{48}{64} = \frac{3}{n}$ n=

$\frac{n}{28} = \frac{30}{105}$ n=

10 In a map of a city drawn to scale, 1 inch represents $2\frac{1}{2}$ miles. If the city's size is a rectangle $13\frac{3}{4}$ miles by $6\frac{1}{4}$ miles, how big is the map?

_____ in. X _____ in.

11 Express this fraction as a %. If necessary, round your answer to the nearest tenth of a percent.

$\frac{23}{32}$

12 Arrange the numbers from least to greatest.

0.825
0.8
0.84
$\frac{5}{6}$

_____, _____, _____, _____

13

0.82 + 23.706 + 520

124.52 – 38.081

14 A runner completed the 200 meter race in 22 seconds.

a What was the speed in meters per second, to the nearest tenth?

m/sec

b What was the speed in kilometer per hour to the nearest tenth?

km/hr

15 63% of 285 is represented by which formula. Fill in the square next to the correct one.

- 285 ÷ 63 = 222.00 ☐
- 63 ÷ 285 = 221.05 ☐
- .63 x 285 = 179.55 ☐
- .63 ÷ 2.85 = .22105 ☐
- .63 x 2.85 = 1.7955 ☐

16 Each of these magazines increased then circulation between 1980 and 1985 by 20,000 copies. By what percent did each magazine increase its circulation? Complete the table.

Magazines	Circulation 1980	Circulation 1985	Percent Increase
Gone Fishing	40,000	60,000	
Reading for Fun	5,000	25,000	
Teen World	120,000	140,000	

17 A calf increased in weight from 10lb. to 50lb. What percent increase was that?

[%]

18 A survey was taken residents of California to determine their views on the construction of a proposed highway. The results of the survey showed that 31% favored construction, 60% opposed construction, and the rest had no opinion.

If only 18 people said they had no opinion, how many people in all were surveyed?

[ ]

19 These four sets of numbers follow the same kind of sequence. (In any sequence, each number differs from the next in the same way.)

- 2 → 8 → 32
- 5 → 10 → 20
- 1 → 6 → 36
- 2 → 6 → 18

These three sets follow a sequence, but not the same kind of sequence, as the ones above.

- 4 → 7 → 10
- 1 → 5 → 10
- 3 → 6 → 9

Tell whether each of these follows the same kind of sequence as the first group. Circle 'Yes' or 'No' for each one.

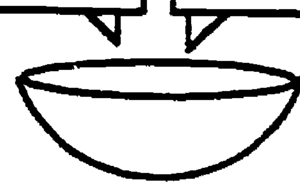
a 3 → 6 → 12	Yes	No
b 2 → 4 → 20	Yes	No
c 4 → 12 → 24	Yes	No

Now write another set of numbers that follows this kind of sequence.

____ → ____ → ____

20 A customer at the health food store where you work wants a mixture of oat and bran cereals in a certain ratio. You have to make the mixture for him using cereal premixed in certain other ratios. Show how many boxes of each premixed cereal would produce the ordered mixture.

Order	Premixes	
order 48 oz. oats 72 oz. bran	premix #1 1 oz. oats 7 oz. bran	premix #2 7 oz. oats 5 oz. bran



boxes of premix #1 ____

boxes of premix #2 ____

21

a What is 250% of 600? [ ]

b 5% of what number is 100? [ ]

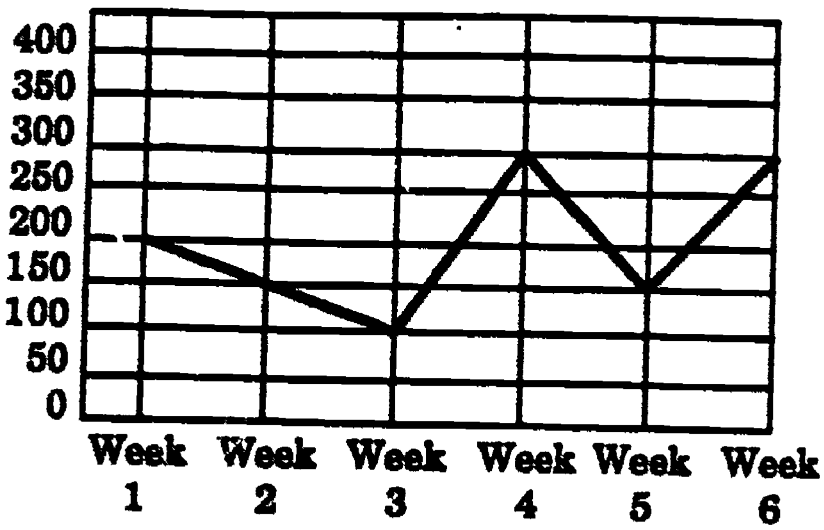
Johns Hopkins University National School Year Review

P43	Wooldridge	Class ID:	Name:

1 The manager of a gift shop is ordering sweatshirts. Each sweatshirt costs $7. She has budgeted $500. How many can she order?

[ ] sweatshirts

2 Company Profits



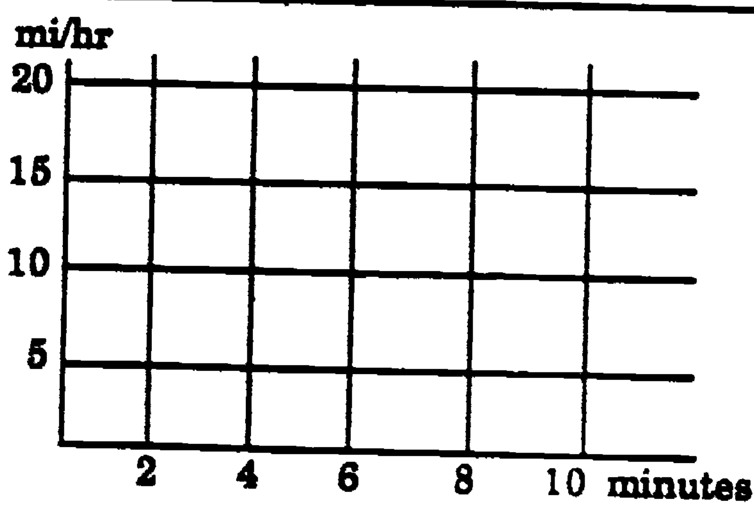
During which week did the company make the <u>least</u> profit?

[week #]

During which week(s) did profit decrease from the preceding week?

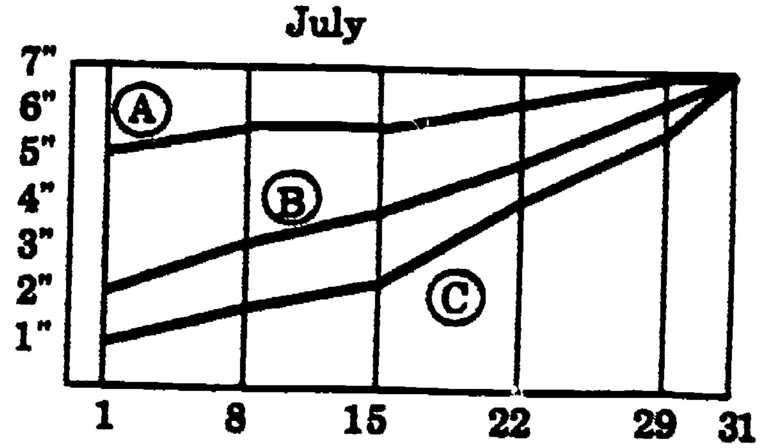
[weeks #]

3



Draw a graph to show the speed of a bicycle going along at 15 mi./hr. for 5 minutes and then coasting to a stop over the next two minutes.

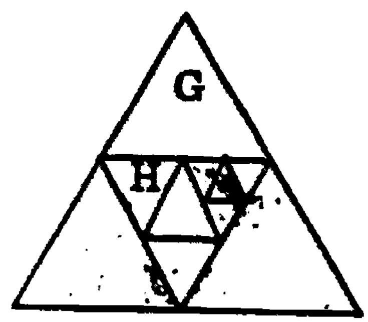
4



Which plant grew more rapidly during July? Circle one choice below.

A B C

5



Part G is $\frac{1}{4}$ of the large triangle. What fraction of the large triangle is ....

Part H? [ ]

The shaded part? [ ]

6

Who am I?

I am a proper fraction. I am equivalent to $\frac{4}{7}$. The sum of my numerator and denominator is 33.

Who am I? [ ]

7

a

$$\begin{array}{r} 4 \\ -1\frac{3}{7} \\ \hline \end{array}$$

b

$$\begin{array}{r} 4\frac{1}{6} \\ -1\frac{3}{10} \\ \hline \end{array}$$

8 Write >, <, or = in each circle to make correct statements.

$\frac{2}{5} \bigcirc \frac{1}{2}$

$\frac{3}{5} \bigcirc \frac{2}{3}$

$\frac{7}{12} \bigcirc \frac{4}{5}$

9 Find the missing values.

$\square = \square$ means that the same number goes in both squares.

a

$\frac{2}{5} \times 6 = \square$

$= \square + \square = 12$

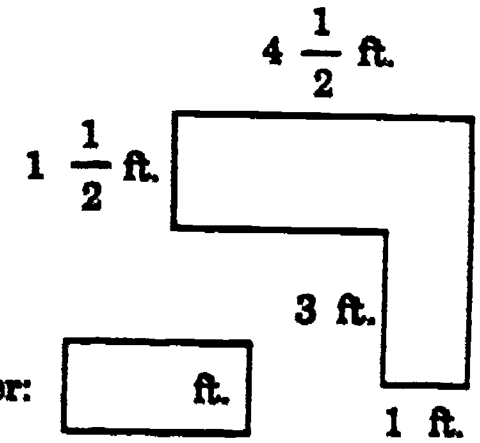
b

$12 \times \frac{2}{3} = \square$

$= \square + \frac{7}{12} = \square$

10 What is the perimeter and what is the area?

(The lengths of some sides must be calculated first)



perimeter: $\square$ ft.

area: $\square$ sq. ft.

11 Check the boxes next to the facts that you need to solve the problem. Then solve.

Problem: How many cans of paint does Kay need to buy?

- ☐ - A can of paint will cover $3\frac{1}{2}$ walls
- ☐ - Each wall is $8\frac{1}{4}$ ft. high.
- ☐ - Each room has 4 walls.
- ☐ - Kay must paint 4 rooms.
- ☐ - It takes Kay 40 minutes to paint each wall.
- ☐ - Kay already has $1\frac{1}{2}$ cans of paint.

$\square$ cans

12 Solve if you have enough information. If not, tell that fact is missing.

One program is on TV for 88 minutes. It is interrupted 8 times for commercials? What fraction of the time for the program is used for commercials?

13 Use the numbers below to write proportions.

a	3,	24,	6,	12	— = —
b	8,	6,	12,	9	— = —
c	9,	2,	6,	3	— = —
d	12,	15,	16,	20	— = —

14 John has 4 red marbles, 2 blue marbles and 3 black marbles in a bag.

What is the probability (fraction) that the first marble he pulls out will be black?

If it is black, what is the probability that the second marble will also be black?

15 Write in order from least to greatest.

8.063, 80.002, 8.603, 80.01, 80.009

________, ________, ________,

________, ________

16 On a circus train, a lion requires 4.5 cubic meters of sand, a monkey requires 10.2 cubic meters. An elephant requires as much space as a monkey and lion combined. How many elephants could be transported if there are 150 cubic meters of space available for elephants.

elephants

17 I have a secret 4-digit decimal number.

Guess my number.

Here are some hints—guesses that were <u>too high</u> or <u>too low</u>.

Guesses that were Too High	Guesses that were Too Low
2.574	.0453
1.742	.0775
0.882	.0987

Another hint: If a digit was correct and in the correct decimal position (tenths, hundredths, etc.), I put a circle around it.

Guess my number (you can figure it out from the hints.)

Your guess

APPENDIX B: Mental Math/Estimation Posttests

(Grade 7-8 Version)

(Reduced to one-fourth normal size)

$\boxed{1}$

$$\begin{array}{r} 35 \\ +\ 13 \\ \hline \end{array}$$

$\boxed{2}$

$$\begin{array}{r} 900 \\ \times\ 60 \\ \hline \end{array}$$

$\boxed{3}$

$$\begin{array}{r} 25 \\ \times\ 15 \\ \hline \end{array}$$

$\boxed{4}$

$$96 \div 16$$

$\boxed{5}$

$$6\frac{1}{4} \text{ X } \frac{1}{3}$$

$\boxed{6}$

$$16\frac{1}{2} \div 4$$

$\boxed{7}$

What is the AVERAGE?

24, 36, 45

$\boxed{8}$

$$\begin{array}{r} 1395 \\ 4795 \\ +\ 1095 \\ \hline \end{array}$$

$\boxed{9}$

$$\begin{array}{r} 420 \\ 370 \\ 280 \\ 130 \\ +\ 50 \\ \hline \end{array}$$

$\boxed{10}$

$$\begin{array}{r} \$\ 2.28 \\ 8.67 \\ 7.25 \\ 5.92 \\ +\ 6.15 \\ \hline \end{array}$$

$\boxed{11}$

29 X 31

$\boxed{12}$

$$\begin{array}{r} 303 \\ \text{X}\ \boxed{} \\ \hline 309{,}386 \end{array}$$

$\boxed{13}$

$$9\overline{)46{,}000}$$

$\boxed{14}$

$$12\overline{)4955}$$

$\boxed{15}$

$$3.1 \times 4.98$$

$\boxed{16}$

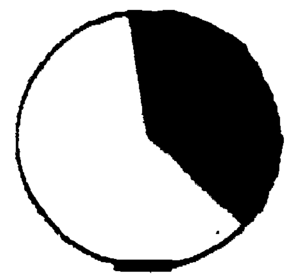
$$\begin{array}{r} 11.03 \\ \times\ 0.51 \\ \hline \end{array}$$

$\boxed{17}$

$$4\frac{3}{4} + 6\frac{9}{10}$$

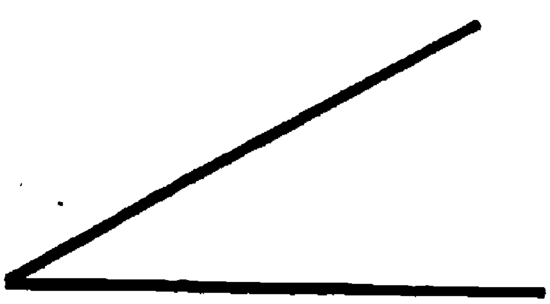
$\boxed{18}$

What fraction of the circle?



$\boxed{19}$

How many degrees?



$\boxed{20}$

What is