Congress Logopedics and Phoniatrics, Paris, September 1968.
15. M. A. Young and R. A. Campbell, *J. Acoust. Soc. Amer.* **42**, 1250 (1967).
16. K. N. Stevens, C. E. Williams, J. P. Carbonell, B. Woods, *ibid.* **44**, 1596 (1968).
17. L. G. Kersta, personal communication, 1969; O. Tosi, personal communication, 1969.
18. P. Ladefoged and R. Vanderslice, in *Working Papers in Phonetics* (Dept. of Linguistics, Univ. of California, Los Angeles, November 1967), p. 126.
19. A. Fourcin and A. W. F. Huggins, personal communication, 1969.
20. L. G. Kersta, *J. Acoust. Soc. Amer.* **34**, 1978 (1962), abstract.
21. F. R. Clarke, R. W. Becker, J. C. Nixon, *Characteristics that Determine Speaker Recognition*, Report ESD-TR-66-636 (Decision Sciences Laboratory, Hanscom Field, Bedford, Mass., December 1966).
22. Research projects on spectrographic voice identification, sponsored by the U.S. Department of Justice, are currently in progress at Michigan State University (see Tosi, *14*) and at Stanford Research Institute (see Hecker, *4*).

# Computer-Assisted Instruction

Two computer-assisted instruction programs are evaluated.

Patrick Suppes and Mona Morningstar

Although computer-assisted instruction has reached the operational stage in a number of places in the United States, very few "hard data" evaluations of student achievement in these programs have as yet been published. The purpose of this article is to report the results of the evaluative testing of students in two programs that have been in progress at Stanford for the past several years. The first is the drill-and-practice program in elementary mathematics. Results are reported for schools in California for the 1966–67 and 1967–68 academic years and for schools in McComb, Mississippi, for the 1967–68 academic year. The second program is the tutorial curriculum in elementary Russian at Stanford University, now in its third year of operation.

We do not attempt, here, to report a wide-ranging evaluation of computer-assisted instruction—one that would include observations of student behavior; the results of student, parent, or teacher questionnaires; or detailed analyses of curriculum performance. Some results of this kind have already been published (see *1*). Nor do we report evaluation of the Stanford tutorial programs in reading. For this the reader is referred to Atkinson's report (*2*). Our purpose here is to concentrate on the classical comparison of experimental groups with control groups and to compare their relative rates of achievement. In the case of the mathematics program, the primary instruments of evaluation were Stanford Achievement Tests (SAT) (*3*), which are not a product of Stanford University but are widely used commercial tests. In the Russian program, which was under the direct supervision of Joseph Van Campen at Stanford University, the evaluation was based on comparative performance on midterm and final examinations in the course.

It should be emphasized that our main purpose in this article is to present without extensive interpretation the evaluative results. We do conclude with some discussion of the results, but the main function of the article is to present, in standard data form, the results of the testing.

## Mathematics Program: Description

The drill-and-practice program in elementary-school mathematics began in the spring of 1965 with 41 fourth-grade children who were given daily arithmetic drills on a teletype machine in their classroom. By the end of the 1965–66 school year, 270 students in grades 3 through 6 in three California elementary schools were participating in the program (see *1*). During the 1966–67 school year, the program was further expanded to include grades 1 through 6, with more than 1500 students involved. Student participation increased again during 1967–68, with approximately 1000 students in California, 600 students in Mississippi, and 1100 students in Kentucky.

Because changes occurred in the curriculum and the computer system as the program developed during the first 2 years, statistical evaluation was not begun until the academic year 1966–67. During 1966–67 and 1967–68, Stanford Achievement Tests were used for evaluation. The primary aim of the program was to provide drill and practice in the skills of arithmetic, especially computation, as an essential supplement to regular classroom instruction. The concepts presented to the students for drill and review at the computer terminal had been previously introduced in the classroom by the teacher.

For the 1966–67 and 1967–68 school years, the curriculum material, for each of grades 1 through 6, was arranged sequentially in blocks to coincide approximately with the development of mathematical concepts introduced in several text series. There were 20 to 27 concept blocks for each grade level. Each concept block included a preliminary test (or pretest), 5 days of drill, a subsequent test (or posttest), and sets of review drills and review posttests. A brief description of the material in each concept block for grades 1, 3, and 6 is given in Table 1.

Parallel forms of a test were prepared for each concept block. The test consisted of equal numbers of problems from each of five levels of difficulty. For a given student, different forms of the test were assigned for the pretest and for the posttest in each block. The form assigned for the pretest was randomly determined, with the restriction that equal numbers of students receive each form. The forms of the test not assigned as a pretest or a posttest for a given student were divided into halves and used as review posttests for that student. For each day of drill, five

Patrick Suppes is director of the Institute for Mathematical Studies in the Social Sciences and professor of philosophy, statistics, and education at Stanford University, Stanford, California. Mona Morningstar is a research associate at the Institute for Mathematical Studies in the Social Sciences.

drills, one at each of the five levels of difficulty, were prepared—a total of 25 drills per block. Several sets of review drills for each block were also prepared at the five defined levels of difficulty. The drills and the review drills for the most difficult level, level 5, and the level-5 problems on the tests were rewritten for the 1967–68 school year to be more difficult than those presented in the 1966–67 curriculum.

Each student responded to problems presented on a teletype located in the school. The model-33 teletypes were connected to the PDP-1 computer at Stanford by means of telephone lines. After the student had signed into the program by typing his assigned student number and his first name, the teletype printed his last name and presented the appropriate set of problems. The pace of the problem presentation was determined by the student's response rate.

The tests and drills presented to the student for the 7 days required for each concept block were as follows. Day 1, pretest; days 2 to 5, drill and review drill; day 6, drill and review posttest; day 7, posttest. The teletype printed each individual problem and then positioned itself in readiness to accept the answer in the appropriate place. The student typed in the answer. If his answer was correct, he proceeded to the next problem. If he gave the wrong answer, the teletype printed "No, try again," and presented the problem again. If he made a second error, the teletype printed "No, the answer is . . .," and presented the problem once more. If the student gave the wrong answer for the third time, he was given the correct answer and the teletype automatically proceeded to the next problem. The student was allowed from 10 to 40 seconds to respond, depending upon the type of problem presented. If a student took more than the allotted time to give his answer, the procedure just described was followed, but the teletype printed "Time is up, try again" in place of "No, try again."

The level of difficulty of the first day of drill was determined by the student's performance on the pretest according to the criteria presented in Table 2. The level of difficulty of each successive drill in the same concept block was determined by the student's performance on the preceding day's drill. Thus, if the student's performance on a drill was 80 percent or greater, his next drill was one difficulty level higher. A score of less than 60 percent branched him down a level for the next drill. If the score was between 60 and 80 percent, the student remained at the same difficulty level for the next drill.

Whereas the drill content was the same for all students in a class, and only the difficulty level changed as a function of the preceding day's performance, the content of the review drills differed for different students as a function of the total-past-performance history of the student. The computer selected the review drills that corresponded to the content of the past block on which the student had had the lowest posttest score, with the restriction that the student was not to be

Table 1. Concept blocks for grades 1, 3, and 6, drill-and-practice program 1966–67.

| Grade 1 | | Grade 3 | | Grade 6 | |
|---|---|---|---|---|---|
| Block | Description | Block | Description | Block | Description |
| 1. | Counting, how many, 0–9 | 1. | Mixed addition and subtraction, horizontal format, sums 0–18 | 1. | Mixed drill: ½ column addition, subtraction; ½ multiplication; some involving decimals |
| 2. | Counting in sequence | 2. | Addition, sums 0–18, horizontal and vertical | 2. | Multiplication: 2's → 12's, level by products, horizontal format |
| 3. | Sums to 4 | 3. | Subtraction, sums 0–18, horizontal and vertical | 3. | Column multiplication: (1 digit) × (2 digit) through (2 digit) × (3 digit) |
| 4. | Sums to 4, vertical, mixed | 4. | Addition, no carry, vertical (2 addends, 3-digit) and (3 addends, 2-digit) | 4. | Division: ladder form, 1-digit divisor |
| 5. | Differences to 4, vertical, mixed | 5. | Subtraction, no borrow, vertical, 2- and 3-digit | 5. | Fractions: factors, reducing, comparing, simple addition, subtraction |
| 6. | Sums to 6, vertical, mixed | 6. | Addition, vertical with carry | 6. | Mixed drill: inequalities, decimals, word problems, exponents, addition, subtraction, multiplication, division |
| 7. | Sums to 7, vertical, mixed | 7. | Subtraction, with borrow | 7. | Division: ladder form to 2-digit divisors |
| 8. | Differences to 7, vertical, mixed | 8. | Mixed addition and subtraction, carry and borrow | 8. | Fractions: add, subtract |
| 9. | Sums to 9, vertical, mixed | 9. | Measure and word problems and inequalities | 9. | Measure: length, time, money, temperature, liquid measure |
| 10. | Sums to 10, vertical only | 10. | Column addition and subtraction; add; subtract | 10. | Ratio: percent |
| 11. | Differences to 10, vertical only | 11. | Measure, inequalities | 11. | Division: ladder form, 2-digit divisor |
| 12. | Sums to 10 with variables | 12. | Multiplication, horizontal, 2's and 3's | 12. | Mixed drill: fractions (addition, subtraction, multiplication), ratio, percent, division, decimals (addition, subtraction) |
| 13. | Differences to 10 with variables | 13. | Mixed multiplication and division, 2's and 3's | 13. | Fractions: decimals, addition, subtraction, multiplication |
| 14. | Sums and differences to 10, horizontal | 14. | Division, ladder form, 1 digit into 2 digit | 14. | CAD laws: days 1–4, apply law; day 5, identify law |
| 15. | Sums and differences to 10, vertical format | 15. | CAD laws: add, subtract, multiply | 15. | Multiplication: multiples of 10, horizontal format |
| 16. | Sums and differences to 10 with variables | 16. | Mixed drill: measure, word problems, inequalities | 16. | Division: ladder form, 2-digit divisors, 3- to 5-digit dividends |
| 17. | Sums to 10, 3-digit numbers | 17. | Fractions | 17. | Mixed drill: fractions (+, −, ×, ÷), (column addition), CAD laws, division |
| 18. | Column addition, sums with 10's, no regrouping | 18. | Multiplication, horizontal 2's → 9's | 18. | Measures: all, including a few metric; area, volume |
| 19. | Column subtraction, no regrouping | 19. | Mixed drill: multiplication, division, fractions | 19. | Ratio, percent |
| 20. | Mixed addition and subtraction in columns, facts to 10 | 20. | Division, ladder form, 1 digit into 3 digit | 20. | Mixed drill: all operations, percent, decimal multiplication |
| 21. | Mixed addition and subtraction, inequalities | 21. | Multiplication, vertical, 1 × 2 digit | 21. | Negative numbers: add, subtract, multiply |
| 22. | Mixed 1- and 2-digit column addition and subtraction | 22. | Achievement tests | 22. | Achievement tests |
| 23. | Sums to 10 with form $a + b = c + d$ | 23. | Mixed drill: column add; subtract; multiply | 23. | Mixed drill: summary |
| 24. | Sums to 10 with variables, form $a + b = c + d$ | 24. | CAD laws | 24. | Estimation of quotients in division |
| 25. | Special addition and subtraction | 25. | Special addition and subtraction drills | 25. | Special addition and multiplication drills |
| | | | | 26. | Special subtraction and division drills |
| 27. | Special mixed drills | 27. | Special mixed drills | 27. | Special mixed drills |

given the same review drill in two 7-day blocks in a row. The level of difficulty of the review drills was determined by the posttest according to the criteria presented in Table 2; the difficulty level remained constant for the 4 days of review. Once a student had received a set of review drills on a given concept block, the score on the review posttest, given on the sixth day, replaced the previous posttest score on that concept block as a basis for determining the concept block and difficulty level for future review drills. The branching structure for a 7-day sequence of problems is shown in Fig. 1. To make up for absences, a student could take more than one drill per day.

### Mathematics Program: Evaluation

To evaluate the effectiveness of the drill-and-practice program, the arithmetic portion of the Stanford Achievement Test was administered to both experimental and control classes. Four different difficulty levels of the SAT were used. Each level had one, two, or three arithmetic sections; these are described briefly in Table 3. Unless otherwise noted, the tests were given in October and again in May, by either the classroom teacher or a member of the staff at Stanford's Institute for Mathematical Studies in the Social Sciences.

Although the publishers of SAT recommend that the test given in May for grades 2 and 5 be one level higher than the test given in October (Table 3), whenever possible we administered different forms of the same test within a given grade—for example, Primary II in grade 2 and Intermediate II in grade 5. However, when the administration of the SAT was an integral part of a testing program within a school system requiring adherence to the SAT manual, we were unable to dictate which tests or forms were to be administered. This fact must be kept in mind when deviations from standard experimental design occur. In the control schools, the experimental design did not include any manipulation of the amount of drill and practice, or of the curriculum.

*California 1966–67.* Tests were given in four California schools for the 1966–67 evaluation. For grades 3 through 6, students in Experimental School A and Control School B were tested; for grades 4 through 6, students in Experimental School C and Control School D were tested. In each case, the control school was located in the same district as the experimental school. In all four schools, the pretest and posttest administered was Primary II for grade 3, Intermediate I for grade 4, and Intermediate II for grade 6. For grade 5, schools A and B administered the Intermediate II test for both the pretest and the posttest; schools C and D administered Intermediate I as a pretest and Intermediate II as a posttest.

The difference between the posttest and pretest grade placement on the SAT Computation Section for each grade for school A relative to school B and for school C relative to school D was examined. The statistical results of *t*-tests and the average pretest and

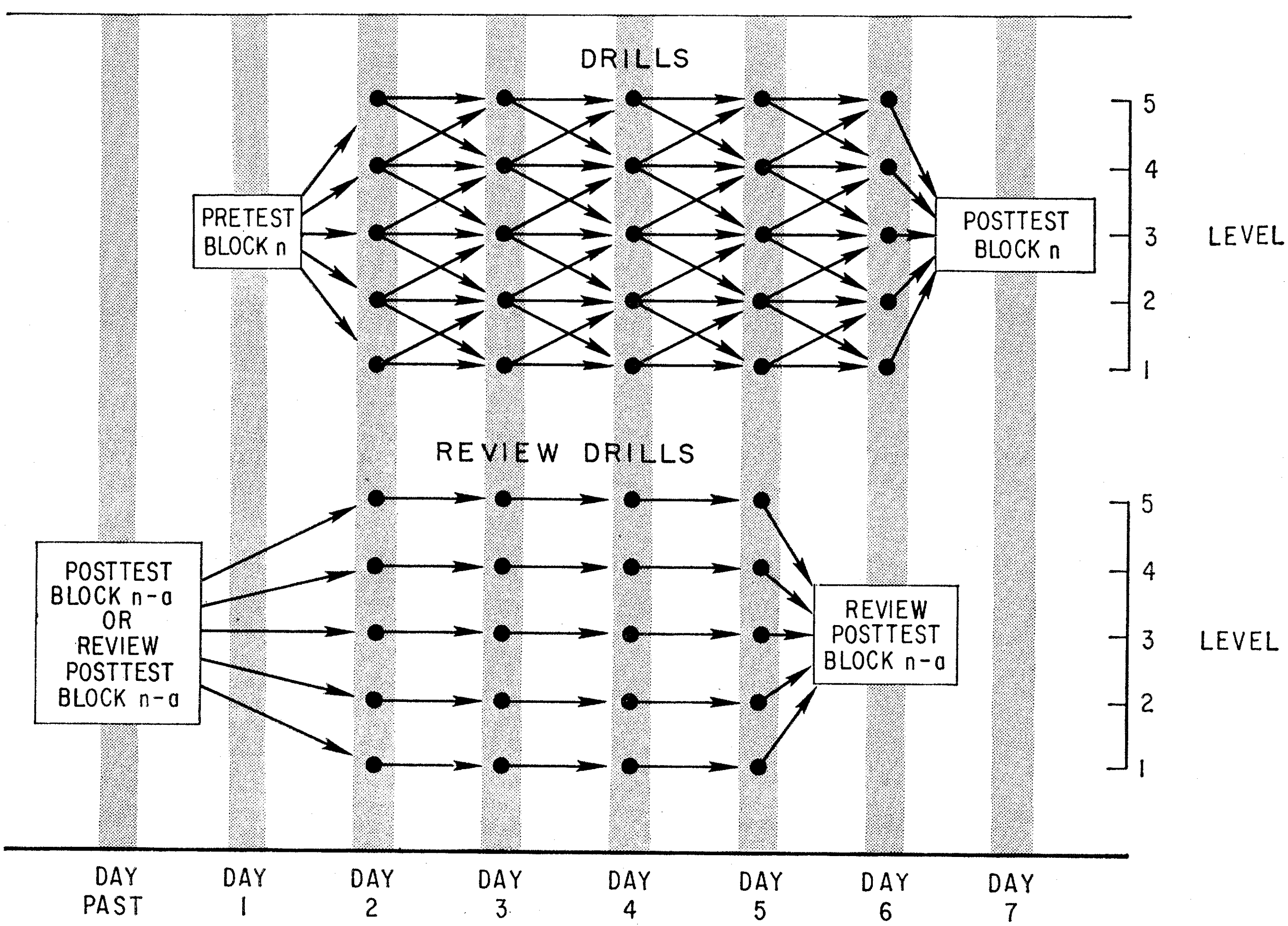


Fig. 1. Branching structure for a 7-day concept block; *n-a*, block with lowest posttest performance.

Table 2. Branching criteria.

| From pretest to drill* | | From drill to drill | |
|---|---|---|---|
| Percent correct | Level assigned for drill | Percent correct on drill $D_i$† | Level assigned for drill $D_{i+1}$ |
| 0–19 | 1 | 0–59 | Next lower level |
| 20–39 | 2 | 60–79 | Same level as $D_i$ |
| 40–59 | 3 | 80–100 | Next higher level |
| 60–79 | 4 | | |
| 80–100 | 5 | | |

* Also from posttest to review. † $D_i$ = day i drill.

posttest grade placement are shown in Table 4. The increase in performance level for students in the experimental school was significantly greater than that for students in the control schools for grade 3 in school A and for grades 4 and 6 in school C.

At the end of the school year we learned that, after examining the results of the pretesting at their school, the teachers and administrators at school B, a control school, instituted an additional 25 minutes per day of classroom instruction and practice in arithmetic for grades 4 and 5. Since the improvement in performance of the students in grades 4 and 5 was significantly greater in Control School B than in Experimental School A, we conclude that 25 extra minutes per day of classroom drill can be more beneficial than 5 to 8 minutes per day of computer-based drill. The fact that gains for Control School B were greater than those for Control School D supports the conclusion that the performance of the students in school B was a function of the extra drill in the classroom. The effect of classroom drill, however, does not detract from the effectiveness of the drill-and-practice program. Whereas the classroom approach required that 25 more minutes of the teacher's time be devoted to arithmetic and that 25 minutes less of the student's time be available for nonmathematical topics, the drill-and-practice program required no extra time from the teacher, and the student lost only 5 to 8 minutes from nonmathematical subjects.

Since the SAT Concepts and Applications Sections do not include many items contained in the drill-and-practice curriculum, they are not as appropriate as the Computation Section for testing the effectiveness of the program. However, on the Concepts Section the increase in performance level was significantly greater for the students in the experimental schools than for those in the control schools for grade 6 [$t(118) = 2.18$, $P < .01$], in school A; for grade 4 [$t(122) = 2.37$, $P < .01$], in school C; and for grade 5 [$t(138) = 4.21$, $P < .01$], in school C. On the Applications Section of the SAT the experimental classes performed significantly better than the control classes in grade 4 [$t(122) = 1.96$, $P < .05$], in school C, and in grade 5 [$t(138) = 2.50$, $P < .01$], in school C.

*California 1967–68.* For evaluation of the 1967–68 computer-assisted instruction program in California, the SAT was administered to students in grades 1 through 6 in seven different schools. Two of these schools had both experimental and control students, two had only experimental students, and three had only control students. Within the experimental group, from 5 to 9 classes were tested at each grade level; within the control group, from 6 to 14 classes were tested at each grade level. Although the testing program for 1967–68 was more advantageous than that for 1966–67 in terms of number of students tested, the distribution of students among schools and classes made it impossible to conduct matched comparisons, as was done in the 1966–67 evaluation. The tests administered as the pretest in October and as the posttest in May were Primary I for grade 1, Primary II for grades 2 and 3, Intermediate I for grade 4, and Intermediate II for grades 5 and 6.

Again a *t*-test was used to determine the difference between the experimental and the control students in terms of change in performance from October to May within each grade. The results of this comparison and the average pretest and posttest grade placement are shown in Table 5. The students who had received computer-based drill and practice had a significantly greater increase in performance level than the control students on the SAT Computation Section in grades 2, 3, and 5.

The performance of students in the experimental group was significantly better than that of students in the control group on the Concepts Section for grade 3 [$t(344) = 4.13$, $P < .01$] and

Table 3. Description of the arithmetic portion of the Stanford Achievement Test batteries.

Primary I: Middle of grade 1 to middle of grade 2
- A. Arithmetic, 63 items
  (Measures, problem solving, number concepts)

Primary II: Middle of grade 2 to end of grade 3
- A. Arithmetic Computation, 60 items
  (Addition, subtraction, multiplication, division)
- B. Arithmetic Concepts, 46 items
  (Numbers, measures, problem solving)

Intermediate I: Beginning of grade 4 to middle of grade 5
- A. Arithmetic Computation, 39 items
  (Addition, subtraction, multiplication, division)
- B. Arithmetic Concepts, 32 items
  (Place value, meanings and interrelationships of operations, average, percent, etc.)
- C. Arithmetic Applications, 33 items
  (Reasoning and problem solving in area, ratio, volume, averages, graphs, etc.)

Intermediate II: Middle of grade 5 to end of grade 6
- A. Arithmetic Computation, 39 items
  (Addition, subtraction, multiplication, division)
- B. Arithmetic Concepts, 32 items
  (Place value, meanings and interrelationships of operations, average, percent, etc.)
- C. Arithmetic Applications, 39 items
  (Reasoning and problem solving in area, ratio, volume, averages, graphs, etc.)

Table 4. Average grade-placement scores on the Stanford Achievement Test: California, 1966–67.

| Grade | Pretest* | | Posttest | | Posttest-pretest | | *t* | Degrees of freedom |
|---|---|---|---|---|---|---|---|---|
| | Experimental | Control | Experimental | Control | Experimental | Control | | |
| | | | *School A versus school B* | | | | | |
| 3 | 2.9 (51) | 3.0 (63) | 3.9 | 3.6 | 1.0 | 0.6 | 2.50† | 112 |
| 4 | 3.9 (60) | 3.9 (75) | 4.7 | 5.3 | 0.9 | 1.4 | −2.93† | 133 |
| 5 | 4.6 (66) | 4.6 (81) | 5.2 | 6.3 | 0.7 | 1.7 | −4.74† | 145 |
| 6 | 4.9 (50) | 5.2 (70) | 7.1 | 7.1 | 2.1 | 1.9 | 0.95 | 118 |
| | | | *School C versus school D* | | | | | |
| 4 | 3.7 (61) | 3.8 (63) | 5.4 | 4.8 | 1.7 | 1.0 | 4.50† | 122 |
| 5 | 5.4 (63) | 4.9 (77) | 6.2 | 5.4 | 0.8 | 0.6 | 1.32 | 138 |
| 6 | 5.8 (58) | 6.0 (56) | 7.4 | 7.1 | 1.6 | 1.1 | 2.19‡ | 112 |

* Values in parentheses are numbers of students. † $P < .01$. ‡ $P < .05$.

on the Applications Section for grade 6 [$t(399) = 2.14$, $P < .05$].

*Mississippi 1967–68.* For evaluation of the first year of drill and practice in Mississippi, the SAT was administered to students in grades 1 through 6 in 12 different schools. Eight of these schools had both experimental and control students, three had only experimental students, and one had only control students. Within the experimental group, from 1 to 10 classes were tested at each level; within the control group, from 2 to 6 classes were tested at each grade level.

The testing of the students in Mississippi was not as consistent, in terms of experimental design, as the testing in California. On the other hand, the computer-assisted instruction program itself operated in an environment that was far removed from the Stanford group responsible for the program, so the lack of detailed control of the testing by the Stanford research group was somewhat compensated by the independent environment in which the program was tested. The Primary I level of the SAT was administered as a pretest to students in the first grade in February rather than in October. The posttest was given in May. For the remaining grades, the pretest was given in October and the posttest in May. For grade 2, Primary I was given as a pretest and Primary II as a posttest. For grades 3 and 4 the pretest and posttest were the same—Primary II for grade 3 and Intermediate I for grade 4. For grade 5, all the control students and two classes of experimental students were given Intermediate I for a pretest and Intermediate II for a posttest; one experimental class was given Intermediate I as both a pretest and a posttest. Although grade 6 was given Intermediate II as both a pretest and a posttest, 2 of the 10 classes in the experimental group and 1 of the 6 classes in the control group were given the same form of the test, rather than different forms, for the two testing sessions.

The $t$-value and the average pretest and posttest grade placement for each grade are shown in Table 6. The performance of the experimental students improved significantly more than that of the control students in all six grades. The difference in degree of improvement between the experimental group and the control group was largest in grade 1, where, in only 3 months, the average increase in grade placement for experimental students was 1.14, as compared with 0.26 for control students.

On the Concepts Section the performance of students in the experimental group was significantly better than that of students in the control group for grade 3 [$t(76) = 3.01$, $P < .01$] and grade 6 [$t(433) = 3.74$, $P < .01$]; on the Applications Section it was significantly better for grade 6 [$t(433) = 4.09$, $P < .01$]. In grade 4, the control group showed more improvement on the Concepts Section than the experimental group did [$t(131) = -2.25$, $P < .05$].

*California and Mississippi results compared.* The average grade placement on the pretest for grades 1 through 3 was similar for the California and the Mississippi experimental groups (Tables 5 and 6). The difference between averages for the two groups increased in grades 4 through 6, with the Mississippi students performing at a lower level than the California students. In spite of the significant gains made by the experimental groups relative to the control groups in Mississippi, the discrepancy in grade placement between the California and the Mississippi experimental groups for grades 4 through 6 was larger on the posttest than on the pretest. Thus, the overall superiority of the experimental program in Mississippi was related more to a lesser increase in performance level for the control schools in Mississippi than to a greater change in performance level for the Mississippi experimental groups relative to the California experimental groups. In this connection it is important to emphasize that the California schools were all located in relatively affluent middle-class neighborhoods, and that the average family income and educational level were undoubtedly higher for the California than for the Mississippi groups, although we did not collect systematic data on economic and social variables.

Table 5. Average grade-placement scores on the Stanford Achievement Test: California, 1967–68.

| Grade | Pretest* | | Posttest | | Posttest-pretest | | $t$ | Degrees of free-dom |
|---|---|---|---|---|---|---|---|---|
| | Experi-mental | Con-trol | Experi-mental | Con-trol | Experi-mental | Con-trol | | |
| 1 | 1.39 (58) | 1.30 (267) | 2.64 | 2.51 | 1.24 | 1.21 | 0.33 | 323 |
| 2 | 2.06 (65) | 2.16 (238) | 3.21 | 2.90 | 1.15 | 0.74 | 5.19† | 301 |
| 3 | 3.00 (136) | 2.85 (210) | 4.60 | 3.89 | 1.59 | 1.05 | 6.28† | 344 |
| 4 | 3.40 (103) | 3.49 (185) | 4.86 | 5.00 | 1.46 | 1.50 | −0.38 | 286 |
| 5 | 4.98 (149) | 4.44 (90) | 6.40 | 5.32 | 1.42 | 0.88 | 4.03† | 237 |
| 6 | 5.42 (154) | 5.70 (247) | 7.44 | 7.61 | 2.02 | 1.91 | 0.93 | 399 |

* Values in parentheses are numbers of students. † $P < .01$.

## The Russian Program

The computer-based Russian program was instituted at Stanford in September 1967 under the direction of Joseph Van Campen, who designed a program to teach first- and second-year courses at the college level. This program included comprehension of written Russian, comprehension of spoken Russian, and mastery of grammar and syntax. Of the three main components of a college-level language course—classroom sessions on a daily basis, time in the language laboratory, and regular homework assignments—only the functions of the classroom sessions were assumed by the computer program. In addition to their time at the computer console, the students spent time in the language laboratory and did home assignments. The language-laboratory tapes, with drill sheets and homework assignments, are prepared by the staff at the Institute.

*First-year course.* Thirty students began the first-year Russian course in the

Table 6. Average grade-placement scores on the Stanford Achievement Test: Mississippi 1967–68.

| Grade | Pretest* | | Posttest | | Posttest-pretest | | $t$ | Degrees of free-dom |
|---|---|---|---|---|---|---|---|---|
| | Experi-mental | Con-trol | Experi-mental | Con-trol | Experi-mental | Con-trol | | |
| 1 | 1.41 (52) | 1.19 (62) | 2.55 | 1.46 | 1.14 | 0.26 | 3.69† | 112 |
| 2 | 1.99 (25) | 1.96 (54) | 3.37 | 2.80 | 1.42 | 0.84 | 5.23† | 77 |
| 3 | 2.82 (22) | 2.76 (56) | 4.85 | 4.04 | 2.03 | 1.26 | 4.64† | 76 |
| 4 | 2.26 (58) | 2.45 (77) | 3.36 | 3.17 | 1.10 | 0.69 | 2.63† | 131 |
| 5 | 3.09 (83) | 3.71 (134) | 4.46 | 4.60 | 1.37 | 0.90 | 3.43† | 215 |
| 6 | 4.82 (275) | 4.36 (160) | 6.54 | 5.48 | 1.72 | 1.13 | 5.18† | 433 |

* Values in parentheses are numbers of students. † $P < .01$.

fall of 1967. Two of the four sections of beginning Russian served as a control group; the other two sections were asked to volunteer for the computer-assisted instruction (CAI) course. None of the students refused to remain in their assigned sections. The CAI students were required to spend about 50 minutes a day, 5 days a week, at the computer console. In all, 135 lessons were presented to the students in a combined audio and teletype format. The students responded on a model-35 teletype with a special Cyrillic-alphabet keyboard.

Although the basic curriculum was the same for all students, there were several remedial branches. At given points in the curriculum, students were tested on several items of a given type and were given remedial instruction on the material covered if their performance on the test block failed to meet a satisfactory standard. Later in the year (1967–68) routines were provided which produced more specific remedial work, based on the type of error the student had made.

During the period prior to the final examination, lesson summaries for each new lesson and a final summary covering the material for the entire quarter were given the students. The computer then assessed the student's performance and told him the areas on which he should concentrate his efforts. At subsequent sessions the student was again tested on the material he had missed and was informed where more study was needed. In addition, he could repeat any lesson or portion of a lesson at the computer console.

Language-laboratory tapes provided material for pronunciation practice and also for testing a student's ability to comprehend spoken Russian. A test at the end of the tape either required the student to transcribe into English a number of Russian sentences or required him to respond in writing to oral questions on a paragraph that he had just heard.

In order to evaluate his own pronunciation, each student made two recordings during each quarter. After each recording session the student was counseled immediately and was told what pronunciation errors he had made and how he could correct them.

*Second-year course.* Instruction began in September 1968 for 19 students enrolled in the second-year course in Russian. Thirty-nine lessons, including review lessons, were available for the quarter. The students were at the console for about 45 minutes, 5 days a week. Homework and study sheets for lessons 1 through 39 were distributed to the students as they progressed through the lessons. The homework involved translating English sentences into Russian, while the study sheets dealt with new grammar and new vocabulary pertinent to the day's lesson. To develop speech and the ability to write correctly what is heard, the students occasionally were allowed to choose at the conclusion of a teletype lesson whether to take dictation or to practice pronunciation. These exercises were recorded at the end of the tape used as the audio portion of the regular teletype lessons. Once every 2 weeks students took written quizzes or read from handwritten or typed scripts. Their pronunciation was corrected, and suggestions were made for improvement.

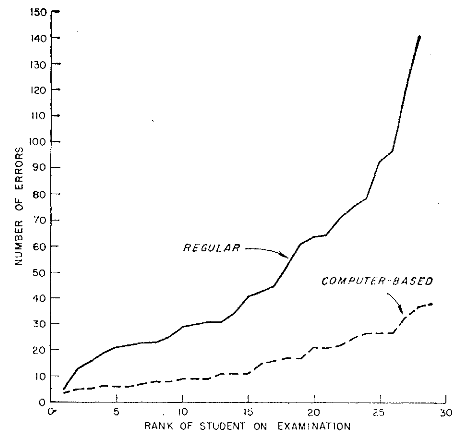


Fig. 2. Student performance for the portion of the fall-quarter final examination in first-year Russian that was common to the computer-based and regular sections.

*Evaluation.* Of the 30 students who started the first-year computer-based course, 1 left during the first quarter, 3 left between the first and second quarters, 1 left during the second quarter, and 3 left between the second and third quarters. Two new students entered the computer-based section at the beginning of the second quarter. Of the 38 students enrolled for the autumn quarter in the regular Russian section, 10 left the course during the first quarter, 13 left between the first and second quarters, and 3 left between the second and third quarters. Four new students entered the regular section at the beginning of the third quarter, one of them having transferred from the computer-based class. Of the 30 students originally enrolled in the computer-based program, 22 (73 percent) finished all three quarters, whereas, of the 38 students in the regular class, only 12 (32 percent) finished the year's curriculum. This finding suggests that the computer-based course held the interest of the students much better than the regular course did. Probably because Russian is more difficult than French, Spanish, or German for American students, the dropout rate in Russian at Stanford and other universities is traditionally quite high.

Approximately 66 percent of the content of the final examinations for the autumn and winter quarters was identical for the computer-based and the regular Russian sections; the complete final examination for the spring quarter was identical for the two groups. The number of errors for each student, when the students are ranked according to their performance on the final examination, is shown in Figs. 2 to 4 for the fall, winter, and spring quarters, respectively. Although the average number of errors was lower for the computer-based students in all three quarters—15.8 relative to 49.0 in the fall quarter, 21.8 relative to 25.8 in the winter quarter, and 53.0 relative to 71.1 in the spring quarter—the difference was statistically significant for the fall quatrer (Mann-Whitney $U$ test, $P < .001$) and the spring quarter ($P < .05$), but not for the winter quarter. Since the selection process resulting from the poorer students' leaving the regular course biases the results on the examination against the computer-based group, the superiority of the computer-based group on the spring examination is more impressive than the difference indicated by the average number of errors.

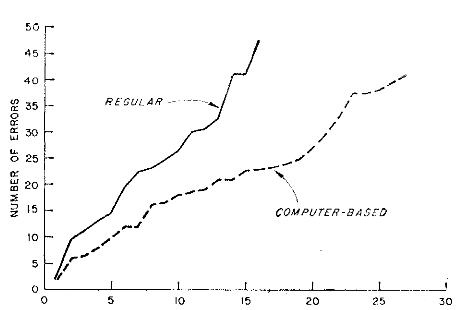


Fig. 3. Student performance for the portion of the winter-quarter final examination in first-year Russian that was common to the two sections.

Figures 2 to 4 indicate three facts about the differences in performance between the two groups. First, in the fall quarter (Fig. 2), as the position of the student's rank on the examina-

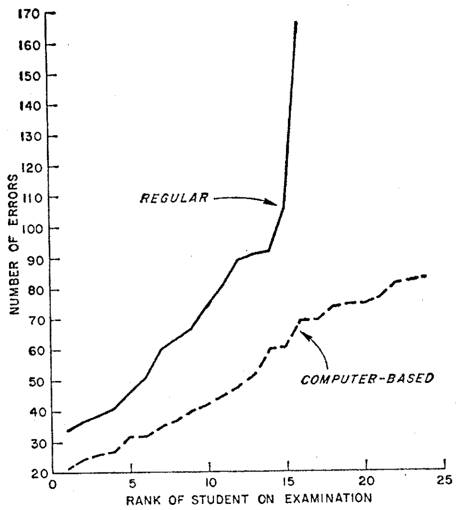


Fig. 4. Student performance for the spring-quarter final examination in first-year Russian.

tion increases, the superiority of the computer-based student over the regular student increases. Second, for the winter and spring quarters (Figs. 3 and 4), for the best students within each group, the superiority of the computer-based group is at first relatively constant; this phase is followed by an increased difference between the two groups as the position of the student's rank increases. Finally, the difference between the best students increases from the winter quarter to the spring quarter, the superiority of the computer-based students being greater in the spring quarter.

Of the 19 students enrolled in the computer-based class of the second-year Russian course, 12 had participated in the first-year computer course and 7 had taken the Stanford placement test to qualify for the second-year course and were new to computer-based instruction. Eleven students were enrolled in the second-year Russian course in the regular classroom.

For both the computer-based class and the regular class, the number of errors for each student, ranked on performance on the final examination for the fall quarter in the second year, is shown in Fig. 5. Seventy-four percent of the computer-based students performed better than the best student in the conventional class. The average number of errors, 53.0 for the computer-based class and 71.1 for the regular class, differed significantly for the two groups (Mann-Whitney $U$ test, $P < .001$); the computer-based students performed better on the examination than the conventional-classroom students did.

## Discussion of the Two Programs

As is the case when any new technology is applied to an area where many skills have been developed and much knowledge already exists, the results of computer-assisted instruction in the early years will necessarily be mixed. From an operational standpoint, the Stanford efforts reflected in the two programs reported here began only in 1965, and so the evaluation reported here covers the first 3 years of effort. On balance, we feel that the results are positive, but it is also important to note that not all the results have been positive. We feel, however, that we have some explanation for some of the negative results. More importantly, we have learned a great deal since 1965, and we believe that the programs we are now developing and beginning to place in schools on an experimental basis will benefit from the work reported here. The technology is complicated, and our understanding of the underlying psychological principles of learning in any major area of curriculum is still rather tenuous. In our judgment, while there is nothing definitive about any of the evaluation results reported in this article, it did seem desirable to report as early as possible, in systematic form and in objective fashion, the evaluation results obtained.

We turn now to some specific remarks about the two programs. Turning to the drill-and-practice program, we observe first that, as the evidence comparing California school A with school B in 1966–67 indicates, teachers can do as well with a good regime of drill and practice in the fundamentals of arithmetic as computers can. We do not find this conclusion at all surprising. We have known for a long time, from studies dating back to the 1920's, that a daily regime of drill and practice, carried out with faithfulness and regularity by the teacher, does improve the performance of students [see, for example, Wilson (*4*)]. What seems evident already is that the use of terminals to bring a drill-and-practice program to schools can bring a kind of quality control that is difficult to achieve in large numbers of schools with large numbers of teachers. Concentrated efforts in single schools with a dedicated staff can certainly do as well as anything that we can currently offer, but it is especially true of the elementary-school mathematics curriculum that many teachers in the upper three grades—that is, grades 4 to 6—are not really interested

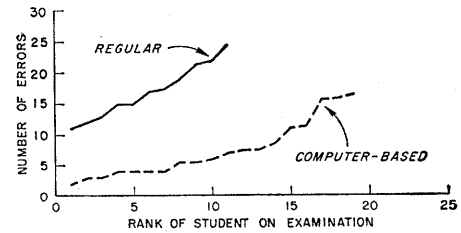


Fig. 5. Student performance for the portion of the fall-quarter final examination in second-year Russian that was common to the two sections.

in mathematics and would much prefer to turn the problem of providing a regime of review and maintenance of arithmetic skills over to a computer-based instructional program.

The results of the data reported here indicate that an individualized drill-and-practice program in elementary mathematics will produce its most impressive results in school environments not educationally and economically affluent. This is evident from the comparison of experimental and control groups in Mississippi and California. This remark is closely connected with the preceding one, for it is in the less affluent areas of the country that, in general, teacher preparation and teacher training are least satisfactory. One way to meet some of these problems of teacher training, as in the case of mathematics, is to bring work to the student directly on computer-based terminals. In the evaluation data cited above, data for the Elliotsville School in rural Kentucky were not included, because there was no control school to provide a basis for assessing the change in achievement data that occurred in the spring of 1967. But striking effects can be achieved in deprived areas; at the Elliotsville School, for example, the average grade-placement increase for a fourth-grade class of 27 students was 7 months after only 1½ months' work at teletype terminals.

It would be a mistake, however, to conclude that it is only with deprived or slower students that computer-assisted instruction will show really effective results. The program in Russian at Stanford University provides clear evidence to the contrary. There is much about the teaching of a foreign language that is particularly well suited to computer-assisted instruction. To keep pace with the programmed exercises, the student must concentrate more directly on the language and not return to an internal monologue in English as he listens to other students respond in a class of 20 or 30. The concentration required of the student at computer-based terminals in the Russian program precludes inattention; thus he achieves a degree of efficiency, it seems to us, that would be difficult to match in even the best-organized classroom. This is not to say that our Russian program is without defects. Van Campen plans a large number of improvements that will further deepen the degree of individualization.

We do feel, however, that, at both the secondary and the college levels, computer-based instruction can take over a good deal of the teaching of a foreign language, especially in languages for which the teaching staff is inadequate. From the standpoint of national interest, we need increasing instruction in Russian, Japanese, and Chinese, and yet the staff for teaching these three languages is not generally sufficient, particularly in secondary schools.

Another example not discussed here, but one that provides clear evidence that the benefits of computer-assisted instruction are not restricted to the deprived or to slower learners, is some of our work in logic and algebra at the elementary- and beginning secondary-school levels. We have not provided a classical evaluation of this program, which was one of our first curriculum efforts and began with demonstrations in December 1963. There is no good direct comparative evaluation of control-group performance, since this body of curriculum material is not offered in ordinary classes. There is no doubt, however, that this program, which is primarily aimed at bright students in grades 4 to 8, has been effective, because a great many mathematical ideas and skills have been learned by students who would not otherwise have been exposed to the material. One of our finest examples (although the evidence is anecdotal) is the rapid progress made in the logic program by students in Mississippi in comparison with students from upper middle-class environments in Palo Alto. We are especially proud of two Mississippi Negro boys in the eighth grade who stood at the top of the first-year logic program during 1967–68. This possibility of bringing enriched programs to students in a variety of environments where such courses cannot reasonably be offered by the teaching staff, either because of lack of time or because of lack of training, is probably one of the most immediately practical aspects of computer-assisted instruction. We want to conclude this article by emphasizing the important role of such enrichment programs, and to stress their importance, in spite of the fact that it is not easy to provide a classical, "hard data" evaluation of such programs.

### References and Notes


1. P. Suppes, M. Jerman, D. Brian, *Computer-assisted Instruction: The 1965–66 Stanford Arithmetic Program* (Academic Press, New York, 1968).
2. R. C. Atkinson, *Amer. Psychol.* **23**, 225 (1968).
3. T. Kelley, R. Madden, E. F. Gardner, H. C. Rudman, *Stanford Achievement Test* (Harcourt, Brace and World, New York, 1963).
4. G. M. Wilson, *Research in Conducting the Elementary School Curriculum* (National Education Association, Washington, D.C., 1925).


5. This study was supported by the National Science Foundation and the U.S. Office of Education. With respect to the drill-and-practice program, we thank M. Jerman and B. Searle of Stanford's Institute for Mathematical Studies in the Social Sciences, J. Prince, Superintendent of Schools, McComb, Mississippi, and E. R. Jones, Eastern Kentucky Education Development Corporation, Ashland, for their cooperation and assistance. With respect to the Russian program, we thank J. Van Campen and E. Belenky of the Institute for Mathematical Studies in the Social Sciences.


**NEWS AND COMMENT**

# DuBridge: Nixon's Science Aide Takes a Swing through Europe

*London.* The Nixon administration's blend of earnest rhetoric and political catalespy is not visible in detail from this distance. But some sense of the peculiar character of the administration is often conveyed to foreign places by one or another of the official delegations dispatched from Washington. With Nixon himself setting the trend, these come in increasing numbers now, thus raising the possibility that if, as was long ago noted by students of professional behavior, meetings are a substitute for work, then perhaps travel has become a substitute for meetings.

In any case, among the recent Washington-based touring companies was one headed by Nixon's science adviser, Lee A. DuBridge, traveling for the purpose of promoting the new era of European-American scientific and technical cooperation of which Nixon spoke during his own recent European visit. The DuBridge tour, which ended here 8 October after 3 weeks of travel through France, Romania, Yugoslavia, Belgium, and the Netherlands, did not set any new marks on the hokum index. That would have required both tastes and skills far beyond those possessed by DuBridge or his entourage, which included former AEC Commissioner Gerald Tape; Lewis Branscomb, director of the National Bureau of Standards; Herman Pollack, the State Department's director of international scientific and technological affairs; and David Beckler and Norman Neureiter, of the Office of Science and Technology. But it may be inferred that, with research budgets being gutted, DuBridge left Washington with instructions that amounted to, "All praise for cooperation, but don't buy anything expensive."

In Paris, for example, after several days of touring and conferring with French scientific and technical leaders, a press conference was called at which DuBridge and the French Minister of Industrial and Scientific Development, François-Xavier Ortoli, announced in a joint statement that they had held discussions "devoted to a review of the status of existing cooperative projects and the means for developing new areas of scientific exchange. . . . They decided, for instance," the statement continued, "to increase exchanges in disciplines and technologies relating to environmental and urban problems [and] that additional agreements should be concluded between interested agencies or institutes of the two countries as soon as the necessary preparatory discussions are completed." DuBridge then stepped forward to praise what he had seen in France, especially the Concorde supersonic airliner and oceanographic work in the Marseilles region. He said, "Our two countries share many concerns and we intend to establish closer scientific and technical relations in these areas." Among the areas cited