DOCUMENT RESUME

ED 111 884 TM 004 854

AUTHOR Nesset, Bonna; Faunce, R. W.

TITLE Selection of a Mathematics Test for a Citywide Testing Program.

INSTITUTION Minneapolis Public Schools, Minn. Dept. of Research and Evaluation.

REPORT NO C-72-36
PUB DATE Jan 74
NOTE 27p.

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS *Achievement Tests; *City Wide Programs; Comparative Analysis; Elementary Education; *Mathematics; Teacher Attitudes; *Testing Programs; *Test Selection
IDENTIFIERS *Minneapolis Public Schools

ABSTRACT

Teacher dissatisfaction with the existing math testing program in the Minneapolis Public Schools (MPS) led to the establishment of the Committee on Citywide Mathematics Testing. The committee was formed to identify tests to be used, grade levels to be tested, and testing schedules. The committee recommended that tests be given in the spring instead of the fall, as before, that tests be given in grades 3, 6, and 8, and that a pilot study be conducted by the Research and Evaluation Department to select one test for each grade from among those tests which had passed the initial screening. Five tests were under consideration at the third grade level (the SRA, the CAT, the Stanford, and the Metropolitan Form F and Primary II), four at the sixth grade level (the SRA, the CAT, the Stanford, and the Metropolitan Form F), and three tests at the eighth grade level (the SRA, the CAT, and the Stanford). The Stanford series was eliminated because testing materials could not be obtained in time for the pilot test. Each of the tests was given in three classrooms. Each teacher administering one of the tests responded to a questionnaire on the test used. Similar ratings were obtained on all the tests for: face validity, reading difficulty, cultural bias, pupil motivation, administration instructions, and interpretation instructions. Based on the limited evidence from this pilot tests, it was recommended that the SRA be selected. (Author/RC)

Minneapolis Public Schools

Selection of a Mathematics Test for a Citywide Testing Program

Bonna Nesset
Administrative Assistant

R. W. Faunce
Director for Research and Evaluation

Ideas expressed in this report do not necessarily reflect the official position of the Minneapolis Public School Administration nor the Minneapolis School Board

January 1974
C-72-36

Research and Evaluation Department
Planning and Support Services
807 N. E. Broadway
Minneapolis, Minnesota 55413

Selection of a Mathematics Test
for a Citywide Testing Program

Summary

See Page

Teacher dissatisfaction with the existing math testing program in the Minneapolis Public Schools (MPS) led to the establishment of the Committee on Citywide Mathematics Testing. The Committee was formed in December 1972 to identify tests to be used, grade levels to be tested, and testing schedules. 1

After examining a number of tests, the committee recommended that tests be given in the spring instead of the fall, as before, that tests be given in grades 3, 6, and 8, and that a pilot study be conducted by the Research and Evaluation Department in June 1973 to select one test for each grade from among those tests which had passed the initial screening. Five tests were under consideration at the third grade level (the SRA, the CAT, the Stanford, and the Metropolitan Form F and Primary II) four at the sixth grade (the SRA, the CAT, the Stanford, and the Metropolitan Form F), and three tests at the eighth grade level (the SRA, the CAT, and the Stanford). The Stanford series was eliminated because testing materials could not be obtained in time for the pilot test. 1-2

Each of the tests was given in three classrooms which were randomly selected within schools which were themselves randomly selected from high, middle, and low income schools. The pilot study involved 590 students and 25 teachers (two of the 27 teachers scheduled to participate did not take part). Each teacher administering one of the tests responded to a questionnaire on the test used. Sampling limitations and the small number of teachers involved dictated that the results of this study be considered suggestive rather than conclusive. 2

With one minor exception, pupils from high income schools scored highest and pupils from low income schools scored lowest regardless of the test used. Errors in test scoring were made by all teachers involved in the pilot test. 6
14

Teachers' ratings did not vary substantially for most tests' characteristics. For practical purposes, similar ratings were obtained on all the tests for: face validity, reading difficulty, cultural bias, pupil motivation, administration instructions, and interpretation instructions. The CAT and the SRA were, however, the most favorably received, overall. The SRA yielded median percentile ranks which most consistently paralleled publishers' norm medians. 11
6

Noting the desirability of using the same test (different forms) at all three grade levels, it was recommended that if a decision can be delayed for another year, further study of the SRA, the CAT, and the Stanford should be made. If a decision must be made now, based on the limited evidence from this pilot test, it was recommended that the SRA be selected. 18

February 1974 Research and Evaluation Department

Contents

Minneapolis Public Schools

Selection of a Mathematics Test
for a Citywide Testing Program

In December 1972 the Committee on Citywide Mathematics Testing[1] was established by the Minneapolis Public Schools to review the mathematics component of the citywide testing program. Teacher dissatisfaction with the existing program, particularly dissatisfaction with the use of the Iowa Test of Basic Skills Modern Mathematics Supplement in inner-city schools, was the impetus for the establishment of this committee. Identifying tests to be used, grade levels at which tests were to be given, and testing times were the specific charges to the committee.

After examining a number of tests, the fourteen member committee recommended that the new tests be given in the spring--rather than in the fall which had been the custom--and that a pilot study be conducted in June 1973 to select one test from the five which had passed the initial screening. Grades 3, 6 and 8 were to be tested.

In March 1973, the Research and Evaluation Department (R & E) was asked to conduct a pilot study of the following tests:

Grade 3

SRA Achievement Series, 1972, Primary Level II
California Achievement Test, 1970, Level 2
Metropolitan Achievement, 1970, Primary II
Metropolitan Achievement, 1970, Elementary Battery, Form F
Stanford Achievement Series, 1972, Primary Level II

[1] Citywide Mathematics Testing Committee members: Ed Anderson, Don Borgeson, Charles Dorniden, Kathy Feller, Ruth Hart, Art Indelicato, Randy Johnson, George Keprios, Mary Lou Knipe, Elmer Koch, Dennis Lander, Sally Sloan, Mike Sundberg, Ross Taylor.

Grade 6

SRA Achievement Series, 1972, Green Level, Form E
California Achievement Test, 1970, Level 3
Metropolitan Achievement, 1970, Intermediate, Form F
Stanford Achievement Series, 1972, Intermediate Level II

Grade 8

SRA Achievement Series, Red Level, Form E
California Achievement Test, 1970, Level 4
Stanford Achievement Series, 1972, Advanced Battery

Initial plans were to contract an independent research organization to conduct a pilot test focussing on teacher reactions, opinions of mathematics department chairmen, a review of technical literature on the test, and test administration to a sample of students. These plans were discarded because of a lack of funds.

The actual study, conducted by R&D E Department personnel, provided only a partial answer to the questions asked. Tests were administered to a sample of students, and teachers administering these tests responded to a questionnaire on the tests they had used. Sampling limitations and the small number of teachers involved dictate that the results of this study be considered suggestive rather than conclusive.

The Pilot Test

Testing

Each of the proposed tests was given in three classrooms: one classroom in a high socioeconomic school, one classroom in a low socio-economic school and the third classroom in a middle income school.

The Stanford series was not used because testing materials could not be obtained in time for the June 1973 pilot test.

Schools were selected randomly from high, middle, and low-income ranges of the Minneapolis Public Schools Management Index for 1972. This Index is a management tool which provides a crude measure of socioeconomic status. Classrooms were selected randomly within each of the schools.

Table 1 lists the schools involved. Table 2 gives the number of students who completed the tests. Two classrooms were not tested due to a mix-up in a mailing of test materials. These classrooms were the third grade, low SE school for the California Achievement Test and the sixth grade low SE school for the Metropolitan-F Test. General test administration instructions are given in Appendix A.

Tests were hand scored by the teachers involved in the pilot test. Scoring was checked in the Research and Evaluation Department and discrepancies between teacher scoring and the R and E scores were reconciled by a third person.

Median percentile ranks for each classroom were computed by converting each student's raw score to a percentile rank score and then taking a median for the class. A similar procedure was used to obtain an overall percentile rank on each test for the three classes combined from the various socioeconomic levels. Appropriate publisher's norms were used for each test.

Teacher Reactions

Teacher perceptions were obtained from a questionnaire which asked about face validity, reading difficulty, cultural bias, ease of administration, ease of scoring, ease of interpretation and results, and student reactions to the tests. This information was obtained after the teacher had administered the test, scored the answer sheets, and interpreted the test results. Appendix B provides a sample of the questionnaire sent to the teachers.

Table 1

Schools Involved in a Pilot Test
of Proposed Mathematics Tests
June 1973

Grade 3

Test	High SES	Middle SES	Low SES
SRA	Armatage	Shingle Creek	Harrison
CAT	Wenonah	Putnam	Harrison
MET II	Armatage	Holland	Hay
MET F	Keewaydin	Shingle Creek	Harrison

Grade 6

SRA	Audubon	Hamilton	Hay
CAT	Audubon	Shingle Creek	Hay
MET-F	Keewaydin	Shingle Creek	Lyndale

Grade 8

SRA	Anthony	Ramsey	Lincoln
CAT	Southwest	Sheridan	Bryant

Table 2

Number of Students Involved in a Pilot Study of Four Mathematics Tests--June 1973

Grade \ SES Status	SRA				CAT				MET II				MET F			
	High	Mid.	Low	Total	High	Mid.	Low	Total	High	Mid.	Low	Total	High	Mid.	Low	Total
3	23	27	19	69	25	13	NA	38	18	22	23	63	25	25	17	67
6	27	30	19	76	25	33	24	82					31	35	NA	66
8	28	22	13	63	26	26	14	66								

NA= Not applicable; students not tested

Results

Table 3 gives median percentile ranks for each test at each grade level for the combined sample of students from middle, low, and high socioeconomic schools.

On the surface, the SRA appears to have yielded median percentile ranks which most consistently paralleled publisher's norms. Median percentile ranks were 58, 50, and 55 for grades 3, 6, and 8 respectively.

Median percentile ranks for the CAT ranged from 38 to 63. This larger range may have been due to sampling bias, since the absence of low income pupils in the third grade sample resulted in a sample of pupils from middle and high income schools. Such a biased sample would produce spuriously high average scores if test scores are correlated with socioeconomic level.

Despite a similar bias in sampling for the MET-F, the expected high median percentile rank did not occur. Sixth grade pupils who took MET-F ranked at the 46th percentile while the third grade pupil sample, which included low income students, ranked at the 23rd percentile.

Third grade pupils who took the MET-2 had a median percentile rank of 65.

Results showing the median percentile rank for pupils from high, middle and low socioeconomic schools are shown in exhibits 1, 2 and 3. With one exception, pupils from high income schools scored highest and pupils from low income schools scored lowest--regardless of the test used. Only with the MET-F was there a minor deviation from this general pattern. At the third grade level, pupils from the middle income schools scored slightly higher on MET-F than did pupils from the high income school. Pupils from the low income school scored substantially below the other two income groups however.

Table 3

Median Percentile Ranks for Minneapolis Pupils Based on Publisher's Norms June 1973

Test	Grade 3	6	8
SRA	58	50	55
CAT	63[a]	50	38
MET II	65	-	-
MET F	23	46[a]	

[a]Low SES sample of students not included.
Third grade CAT=38 students; sixth grade MET F=66 students.
Other tests ranged from 63 to 82 students at each grade level.

Median Percentile Ranks on Four Mathematics Tests for 3rd Grade Pupils in High, Middle and Low Income Schools

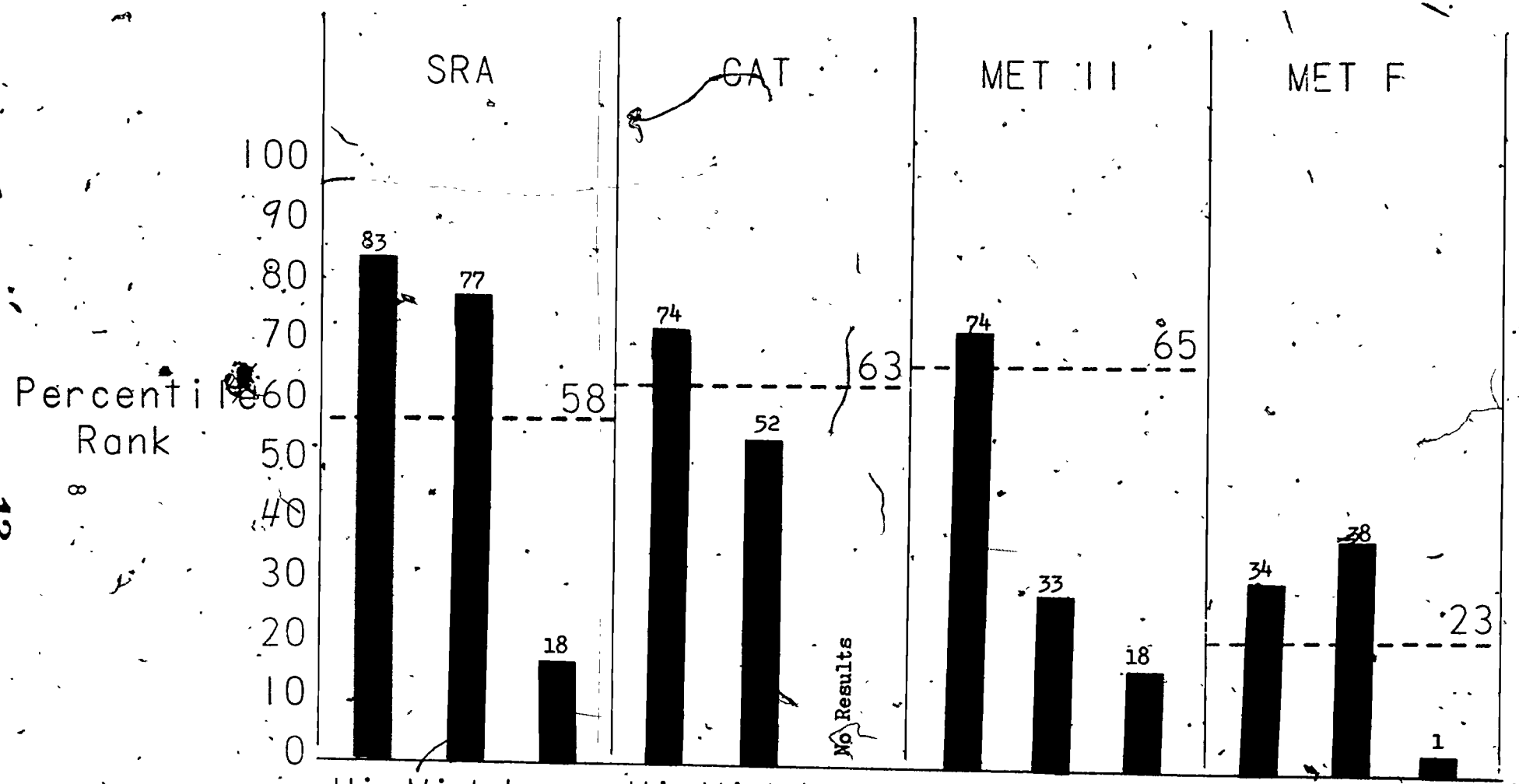


Exhibit I

Median Percentile Ranks on Three Mathematics Tests for 6th Grade Pupils in High, Middle and Low Income Schools,

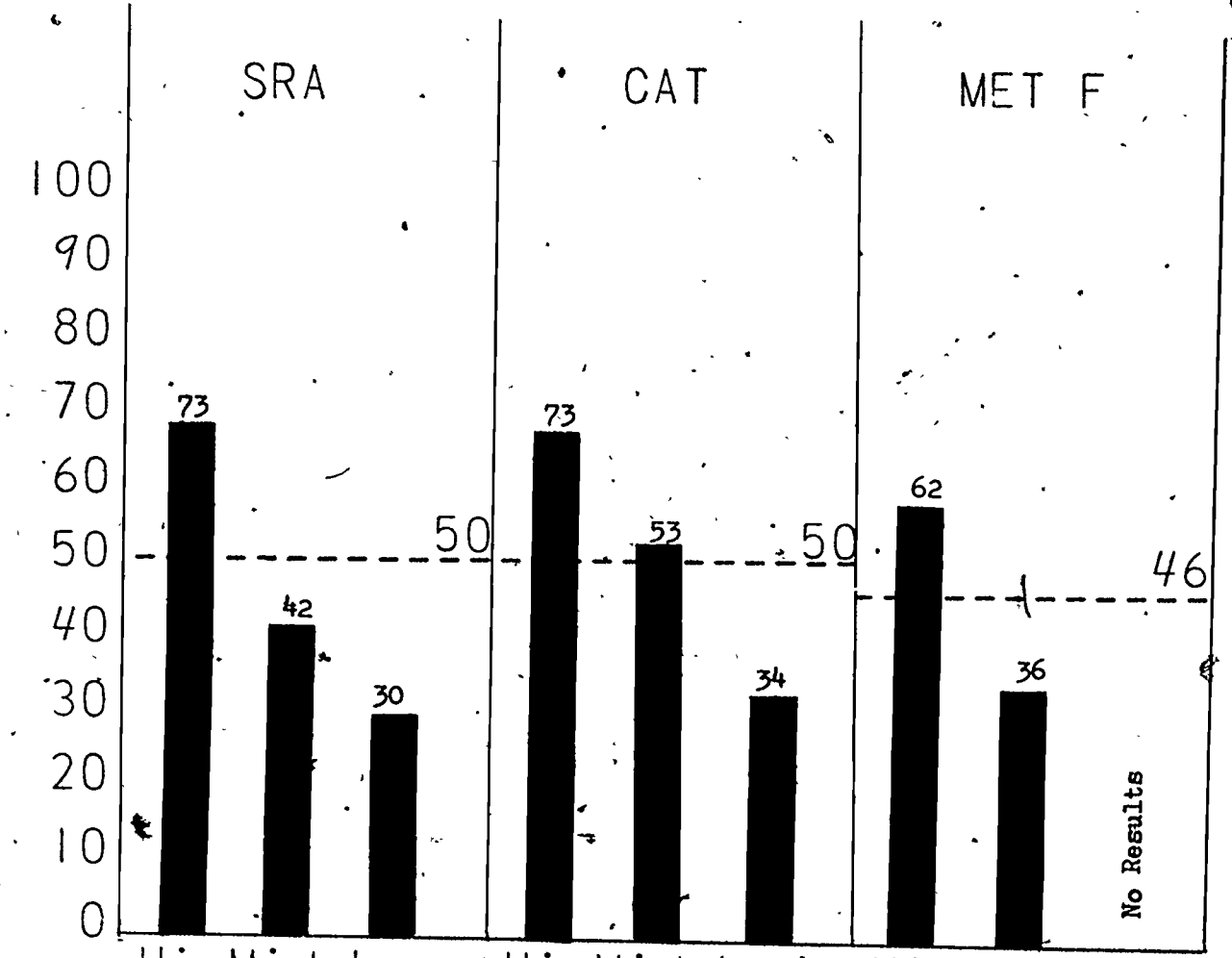
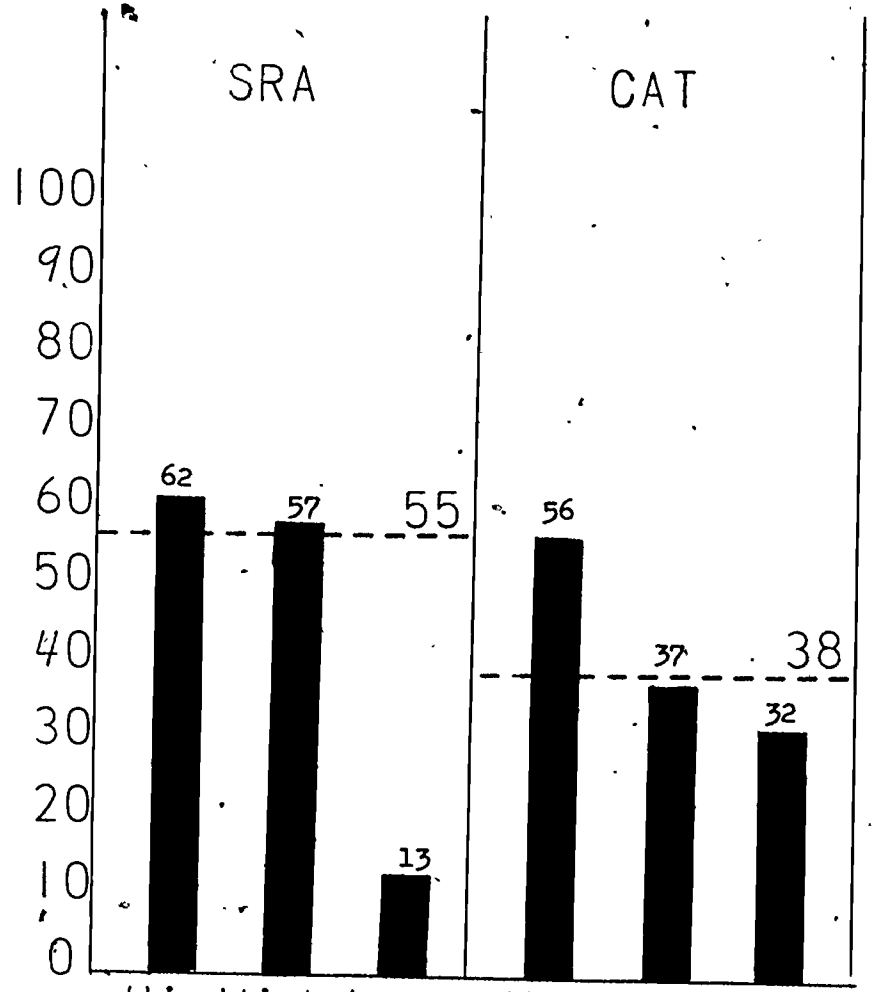


Exhibit 2

----- Class Median

Median Percentile Ranks on Two Mathematics Tests for 8th Grade Pupils in High, Middle and Low Income Schools



Percentile Ranks

Exhibit 3

----- Class Median

Teacher Ratings

Exhibit 4 summarizes the ratings of various test characteristics made by the teachers participating in the pilot study. A more complete picture of the exhibit entries may be obtained by referring to the teacher questionnaire in Appendix B.

Note that only two or three teachers are involved in making these ratings for each test at each grade level. Exhibit 4 attempts to present a composite picture of these ratings.

Teacher's ratings did not vary substantially for most test characteristics. For practical purposes, similar ratings were obtained on the four tests for: face validity, reading difficulty, cultural bias, pupil motivation, administration instructions, and interpretation instructions.

Ratings on a few characteristics did differ among the four tests.

SRA: No "unusual" ratings.

CAT: Some question about the adequacy of practice problems was raised. Hand scoring procedures were rated poor by eighth grade teachers but this may have been due to the fact that the appropriate scoring template was not available. Fewer teachers found it necessary to deviate from administration instructions on the CAT than on the other tests.

Met II: Test norm information was rated poor. Hand scoring rated poor to adequate. Two of the three teachers trying this test found it necessary to deviate from the administration instructions.

MET F: Practice problems were adequate, at best, according to teacher ratings. Three of five teachers deviated from administration instructions. A number of low income, third grade students were reported discouraged by the test. In addition students had a hard time grasping the directions concerning NG (not given) and DK (don't know).

Exhibit 4

Summary of Teacher Ratings of Various Aspects of the Tests

Item	Grade	SRA	CAT [b]	MET II	MET F [b]
Face Validity	3	Good	Adequate to poor	Good	Adequate
	6	Good	Adequate to good	-	Good
	8	Adequate to good	Adequate to good	-	-
Reading Difficulty	3	Easy for half to most	Easy for half to most	Easy for half to most	Easy for half to most
	6	Easy for most	Easy for most to all	-	Easy for most to all
	8	Easy for half to most	Easy for half to most	-	-
Cultural Bias	3	None reported	None reported	None reported	None reported
	6	" "	" "	- -	" "
	8	" "	" "	- -	- -
Practice Problems	3	Adequate	Questionable	Adequate	Instruction not clear
	6	Questionable	Questionable	-	Adequate
	8	Adequate	Adequate	-	-
Pupil Motivation	3	Most motivated	Most to all motivated	Half to most motivated	Half motivated
	6	Half to most	Most to all	-	Almost all
	8	Half to most	Half to most	-	-
Administration Instruction	3	Good to excellent	Good	Good	Good to excellent
	6	Good	Excellent	-	Good to excellent
	8	Good	Adequate to good	-	-
Deviations from Administration Instructions	3	Two of three teachers	None	Two of three teachers	One of three
	6	None	None	-	One of two
	8	One of three	One of three	-	-
Test Interpretation Instructions	3	Good to excellent	Good	Adequate to good	Good
	6	Adequate to good	Good	-	Adequate to good
	8	Good	Adequate	-	-

(continued)

Summary of Teacher Ratings of Various Aspects of the Tests (continued)

		SRA	CAT	MET II	MET F
Test Norm Information	3	Good	Good	Poor	Good
	6	Adequate	Adequate to good	-	Good
	8	Good	Adequate	-	-
Hand Scoring	3	Good	Excellent	Poor to adequate	Adequate to good
	6	Adequate to good	Adequate[a]	-	Adequate to good
	8	Adequate	Poor[a]	-	-
Students Discouraged	3	A Few	A Few	A Few	A Few
	6	A Few	Few or none	-	None
	8	Few or none	Few or none	-	-

[a]Proper scoring template not available; substitute used.

[b]Only two teachers rated the third grade CAT and the sixth Grade MET

Test Scoring

Errors in test scoring were made by all teachers involved in the pilot test (N=25). While 85% of all scores were within five percentile ranks of the correct rank, some errors were substantial. About one out of nine pupils (11%) received scores which were more than ten percentile ranks above or below their correct rank.

All pupils in one eighth grade class received teacher scores which were higher than their correct scores. All third grade pupils in one class, except one, received scores lower than their correct scores.

One pupil whose correct percentile rank was 18 received a teacher score of 62. Another pupil with a correct rank of 91 received a rank of 41.

Discussion

It should be clear that this rather cursory analysis of the various tests provides evidence which permits us to make only an educated guess as to which one might be most appropriate for use in the Minneapolis Schools.

The most substantial evidence we have for the various tests is the median percentile ranks for the various samples of students on each of the tests. However, we cannot conclude that the combination of the high, middle and low income samples used in this pilot study gives us an accurate reflection of the results we would obtain if the tests were used citywide. A more accurate determination would have to be made using a sample which had the same proportion of high, middle and low income children as in the total city population. The samples used in this test consisted simply of one classroom in each of the schools from the three income levels. If students from low income schools are more highly represented in the general population than they were in this sample, then the overall median rank for any of the tests would probably be lower.

The socioeconomic makeup of the pupil samples for each test may not be similar. Although the selection procedures for selecting the samples were the same for each test a determination of each child's economic status was not attempted.

The fact that students from high income schools consistently scored higher than students from middle income schools and that the middle income students, in turn, scored consistently higher than the low income students gives a strong indication that whatever is measured by the tests is highly correlated with socioeconomic status. The few teachers who used the tests, and the screening committee, reported no obvious evidence of cultural bias related to racial characteristics.

Screening committee members indicated the desirability of using the same test (different forms) at the three grade levels. This would not be possible if either of the Metropolitan tests were selected.

The selection of the Metropolitan tests does not appear desirable for other reasons. Median percentile ranks for the Minneapolis samples appear too high on MET II (65) and too low on MET F (23) at the third grade level. The Minneapolis percentile rank on MET II was extremely low even though the sample of low income pupils was not included!

Teachers rated several aspects of these two tests as poor (e.g. norm information and scoring for MET II; practice problems and students discouraged on MET F).

MET II and MET F should be eliminated from further consideration based on the information obtained from this limited pilot study.

A choice between the SRA and the CAT is difficult. Based on pilot study information, alone, the SRA appears to be somewhat more desirable. Median percentile ranks for Minneapolis pupils are closer to the 50th percentile on the SRA than they are on the CAT.

	Grade		
	3	6	8
SRA	58	50	55
CAT	63	50	38

These differences may be due to sampling differences. The low income pupils were not included in the third grade CAT. Their inclusion would probably lower the median percentile rank. The low percentile rank for the eighth grade CAT may be due to a sampling fluctuation.

Teacher ratings provide little credence for making a decision between those two tests. Only a few teachers were involved and no clear superiority for either test was indicated.

The very limited information gained from this pilot study should be seen in perspective. We do not know how Minneapolis teachers will react to either of these tests. We have had some screening of the tests by a committee representing teachers and we have had a very limited try-out in the field. On the basis of this information either of these two tests is worth further consideration.

Nevertheless, one can have no strong feeling of confidence that the selection of either of these tests will dispell all the problems encountered with previous tests. Nor can we be certain that the selection of either of these tests will necessarily be the most appropriate tests for our purposes. The Stanford Test which was not included in the pilot study, because materials could not be obtained in time, could be given further consideration if time permits. However, if the selection must be made now our best guess is, and its little more than a guess, that the SRA Test would be most suitable for Minneapolis purposes.

The errors observed in teacher scoring may not be indicative of errors which would occur once a test is selected for the citywide program. All tests in the pilot were new to the teachers involved. Staff training for test administration was not given. About 30% of the teachers deviated from the manuals' administration instructions. Teachers may not have been as careful in test scoring for a pilot study as they would be under conditions where the results would directly influence children.

Nevertheless, the errors made were so pervasive, and of such a magnitude, that caution and training will be required regardless of the test selected. Machine scoring will help eliminate some errors, but errors of administration, interpretation, recording and transferring could still occur.

Recommendations

1. If a decision must be made now, the SRA test should be selected.
2. If a decision can be delayed for another year, further study of SRA, CAT, and the Stanford should be made.
3. Additional study, if attempted, should include reactions of a larger number of teachers. Each teacher should have the opportunity to rate all tests being considered.
4. The Citywide Testing Committee should direct its attention to the problem of test scoring. If tests are not accurately scored and recorded then it will matter little which test is selected. It appears desirable to have tests machine scored at all grades.
5. The Committee should note three major assumptions underlying this pilot test and the resulting recommendations:
   a. Any test in the pilot study would be preferable to the test currently in use.
   b. It is desirable to use the same test (different forms) at grades 3, 6 and 8.
   c. Median scores by Minneapolis pupils should approximate the median scores in the publisher's norms.

Minneapolis Public Schools

INSTRUCTIONS

Enclosed are the necessary materials for administering the mathematics section of _____ to your class: _____ copies of the test booklet, _____ answer sheets, and the examiner's manual and scoring key.

Please familiarize yourself with the directions for administering the test ahead of time, and do not deviate from the directions in the manual during the testing session.

The tests are to be administered to all students on whichever day is most convenient for you during the week of June 4. We would prefer that they be given in the morning. Those eighth grade teachers who have more than one eighth grade standard math class are requested to administer the test to whichever standard class meets during the earliest hour of the day. No basic or enriched math classes are to be involved in this study.

Eighth grade teachers administering the SRA should give the CONCEPTS section one day and the COMPUTATION section on the following day.

Eighth grade teachers administering the California Achievement Test should give the COMPUTATION section on one day and the CONCEPTS and the PROBLEMS sections on the following day.

Please hand score the tests for your class with the scoring key provided. Ultimately, the tests used in the citywide testing program will be machine scored, but we are evaluating the tests' hand scoring procedures because the Committee on Citywide Mathematics Testing felt that teachers should, if they desire, be able to obtain immediate results for certain of their students by hand scoring their tests. Please score these tests according to the directions provided and ignore any instructions which refer to preparation of answer sheets for machine scoring.

Scoring the tests for your class should take about an hour or less of your time. You will be reimbursed by the Research Department for this time. Please send a service report to the Research and Evaluation Department, 807 N. E. Broadway, Minneapolis 55413. Don't forget to put your address on the service report so we can mail you a check.

Return <u>all</u> these testing materials and your <u>signed</u> service report to the Research Department by Thursday, June 14.

5/73

Research and Evaluation Department

Appendix B

Teacher Questionnaire

Minneapolis Public Schools

TEST _____

SCHOOL _____

GRADE _____

TEACHER _____

NUMBER OF CHILDREN TESTED _____

Please read the following questions and check the appropriate response. Feel free to comment in the space provided after each question.

1. How well do the items in this test reflect the content of the curriculum taught?

_____Excellent coverage of curriculum

_____Good coverage of curriculum

_____Adequate coverage of curriculum

_____Poor coverage of curriculum

_____Very poor coverage of curriculum

Comments: _____

_____

_____

2. How appropriate was the reading difficulty level of this test for your class? (Third grade teachers: were the directions read to the students easy for them to understand?)

_____Easy enough for everyone in the class

_____Easy enough for most of the class

_____Easy enough for about half of the class

_____Too difficult for most of the class

_____Too difficult for everyone in the class

Comments: _____

_____

3. Do you think this test is biased against minority or low income students?

_____Yes

_____No

Comments: _____

_____

4. Do you think adequate practice problems were provided when necessary?

_____Yes

_____No

_____Weren't necessary

Comments: _____

_____

_____

	All	Most	About Half	A Few	None
5. For how many of your students was the form or layout of the test booklet confusing or hard to follow?	____	____	____	____	____
6. For how many of your students was the form or layout of the answer sheet hard to follow? (leave blank if none were used)	____	____	____	____	____
7. How many of your students appeared to be motivated and interested in doing their best during the testing?	____	____	____	____	____
8. How many of your students do you think this test left feeling discouraged or defeated?	____	____	____	____	____
How many of your students did you notice engaged in the following kinds of activities during the testing session?					
9. Asking questions of the teacher during the timed session	____	____	____	____	____
10. Talking with other students despite instructions to the contrary	____	____	____	____	____
11. Being generally inattentive to the testing--staring out the window, doodling, etc.	____	____	____	____	____
12. Crying	____	____	____	____	____
13. Refusing to take test	____	____	____	____	____

14. Did you find it necessary to deviate at all from the manual's directions for test administration at any time?

_____No

_____Yes If Yes, why?_____

_____

	Excellent	Good	Adequate	Poor	Very Poor
15. How would you rate the directions for administering the test in the examiner's manual?	____	____	____	____	____
16. How would you rate the adequacy of the information presented on test interpretation?	____	____	____	____	____
17. How would you rate the manual's discussion of test norms in terms of its helpfulness in interpreting results?	____	____	____	____	____
18. How would you rate the ease of the hand scoring procedures presented?	____	____	____	____	____

May 1973

Research and Evaluation Department

The particular test which you tried out was one of the possible alternatives to the ITBS Modern Math Supplement. It has not been predetermined that the ITBS should necessarily be replaced with another standardized test of this form. The K-12 Committee on Citywide Math Testing is considering all alternatives. Do you personally have any recommendations as to what form of assessment would be most appropriate?