

1987

## Using standardized tests scores for better education of students

Linda W. Wackwitz  
*University of Northern Iowa*

*Let us know how access to this document benefits you*

Copyright ©1987 Linda W. Wackwitz

Follow this and additional works at: <https://scholarworks.uni.edu/grp>



Part of the [Education Commons](#)

---

### Recommended Citation

Wackwitz, Linda W., "Using standardized tests scores for better education of students" (1987). *Graduate Research Papers*. 3503.

<https://scholarworks.uni.edu/grp/3503>

This Open Access Graduate Research Paper is brought to you for free and open access by the Student Work at UNI ScholarWorks. It has been accepted for inclusion in Graduate Research Papers by an authorized administrator of UNI ScholarWorks. For more information, please contact [scholarworks@uni.edu](mailto:scholarworks@uni.edu).

**Offensive Materials Statement:** Materials located in UNI ScholarWorks come from a broad range of sources and time periods. Some of these materials may contain offensive stereotypes, ideas, visuals, or language.

---

## Using standardized tests scores for better education of students

### Abstract

The push for excellence in education has accelerated during the 1980's. As a result, there is renewed emphasis on using standardized achievement tests in American schools. Every school district in the United States requires some sort of standardized achievement test (Sproull & Zubrow, 1981). In a report to the 1980 National Invitational Conference, Jennie Yeh, Joan Herman, and Lawrence M. Rudner stated that every child will complete at least six full batteries of standardized achievement tests before graduating from high school. The results and use of these results has become a national concern to parents, teachers and others involved in educating children in the United States.

# Using Standardized Tests Scores for Better Education of Students

A Graduate Project  
Submitted to the  
Department of Curriculum and Instruction  
In Partial Fulfillment  
of the Requirements for the Degree  
Master of Arts in Education  
UNIVERSITY OF NORTHERN IOWA

by

Linda W. Wackwitz

Summer, 1987

This research paper by: Linda W. Wackwitz  
Entitled: Using Standardized Test Scores for  
Better Education of Students

has been approved as meeting the research  
paper requirement for the Degree of Master  
of Arts in Education.

May 5, 1987  
Date Approved

Mary Nan Aldridge

Director of Research Paper

May 5, 1987  
Date Approved

Mary Nan Aldridge

Graduate Faculty Adviser

July 13, 1987  
Date Approved

Marvin Heller

Graduate Faculty Reader

July 15, 1987  
Date Approved

Greg P. Stefanich

Head, Department of  
Curriculum and Instruction

## Table of Contents

Chapter 1: Introduction.....	4
Chapter 2: Review of Literature and Research.....	8
Use of Standardized Achievement Tests.....	8
Attempts to Improve Standardized Achievement Test Scores.....	12
Reanalysis of Achievement Test Scores.....	15
Research Involving Teachers.....	16
Research Involving Parents.....	19
Uses and Misuses of Standardized Achievement Test Scores.....	22
Use of Standardized Achievement Test Scores.....	22
Misuse of Standardized Achievement Test Scores.....	24
Chapter 3: Procedures.....	28
Chapter 4: Summary and Conclusions.....	30
Awareness of Tests and Scores.....	31
Parents and Test Use.....	33
Teachers and Test Use.....	34
More Usable Test Results Are Available.....	36
Appropriate Uses of Standardized Achievement Tests.....	36

Limitations of Standardized	
Achievement Tests .....	39
Standardized Testing Narrows Curricula....	40
Other Measures of Student Progress.....	41
Discussion.....	42
Conclusions .....	43
Limitations.....	44
Recommendations.....	45
Reference.....	46

## Chapter 1: Introduction

The push for excellence in education has accelerated during the 1980's. As a result, there is renewed emphasis on using standardized achievement tests in American schools.

Every school district in the United States requires some sort of standardized achievement test (Sproull & Zubrow, 1981). In a report to the 1980 National Invitational Conference, Jennie Yeh, Joan Herman, and Lawrence M. Rudner stated that every child will complete at least six full batteries of standardized achievement tests before graduating from high school. The results and use of these results has become a national concern to parents, teachers and others involved in educating children in the United States.

The Nation At Risk report, published in 1983, states that:

Standardized tests of achievement (not to be confused with aptitude tests) should be administered at major transition points from one level of schooling to another and particularly from high school to college work. The purposes of these tests would be to : a) verify the student's credentials; b) identify the need for remedial intervention; and c) identify opportunity for advanced or accelerated work. The tests should be administered as part of a nationwide (but not Federal)

system of State and local standardized tests. (p. 28)

The writers of this document place so much importance on standardized achievement tests, that 11 of the 14 "indicators of risk" (p. 5) deal with tests scores.

Kellaghan, Madaus, and Airasian (1982) state that there are currently two schools of thought on standardized achievement tests. The first position is that testing is useful to teachers. It provides information that can be used in determining classroom organization, selection of curriculum materials, and teaching methods. These tests should improve achievement. The second position is that testing has negative effects on classroom methods and students. They result in a narrowing of the curriculum, unchangeable grouping situations, labeling, and greater pupil failure rates.

The concerns of many teachers about standardized tests are similar to those expressed by Walt Haney (1985) in his article "Making Testing More Educational". He states that standardized achievement tests do not promote learning because 1) students do not receive immediate feedback, 2) standardized tests do not have much relationship to what is taught and what is learned, and 3) test programs, which are a threat to kids, teachers, and schools, put pressure on them to achieve - no matter what method is used.

Parents, if not now concerned, should be. Standardized



tests and their results could have an adverse effect on their children's education and future life.

Standardized achievement tests take a great deal of time to administer, and teachers are required to report the results to parents, but little is done to use the test results. Central office administrators feel that test results are useful to teachers and principals (Sproull & Zubrow, 1981). Teachers feel that the results are important, but very few use them for diagnostic purposes (Salmon-Cox, 1981). Parents find results useful (Resnick, 1981) but often misunderstand them (Hopper, 1977).

Precisely how these tests are used and how they should be used is a dilemma currently facing educators and the general public. Three things must be done to make test results more usable to parents and teachers. First, test results must be of some diagnostic value to teachers. Scores need to be usable for improving instructional methods of teachers and therefore improving achievement of students. Second, there needs to be a simple, comprehensible way to report standardized achievement test scores to parents so that they can be part of the process to improve achievement test scores. Third, more care must be given when interpreting test results. Analysis of test scores needs to be more thorough so that score trends can be understood by parents, teachers, and the

general public.

In this paper, a standardized achievement test is defined as a test which is norm-referenced and given to students at regular intervals in order to document academic progress. The intent of this paper is to investigate what is currently being done to improve test scores, to discover how these scores are used, and in some cases, misused. Also, it is important to know if test publishers and educational experts are in agreement as to how test scores should be used. Another area to be studied is teacher and parent attitude toward standardized testing. Finally, it is important to know the effect of standardized testing on curricula.

## Chapter 2: Review of Literature and Research

Literature on improving standardized test scores falls into several categories: Several studies report what is currently being done in the area of standardized testing and what is being done to directly influence achievement test scores. Other studies explain re-examination of test scores and how more in-depth study can show different outcomes than when simply using percentile rank and grade equivalent scores. Several researchers have studied more effective reporting procedures to make test results more useful to teachers and parents.

Much of the literature on this topic is expert opinion, little reflects basic research. Issues of use and misuse of standardized achievement tests and results have been addressed by many writers.

### Use of Standardized Achievement Tests

Concerned about how teachers used standardized achievement tests, Yeh et al. (1980) conducted a study in five California school districts in various locations and with different socio-economic populations. Kindergarten through sixth grade teachers were sent questionnaires. Two hundred, sixty were completed.

The first finding of this study was that though some type of test was used at the beginning of each school year to

determine reading and math placement, standardized tests and curriculum embedded tests were used the least. Observation, oral quizzes, oral reading, or teacher-made tests were used more frequently. Standardized tests were least used by responding teachers.

Some type of test was used for reporting to parents and staff and for evaluating teaching methods and materials. Standardized tests were used only as a comparison while other types of tests were used to make instructional decisions.

Teachers felt that the greatest influences on standardized test scores were test-taking skills, motivation, quality of the test, and outside influences (Yeh et al., 1980). Teachers felt that parents had little influence on pupil performance. Yeh et al. (1980) discovered through this study that only 50 percent of the teachers questioned could correctly interpret percentile and grade equivalent scores.

Haney (1985) studied 30 schools that used standardized tests for curriculum modification. Of the 30 programs, four were identified as being exemplary of testing practices. The Portland, Oregon, Public School System has developed its own tests. These tests demonstrated student knowledge of the required curricula. The major use of the test results was for instructional remediation and adaptation. Teachers had input into developing tests and advise changes. The Orange County,

Florida, Public School System used a standardized achievement test which focused on reading comprehension. Test information guided subject-matter teachers to evaluate and improve instruction.

In Pittsburgh, Pennsylvania, the improvement of student achievement was a priority in a needs assessment. A series of three to six tests were administered each year in math, grammar, composition, reading, critical thinking, and science. This testing program was viewed as instructional, not administrative in nature. Teachers and parents have become more involved. Prospect School, a small private school in Vermont, uses no standardized testing. Extensive records are kept on each student. Student data includes observation by teachers and samples of children's writings and drawings. It is felt that these are "more comprehensive" (Haney, 1985, p. 11) ways of evaluating a student's progress. Records are kept in narrative form rather than in checklist form. When a child leaves the school, he/she is regarded a personality rather than a series of numbers and test scores.

Fifty-eight administrators in eighteen school systems were interviewed in Pennsylvania (Sproull & Zubrow, 1981). Administrators considered tests to be regarded as important to individual building personnel. Teachers were believed, by these central office administrators, to make the greatest use

of test scores. Central office administrators believed that teachers would miss these tests the most if they were discontinued, followed by students, parents, and the general community. Most central office administrators use standardized test scores as a "snapshot" (Sproull & Zubrow, 1981, p. 630) of the school system.

## Attempts to Improve Standardized Achievement Test Scores

In San Diego, California, in 1980, Judge Louis M. Welsh ruled that the San Diego Unified School District had to raise achievement test scores (Nagel, 1986). The target of this ruling was minority isolated schools. He believed that the long term solution to desegregation was to improve children's academic achievement in basic skills. He mandated that 50 percent of the students in these schools must score at or above grade level.

Comprehensive Tests of Basic Skills (CTBS) were reviewed for fifth-graders from 1975 to 1985. A time-series design of teaching and testing was used to determine the effects of the program. Total reading, total math, and total language scores were used for comparison of pre- and post-intervention data. Every student received 30 minutes of direct teaching instruction on a daily basis. As a result, 90 minutes of the school day was devoted to reading, 30 minutes to language, and 60 minutes to math. There was a deliberate decrease in time spent on social studies, science, and other areas of the curriculum which were considered non-basic. If a student did not receive 80% or better on the post-unit test, reteaching was done. Scores of 80% or better allowed a student to participate in enrichment activities.

Nagel (1986) pointed out that along with reconstruction of academic focus in the school district, a number of other essentials were necessary to improve basic skills scores. These included maximum use of time, definite objectives, use of materials and practices that were in line with the objectives, homework, direct teacher contact with students, regular assessment, student responsibility and involvement, high standards, staff in-service, safe environment within the school, recognition, sense of community, and home-school involvement.

As a result of this program, scores on CTBS rose 25 percentile points while the total San Diego Unified School District rose only nine percentile points in reading. An increase of 38 percentile points was found in language compared to 24 district-wide. Math scores also rose after intervention. There was a 29 percentile point improvement in minority-isolated schools compared to 16 in the entire district. In order to make this type of intervention succeed, Nagel (1986) calls for the best possible materials and teachers trained to use them.

The New Jersey State Department of Education has developed a Basic Skills Improvement Program to strengthen the development of basic skills at the elementary school level. Children who do not meet minimum proficiency standards on



yearly mandated basic skills tests are eligible for extra assistance. Funds for this program come from federal, state, and local resources.

In a booklet published by the New Jersey State Department of Education, parents are encouraged to help their child. They are given suggestions for increasing opportunities to learn and for improving self-concept. Also listed in the booklet are many activities for parents to engage in with their children. Finally, suggestions are made concerning how parents can become involved in their children's education at school.

In an article by Linda Lantor, published in the Des Moines Register on October 11, 1986, cash awards were available in some schools for students who received the highest scores on the Iowa Test of Educational Development. At Maquoketa Valley High School in Delhi, Iowa, a \$10 cash award was to be given to the six highest scoring freshmen and to the three highest scoring sophomores, juniors, and seniors. The article goes on to say that other incentives, such as a day off from school or cutting in front of others in the cafeteria line, are being offered to students in South Tama Community Schools if they have high composite score increases on achievement tests.

### Re-analysis of Achievement Test Scores

In examining the "average" achievement test score, Ligon and Wilkinson (1985) made several discoveries. They pointed out that though, as a whole, the Austin, Texas, School District was considered to be average, individual school composition must be considered. Several case studies were presented to illustrate this point.

In one elementary school in Austin, 31% of the students were high-achievers (above the 75th percentile on the Iowa Tests of Basic Skills), while 30% were low-achievers (below the 25th percentile). Reporting a 50th percentile median score did not give a true picture of the make-up of the school.

In another case study, the median decreased from the 47th to the 46th percentile while the number of students scoring below the 25th percentile decreased by six percent. Therefore, the school was not given credit for these gains. A one percentile drop for the entire school did not seem significant. However, raising test scores of six percent of students above the 25th percentile deserved more attention.

Ligon and Wilkinson (1985) concluded that average scores do not present a complete picture and are frequently misleading. Care must be taken so the wrong conclusions are not reached.

### Research Involving Teachers

Kellaghan, Madaus, and Airasian (1982) summarized a study by Beck and Stetz which was conducted in 1978 and 1979. Thirty-five hundred teachers in kindergarten through twelfth grade were interviewed. Two-thirds of those felt that the amount of standardized testing was about right. Sixty percent thought that additional training of school personnel was needed to provide correct interpretation of the scores. Twenty-one percent favored a moratorium on standardized testing.

Teachers found standardized tests useful for determining educational growth of an individual (77%), detecting system-wide weaknesses (73%), planning class instruction (65%), reporting to parents (63%), and planning instruction for the individual (61%). Other ways in which teachers thought the tests could be used were for: comparing students with their national peers (54%), screening for special education (51%), help in evaluating teaching procedures or methods (36%), comparing schools within a system (33%), comparing classes in a school (26%), evaluating teachers (19%), and reporting to newspapers (8%).

In a study by Salmon-Cox (1981) 68 teachers were surveyed. She found that their use of standardized achievement tests fell into several categories. First, just

less than half of the teachers reported that the results helped them to confirm information that they already possessed about the student. Second, information gained from the test results was used as a guide to what needed to be taught. Third, teachers felt that the scores were used to confirm that children had been grouped correctly. For example, grouping appropriately for reading, was checked. In general, standardized achievement tests were used as a supplement, to confirm what teachers already knew about the student.

Some educators do not feel that enough information is provided by test scorers to enable them to use standardized achievement test results for valid decision-making (Singer and Dreher, 1983). In one study, educators were asked which of four types of test results they would prefer. The choices were (a) norm-group comparison only, (b) self-comparison only, (c) norm-group and self-comparison, and (d) norm-group and self comparison with some content information. The preferred type of test result was norm-group and self-comparison with some content information. They felt that this type of report could help to explain the child's reading performance to parents, show parents that their child had made progress since the previous year, and help parents choose reading material for their children to use at home. It could help teachers plan instruction for children and aid in appropriate ability grouping

within classrooms.

According to Yeh et al. in their report to the NIE Invitational Conference on Testing in 1980, many teachers do not know how to explain test results. In a study of 260 teachers in five California school districts, only 50% were able to correctly interpret percentile rank and grade equivalent scores, the two most common ways of reporting test scores to parents.

In interviews of 68 teachers (Salmon-Cox, 1981), it was found that observation, teacher-made tests, and interaction with students were used to determine student skill level. Even though the teachers were aware of the interviewer's emphasis on standardized achievement tests, only 3 of 68 mentioned use of these tests as a useful assessment tool.

Teachers would be willing to do more testing if useful information were supplied (Singer & Dreher, 1984). The types of information currently furnished to teachers, provide no suggestions for the improvement of instruction (Dreher & Singer, 1984). Test results tend to focus on weaknesses rather than strengths (Haney, 1985).

"The Florida State Merit Pay Commission flirted with creating an educational Frankenstein by grafting merit pay to student performance as measured by test scores" (Madaus,

1985, p. 614). According to Madaus (1985), using test results to reward teachers is nothing new. Scores were used in 15th century Italy, in 18th century England, and again the 19th century America. Yeh et al. (1980) noted that, "Although teachers felt test scores to some extent reflected their instructional efforts, they felt test-taking skills and motivation and test quality were more influential" (p. 54).

Linking test scores to teacher rewards has some inherent flaws that can erode educational quality. The pressure for high scores can distort teacher judgment regarding sound curriculum.

#### Research Involving Parents

Gallup (1979) reported that 75% of parents interviewed found testing programs to be useful. Using tests as a graduation requirement was supported by 65% of those questioned. Singer and Dreher (1985) found that 25% of parents felt that typical test results were useless.

In order to determine if parents understood and retained information presented to them at parent-teacher conferences, Hopper (1977) sampled parents from a small, Iowa city with a higher socio-economic population. Parents were asked to recall stanine scores and groups to which their children were compared. Only one-third of the parents correctly recalled their child's stanine score. However, 40% were only one

stanine off. Parents of high- achievers underrated their child's performance, while parents of low-achievers overrated performance. Only 20% of parents correctly identified the groups with which their child was compared. Hopper (1977) concluded that parents are not accurate in their recall of their child's test performance or with whom their child was compared.

Singer and Dreher (1985) conducted a study to investigate which type of test result was preferred by parents. They were asked to choose between (a) norm-group only, (b) self comparison only, (c) norm-group and self comparison, and (d) norm-group and self comparison with some content information. Of 24 parents, 20 agreed that they would prefer a report containing norm-group and self comparison with some content information.

In this study (Singer & Dreher, 1985), an 18-item questionnaire was completed by the parents. Parents were asked to complete the questionnaire twice: first, keeping in mind the most preferred, and second, the least preferred type of test result. Parents felt that the norm-group, self comparison report helped them to understand their child's reading performance, told them how successfully their child was learning to read, and explained how their child's reading performance had improved compared with past scores and

other children's scores. This type of report, with some content information, would help parents determine what instruction their child would need, tell if their child could read more difficult material than last year, and help them to select materials for the child to read at home. Singer and Dreher (1985) found that the more informative reports generated favorable opinions from parents about the performance of schools relative to teaching reading.



## Uses and Misuses of Standardized Achievement Test Scores

Experts categorize test use into several areas. These general categories are individual analysis of students, group analysis of students, administrative procedures, teacher evaluation, and score reporting. Examples cited earlier in this chapter fall into one or more than one of these categories.

### Use of Standardized Achievement Test Scores

Individual analysis of student scores is a major use of standardized achievement test scores. Scores determine level of development (Hieronymous, Lindquist, & Hoover, 1982), extent of learning (Ebel & Frisbie, 1986), or gauge academic growth (Ravitch, 1983-84). Scores indicate trends individual strengths and weaknesses (Hieronymous, Lindquist, & Hoover, 1982; Ravitch, 1983-84) and serve as a warning of possible learning difficulties (Ravitch, 1983-84). Scores can be used to provide feedback to teachers and pupils (Hieronymous et al., 1982) and to help establish individual goals.

Tests can be diagnostic if students are told what they missed (Ebel & Frisbie, 1986). Standardized achievement tests should be used to guide and improve individual instruction (Hieronymous et al., 1982; Salmon-Cox, 1981; Singer & others, 1983). Instructional materials can be chosen using test scores as a reference (Hieronymous et al., 1982; Ebel & Frisbie, 1986). These tests can be used to reveal differences between ability

and performance (Hieronymous et al., 1982; Salmon-Cox, 1981; Singer & others, 1983). These data, when used by teachers, can strengthen attempts to individualize instruction.

Experts also list several uses of standardized achievement test scores in dealing with groups. Test scores can be used to evaluate curricula (Sproull & Zubrow, 1981; Ebel & Frisbie, 1986). Scores are one part of the evaluation process (Ebel & Frisbie, 1986). Compiled test scores can determine group strengths and weaknesses (Hieronymous et al., 1982). Therefore, scores can be used as guides to improve group instruction (Ebel & Frisbie, 1986; Hieronymous et al., 1982).

Administrative uses of standardized achievement test scores is another frequently mentioned category. Scores can be used to make decisions about grouping students (Hieronymous et al., 1982; Ravitch, 1983-84; Salmon-Cox, 1981). Another use is for admission to programs and schools (Singer & others, 1983).

Determining instructional effectiveness (Hieronymous et al., 1982) is a particularly controversial use of test scores. Experts (Hieronymous et al., 1982; Singer & others, 1983) suggest that this should only be used if other factors, such as learning ability of pupils, school environment, quality of materials, and general morale are taken into account.

Test scores can also be used to report to parents and

inform the general public (Hieronymous et al., 1982; Ebel & Frisbie, 1986). Parents can use these scores to make educational decisions, such as in which school to enroll children (Resnick, 1981). There is an associated danger in this practice as personnel strive to create elitist schools.

### Misuse of Standardized Achievement Test Scores

Gardner (1985) discusses many ways in which tests and test scores can be misused. Some of these uses deal with the test itself, while others deal with the ways in which test results are allowed to influence curricula and public opinion.

One misuse is to accept a test on the basis of the title alone (Gardner, 1985). Basic skills cover only limited skills and ignore learning areas such as decision making, development of self-concept (Stedman & Kaestle, 1985) and higher order thinking skills (Ravitch, 1983-83; Stedman & Kaestle, 1985). According to Salmon-Cox (1981), these tests measure some of the academic goals, but none of the social ones which are taught in classrooms. Standardized achievement tests do not measure enough of the skills being taught. Berlak (1985) is in agreement with this when he states:

Virtually all standardized achievement and competency test batteries ignore significant aspects of the school curriculum - most notably writing, the graphic and

performing arts, and the ability to use critical thought and to engage in subtle and complex interpretation and analysis. (p. 17)

Another misuse deals with bias (Gardner, 1985); the drive for high test scores has been very hard on low socio-economic children (Stedman & Kaestle, 1985). Tests are also biased against children who are not familiar with the language or are not of the majority culture (Ravitch, 1983-84).

Once test scores are obtained, many misuses may occur. This misuse may stem from lack of understanding of what test scores mean (Gardner, 1985). In one study, only 50% of teachers could define grade equivalents and percentile rank correctly (Yeh et al., 1980). As noted, parents could not recall their child's stanine score as explained to them by teachers (Hopper, 1977) and therefore, the report was meaningless.

Misuse occurs when statistical error of measurement is ignored (Gardner, 1985). Errors can also be made in administration and interpretation of test scores (Cryan, 1986). This can lead to decisions based on erroneous information and could have long range detrimental impact on the child.

Using a single score for decision-making is considered a serious misuse by Gardner (1985). This is especially harmful when students are labeled (Salmon-Cox, 1981; Cryan, 1986). Ravitch (1983-84) states, "Over reliance on a single test score

is dangerous to the health of education" (p. 26). Mislabeled damages the credibility of the school program.

Tests can be misused when they are allowed to determine the curriculum. If a text is matched to a test, the text with greater similarities to the test will be chosen (Leinhart & Seewald, 1981). This text may not be the best choice for meeting the needs of students. Test publishers, in this way, are dictating curriculum (Madaus, 1985).

When tests are allowed to become the end product of education rather than one means to evaluate learning (Ravitch, 1983-84), a danger is posed. A test is a sample and areas not included may be important for program balance but will likely receive little emphasis.

When "ordinary classroom practice is dominated by fill-in-the-blanks mentality" (Ravitch, 1983-84, p. 26), misuse occurs. In a case reported by Madaus (1985), a principal in a New York City public school said that reading material used is very much like the reading samples which would be found on a basic skills test. Materials were even designed to look like tests which are taken by students in the spring.

Test scores are misused when competition occurs between teachers or buildings. There is danger in generalizing without considering circumstances surrounding test score differences (Hieronymous et al., 1982). Test scores should

never be used in isolation to compare one teacher to another (Ebel & Frisbie, 1986). Programs are strengthened when teachers collaborate cooperatively.

Finally, when information is released to the press and the general public without explanation (Hieronymous, Lindquist, & Hoover, 1982) or when uses are exaggerated (Salmon-Cox, 1981), there is danger of misinterpretation.

Many uses and misuses reported in this review of literature are similar. Standardized achievement tests can be educationally beneficial if they are not over-emphasized. Misinterpretation of the significance of test scores damages the credibility of schools and inhibits legitimate use of results.

### Chapter 3: Procedures

Since the "Nation at Risk" report was published in 1983, concern about the emphasis on standardized achievement tests has increased. Dreher and Singer's (1985) article, dealing with parents' attitudes toward standardized reading test results, presented a positive format for use of test scores in improving reporting procedures to parents. A companion article used this same type of test report to give teachers more information about students.

To expand on this subject, and test use in general, references from this article were consulted. These references served as a starting point to obtain more information.

Searches in Resources in Education (RIE) and Current Index to Journals in Education (CIJE) were conducted. Two computer searches were made using 1980 as a cut off date so that only the most recent articles and research would be consulted. The most useful indicators in these searches were "achievement tests" and "test use", cross-referenced to "elementary education". Several books from Rod Library, University of Northern Iowa, were also utilized.

Much of the available research dealt with current practices by individual school districts to improve scores or use of standardized achievement scores. These were summarized. Journal articles on this topic dealt with expert

opinion. Notes were taken on these articles, categorized, and compiled according to content agreement.



#### Chapter 4: Summary and Conclusions

Standardized achievement tests consume a great deal of time. Teachers are required to report test results to parents, but little use is made of these results (Wackwitz, 1985). Central office administrators believe test results are useful to teachers and principals (Sproull & Zubrow, 1981). Teachers feel that results are important, however, very few use them, diagnostically (Salmon-Cox, 1981). Parents find results useful (Resnick, 1981) but often misunderstand them (Hopper, 1977). There remains a question, then, as to whether the time, effort, and expense is justifiable.

There is no consensus concerning how much weight standardized achievement tests should carry. Opinions vary from the National Education Association's call for a moratorium on standardized testing in the 1970's to the call for more testing in the "Nation at Risk" report in 1983.

The general public needs to know more about standardized achievement testing. Test publishers need to know what type of test results parents and teachers want and can use. Parents and teachers need to be aware of the types of test result reports which are currently available and the recommended uses of test results. There needs to be more awareness of the effects of standardized achievement tests on curriculum. Parents need to know that there are ways to judge

children's academic growth other than reliance on standardized test scores.

### Awareness of Tests and Scores

The public is constantly told that test scores have declined, but are seldom told possible reasons for this decline other than that the educational system is in a state of deterioration. Stedman and Kaestle (1985) offer several other explanations for test score decline. There was a time when only the very brightest students took college entrance exams. Now, there are more students of lesser ability who take the Scholastic Achievement Test (SAT) which has caused a 20 to 30% decline in scores. Family size may be a contributing factor to the score decline. First and second children score well on standardized tests, while those born later score lower. In the 1970's, children being tested were often fourth or fifth children in birth order. Today, scores seem to be rising. This may be due to a change in traditional family size. Today many students taking tests are from two children families, therefore, they score higher.

In schools, there is a concerted effort to keep children enrolled for more years. Drop out rates have declined among black students, and there has been increased immigration of Hispanics and Asians (Stedman & Kaestle, 1985). These types of students, especially those from low socio-economic groups

and those who do not speak English as a first language (Ravitch, 1983-84) do not perform as well on standardized achievement tests. Another possible reason for achievement score decline could be automatic promotion policies. Children who are simply not ready to for such tests are taking them, and therefore, score lower.

Government agencies find test scores an easy way to distribute educational funds. Madaus states, " No longer merely tools used by local school district administrators, the tests have assumed a central role in establishing and implementing state and federal education policy" (p. 613). Use of test scores by government agencies include looking for equality in schools and to see if funds invested had paid off. Tyler (1970) states that test scores were used as a basis for discussion of educational progress and problems.

Grade equivalent scores, one of the most common ways to report student progress, is often misleading. The general public needs to be aware of this fact. "Standardized tests are constructed in such a way that small shifts in performance produce large changes in percentile ranks and grade equivalent scores" (Stedman & Kaeslte, 1985, p. 206). Stedman and Kaestle cite an example of reading scores on one standardized test. Grade equivalent scores dropped one-half to one full grade level between 1971 and 1978. However, percentage of

items missed dropped only from 72% to 68%.

The general public needs to be aware that many educational experts feel that even though scores may rise, the overall quality of education could decline (Madaus, 1985; Berlak, 1985). "When emphasis is placed on test results, the beneficiary may be test scores rather than more general learning" (Haney, 1985, p. 4). This has long been recognized by teachers who realize that social progress cannot be measured easily.

### Parents and Test Use

There are three reasons why Hardy (1982) feels that standardized test scores should be important to parents. First, schools make decisions about a child's educational program based on test scores. Scores indicate, for example, whether a child needs further testing for remedial assistance or if the child would benefit from a gifted program. Second, and most important, is that parents may see exactly how their child is progressing. Boehm and White (1982) concur with this statement and feel that testing should be taken very seriously by parents.

In order for parents to effectively use test results, these results must be reported in the language of the laity (Hopper, 1977). Singer and Dreher's (1985) study found that, when given a choice, parents prefer to be given test results

with norm-references and some information about the test's content. They also need more interpretation provided by teachers. Parents felt that this type of report would indicate that their child was learning to read and how performance had improved. Parents must be given this type of report if they are expected to help improve reading achievement scores (Hopper, 1977).

### Teachers and Test Use

According to Singer and others (1983), teachers realize that testing is necessary but are confused and annoyed by the conflicting demands that local, state, and federal authorities have imposed upon them. If test scores are reported "haphazardly" (Ebel & Frisbie, 1986, p. 297) and use of test scores is not well defined, there is little chance that teachers will ever accept them completely.

In Salmon-Cox's (1981) interviews with teachers, she found that test results were used to confirm what they thought they already knew about students. Scores provided a guide to what should be taught, and confirmed if children were grouped correctly. Generally, standardized tests were used in a supplemental role to confirm teacher judgments.

Singer and Dreher's (1983) study shows that teachers would prefer test result reports with norm-group and self comparisons with addition of some content information about questions which were missed. This concurs with what parents prefer.

### More Usable Test Results Are Available

Teachers and parents want more usable information from standardized achievement tests. Iowa Tests of Basic Skills (Hieronymous et al.,1982) has several different types of reporting systems available which can provide more than norm-group and percentile rank scores. The profile narrative report lists each subtest and its percentile rank and grade equivalent for these. It also provides a printed statement of strengths and weaknesses. The student criterion-referenced skills analysis provides more specific results. It lists each subtest, the number of examples of each skill, number of correct answers, individual percentages and national percentages. The item analysis lists each item on the test. Computer print-outs indicate if an item was answered correctly or reveals the wrong answer that was marked. By referring to the test manual, a determination can be made concerning why the child possibly answered as he/she did.

Ebel and Frisbie (1986) suggest that teachers use a narrative report when talking with parents because this type of report is free of numbers and jargon and is easier for parents to understand.

### Appropriate Uses of

### Standardized Achievement Tests

Parents and teachers need to be aware of the

recommended uses of standardized achievement tests in order to better utilize results for making curricular decisions. Individual analysis is a major use of standardized tests. They can be used to determine what a student knows (Hieronymous et al., 1982; Ebel & Frisbie, 1986) and how students have progressed (Ravitch, 1983-84). However, test scores should never be used in isolation to make decisions about individual students (Gardner, 1985). Test scores can be used to improve instruction for individuals (Hieronymous et al., 1982; Salmon-Cox, 1981; Singer & others, 1983), but should not be allowed to determine entirely what students are taught (Leinhart & Seewald, 1981; Madaus, 1985). They contribute to a profile for evaluation that should also include subjective judgment.

Group analysis is another major use of standardized achievement tests (Ebel & Frisbie, 1986; Hieronymous et al., 1982). Group strengths and weaknesses can be determined (Hieronymous et al., 1982) and teaching methods and materials adjusted to meet the needs of the group. This should be of great interest to teachers who must make decisions about appropriate distribution of instructional time.

Administrative uses of standardized test scores include grouping decisions (Hieronymous et al., 1982) and admission to schools and programs (Singer & others, 1983). Though scores can be used partly for these purposes, caution must be used so



as not to place too much emphasis on a single test score (Gardner, 1985; Ravitch, 1983-84).

Test scores can be used partially in evaluating instruction but should only be used in combination with other factors such as ability of pupils, home and school environment, materials available, and general morale (Hieronymous et al., 1982; Singer & others, 1983; Ebel & Frisbie, 1986). Test scores should never be used to create competition between students or teachers (Hieronymous et al., 1982). Such competition is demoralizing and damages relationships which make positive contributions to the dynamics of learning.

Popham (1980) states, "One of measurement's most beneficial roles is to improve instruction" (p. 531). People who make policies know that they can do little to improve instruction directly, so tests are used as a threat (Madaus, 1985). Students fear that they will not pass or graduate if test performance is poor. Teachers fear that they will lose their jobs. Scores are used as a weapon to explain to parents why their child is placed as he/she is or why he/she will not be promoted. In spite of these fears, the general public views testing as "an objective, scientific, logical totem" (Madaus, 1985, p. 615). Most standardized achievement tests can provide useful diagnostic information to help improve instruction. It has been found, however, that little use is made

of the information for this purpose.

Finally, test scores can be used to report to parents and the general public on the educational state of schools (Hieronymous, Lindquist, & Hoover, 1982; Ebel & Frisbie, 1986). Raw scores should not, however, be released indiscriminately without explanatory information (Hieronymous, Lindquist, & Hoover, 1982).

#### Limitations of Standardized Achievement Tests

Ebel and Frisbie (1986) list several limitations of standardized achievement tests as measures of student progress. These include lack of test-curriculum match between what is taught and what is tested, lack of reliability because there are so few items for each objective, possibility of coaching by test administrator, loss of validity with repeated use, possible improper administration or scoring errors, and possible use of wrong norms (p. 296).

Teachers also feel that tests have limited usefulness. Some problems that teachers cite are a lack of match between test and curriculum, poor use of valuable instruction time, unavailable measures of student attitude at the time of testing, group scores invalidated by one or two students, and no immediate feedback to students (Ebel & Frisbie, 1986). Tests do not differentiate between lack of background and lack of skill (Singer & Dreher, 1983). Teachers feel that test

results do not provide guidance for improving instruction (Ebel & Frisbie, 1986; Singer & Dreher, 1983).

### Standardized Testing Narrows Curricula

A prevalent theme throughout much of the literature on standardized testing is that producers of these tests are dictating curricula in school districts (Madaus, 1985). If this trend continues, the ultimate result will be a universal curriculum where children in large cities will have the same educational objectives as children from small farm communities, even though their interests and social need may vary greatly.

Ebel and Frisbie (1986) state that schools have two choices when adopting testing programs. A school district may set up a testing program to fit its needs or take part in a testing program, such as ITBS, which is nationally available. The first choice is popular with teachers, but requires time, expert preparation, and lacks external comparisons. Packaged testing programs may not perfectly fit a school's curriculum, but are expertly prepared, economical, and are normed for easy comparison outside of the district.

Testing pressures teachers into teaching for the test and leads to a narrowing of the curriculum (Salmon-Cox, 1981; Kellaghan, Madaus, & Airisian, 1982; Singer & others, 1983; Madaus, 1985). Nagel (1986) cites the example of one

California school district which was ordered by the courts to raise test scores. In an effort to do this, a highly specific time limitation was prescribed for the testable areas of the curriculum. Time spent in other areas such as art, science, and social studies was reduced resulting in a narrowing of curriculum.

Many materials now in use resemble standardized achievement tests in content and structure (Madaus, 1985). A principal in a Manhattan public school reported in 1981 that "reading instruction in New York City public schools closely resembles practice in taking standardized reading tests" (Madaus, 1985, p. 616). Materials were designed to look like tests given each spring. This back to the basics movement discourages innovative teaching practices (Ravitch, 1983-84) and narrows the curriculum.

#### Other Measures of Student Progress

Haney (1985) states that in one Vermont school, standardized achievement tests were not used. Other types of student generated work are used for evaluation. These included children's writings and drawings and teacher observation.

Other materials which might be examined by parents and teachers to determine student progress are teacher guides that list objectives, materials, and experiences, diagnostic tests

measuring achievement of stated objectives, mastery tests, student study guides that stressing language and practice items, and self-instructional booklets which practice basic skills (Singer & others, 1983).

Academic achievement of total school populations can be judged by looking at numbers of college-bound seniors, honor roll students, percentage of students not failing any courses, students making one or more years growth on standardized achievement tests, gifted and talented students, and average daily attendance (Ligon & Wilkinson, 1985). These types of measures will provide a more accurate picture of educational achievement than test scores alone.

### Discussion

Teachers and parents seem to have many of the same beliefs about standardized test scores (Wackwitz, 1985). There are several steps which need to be taken to ensure better administration, use, and understanding of standardized achievement tests and scores.

First, since it is the teacher's responsibility to administer and interpret test results, Hopper (1977) suggests that the initial step is to give better training to teachers in the administration and interpretation of these tests. If teachers, themselves, do not understand results, there is no way that they can explain results to parents. This step could

be accomplished through undergraduate training in use and interpretation of standardized tests, or by making in-service training available to teachers already in the work force.

Second, there needs to be an adjustment in the attitude of both parents and teachers toward test scores. The prevalent attitude is that if scores are lower than expected, they are not valid, and if higher than expected, rethinking is done about the child's ability. Questions need to be asked. Why are the child's scores higher or lower than expected? Did the child seem anxious, upset, or ill on the day of the test? Does the child know the skill, or did he/she not deal well with the mechanics of taking the test? Are the skills covered on the test compatible with those taught in the classroom?

Third, a different type of test report needs to be given to parents and teachers. Singer and Dreher (1985) suggest that a norm-referenced and self comparison report with some content information may solve this problem.

Fourth, a public information program needs to be started to inform everyone involved with standardized achievement tests of the possible uses of these tests. Part of this program should relate how tests can be misused and their limitations.

### Conclusions

More information about standardized achievement tests needs to be made available to parents and teachers. More

information would help teachers use tests diagnostically (Salmon-Cox, 1981). Tests can be used as one objective way to check a student's skill development. More information given to parents would leave less room for misinterpretation (Singer & Dreher, 1985). More information given to parents and teachers would create better understanding and acceptance of standardized achievement tests and results gained from them.

It should always be remembered that schools are a social as well as academic learning atmosphere. Stedman and Kaestle (1985) summarize:

Schools must impart more than basic skills; they must become better places for teachers to work and for children to learn about themselves and their society. This cannot happen with top-down, test-based solutions. The challenge, then, is for each community to find a philosophically appealing and educationally effective balance between common experiences and cultural diversity, between a supportive atmosphere and standards of excellence, between student initiative and transmission of uplifting knowledge. In this process, tests can lay only a limited role. (p. 210)

### Limitations

This study is limited by the absence of longitudinal studies that show the long-term effects of reliance on

standardized achievement test scores as a basis for children's educational programs. Educators need to know if the efforts to improve basic skills test scores have indeed narrowed the curriculum and the knowledge base of today's children.

A second limitation is lack of consensus concerning how standardized achievement test scores should be used. It seems that every user of these tests has a different idea about how to use tests and test results. Until agreement is reached, little can be done to use these tests to the benefit children.

### Recommendations

Educators need to be more knowledgeable about standardized achievement tests. Teacher training programs must place more emphasis on learning how to use these tests, interpreting scores, and reporting scores to parents in an understandable way. When teachers understand these tests better, they can be more vocal in objecting to the misuse of test scores by local, state, and federal officials as a measure of educational excellence. Meanwhile, caution needs to be used in putting too much or too little emphasis on standardized achievement tests and scores.



## References

- Beck, M.D. & Stetz, F.P. (1979, April). Teachers opinions of standardized test use and usefulness. Paper presented to American Education Research Association, San Francisco, CA.
- Berlak, H. (1985). Testing in a democracy. Educational Leadership, 43(2), 16-17.
- Boehm, A.E. & White, M.A. (1982). The parents handbook on school testing. New York & London: Teachers College Press.
- Cryan, J.R. (1986). Evaluation: Plague or promise? Childhood Education, 62(5), 344-350.
- Dreher, M.J. & Singer, H. (1984). Making standardized tests work for you. Principal, 63(4), 20-24.
- Ebel, R.L. & Frisbie, D.A. (1986). Essentials of educational measurement. Englewood, N.J.: Prentice-Hall, Inc.
- Gallup, G.H. (1979). Attitudes on educational testing. Princeton, N.J.: Gallup International.
- Gardner, E.F. (1985, April). How can tests be misused: Let me count the ways. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL. (ERIC Document Reproduction Service No. 262 045)
- Haney, W. (1985). Making testing more educational. Educational Leadership, 43(2), 4-13.

- Hardy, R.A. (1982). Parents' concerns about standardized tests. S. Anderson and L. Coburn (Eds.), Academic testing and the consumer (pp. 89-92). San Francisco, CA: Jossey-Bass.
- Hieronymous, A.N., Lindquist, E.F., & Hoover, H.D. (1982). Iowa tests of basic skills manual for administrators. Chicago: Riverside Press.
- Hopper, G. (1977). Parental understanding of their child's test results as interpreted by elementary school teachers. Measurement and Evaluation in Guidance, 10(2), 84-89.
- Kellaghan, T., Madaus, G.F., & Airisian, P. W. 1982). The effects of standardized testing. Boston: Kluwer-Nijhoff Publishing.
- Lantor, L. (1986, October 11). School uses cash to boost test scores. Des Moines Register, pp. 1,2.
- Leinhart, G. & Seewald, A. (1981). Overlap: What's tested? What's taught? Journal of Educational Measurement, 18(2), 85-96.
- Ligon, G. & Wilkinson, D. (1985). The average achievement test score: A demogogue statistic. Austin, TX: Office of Research and Evaluation. (ERIC Document Reproduction Service No. ED 261 066)
- Madaus, G.F. (1981). Reaction to the "Pittsburg papers". Phi Delta Kappan, 62, 634-636.
- Madaus, G.F. (1985). Test scores as administrative mechanisms in educational policy. Phi Delta Kappan, 66(9), 611-617.

- Nagel, T. (1986). A longitudinal study of systematic efforts to raise standardized test scores using factors from school effectiveness research. Proceedings of American Educational Research Association Meeting. (ERIC Document Reproduction Service No. ED 269 432)
- National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington, D.C.: Author.
- Popham, W.J. (1980). Educational measurement for the improvement of instruction. Phi Delta Kappan, 61(8), 513-535.
- Ravitch, D. (1983-1984). The uses and misuses of tests. The College Board Review, 130, 23-26.
- Resnick, D.P. (1981). Testing in America: A supportive environment. Phi Delta Kappan, 62(9), 625-628.
- Salmon-Cox, L. (1981) Standardized achievement tests: What's really happening? Phi Delta Kappan, 62(9). 631-634.
- Singer & others. (1983). Testing achievement in basic skills. Volume II: Description, analysis, & final report. Riverside School of Education: California University. (ERIC Document Reproduction Service No. ED 233 326)
- Singer, H. & Dreher, M. (1983). Attitudes towards testing and test results of reading achievement. Journal of Reading Behavior, 15(4), 19-33.

- Singer, H. & Dreher, M. (1985). Parents' attitudes toward reports of standardized reading tests. The Reading Teacher, 38(7), 624-632.
- Sproull, L. & Zubrow, D. (1981). Standardized tests from the administrative viewpoint. Phi Delta Kappan, 62(9), 628-631.
- Stedman, L.C. & Kaestle, C.F. (1985). The test score decline is over: Now what? Phi Delta Kappan, 67(3).
- Tyler, R. (1970). National assessment: A history and sociology. School and Society, 98, 474.
- Wackwitz, L.W. (1985). Making test results more usable to teachers and parents. Iowa Educational Leadership, 2(1), 14-16.
- Yeh, J., Herman, J., & Rudner, L.M. (1980). A survey of the use of various achievement tests. Washington, D.C.: National Institute of Education. (ERIC Document Reproduction Service No. ED 237 579)