Should Instruction Be Measurement-Driven?: A Debate

Opponent: Lorrie A. Shepard
University of Colorado, Boulder

Measurement-driven instruction occurs when an achievement test has such serious consequences for teachers or students that the test determines what is taught. Examples are grade-to-grade promotional tests or standardized tests used to rank schools in the local newspaper. In coining the phrase, Popham took a long-standing complaint against accountability tests, that tests drive the curriculum, and turned it into a virtue. If the learning goals represented by the test are central and important, then there is no harm in forcing teachers to teach to it, or so the argument goes.

In his 1987 Kappan article, Popham went on to specify several attributes of a good measurement-driven instruction (MDI) program. The instructional targets must be clearly described by means of criterion referenced tests, the tests must assess defensible content, and be limited to a manageable number of targets. For example, "a number of high-stakes tests in reading assess only five or six truly essential skills, such as finding the 'main idea' and drawing the correct inference." (Popham, 1987, p. 680). In addition the criterion-referenced test items must be crafted in such a way that they convey to teachers what discriminations their students must be able to make. Finally, MDI must be accompanied by instructional support. As an example, Popham points to the Instructional Strategies Guides distributed in Texas to prepare students for essential skills testing in grades 1,3,5,7,9, and 11. "For each skill to be tested, the guides contain 1) a description of the skill, 2) an illustrative test item, 3) an analysis of how the skill might be taught, and 4) potentially useful instructional activities and exercises." (p. 681). Popham claims that measurement-driven instruction is a highly effective educational reform, specifically he says it is more cost effective than hiring "a host of well-paid, highly skilled teachers." (p. 679) As evidence of the success of MDI Popham cites increases in passing rates on MDI tests in six states ranging from 1 to 25%.

Despite these lofty aims, measurement-driven instruction does harm both to what is taught and what is not taught. By focusing on item formats it cheapens teaching of even the essential skills represented on the test. Because

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 1988.

of the high-stakes that are a defining characteristic of MDI, apparent test score gains cannot be believed. Furthermore, MDI diverts attention and instructional time away from important instructional goals that cannot be captured by multiple-choice formats or that do not get high enough agreement in the political negotiation of test content. And, finally MDI has negative side effects for both teachers and learners that are persistently denied by its advocates.

In advancing these arguments it is important to emphasize that the negative aspects of measurement-driven instruction are inherent in its conception. In the past Popham has dismissed criticisms of minimum competency testing, for example, by disavowing the negative effects of poorly implemented programs. Thus he champions high-quality minimum-competency tests (Popham, 1982) and measurement-driven instruction properly conceived and implemented. I will occasionally cite examples from the Texas TEAMS test but not because I enjoy bashing Texas. To the contrary, my previous research in Texas (Shepard & Kreitzer, 1987) has acquainted me with remarkably competent and dedicated measurement specialists in the Texas Education Agency who spare no expense in implementing the ideal that Popham has recommended. It is important to recognize that negative consequences from MDI accrue from even its most admirable installations.

Teaching to the test

Popham is always careful to say that high-stakes tests will motivate teachers to teach the skills and knowledge measured by the test. On the face of it he does not recommend teaching specific test items. He advocates teaching the test in the good, not the bad sense. But this rhetorical stance is not a true representation of MDI. In fact, MDI delivers two messages. High-minded rhetoric is followed by high stakes and item specifications--both fundamental to the definition of MDI--that invite teaching to the test in the more narrow sense.

The sample items provided for the Texas TEAMS, one for every skill, are not vague--suggesting that items might look like this or some other multiple-choice question. They are explicit templates specify the form of the item stem, admissible correct answers, and admissible wrong answer choices. These item forms are like contracts between teachers and test makers. It is guaranteed that there will be no deviations, no surprises on the the test.

Attachments 1 and 2 are examples of materials provided in the TEAMS Instructional Strategies Guides for reading skills at grades 1 and 11 (Texas Education Agency, 1986, &

no date). Teachers of 11th graders are told, for example, that when opinions are used as distractors for fact statements they will always include "at least one concept from the reading selection." And first grade teachers are told that incorrect answers to sequence questions will be either an event from the reading passage that does not answer the question or an event not contained in the passage that appears to answer the question. Children can then be coached to look for these specific kinds of wrong answers. On page 57 of the Exit guide (see Attachment 1), teachers are encouraged to provide practice with textbook and magazine articles of their own choosing by constructing multiple-choice questions in the prescribed format. In an informal survey of Texas teachers representing 10 districts, all reported providing practice for their students on comparable multiple-choice questions over time periods ranging from 3 weeks to 6 months.

In virtually all cases the TEAMS item specifications I have examined are of high quality. But by providing them to teachers and promising not to subject students to other equally good ways of assessing the same skills, the item forms no longer represent a broader conception of each skill. If this evidence is not credible as evidence of teaching to the test in the narrow sense, I propose the following experiment: Select a large sample of Texas teachers and present to them copies of a test keyed to the same objectives but couched in different formats, then record how many of them are willing to have their students tested by the alternate version of the test.

High-stakes tests have spawned a variety of teaching-the-test endeavors. There are prevalent examples of commercial materials designed to help educators improve standardized test scores. They start by aligning curriculum ostensibly to ensure that there are not gaps between what is tested and what is taught. For example, the advertisement for Control Data's TargetTeach reads as follows:

> With TargetTeach Curriculum Alignment and Instruction Management, your students will be well prepared to answer every question on every District, State and Standardized Test you use.
>
> The reason: TargetTeach correlates the Objectives and Skills represented by every test question to the Objectives and Skills in your textbooks and other instructional resources.
>
> The result: you're virtually guaranteed a dramatic improvement in student test scores!

Logically, such systems also help teachers set priorities so that there will be no wasted effort on learnings not measured by the test.

Then we have materials designed to improve students test taking skills. For example, Random House offers a different version of Scoring High tailored to each of the five major standardized test batteries. In a symposium of the National Council on Measurement in Education earlier today, Mehrens and Kaminski (1988) reported on their content analysis of Scoring High-CAT along with other commercial preparation materials. They found a match of 64.5 points out of a possible 69 between the actual CAT tested subskills and those practiced in the preparation program. Their method of awarding points required that subskills match on both content and format before being awarded a full point. Their conclusions about Scoring High-CAT were as follows:

> We did not find any discussion or emphasis on subskills not tested on the CAT. Each practice page has the same format as the CAT. The program also includes instruction on 33 test-taking skills and practice on timed tests. The program's design requires students to do a page and then the teacher is instructed to review the answers and work through the answer choices with the students....It is our belief that this program serves as a pretest for the CAT in the same manner as if one actually used the CAT as a pretest prior to giving the same CAT at a later time (Mehrens & Kaminski, 1988, pp. 29-30).

Rhetoric from advocates of measurement-driven instruction has blurred the distinction between legitimate and illegitimate test preparation activities. Afterall, if students genuinely know the answer to the question how is it cheating to let them practice on highly similar items beforehand? Because, says measurement theory, items are intended to be samples from a domain of knowledge. To make valid inferences about student achievement on the basis of sampled items results must generalize to the wider domain. Most specially apparent mastery or understandings must not depend on the format of the test question.

There is ample evidence, however, that knowledge taught by drill and practice in one format does not transfer. My favorite example is from the New Jersey assessment of 1975-76. When students were asked to add decimals in vertical format, the state percent passing the items was 86%. In horizontal format for identically difficult decimals, the percent passing was 46%. For subtraction of decimals in the two formats the passing rates were 78% and 30%, respectively (New Jersey Department of Education, 1976). Which result is true? My conception of learning requires that students be able to add or subtract in either of these ways and a half-dozen other ways besides.

Perhaps the best evidence of the limited generality of taught-to tests comes from Cohen, an advocate for instructional alignment. In his 1987 article in the **Educational Researcher** he summarizes a dissertation by Koczor (1984) in which the manipulated variable was the degree of instructional alignment represented by varying posttests formats. For example, in a lesson designed to teach the correspondence between Roman and Arabic numerals, "the Arabic was always presented or written after the Roman numerals. One group's posttest was aligned on this factor. In contrast, the misaligned treatment group received a test in which the Arabic numeral came first, and the student had to write the Roman numeral. Most teachers would consider this a minor variation of the instruction's stimulus conditions. That minor misalignment accounted for a 40% difference in posttest raw scores." (p. 17).

That our interpretations of these data could differ so profoundly reveals the presence of powerful belief systems. For me to say that fourth graders know Roman numerals would require that they be equally adept in going from Arabic to Roman, Roman to Arabic, whether the numbers were presented orally or chiseled in stone. Cohen, if I understand him, believes that the aligned condition reflects "near perfect learning."

Cohen's conception of learning where each tiny variant must be specified if it is to be learned reminds me of an apocryphal story told by physicists about teaching Ohm's law in the military. There students are taught three Ohms'laws instead of one:

$v = i\,(r)$, $i = v/r$, and $r = v/i$,

thus not leaving to chance their understandings of the algebraic relations.

Spurious score gains

Cohen's (1987) evidence across studies that instructional alignment leads to effect sizes of 1.2 to 3 sigma should serve as a warning that impressive test score gains may not mean what they appear to mean. It is a basic principle of social science research that indices are corruptible. Police officers are taught, for example, that conscientious efforts to treat rape victims with greater sensitivity might actually increase the number of reported rapes, but that such increases would not necessarily reflect real increases in incidence of the crime. No such admonition has been provided to superintendents and legislators sold on measurement-driven instruction.

Apparent test score gains undoubtedly have multiple explanations. Any serious technical efforts to look at

unusual gains must acknowledge pervasively rising scores across the nation. Contrary to political rhetoric that still has us at the nadir of the SAT test score decline, downward trends turned around in approximately 1976 across the major data sets that were originally used to document the decline, e.g., SATs, ACTs, National Assessment, state testing from Iowa, California, Virginia, etc. In some subject areas and grades, achievement is as high today as it was before the great decline. These large-scale trends are analyzed in great detail by Daniel Koretz in two reports published by the Congressional Budget Office (1986, 1987). According to the CBO reports, the turn-around in achievement indicators predated minimum-competency testing of the late 1970s as well as reforms in the 1980s. Thus any attribution of gains to these interventions in specific locales must be substantially greater than the secular trends and must not be an artifacts of the measuring instruments.

In many cases, gains from high-stakes tests are more melodramatic than the steadily rising trends found nationally. It is just these extraordinary gains, which Popham cites as proofs of success, that should be regarded with suspicion. These gains are more likely attributable to teaching the test that will not generalize and sometimes to out-and-out cheating.

Although flagrant cheating probably accounts for only a small part of the dramatic test-score gains and certainly is not intended by the rhetoric of measurement-driven instruction, it is nonetheless a consequence of the climate created by high-stakes tests. Ethical people do not cheat just because they are under pressure. But personal ethics may be eroded under the joint conditions of high pressure and seemingly unfair or mindless requirements. Consider, for example, the number of otherwise respectable persons who cheat on their income tax.

Conversations with directors of assessment in a dozen states suggest that there are increasing problems with educator-initiated cheating on standardized tests compared to 10 years ago. For example, most major test-scoring agencies now offer computer-scanning for significant blocks of wrong-to-right erasures within schools. In some cases the quantity of such erasures has been so great that the state assessment program refused to release the scores for those schools. A more subtle way to cheat that probably elicits fewer pangs of conscience is to encourage certain students to stay home or be involved in a field trip on the day that standardized tests are administered. Three years ago at a meeting of the American Educational Research Association, methods for detecting irregularities at the classroom level were presented (Stringfield & Hartman, 1985); it is a topic of on-going concern for large-city test directors. In one city, for example, 23 schools with

questionable test results were retested a week later; 70% experienced drops of 7 months or more on a parallel form of the same test (Perlman, 1985).

The lion's share of remarkable test score gains is more likely caused by teaching the test, including both instructional alignment and practice on highly similar items. The Cannell (1987) report released last fall found that every state in the nation was above national norms. For example, Arkansas fourth graders are at the national 60th percentile in reading, California third graders are at the national 55th percentile in reading, a so forth. (Table II from Cannell's report is included as Attachment 3.) The southeastern region of the United States is consistently below the national average on National Assessments. Here, with the exception of a rare grade level or subject area, all of the southern states find themselves about average on their own administrations of nationally normed tests. While I disagree with Cannell's ascription of blame to the profit motives of test publishers, he is quite right to say that these results are not credible. It is quite likely that more effort and attention went into the state administrations than had been given to the tests when they were administered to the norming sample.

In the two years since its inception, performance on the Texas TEAMS has shown steady gains so that in 1987 performance ranged from a low of the 49th percentile on national norms in seventh grade reading to as high as the 81st percentile in first grade math. The complete normative results for both the 1986 and 1987 TEAMS are shown below:

Predicted National Percentile Ranks

	Mathematics		Reading		Writing	
Grade	1987	1986	1987	1986	1987	1986
9	61st	57th	57th	59th	66th	60th
7	64th	54th	49th	45th	62nd	54th
5	72nd	62nd	58th	53rd	75th	63rd
3	76th	62nd	55th	47th	71st	61st
1	81st	70th	71st	63rd	76th	67th

(Texas Education Agency, 1987b, p. 3). Interestingly, the rise in TEAMS scores at each grade level occurred for computation and punctuation rules but not for applications and higher-order thinking skills. For example, the following conclusion was reported by the Texas Education Agency about the 1985 to 1986 results on the eleventh grade exit test:

> In comparing the performance results of eleventh grade students tested for the first time in October 1985 with those tested for the first time in October 1986, mastery of individual objectives

> increased an average of 5.4 point in mathematics and 2.2 points in English language arts. This upward move in objective performance for first time tested students indicates an increased awareness of the skills measured on the exit level TEAMS test and the likelihood that additional instruction in these skill areas occurred prior to students' initial exposure to the examination. Performance gains were greatest on objectives that can be easily instructed, such as sequencing of numbers and equivalencies among percents, decimals, and fractions in mathematics and capitalization and English usage in language arts. No gains were evident in the language arts objectives measuring main idea, drawing conclusions, and fact and opinion, all of which test reading comprehension skills, proofreading, which assesses the application of mechanics, and spelling, which is based on the set of words selected for testing. In future years as students improve in reading comprehension, their skills in spelling will likely improve as their familiarity with these words in context increases. (Texas Education Agency, 1987a, p. 26).

To appraise Popham's claim that six states had experienced impressive learning gains as a consequence of MDI, we collected Chapter I achievement data from these same states over the same time periods. Table 1 is reproduced from Popham's 1987 **Kappan** article and included as Attachment 4. Connecticut's data is not available as yet. In addition, several states had posttest data on too few students at the higher grade levels to warrant the reporting of trends. The available data are summarized in graphs and enclosed as attachment 5. The Chapter I data are consistently flat or show only slight rises on the order of five percentile points, consistent with rising achievement trends nationally. Furthermore, the patterns within states are consistent across grade levels making it difficult to believe that there is any special increase in achievement in the grade levels with high-stakes tests. These data fail to corroborate Popham's claim that competency tests have produced a significant increase in basic skills. These data are an example, however, of the type of evidence that would be needed to provide convincing evidence that score gains from high-stakes tests generalize beyond the confines of specifically learned test questions.

Negative side-effects

Measurement-driven instruction has its roots in behaviorism and mastery learning. According to these

theories, learning is facilitated when knowledge is broken down into its constituent bits and then taught bit by bit. As many critics of MDI have said, this conception leads to fragmentation and narrow skill development at the expense of conceptual understandings. In support of this conclusion, Bracey (1987) cited observational studies by both Richardson (1985) and Goodlad (1983). In community college settings Richardson found students copying bits of information from chalkboards and skimming textbooks to find other bits of information that would answer questions in study guides. The study guides in turn prepared students to answer what Bracey calls "bits-of-information" (multiple-choice) tests. The following quotation from Goodlad (1983) illustrates the "bitting" of knowledge present in elementary and secondary classrooms:

> Children listened; they responded when called on to do so; they read short sections of textbooks; they wrote short responses to questions or chose from alternative responses on quizzes. But they rarely planned or initiated anything of length or created their own projects....Writing even short essays is conducive to organization of thought and learning. Students were not doing much of this. (Goodlad, 1983

When Bracy (1987) complained that "main idea" passages on the Detroit High School Proficiency Exam were specified to be no longer than 125 to 200 words, Popham (1987b) answered that "Bracey appears oblivious to the practical constraints of large-scale assessment." (p. 688) But this is precisely the problem, then, when multiple-choice tests become the template for learning.

In an ethnographic study Linda McNeil (1988) describes the institutional arrangements in schools that foster or deter good teaching. She was studying exemplary magnet schools when the district instituted a proficiency system based on testing. Her account of what excellent teachers did to cope with the demands of the proficiency tests is poignant, indeed:

> To blunt the effects of the proficiency-based curricula--which official district policy claimed to be the minimum standard, not the entire curriculum, but which in fact overwhelmed class time--the magnet school teachers began to deliver "double-entry" lessons. The biology teacher who was working with medical college researchers refused to dumb down her lessons to match the proficiencies; one day she wrote a simplified formula for photosynthesis on the board, telling students to write it in their notebooks and learn it for the proficiency test. For the remaining

> two weeks' lessons, the teacher provided another, more complex version of the formula that students helped derive through lab activities.
>
> Frequently, the magnet school teachers would have to put their "real" lessons on hold for a few days in order to lecture on the proficiencies, or they would have to continually point out proficiencies during each lesson, so that students could keep them separate from the "real" information. The students were not unaware that often the "official" content contradicted the complicated interpretations and reflected only partial information and oversimplified processes (McNeil, pp. 483-484)

In our own research we have also seen the effects of what we have come to call an "accountability culture" in some schools. For the last several years, Mary Lee Smith and I have been investigating the phenomenon of kindergarten retention, its effects but also its origins in the belief-systems of teachers. Although nationally the rates of retaining children in kindergarten are only 8-10%, in some local districts the rates are 20-30% or even greater than 50% in some unusual cases. Based on both systematic interviews and subsequent discussions with teachers about why it is necessary to hold children back in kindergarten, there is a link between the practice of retention and highly academic demands in first grade. Often both first-grade and kindergarten teachers say that the curriculum that is being force-fed in first grade is inappropriate for the developmental level of normal six year olds but that they are unable to make changes because of proficiency standards in grades 1 through 3. Several national associations, including the International Reading Association and the National Association for the Education of Young Children have issued policy statements against narrow literacy and numeracy demands in the earliest grades.

Finally, test-dominated instruction is based on a punitive view of education. It creates a degrading climate both for pupils and teachers. Poor students who are intended to be the beneficiaries of basic skill requirements are made to perseverate on boring tasks until many leave school. We have recently analyzed data sets from several large cities, representing both low and high socio-economic communities. Once prior achievement and family background factors are controlled there is still a substantial increase (20-30%) in the likelihood that a student will dropout of school if he is made to repeat a grade. When school systems have bothered to collect the data, there is a demonstrable increase in the dropout rate associated with the institution of grade-to-grade promotional gates. In Chicago, for example, Rice, Toles, Schulz, Harvey, and Foster (1987), estimated that the repeat year would have to produce

achievement gains of 30 months to compensate for the negative effects of being made a year too old.

For teachers the effect of measurement-driven instruction, is to rob them of their role as professionals and force them to read from longer and longer scripts:

> "We are now ready to begin work on the first test. This is a vocabulary test like the practice test we just studied. Find the sections for Test V: Vocabulary on your answer sheet (Pause.) Now place your answer sheet beside page 4 of your test booklet. (Pause.) We will read the directions at the top of page 4 to remind you of what you are to do. Read them silently while I read them aloud. They say:
>
> > In each exercise, you are to decide which one of the four answers has most nearly the same meaning as the word in heavy type above it....

REFERENCES

Bracey, G.W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. Phi Delta Kappan, 68, 683-686.

Cannell, J.J. (1987). Nationally normed elementary achievement testing in America's public schools. Daniels, WV: Friends for Education.

Congressional Budget Office. (1986, April). Trends in Educational Achievement. Washington, DC: Congressional Budget Office, Congress of the United States.

Congressional Budget Office. (1987, August). Educational Achievement: Explanations and Implication of Recent Trends. Washington, DC: Congressional Budget Office, Congress of the United States.

Cohen, S.A. (1987). Instructional alignment: Searching for a magic bullet. Educational Researcher, 16 (November), 16-20.

Goodlad, J.I. (1983). A study of schooling. Phi Delta Kappan, 64.

Koczor, M.L. (1984). Effects of varying degrees of instructional alignment in posttreatment tests on mastery learning tasks of fourth grade children. Unpublished doctoral dissertation, University of San Francisco.

Mehrens, W.A., & Kaminski, J. (1988, April). Using commercial test preparation materials for improving standardized test scores: Fruitful, fruitless, or fraudulent? Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

McNeil, L.M. (1988). Contradictions of reform. Phi Delta Kappan, 69, 478-485.

New Jersey Department of Education. (1976). Educational Assessment Program: State Report 1975-76. Trenton, NJ: Author.

Perlman, C.L. (1985, April). The Chicago test audit: A case study. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Popham, W.J. (1982). Melvin Belli, Beware! Educational Researcher, 11(January), 5, 11-15.

Popham, W.J. (1987). The merits of measurement-driven instruction. Phi Delta Kappan, 68, 679-682.

Popham, W.J. (1987b). Muddle-minded emotionalism. Phi Delta Kappan, 68, 687-688.

Rice, W.K., Toles, R.E., Schulz, E.M., Harvey, J.T., & Foster, D.L. (1987). A longitudinal investigation of effectiveness of increased promotion standards at eighth grade on high school graduation. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C.

Richardson, R.C. (1985). How are students learning? Change, 43-49.

Shepard, L.A., & Kreitzer, A.E. (1987). The Texas teacher test. Educational Researcher, 16, 22-31.

Stringfield, S.C., & Hartman, A. (1985, April). Irregularities in testing: Ethical, psychometric, and political issues. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Texas Education Agency. (no date). TEAMS Instructional Strategies Guide: Reading Grade 1. Austin, TX: Author.

Texas Education Agency. (1986). TEAMS Instructional Strategies Guide: English Language Arts Exit Level Austin, TX: Author.

Texas Education Agency. (1987a). Texas Educational Assessment of Minimum Skills: Student Performance Results for 1985 and 1986. Austin, TX: Author.

Texas Education Agency. (1987b). Texas Educational Assessment of Minimum Skills: Student Performance Results, 1987, Volume 1. Austin, TX: Author.

ATTACHMENT 1: One of two inference skills from the English-Language Arts Exit Level TEAMS. (Texas Education Agency, 1986)

Fact and Opinion: Distinguish between fact and opinion.

This objective requires the student to distinguish between facts and opinions by correctly identifying a fact or an opinion stated in a reading selection from answer choices based on the reading selection.

Sample Item

> At one time or another most people have waited until the last minute to finish something they've had enough time to do. That's called procrastination. Procrastination, or constantly putting things off, is one way to deal with pressure. It's not the best way, but it is a method used by many people.
>
> Many teenagers say that they procrastinate because they can get away with finishing homework or chores in minimal time. If nothing really bad happens, procrastination continues.
>
> In a recent survey people were asked to indicate why they procrastinated. Most of the people responding indicated one of three reasons--fear of failure, fear of success, or rebellion. The myth that people who procrastinate are lazy, disorganized, or simply don't care about their work is not true. However, people who procrastinate should try to identify why they continually put things off. That is the key to eliminating this behavior.

Which of the following is a FACT presented in the selection?

A. Many teenagers say that they procrastinate because they have too much homework.

B. Procrastination is not the best way to deal with pressure.

C. Some people reported using procrastination because they feared they would fail.

D. Putting things off until the last minute may create stress and anxiety.

Test Questions

1. A test item will consist of a reading selection followed by the question, "Which of the following statements is a FACT presented in the selection?" or "Which of the following is an OPINION presented in the selection?"

Answer Choices

1. For items requiring identification of a fact:

   a. The correct answer choice will be a fact stated in the reading selection.

   b. The incorrect answer choices will be one of the following types:

      -- Contradicted Fact: A statement presented as a fact, but contradicted by information in the selection.

      -- Irrelevant Fact: A statement presented as a fact that introduces information not included in the reading selection. Although irrelevant, this type of answer choice must include at least one concept that is contained in the reading selection.

      -- Opinion: A judgment or belief that contains at least one concept from the reading selection.

2. For items requiring identification of an opinion:

   a. The correct answer will correctly restate or paraphrase an opinion from the reading selection.

   b. The incorrect answer choices will be one of the following types:

      -- Contradicted Opinion: A statement of opinion that is contradicted by information in the reading selection.

-- Irrelevant Opinion: A statement of opinion that introduces information not included in the selection. Although irrelevant, this type of answer choice must include at least one concept that is contained in the reading selection.

-- Fact: A statement from or related to the reading selection that is a fact rather than a view or belief.

Answer Analysis for Sample Item

1. Answer choice A is a contradicted fact because in the second paragraph it is asserted that teenagers feel they can finish their homework in a minimal amount of time.

2. Answer choice B is an opinion (not a fact) that is presented in the last sentence of the first paragraph. However sensible, this belief cannot be conclusively proven or disproven.

3. Answer choice C is correct. This statement paraphrases concrete information from the second sentence of the third paragraph. Fear of failure is listed as one of three reasons cited by people who were asked to indicate why they procrastinate.

4. Answer choice D is an irrelevant fact because the effects of procrastination, such as stress and anxiety, are not dealt with in the selection. The selection more narrowly concerns itself with the causes of procrastination.

Instructional Emphases

The development of critical thinking skills is an educational goal whose realization is at once essential and elusive. An integral part of critical thinking is the ability to distinguish between statements of fact and of opinion. This capacity is necessary if students are to be expected to form their own logical, well-reasoned opinions concerning the many issues facing them in today's complex world. Bombarded by data from many sources, students must be shown how to discern facts from opinions in order to arrive at their own rational conclusions. Far too many students believe that if it's on television or in print, it's true!

A fact is a piece of information that is known to be true. Because facts can be proven, they have objective reality. Facts are supported by evidence gathered through direct observation or measurement.

An opinion is a statement that expresses a belief, a value judgment, or a prediction. Although opinions are often stated as fact, they rely on the acceptance of an idea that cannot be proven. The use of certain words may indicate that a statement is an opinion. The words "must," "can," "should," and "may" suggest the expression of a belief or expectation. "Generally," "frequently," "often," and "usually" suggest that the idea expressed may not always be true. Adjectives such as "important," "beautiful," "necessary," "best," "good," and "bad" indicate personal judgments. A prediction may be indicated by words such as "will," "shall," "probably," "possibly," and "likely." You may want to alert your students that these "clue words," if they are present in a selection, can help one identify statements of opinion.

Because clue words are often not present in a reading selection, emphasize the importance of the possibility of objective verification of a factual statement, something not possible with an opinion.

Readers often mistake an opinion for a fact when it is an opinion that they happen to agree with or when it is stated in a very definite way. Emphasize to your students that terms like "definitely," "absolutely," and "positively" do not make a statement a fact. Also, an opinion with which a reader agrees, no matter how strongly, is nevertheless an opinion.

A particularly important aspect of the TEAMS fact-and-opinion items is the way that the test item's question is phrased. Note the stressed portion of that question: "Which of the following is a FACT presented in the selection?" Your students must be sensitized to the idea that the facts or opinions they seek will be included in the selection. Thus, if the student is told to identify an opinion in the selection, not any opinion will do. Indeed, students must be certain that the fact (or opinion) they are considering is present in the selection.

Possible Instructional Activities

Masquerade

Provide a list of sentences that includes fact statements and opinions "masquerading" as facts. Here are some examples.

The Scarlet Letter is an interesting novel.
(opinion--clue word: interesting)

Daniel Defoe wrote Robinson Crusoe.
(fact)

Jane will probably be here by 5:00.
(opinion--clue words: will probably)

Mark was the first to arrive at the party.
(fact)

Texas is definitely the best state in the country.
(opinion--clue words: definitely, best)

Have students work independently, underlining clue words and labeling each sentence as fact or opinion. Go over each statement and discuss what makes it a fact or an opinion. Try to have students think of facts or opinions that are identifiable without clue words, for it is possible that, in a given TEAMS test item, no clue words will be present.

Assertions

Ask students to write 10 sentences, five under the heading "I Know (FACTS)" and five under the heading "I Think (OPINIONS)." Allow several minutes for independent work as they generate their assertions.

Call on several students for examples of facts, encouraging other students to raise their hands if they can demonstrate that a statement is actually an opinion. As you obtain clear examples of fact statements, write them on one side of the board under "I Know (FACTS)."

When you have obtained five fact statements, ask for opinion statements. Under "I Think (OPINIONS)," write these on the board until you have five opinion statements.

Paraphrasing Facts and Opinions

Answer choices in fact-and-opinion items are either restatements or paraphrases of information presented in the selection. Thus, students must be able to identify paraphrased information in order to discriminate among answer choices. The focus of this activity is recognizing paraphrased facts and opinions.

Have the class read a brief selection that contains both fact and opinion statements. Ask for a student to suggest a statement of fact from the selection. Write the sentence on the board. Ask how the idea expressed in the sentence can be stated in another way. Write on the board two or

three acceptable paraphrases of the original statement. Show how a sentence can be paraphrased by rearranging the word order or by substituting different words having the same meaning. Repeat this procedure with additional fact, then opinion, statements provided by students.

To provide further practice in paraphrasing, have all students read the same selection. Call on a student to give a paraphrase of one sentence in the selection. Ask the rest of the class if they can identify the sentence that was paraphrased and label the statement as a fact or an opinion.

Proving a Point

Provide copies of an article, editorial, or advertisement that uses facts to back up an opinion. Have students list the facts and the opinions they find in the selection. Discuss how the facts are used in an attempt to prove that a certain opinion is "right." Decide what an opposing opinion might be and what kinds of facts might be used to support this contrary point of view.

As a homework assignment, you might have each student select a similar passage and go through the above process in written form, to be handed in with the passage.

Book Blurbs

Have students choose a paperback book with information and promotional copy on the front and back cover. Students should then determine which statements are fact ("has sold over 200,000 copies") and which are opinion ("the most refreshing new author in 30 years"). Discuss how the facts might be proven and why they are included. Then discuss how the opinions could not be proven, but why they are included. Students can then design their own book jackets using both fact-and-opinion statements. The books they refer to may be real or imaginary.

Facts and Opinions in the News

Have students study newspapers to determine which sections are most likely to contain contents that are dominantly factual versus those dominantly opinion-oriented.

Find the Fact

Select a short article from a textbook, magazine, or newspaper, and give the students a multiple-choice question. The question requires the student to identify a fact that was presented in the article. The three distractors should include one factual statement which is irrelevant to the article, one factual statement which is contradictory to the information presented in the article, and one opinion statement. Discuss the correct answer and incorrect answers with students after they have attempted to answer the question.

Find the Opinion

Prepare a set of multiple-choice questions regarding a magazine article or newspaper editorial that students are to read. The questions ask students to identify an opinion supported by the selection. The distractors should include: an opinion statement which contradicts what is in the selection, an opinion statement which contains one concept from the article but which is irrelevant to the selection in that it contains information not mentioned in the article, and one factual statement which contains at least one concept from the selection.

ATTACHMENT 2: Main Idea Skill Area. Texas Education Agency, Instructional Strategies Guide, Grade 1 TEAMS. circa 1986.

Objective 8: Identify the sequence of events.

This objective requires students to identify, based on a reading selection, the first or last event.

Sample Item

Teacher says: "Read the story in the box to yourself. Then answer the question about what happens first or what happens last. Fill in the space next to your answer."

> Marco swims across the pool. He pulls himself out of the pool. He shuts his eyes and holds his nose. Then Marco jumps back into the pool.

What does Marco do last?

- ○ Marco jumps back into the pool.
- ○ Marco splashes water out of the pool.
- ○ Marco swims across the pool.

Test Questions

1. A test item will present a reading selection followed by a question that asks, "What does [character in story] do [first/last]?"

Answer Choices

1. The correct answer choice will be the event from the reading selection that correctly answers the question.

2. Incorrect answer choices will be either of the following types:

   a. Incorrect Event: An event from the reading selection that does not answer the question.

   b. Irrelevant Event: An event not contained in the reading selection that appears to answer the question.

Answer Analysis for Sample Item

1. Answer choice A is correct because the last action that Marco takes is to jump back into the pool.

2. Answer choice B is irrelevant. Splashing water out of the pool is not described in the selection.

3. Answer choice C is an incorrect detail. Swimming across the pool is the first thing that Marco does, according to the selection, not the last.

Instructional Emphases

The skill of determining sequence requires students to recognize that stories are composed of different events and that these events happen in a certain order. Knowing the sequence in which events happen helps students to understand and make sense of the story.

The sequence steps eligible for testing in this objective are first and last. To help students understand the concept of ordering events, encourage students to think about the steps they take in doing ordinary activities such as brushing their teeth or making peanut butter sandwiches. Have students identify the steps that they take during these activities and then put them in the order in which they happen. For example, first students put toothpaste on the brush, next they brush their teeth, last they rinse their mouth.

Students may initially need some assistance determining the difference between information that constitutes an event (for example, "The dog jumped over the wall.") and a description of the setting or time (such as "It is early in the morning.") Emphasize that events are characterized by action. They describe something that is happening to or something that is being done by the character(s) in the story.

You may wish to have students first practice sequencing events within a pictorial context. For example, show students a two-to-three picture sequence that clearly depicts the first and last steps of an activity. Once students understand the concept of sequence, provide opportunities for students to identify the order of events in reading selections. You may wish to have

Attachment 3: Cannell, 1987

TABLE II

All data were twice sent to the State Departments of Education for verification and corrections. National percentile ranks were used for reporting whenever possible, but many states used alternative statistical reporting methods. The percentage of students testing above the 50 percentile was not available (N/A) from all states. The percentage of districts testing above average was gleaned from the state's publication when available.

Unless otherwise stated, the national norm is fifty. This study included only elementary grades one through six, except in the case of Idaho, where grade eight is the lowest grade tested.

ALABAMA

APRIL 1987

STANFORD ACHIEVEMENT TEST
FORM F NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
1	57,399	54	N/A	53	54	55	70.5%
2	52,146	51	N/A	63	56	57	78.3%
4	51,467	50	58	59	54	53	78.5%
5	48,471	48	55	60	52	50	58.5%

Mean Individual Percentile Ranks

SOURCE: Chief State School Officer's Report

ARKANSAS

SPRING 1987

METROPOLITAN ACHIEVEMENT TEST
6th EDITION NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
4	29,491	60	63	67	64	65%	86.5%

National Percentile Rank from Mean Normal Curve Equivalents

SOURCE: Arkansas Standardized Testing Program Report

ARIZONA

APRIL 1987

IOWA TEST OF BASIC SKILLS
FORM H NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
1	49,449	46	64	55	55	N/A	N/A
2	44,987	58	67	57	61	N/A	N/A
3	43,118	51	66	50	56	N/A	N/A
4	42,490	50	62	51	55	N/A	N/A
5	40,087	57	59	53	56	N/A	N/A
6	39,871	53	57	54	55	N/A	N/A

Percentage Of Children At Or Above 50th Percentile

*Number of Students Tested in Reading

SOURCE: Arizona Department of Education

CALIFORNIA

SPRING 1986

CALIFORNIA ASSESSMENT PROGRAM
(State Developed Test)
Scores Normed With CTBS/U NATIONAL NORMS
Matrix Sampling Technique 36 test forms for each grade level

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
3	310,008	55	51	63	N/A	N/A	N/A
6	287,478	54	54	66	N/A	N/A	N/A

Estimated National Percentile Ranks

SOURCE: California Assessment Program Annual Report 1985-86

COLORADO
APRIL 1986

IOWA TEST OF BASIC SKILLS
FORM G NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average*	% of Districts Above Average
3	39,251	54	59	52	N/A	54.9%	77.5%
6	37,351	54	48	52	N/A	56.8%	69.7%

National Percentile Ranks
*In Reading

SOURCE: Results of the 1985-86 Colorado Student Testing Program

DELAWARE
APRIL 1986

COMPREHENSIVE TEST OF BASIC SKILLS
FORM V NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
1	N/A	52.6	N/A	58	N/A	N/A	100%
2	N/A	57	63.2	68.3	62.2	N/A	100%
3	N/A	55	66	63	63.3	N/A	100%
4	N/A	56	57.5	60.1	57.6	N/A	100%
5	N/A	52.7	57.4	61.7	56.1	N/A	100%
6	N/A	54.1	59.8	61.6	59	N/A	100%

Normal Curve Equivalents

SOURCE: Delaware Educational Assessment Program 1986 Statewide Test Results Summary Report

GEORGIA
SPRING 1987

IOWA TEST OF BASIC SKILLS
FORM G NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
2	81,769	61	71	73	68	65.8%	89.8%
4	76,627	53	59	62	57	71.9%	67.2%

National Mean Percentile Scores

SOURCE: 1986-87 Student Assessment Test Summary

HAWAII
SPRING 1987

STANFORD ACHIEVEMENT TEST
FORM E-7th EDITION NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
3	12,500	77 (42)	81 (46)	81 (75)	81 (N/A)	N/A	N/A
6	11,500	77 (49)	78 (55)	79 (76)	80 (N/A)	N/A	N/A

Percentage of children average and above average (Stanine 4 or above)
National Group Percentile Ranks in parentheses

SOURCE: Administrator of Testing--Hawaii Department of Education

IDAHO
MARCH-APRIL 1986

IOWA TEST OF BASIC SKILLS
FORM G NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
8	14,072	57	55	55	54	59.5%	80.2%

National Percentile Rank

SOURCE: Coordinator for Guidance, Assessment and Evaluation
Idaho State Department of Education

INDIANA
1983-84 (Last Year Results Collected)

Various Nationally Normed Tests
ITBS Most Common

Grade	Number Tested	Reading	Language (Composition)	Math	Total Battery	% of Students Above Average	% of Districts Above Average
Primary	N/A	87.9	89	86.2	85.1	N/A	N/A

Scores Reported as Percentage of Students in Stanine 4 or Above
(All scores indicate above-average achievement)

SOURCE: School Achievement in Indiana

IOWA
1986-87

IOWA TEST OF BASIC SKILLS
FORM G & H NATIONAL NORMS

Grade	% of Students Above Average	% of Iowa Schools Above Average
K	68	86
1	74	90
2	72	91
3	71	90
4	72	92
5	73	94
6	73	96

SOURCE: Score Reports and Norms - 1986-87

KENTUCKY
APRIL 1986

KENTUCKY ESSENTIAL SKILLS TEST
MODIFICATION OF CTBS - FORM U

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
1	53,527	52.6	N/A	65.9	N/A	N/A	N/A
2	47,337	54.4	51.8	71.3	57.8	64.1%	93.3%
3	45,855	60.8	69.3	59.3	66.1	79.6%	100%
4	43,035	58.2	56.4	59.2	57.5	61.9%	92.7%
5	42,400	53	60.8	59.5	56.5	60.1%	92.7%
6	40,403	56.8	54.6	62.3	58.1	64.1%	95.5%

Scores in Normal Curve Equivalents

SOURCE: KEST Statewide Testing Results

MAINE
1985-86

MAINE EDUCATIONAL ASSESSMENT (State Developed Test)
Matrix Sampling and Common Testing — 16 Different Tests

Grade	Number Tested	Reading	Math	Social Studies	Science	% of Students Above Average	% of Districts Above Average
4	14,012	+13.5	+6.9	+0.4	+2.2	N/A	N/A

Scores reported in terms of being above average in average percent correct

SOURCE: Maine Educational Assessment - 1985-86 State Summary and Interpretations Report

MARYLAND
FALL 1985

CALIFORNIA ACHIEVEMENT TEST
FORM C — NATIONAL NORM

Grade	Number Tested	Reading Comprehension	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
3	37,458	3.6 (70)	3.8 (79)	3.5 (75)	N/A	N/A	95.8%
5	37,644	6.0 (83)	7.0 (98)	5.9 (90)	N/A	N/A	87.5%

Grade Equivalent Scale
National Percentile Rank in parentheses

SOURCE: Maryland Accountability Testing Program - Annual Report 1985-86

MISSISSIPPI
APRIL 1987

STANFORD ACHIEVEMENT TEST
FORM E — STANFORD NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
1	43,560	53.2	N/A	49.9	51.7	N/A	N/A
4	37,017	47.9	52.0	50.7	49.5	N/A	N/A
6	33,477	46.6	51.7	50.7	49.1	N/A	N/A
					50.1 (Average)		

Mean National Normal Curve Equivalents

SOURCE: Skills Analysis for Mississippi 1987 Results

NEVADA
1985-86

STANFORD ACHIEVEMENT TEST
7th EDITION — NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
3	10,000	93	94	93	N/A	N/A	N/A
6	10,000	90	89	90	N/A	N/A	N/A

The percentage of students who scored in Stanine 4 or Above
(Only information available)

SOURCE: The Nevada Proficiency Examination Program - Results of 1985/1986 Examinations

NEW HAMPSHIRE
OCTOBER 1986

CALIFORNIA ACHIEVEMENT TEST
FORM E — NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
4	11,381	65	61	61	62	62.7%	100%

Median National Percentile Ranks

SOURCE: Highlights of Results - 1986

NEW MEXICO
MARCH 1987

COMPREHENSIVE TEST OF BASIC SKILLS
FORM U — NATIONAL 1981 NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
3	20,862	50.2	55.8	58.4	57.1	60%	80.7%
5	19,139	54.0	56.0	62.3	55	59%	79.6%

National Percentile Ranks

SOURCE: New Mexico Standardized Testing Program Report - 1985-86 School Year
New Mexico Department of Education

NORTH CAROLINA
SPRING 1987

CALIFORNIA ACHIEVEMENT TEST
FORM E — NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
1	86,050	57	53	71	N/A	N/A	N/A
2	80,946	53	60	69	62	N/A	88.7%
3	79,892	54	60	63	58	N/A	87.9%
6	80,413	50	56	59	55	N/A	79.4%

Median National Percentile

SOURCE: Report of Student Performance, Division of Research

NORTH DAKOTA
SEPTEMBER 1986
(Both SRA and ITBS Used in North Dakota)

IOWA TEST OF BASIC SKILLS
FORM G NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
2	214	77	83	71	77	N/A	N/A
3	2,774	67	65	60	65	N/A	N/A
4	694	69	66	59	64	N/A	N/A
5	2,008	68	64	64	66	N/A	N/A
6	603	68	67	63	67	N/A	N/A

Percentile Rank of Average Grade Equivalent

SEPTEMBER 1986

SCIENCE RESEARCH ASSOCIATES
FORM P NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
1	785	63	N/A	69	55	N/A	N/A
2	1,033	69	64	66	69	N/A	N/A
3	2,495	72	68	56	65	N/A	N/A
4	1,277	74	72	62	72	N/A	N/A
5	2,286	71	68	66	70	N/A	N/A
6	1,055	73	68	71	72	N/A	N/A

National Percentile Ranks

SOURCE: Director of Testing, North Dakota - Department of Public Instruction

OKLAHOMA
FEBRUARY 1987

METROPOLITAN ACHIEVEMENT TEST
6th EDITION NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
3	41,100	62	60	61	64	N/A	79%

National Percentile Rank of Mean Normal Curve Equivalents

SOURCE: Oklahoma School Testing Program Summary Report - 1986

OREGON
FEBRUARY 1987

OREGON ASSESSMENT PROGRAM
(State Developed Test)
Equating Study with California Achievement Test
FORM E NATIONAL NORMS

Grade*	Number Tested**	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
8	4,513	58	N/A	49	N/A	N/A	N/A

National Percentile Rank

*Earliest Grade Tested

**Sample Of Students Tested

SOURCE: 1987 Eighth Grade Assessment of Reading, Mathematics and Writing Summary Report, July 1987

PENNSYLVANIA
MARCH 1987

TESTING FOR ESSENTIAL LEARNING AND LITERACY SKILLS (TELLS)
Equating Study with ITBS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
3	105,564	79.4*	N/A	84.6*	N/A	N/A	N/A
5	100,724	78.3*	N/A	81.6*	N/A	N/A	N/A

Mean Percent of Items Correct

*All scores are above the Estimated National Mean, but national norm is not 50 percentile

SOURCE: Testing for Essential Learning and Literacy Skills 1986-87 Statewide Test Results

RHODE ISLAND
SPRING 1986

RHODE ISLAND STATEWIDE ASSESSMENT PROGRAM Scores Normed with METROPOLITAN ACHIEVEMENT TEST (MAT), Edition 6

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
3	8,310	64	64	59	63	63.9	94.7%
6	8,381	64	65	63	65	65.4	91.9%

Percentile Rank of Average Scores

SOURCE: Rhode Island State Assessment Program 1985-86 Basic Skills, Health Knowledge and Physical Fitness Testing Results

SOUTH CAROLINA
MARCH 1987

COMPREHENSIVE TEST OF BASIC SKILLS
FORM U NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
4	44,946	56.1	64.3	67.6	61.2	62.9%	82.4%
5	42,831	50.9	59.4	64.2	54.4	54.5%	52.8%

Median National Percentile Scores

SOURCE: Preliminary Report of the 1987 Statewide Testing Program

SOUTH DAKOTA
APRIL 1987

STANFORD ACHIEVEMENT TEST
FORM F NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
4	9,562	58.8 (85)	59.7 (90)	60.1 (91)	57.9	72%	N/A

Normal Curve Equivalents

National Percentile Ranks in parentheses

SOURCE: Skills Analysis for South Dakota State

TENNESSEE
SPRING 1987

STANFORD ACHIEVEMENT TEST
7th EDITION NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
2	62,456	64	86	86	N/A	N/A	82.9%
5	57,717	52	66	68	N/A	N/A	61.9%

Group Percentile Rank of the Mean Scaled Scores

SOURCE: Tennessee Test Results - 1986-87

TEXAS
FEBRUARY 1986

TEXAS EDUCATIONAL ASSESSMENT OF MINIMUM SKILLS Equated with Metropolitan Achievement Test 6th Edition

Grade	Number Tested	Reading	Writing	Math	Total Battery	% of Students Above Average	% of Districts Above Average
1	268,656	62	66	70	N/A	N/A	N/A
3	236,592	47	61	62	N/A	N/A	N/A
5	225,601	53	63	62	N/A	N/A	N/A

SOURCE: Student Performance Results for 1985 and 1986

UTAH
APRIL 1984

COMPREHENSIVE TEST OF BASIC SKILLS
FORM U NATIONAL NORMS

Grade	Number Tested*	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
5	4,500	59	55	65	N/A	N/A	N/A

*Sample of students tested

Utah administers the test to students under strict security (no regular teachers administer the test)

SOURCE: Utah Statewide Educational Assessment General Report 1984

VIRGINIA
SPRING 1987

SCIENCE RESEARCH ASSOCIATES TEST
LEVEL E - FORM 1 NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
4	67,567	58	62	61	N/A	N/A	60.5%

National Percentile Ranks

SOURCE: Virginia State Assessment Program 1986-87

WASHINGTON
OCTOBER 1986

METROPOLITAN ACHIEVEMENT TEST
6th EDITION NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
4	52,779	56	52	53	56	N/A	65.1%

Median National Percentile Rank

SOURCE: Washington Statewide Assessment Grades 4, 8, 10-Fall 1986

WEST VIRGINIA
MARCH 1987

COMPREHENSIVE TEST OF BASIC SKILLS
FORM U NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
3	23,961	61	72	54	65	69.2%	98.2%
6	23,371	58	62	66	62	64.5%	92.7%

Mean Percentile Scores

SOURCE: Coordinator, West Virginia State-County Testing Program

WISCONSIN
FEBRUARY-MARCH 1986

COMPREHENSIVE TEST OF BASIC SKILLS
FORM U NATIONAL NORMS

Grade	Number Tested	Reading	Language	Math	Total Battery	% of Students Above Average	% of Districts Above Average
4	1,500*	68	63	63	N/A	N/A	N/A

Median National Percentile Rank
*Sample Of Students Tested
SOURCE: Wisconsin Department of Public Instruction

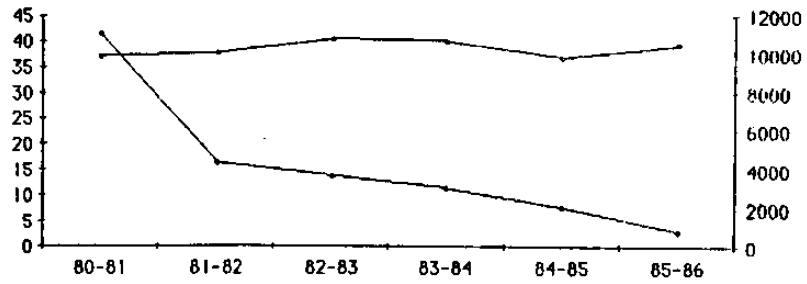
Attachment 4: Popham, 1987

TABLE 1.
Improvements in Student Achievement Associated with Measurement-Driven Instruction

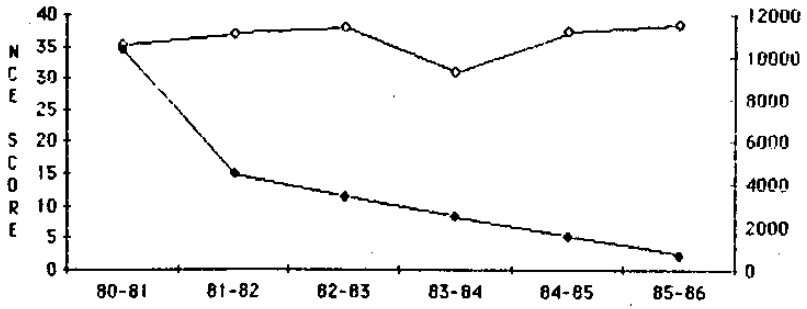
Locale	Subjects	Grade(s)	Period	Improvement (%)*
Ala.	3 R's	3,6,9	1981-86	1-13
	3 R's	11	1983-85	4-8
Conn.	3 R's	9	1980-84	6-16
Detroit	3 R's	12	1981-86	19
Md.	3 R's	9	1980-86	13-25
	Citizenship	9	1983-86	23
N.J.	Reading and Math	9	1977-85	16-19
	Reading and Math	10	1982-85	8-11
S.C.	Readiness	1	1979-85	14
	Reading and Math	1,2,3,6,8	1981-86	12-20
Tex.	Reading and Math	3,5,9	1980-85	7-14

*Figures for improvement represent the increased percentage of students who have mastered standards of quality during the period in question. (For Alabama, in grades 3, 6, and 9, the percentage of improvement reflects test-score percentage improvements rather than improvements in mastery of standards.)

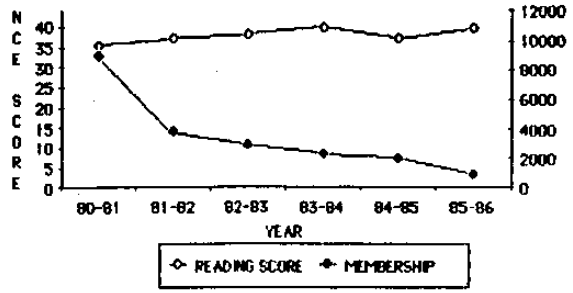
ALABAMA GRADE 2 READING & MEMBERSHIP



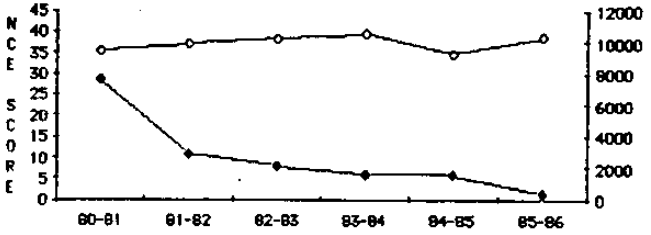
ALABAMA GRADE 3 READING SCORES & MEMBERSHIP



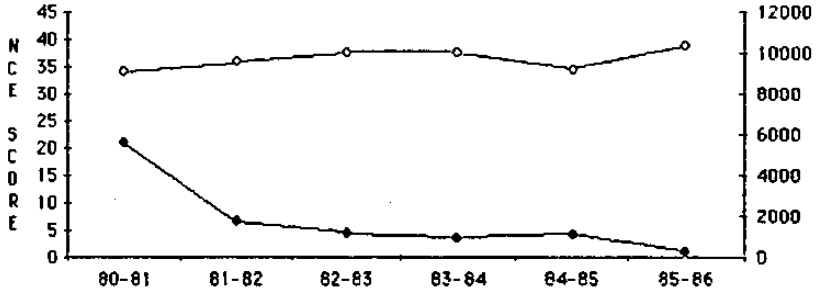
ALABAMA GRADE 4 READING SCORES & MEMBERSHIP



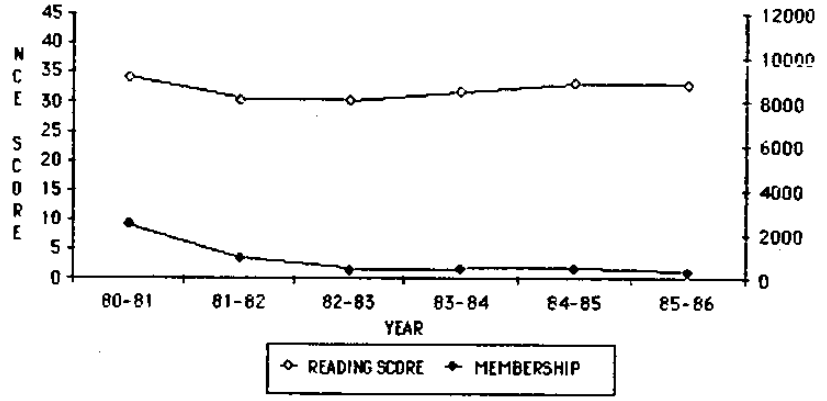
ALABAMA GRADE 5 READING SCORES & MEMBERSHIP



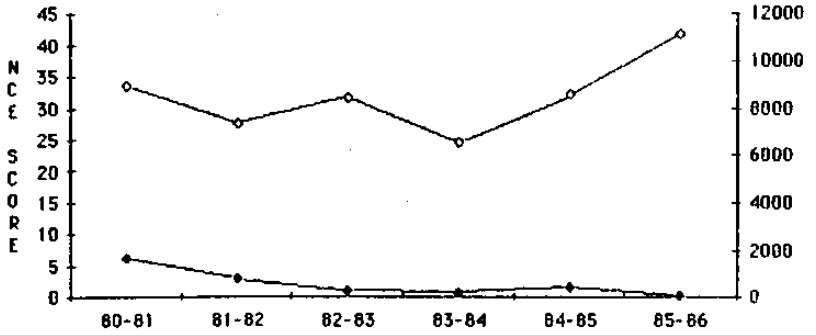
ALABAMA GRADE 6 READING SCORES & MEMBERSHIP



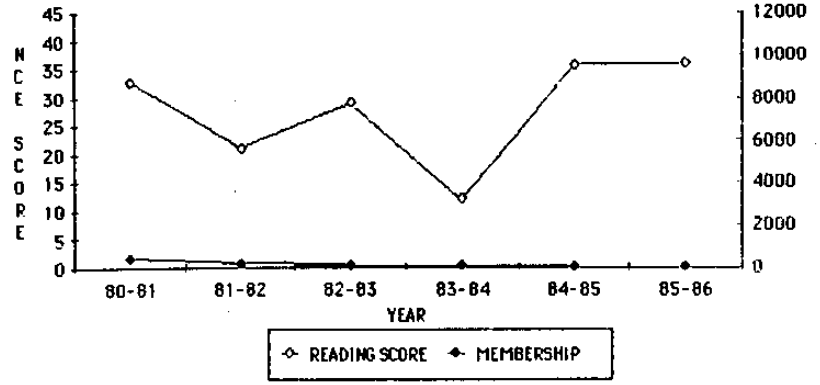
ALABAMA GRADE 7 READING SCORES & MEMBERSHIP



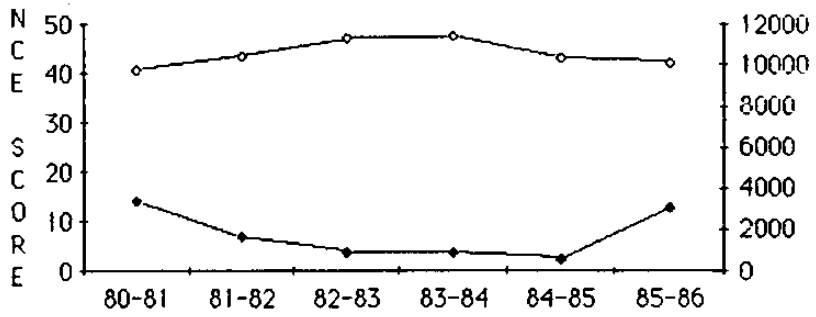
ALABAMA GRADE 8 READING SCORES & MEMBERSHIP



ALABAMA GRADE 9 READING SCORES & MEMBERSHIP

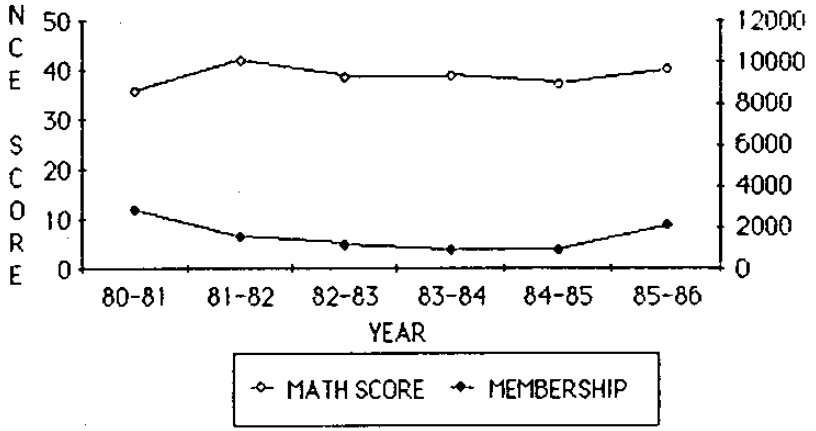


ALABAMA GRADE 2 MATH SCORES & MEMBERSHIP

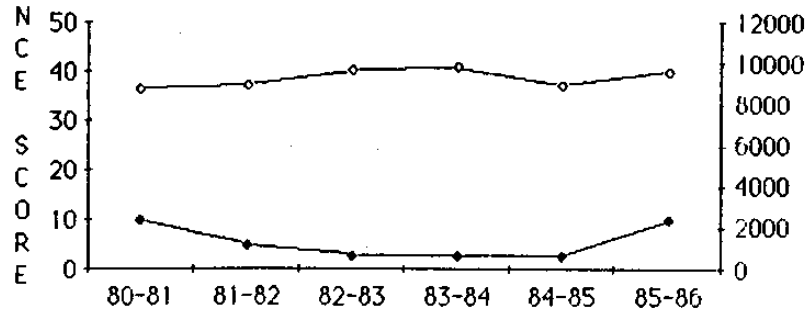


ALABAMA GRADE 3 MATH SCORES & MEMBERSHIP

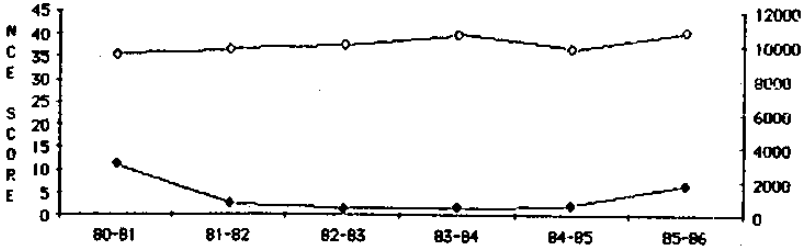
ALABAMA GRADE 4 MATH SCORES & MEMBERSHIP



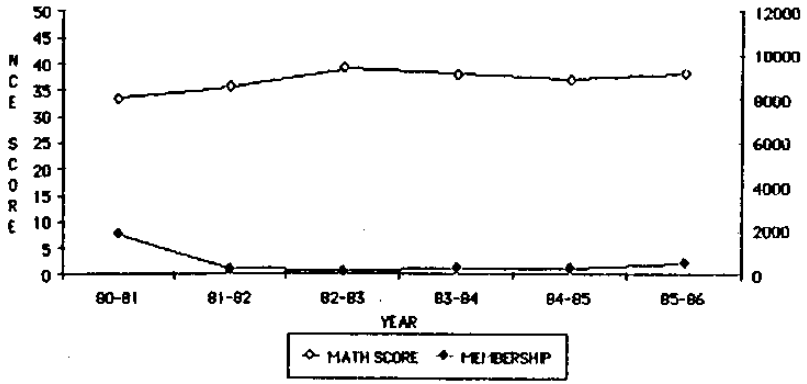
ALABAMA GRADE 5 MATH SCORES & MEMBERSHIP



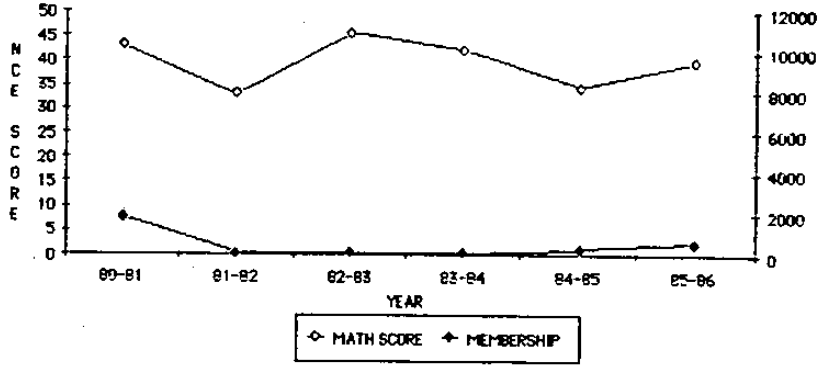
ALABAMA GRADE 6 MATH SCORES & MEMBERSHIP



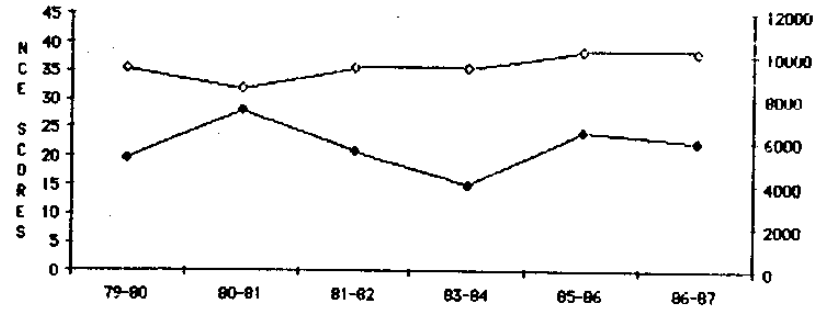
ALABAMA GRADE 7 MATH SCORES & MEMBERSHIP



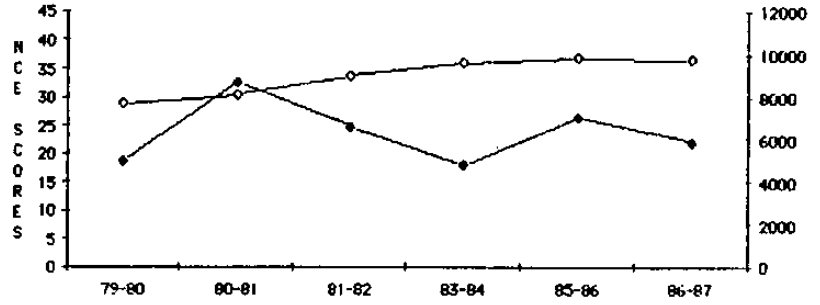
ALABAMA GRADE 8 MATH SCORES & MEMBERSHIP



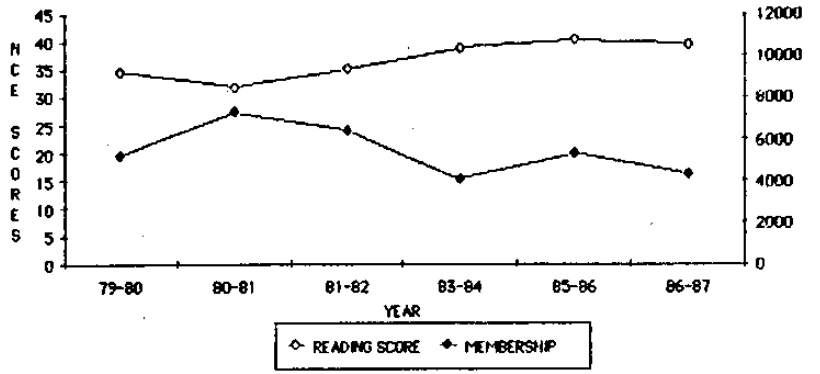
MARYLAND GRADE 2 READING SCORES & MEMBERSHIP



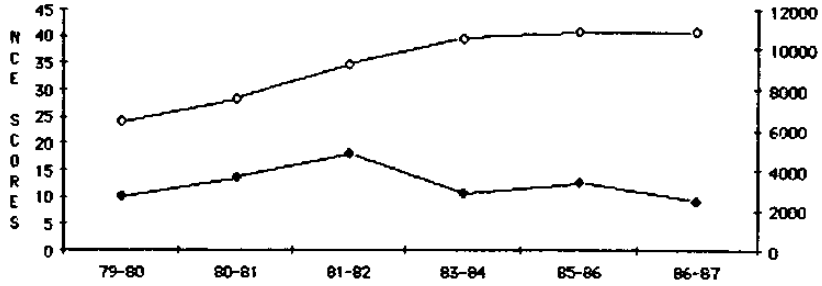
MARYLAND GRADE 3 READING SCORES & MEMBERSHIP



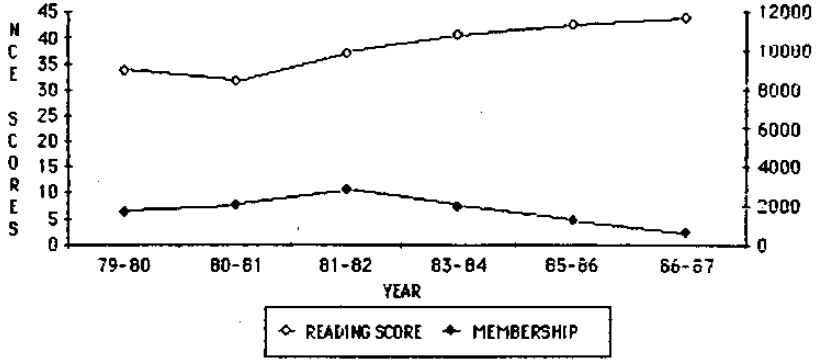
MARYLAND GRADE 4 READING SCORES & MEMBERSHIP



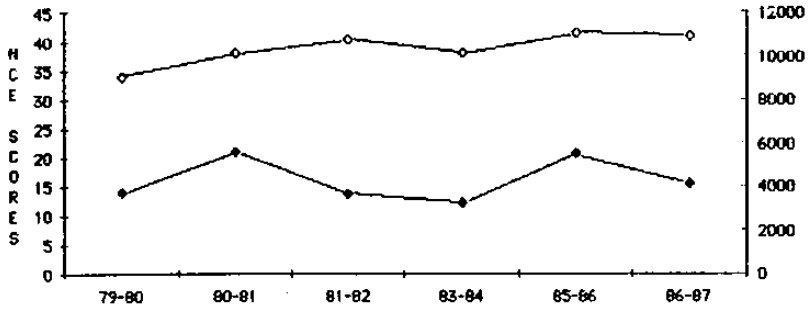
MARYLAND GRADE 5 READING SCORES & MEMBERSHIP



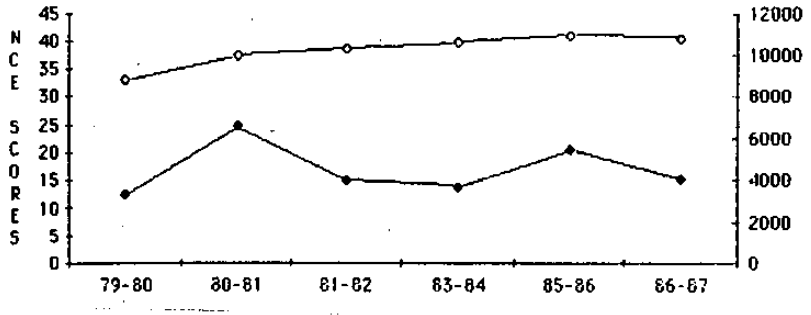
MARYLAND GRADE 6 READING SCORES & MEMBERSHIP



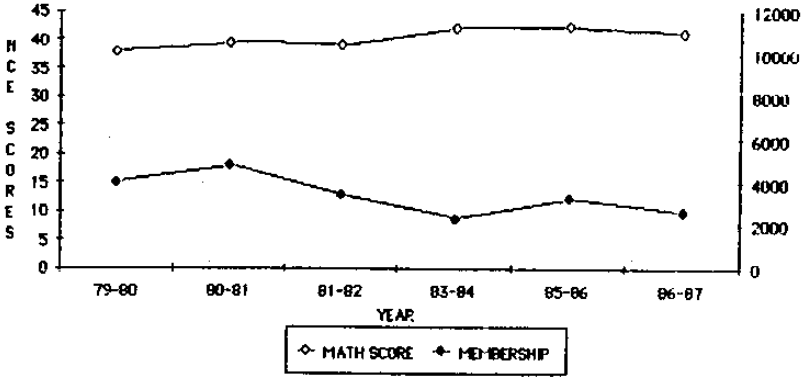
MARYLAND GRADE 2 MATH SCORES & MEMBERSHIP



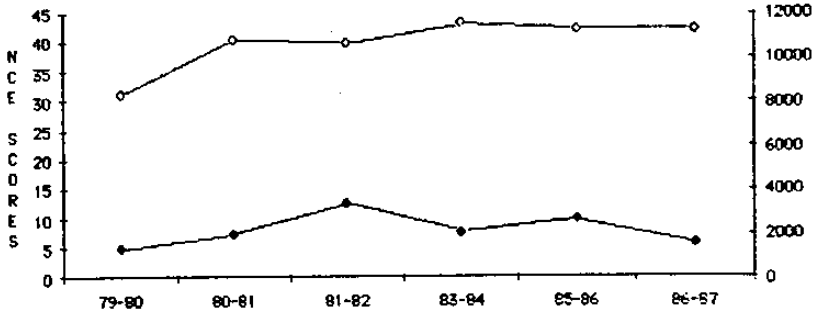
MARYLAND GRADE 3 MATH SCORES & MEMBERSHIP



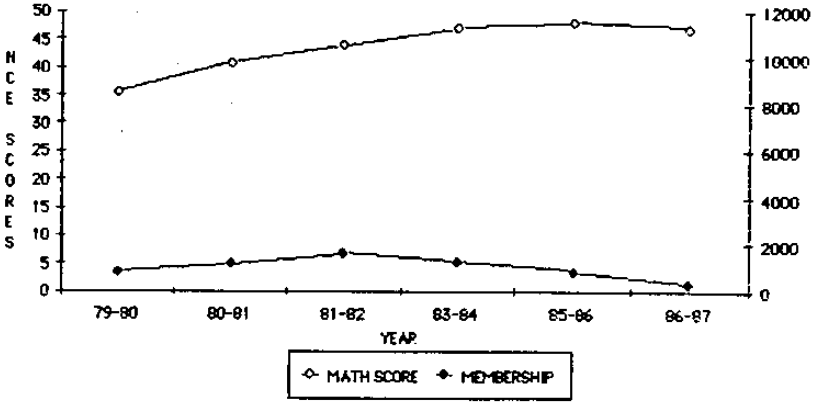
MARYLAND GRADE 4 MATH SCORES & MEMBERSHIP



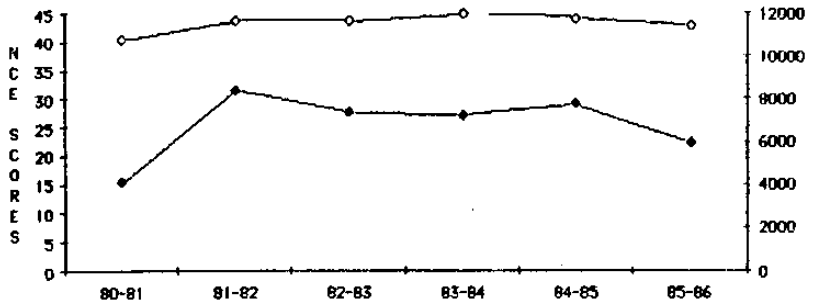
MARYLAND GRADE 5 MATH SCORES & MEMBERSHIP



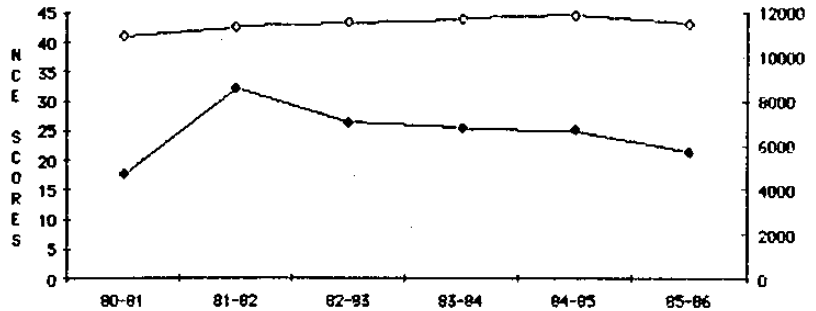
MARYLAND GRADE 6 MATH SCORES & MEMBERSHIP



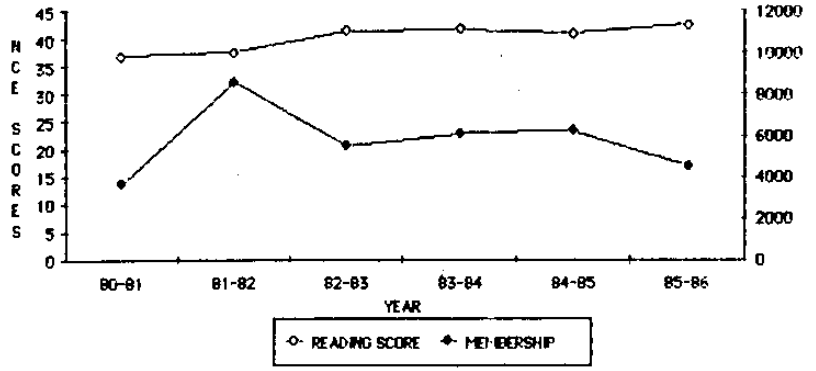
NEW JERSEY GRADE 2 READING SCORES & MEMBERSHIP



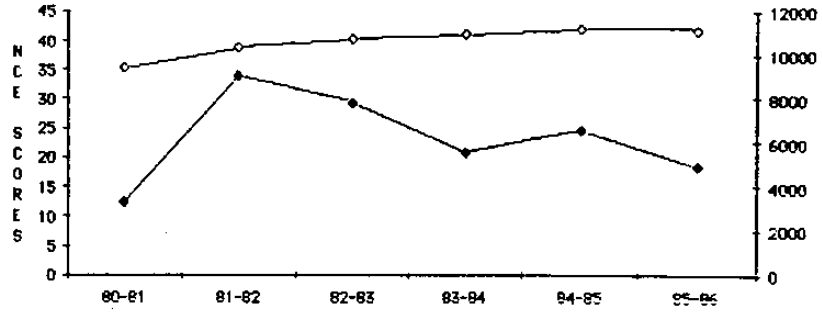
NEW JERSEY GRADE 3 READING SCORES & MEMBERSHIP



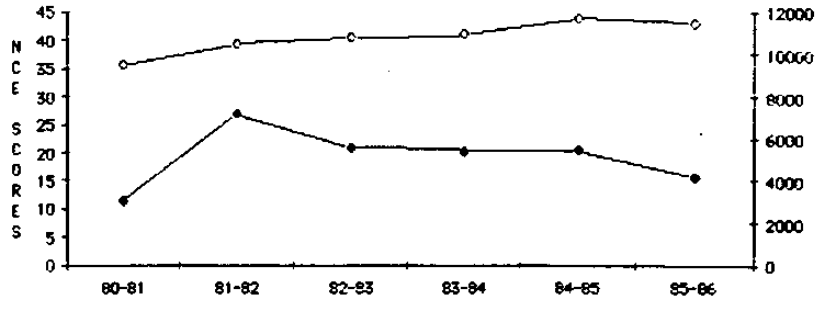
NEW JERSEY GRADE 4 READING SCORES & MEMBERSHIP



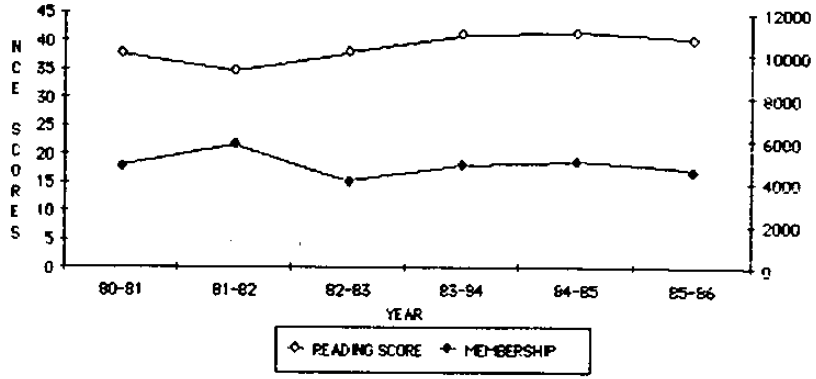
NEW JERSEY GRADE 5 READING SCORES & MEMBERSHIP



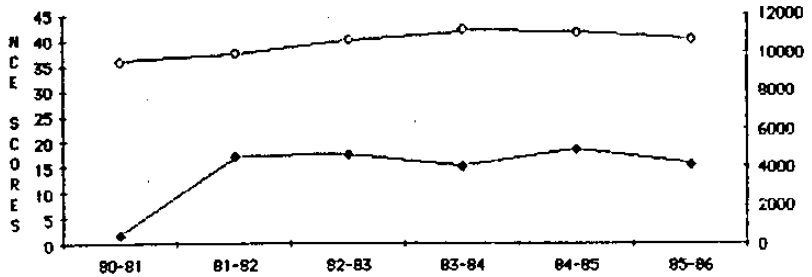
NEW JERSEY GRADE 6 READING SCORES & MEMBERSHIP



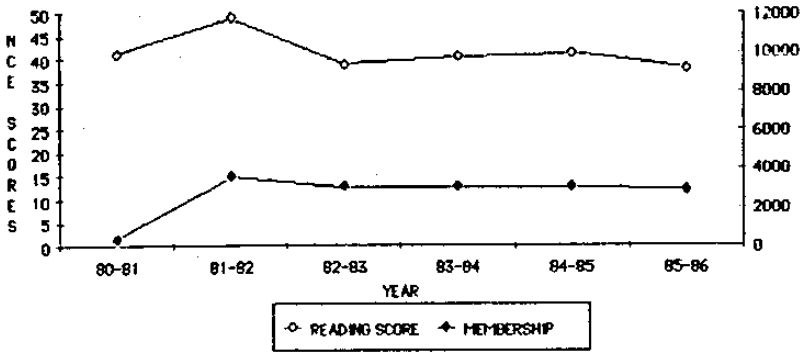
NEW JERSEY GRADE 7 READING SCORES & MEMBERSHIP



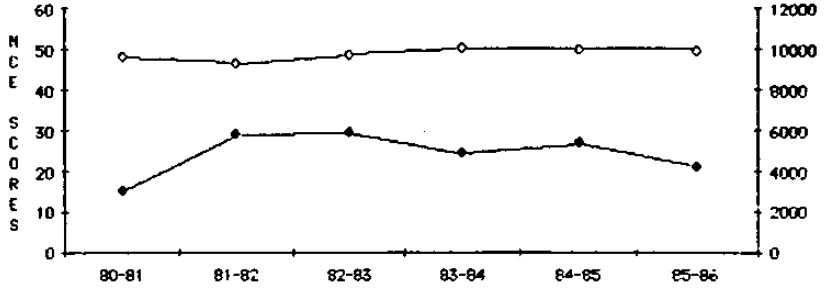
NEW JERSEY GRADE 8 READING SCORES & MEMBERSHIP



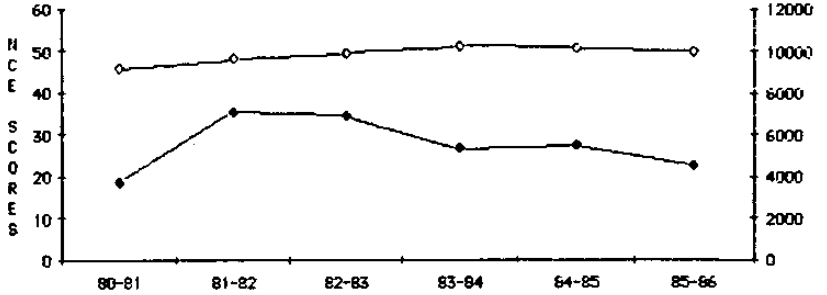
NEW JERSEY GRADE 9 READING SCORES & MEMBERSHIP



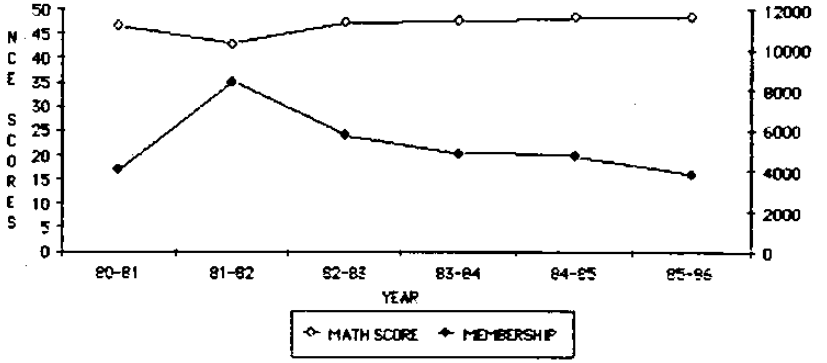
NEW JERSEY GRADE 2 MATH SCORES & MEMBERSHIP



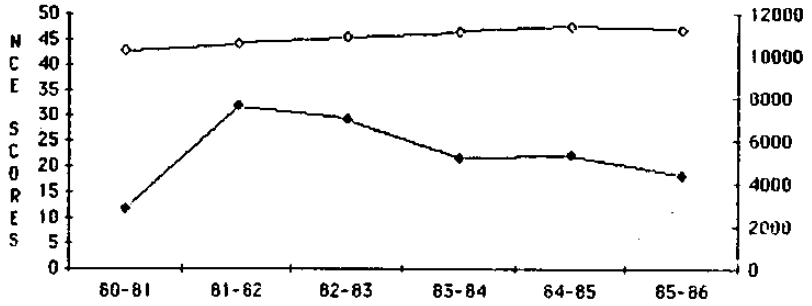
NEW JERSEY GRADE 3 MATH SCORES & MEMBERSHIP



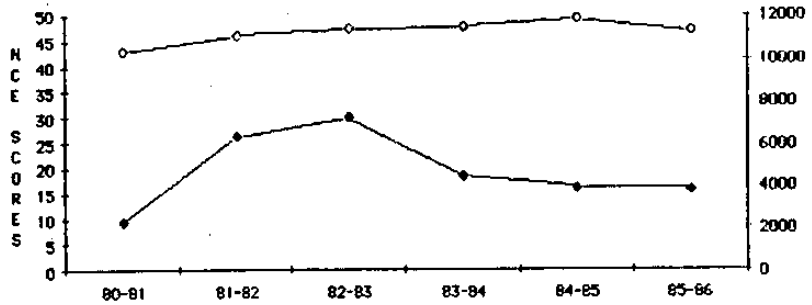
NEW JERSEY GRADE 4 MATH SCORES & MEMBERSHIP



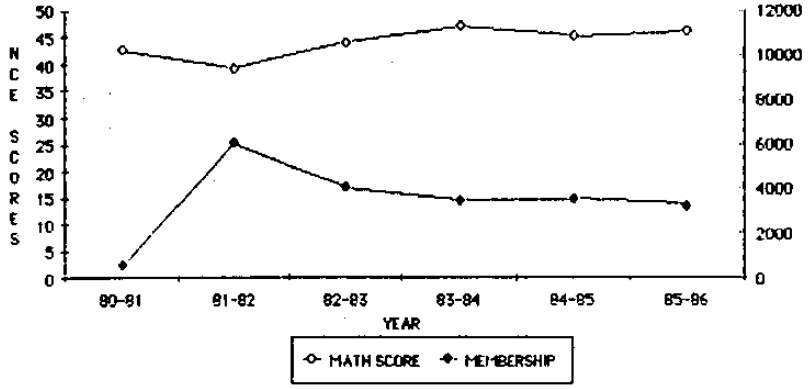
NEW JERSEY GRADE 5 MATH SCORES & MEMBERSHIP



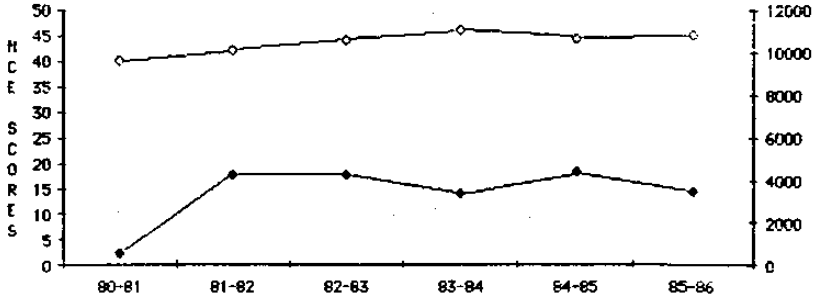
NEW JERSEY GRADE 6 MATH SCORES & MEMBERSHIP



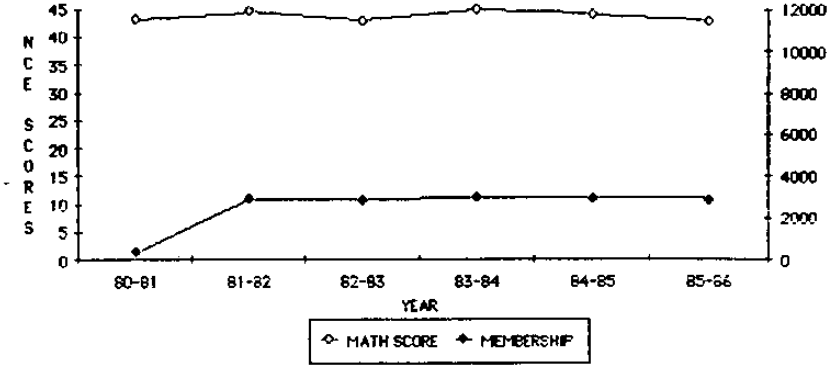
NEW JERSEY GRADE 7 MATH SCORES & MEMBERSHIP



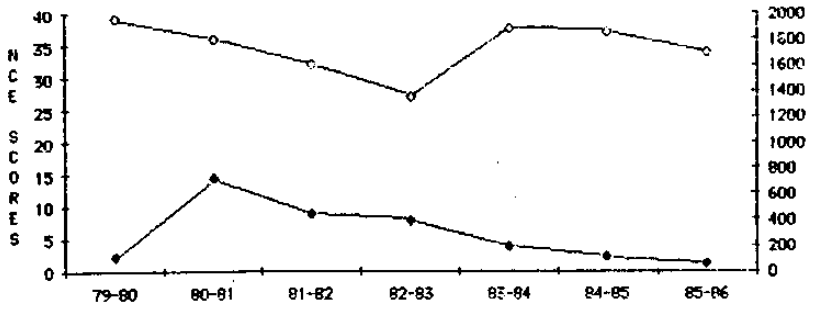
NEW JERSEY GRADE 8 MATH SCORES & MEMBERSHIP



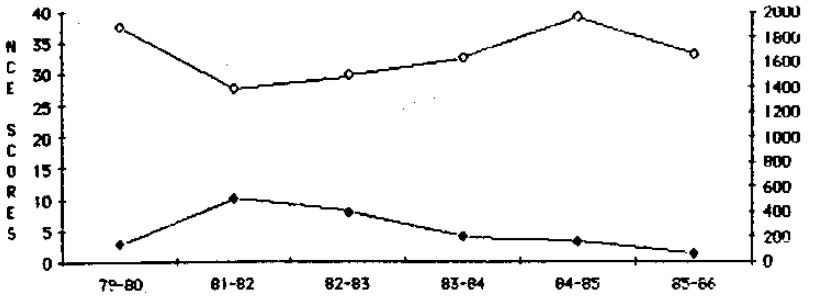
NEW JERSEY GRADE 9 MATH SCORES & MEMBERSHIP



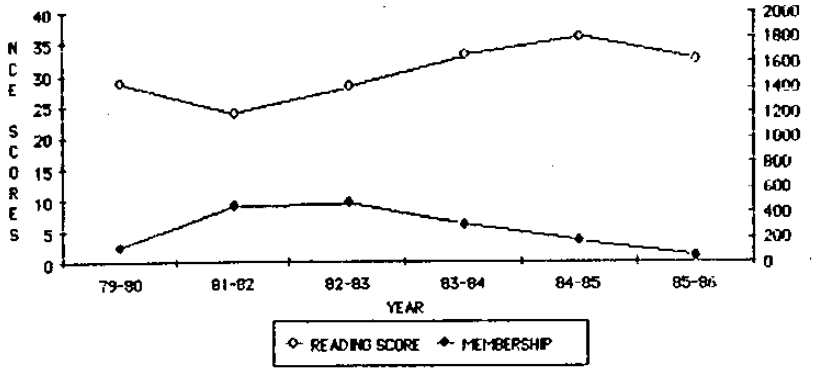
SOUTH CAROLINA GRADE 2 READING SCORES & MEMBERSHIP



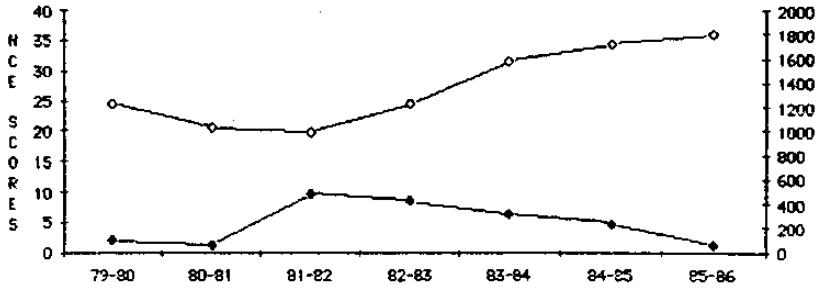
SOUTH CAROLINA GRADE 3 READING SCORES & MEMBERSHIP



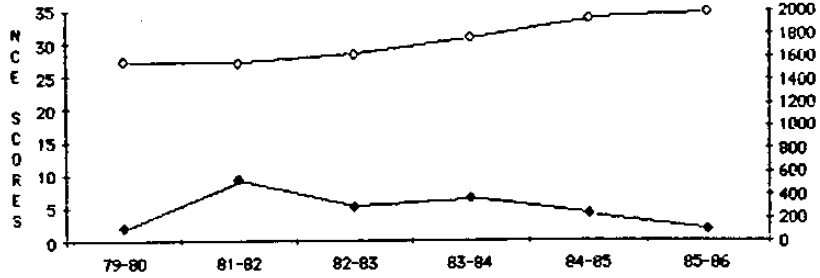
SOUTH CAROLINA GRADE 4 READING SCORES & MEMBERSHIP



SOUTH CAROLINA GRADE 5 READING SCORES & MEMBERSHIP

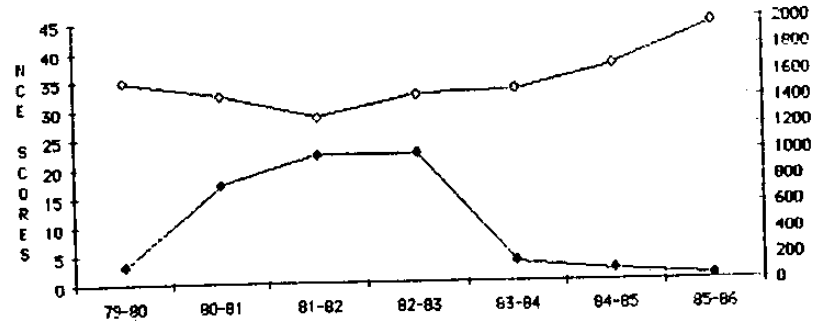


SOUTH CAROLINA GRADE 6 READING SCORES & MEMBERSHIP

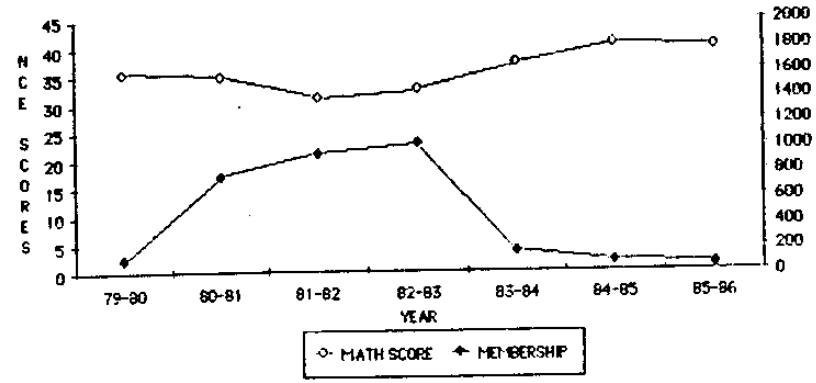


SOUTH CAROLINA GRADE 2 MATH SCORES & MEMBERSHIP

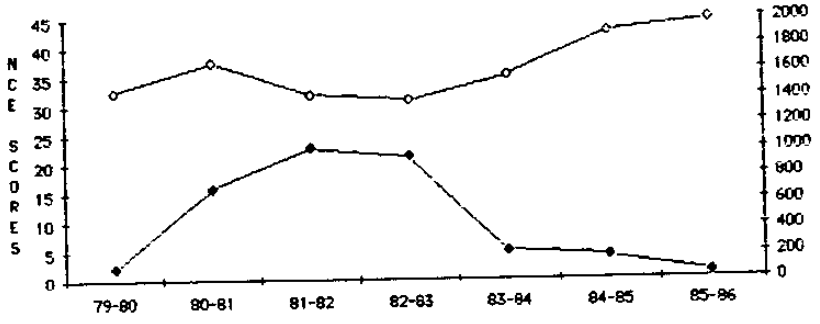
SOUTH CAROLINA GRADE 3 MATH SCORES & MEMBERSHIP



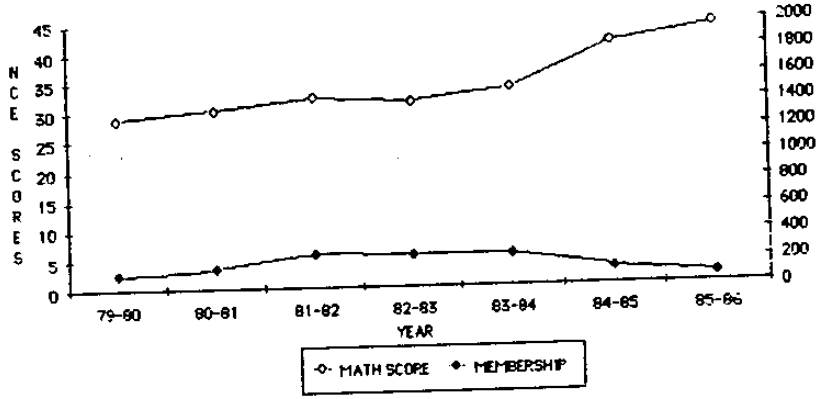
SOUTH CAROLINA GRADE 4 MATH SCORES & MEMBERSHIP

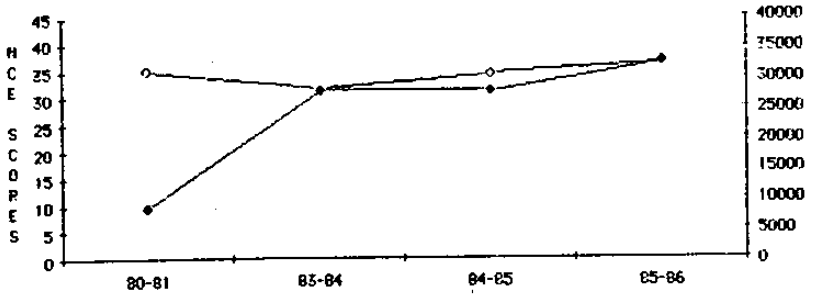


SOUTH CAROLINA GRADE 5 MATH SCORES & MEMBERSHIP

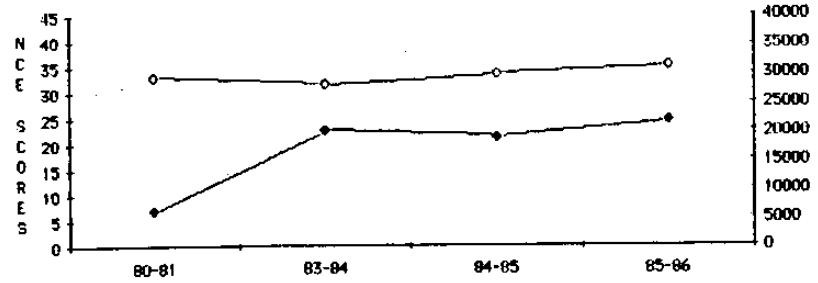


SOUTH CAROLINA GRADE 6 MATH SCORES & MEMBERSHIP

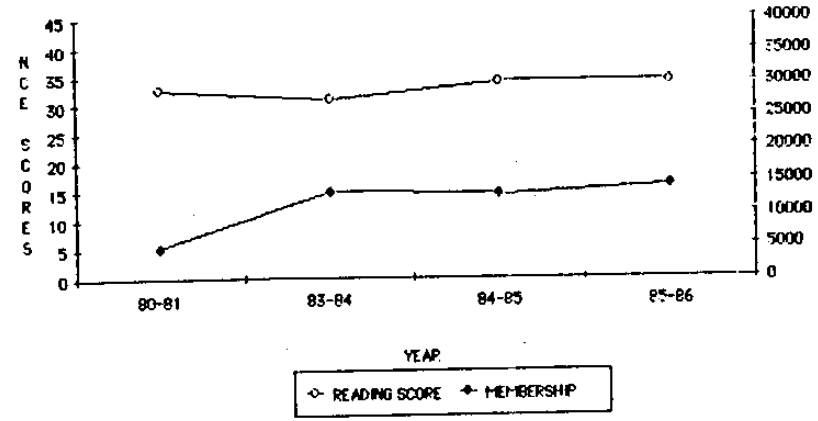




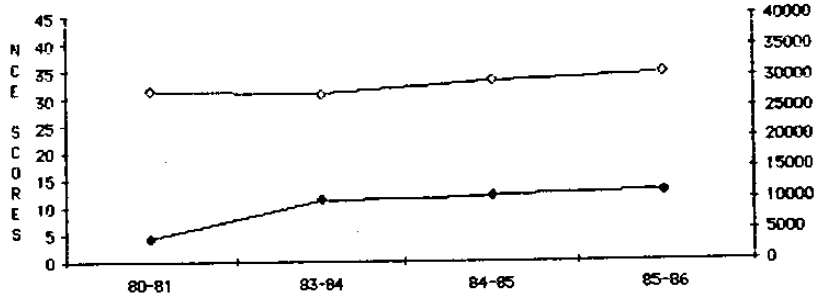
TEXAS GRADE 6 READING SCORES & MEMBERSHIP



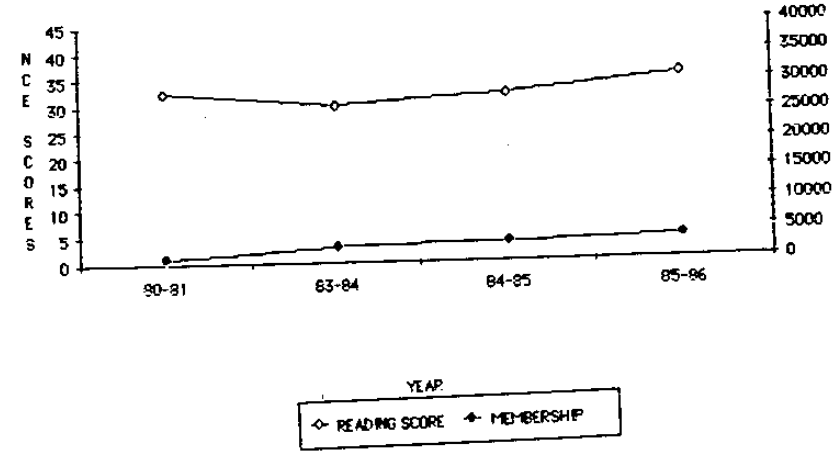
TEXAS GRADE 7 READING SCORES & MEMBERSHIP



TEXAS GRADE 8 READING SCORES & MEMBERSHIP

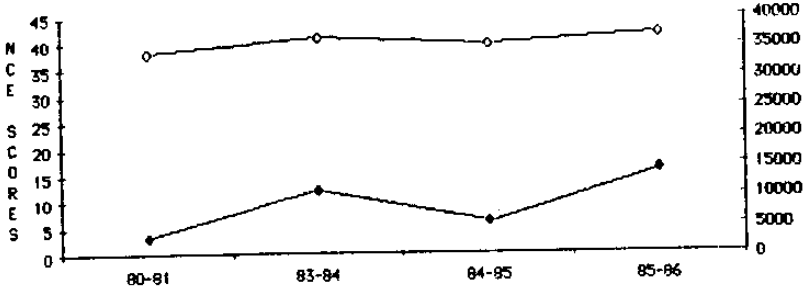


TEXAS GRADE 9 READING SCORES & MEMBERSHIP

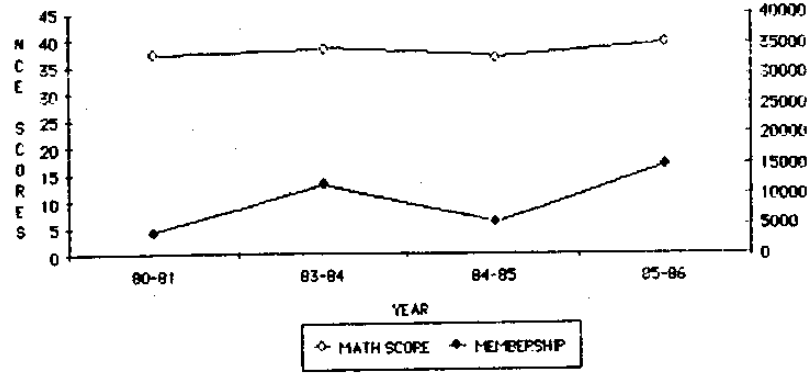


TEXAS GRADE 2 MATH SCORES & MEMBERSHIP

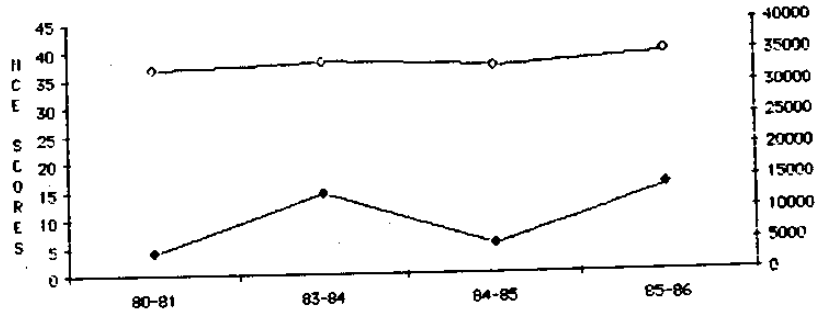
TEXAS GRADE 3 MATH SCORES & MEMBERSHIP



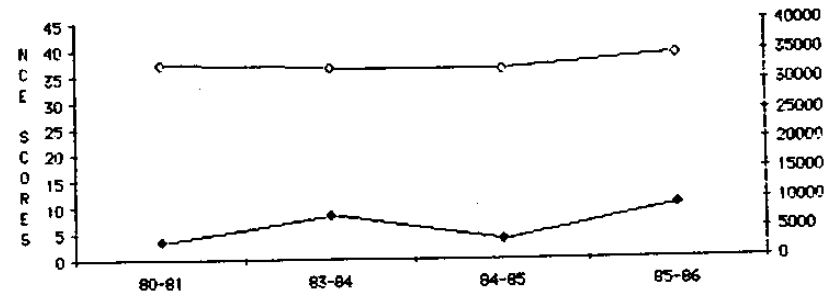
TEXAS GRADE 4 MATH SCORES & MEMBERSHIP



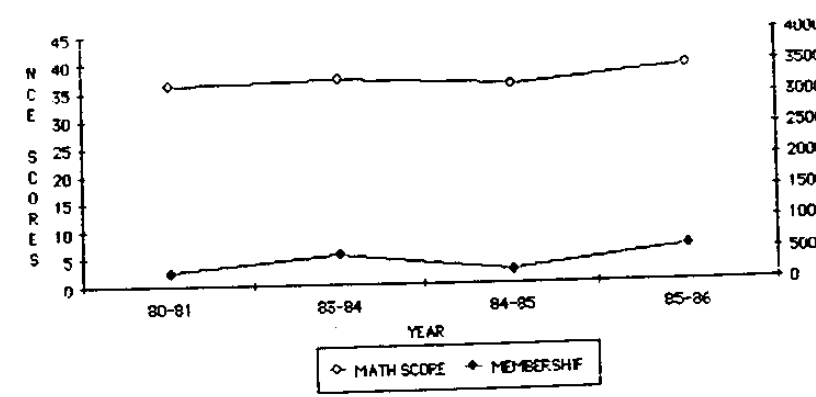
TEXAS GRADE 5 MATH SCORES & MEMBERSHIP



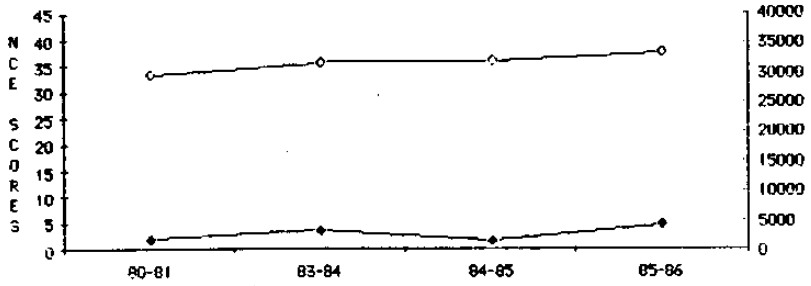
TEXAS GRADE 6 MATH SCORES & MEMBERSHIP



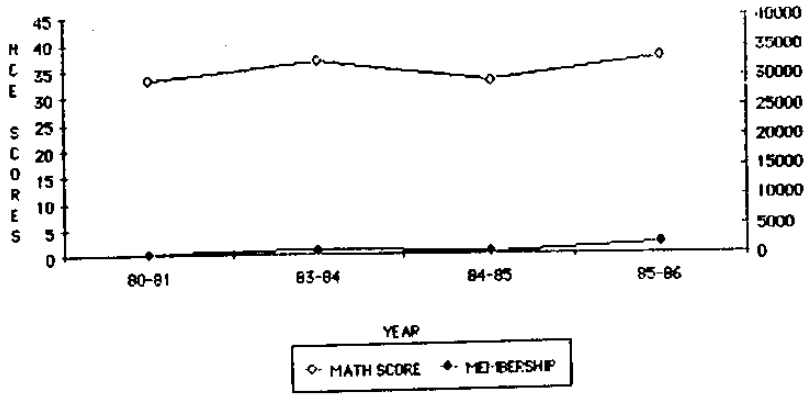
TEXAS GRADE 7 MATH SCORES & MEMBERSHIP



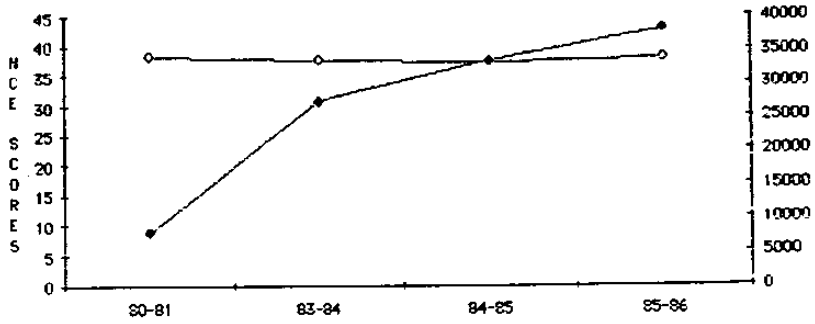
TEXAS GRADE 8 MATH SCORES & MEMBERSHIP



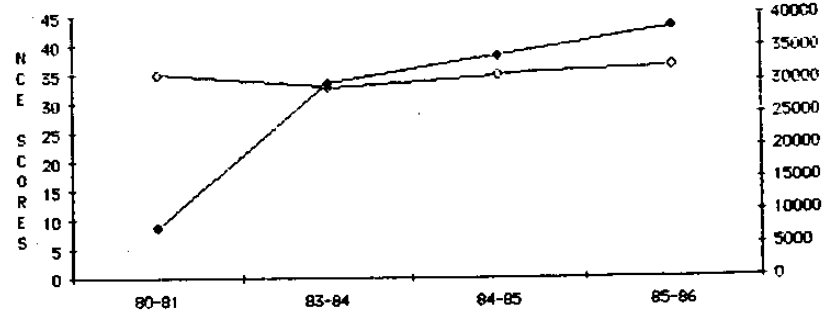
TEXAS GRADE 9 MATH SCORES & MEMBERSHIP



TEXAS GRADE 2 READING SCORES & MEMBERSHIP



TEXAS GRADE 3 READING SCORES & MEMBERSHIP



TEXAS GRADE 4 READING SCORES & MEMBERSHIP