

**IS THE STANDARDIZED TEST SCORE AN ACCURATE PREDICTOR  
OF  
ACADEMIC PERFORMANCE?**

THESIS

Submitted to the Graduate Committee of the Department of  
Curriculum and Instruction

Faculty of Education

State University College at Brockport

In Partial Fulfillment of the Requirements for the Degree of  
Master of Science in Education

by

Barbara E. Johnson

State University College at Brockport

Brockport, New York

March 1994

SUBMITTED BY

Barbara E. Johnson

APPROVED BY:

Dee-Ann L. Rigg  
Project/Thesis Advisor

5/10/94  
Date

Arthur E. Smith  
Second Faculty Reader

5/12/94  
Date

Barbara E. Baker  
Chairman, Graduate  
Policies Committee

5/10/94  
Date

## **Acknowledgments**

Many thanks to the following people who helped make this work possible:

- My professors, Dr. Begy, Dr. Smith and Dr. Ribble for the time which they contributed towards the production of this thesis.
- My friend, Linda Cieminis, for helping me formulate the idea for this paper.
- My co-workers for their encouragement.
- And most of all my family. My husband, Tom, and children, Chad, Seth and Katelyn, who provided the love and support necessary for me to complete this project.

Thank You

## Table of Contents

	<b>Page</b>
<b>Chapter I Statement of the Problem</b>	<b>6</b>
Purpose	9
Definition of Terms	10
<b>Chapter II Review of the Literature</b>	<b>13</b>
What is standardized testing?	13
How are test data used?	15
What attitude is held toward test usage?	18
What affects the test scores?	20
<b>Chapter III Design of the Study</b>	<b>24</b>
Hypothesis	24
Methodology	25
Analysis of Data	26
Summary of the Chapter	27
<b>Chapter IV Analysis of Data</b>	<b>28</b>
Findings and Interpretations	28
<b>Chapter V Conclusions and Implications</b>	<b>34</b>
Implications for Research	35
Summary of the Chapter	41
<b>References</b>	<b>42</b>
<b>Appendix</b>	<b>47</b>
Data from PEP, Otis Lennon, Regents English exam, unweighted and weighted grade point averages, SAT-Reading and ACT-Reading scores	

## **Abstract**

The purpose of this study was to determine whether the third grade Pupil Evaluation Program Test or the Otis Lennon School Ability Index or both together were valid predictors of future academic performance as evidenced by high grade point averages, high eleventh grade English Regents exam grades and high SAT and ACT Reading scores. Ninety-one graduates of 1993 who attended school in a suburb of Rochester, New York were identified. All had taken the Pupil Evaluation Program Test in the third grade and the Otis Lennon School Ability Index in grade ten. They had remained at the designated school continuously from kindergarten through grade twelve. The eleventh grade English Regents exam grade, unweighted high school grade point average, weighted high school grade point average, SAT Reading and ACT Reading scores were analyzed to see where the strongest relationships existed.

After testing it was found that the third grade Pupil Evaluation Program Test score was an unusually

strong predictor for performance on the eleventh grade Regents English exam ( $r^2 = .39$ ), the SAT Reading score ( $r^2 = .30$ ) and the ACT Reading score ( $r^2 = .33$ ) and a very strong predictor for performance as evidenced by a high unweighted ( $r^2 = .27$ ) and weighted grade point average ( $r^2 = .29$ ). Together with the Otis Lennon score the predictive relationship was even stronger. The Otis Lennon score alone was a minimal predictor for performance on the Regents English exam ( $r^2 = .14$ ), a moderate predictor for the unweighted grade point average ( $r^2 = .18$ ) and SAT-Reading score ( $r^2 = .18$ ), a strong predictor for the weighted grade point average ( $r^2 = .21$ ) and a very strong predictor for the ACT-Reading score ( $r^2 = .30$ ).

## **Chapter I - Statement of the Problem**

Formal assessment needs to be closely examined. The existing system can not be discarded until a better assessment mechanism is found. The demand for accountability means that any system for evaluating educational practice be acceptable to the public. The method must be able to assure the public that their children are getting a quality education, that their money is being used wisely, and must be fair and objective. If assessment is to be meaningful it must reflect current knowledge about the reading dynamic, give results from real tasks, be a natural part of classroom activity and beneficial to both student and teacher.

Schools must not relinquish their responsibilities for the educating of children to a standardized test that might be based on theories of reading and writing that are different. At the very least, the district would best clearly articulate its philosophy of reading and writing, and then examine whether the standardized tests they use conform to those principles.

Has reading assessment kept up with new knowledge gained in reading theory, research, or practice? Valencia and Pearson (1987), Levande (1993) and Heithaus (1989) argued not. On the one hand more sophisticated instructional materials and practices are sought, while at the same time the same standardized tests are passively accepted and used. They argued that since the influence of testing is greater now than ever before due to the accountability movement, national reports, and effective schools research, the time is ripe to assess the way reading is assessed. In so doing the knowledge gained in the last 15-20 years regarding basic reading processes could be incorporated.

Valencia and Pearson (1987) cited recent research that views reading as the strategic use of available resources to make sense of the text, as opposed to the formerly held view that expert reading is the culmination of the acquisition of component skills. They felt that current assessment methods have not been influenced by this strategic view of reading. Therefore, as long as reading instructional innovations are based upon one understanding of reading and the testing on another, there will be frustration among teachers and administrators who are trying to cover both

conflicting views.

There are four hidden dangers involved in using two different definitions in reading and testing according to Valencia and Pearson (1987). They are a false sense of security, a failure to note changes in test scores when strategic views of reading are taught, little change in curriculums that teach to the test, and the hesitancy of teachers to trust their judgment because standardized test scores seem objective.

Researchers agreed that central to any development of reading assessment is the acknowledgement that comprehension is influenced by the type of material being read, the purposes or goals for reading, and the characteristics of the readers including their attitudes, knowledge, understanding, and ability to use reading strategies appropriately. Reading is a multifaceted process involving interactions between readers, the text, and the purposes for reading.

This study investigated whether currently used methods of assessment used for grouping and placement are valid predictors of future academic performance. This is a vital step before adopting more fashionable theoretical claims and testing tools is considered.

Completely individualized instruction could be achieved only with one teacher working individually with one student. Little is lost with a ratio of six students to one teacher and some persons would argue that the gains due to mutual reinforcement exceed the losses to individualization. However, when schooling is supported by public funds the ratio of students per teacher tends to increase to 10:1, then 20, then 30 and sometimes more than 30 students to one teacher. In order to still provide some individualization, assessment techniques have been in use for the last seventy years in American schools. This study investigated two such instruments, the Pupil Evaluation Program Test and the Otis Lennon School Ability Index.

#### Purpose

The purpose of this study was to determine whether the third grade Pupil Evaluation Program Test or the Otis Lennon School Ability Index or both together are valid predictors of future academic performance as demonstrated by high scores on the eleventh grade English Regents exam, SAT or ACT reading tests or a high unweighted or weighted grade point average.

More needs to be known about the strength of the prediction. If the relationship is strong then teachers need to pay close attention to the scores and to providing remediation services to any student not already receiving them. If an incredibly weak correlation is noted then the ecological validity of either test would be in question.

### Variables

The predictor variables are the Pupil Evaluation Program Test and the Otis Lennon School Ability Index. The criterion variables are the eleventh grade English Regents exam grade, the unweighted grade point average, the weighted grade point average, the SAT reading comprehension score and the ACT reading score.

### Definition of Terms

In the course of surveying literature for this thesis, it was evident that certain terms needed to be defined.

Standardized tests commercially reproduced and relying on multiple choice and single correct answer formats. Containing concise and fact-filled stories followed by a series of questions asking the students to make appropriate inferences.

Pupil Evaluation Program (PEP) test using the Degrees of Reading Power (DRP) methodology involving the untimed criterion referenced modified cloze test. It consists of eight nonfiction connected prose passages which are arranged in order of increasing difficulty; each passage contains about 300 words. Test items are made by the deletion of seven words in each passage and each deleted word is shown by an underlined blank space. The state reference point is 28 at which point students are to be provided with remedial instruction.

Otis Lennon School Ability Index an index of the pupil's ability, in comparison with pupils of similar chronological age regardless of grade placement, to deal with relationships expressed in verbal, numerical, figural, or symbolic form. The test has a mean of 100 and a standard deviation of 16 points.

Unweighted high school grade point average determined by dividing the number of courses into the sum of the grade points obtained (A=4.00).

Weighted high school grade point average determined by dividing the number of courses into the sum of the grade points obtained and also involving three levels of weighting including Honors/Advanced Placement (A=4.33),

Regents (A=4.0) and General (A=3.66).

Scholastic Aptitude Test (SAT)-Reading Comprehension

multiple choice test designed to measure reading comprehension skills that are related to academic performance in college such as analogies, sentence completions and critical reading questions. The mean scores are 430 for males and 420 for females.

American College Testing (ACT)-Reading a test that measures the student's level of reading comprehension as a product of skill in referring and reasoning. The tests items ask students to glean meaning from several selections by referring to what is directly stated and by reasoning to determine implied meanings and to make conclusions, comparisons, and generalizations. ACT test scores are reported on a scale that ranges from 1-36, the mean score is 22 for both men and women.

## **Chapter II - Review of the Literature**

### What is Standardized Testing?

Standardized tests are the most common form of assessment in the United States according to Pearson and Stallman (1993). Historically, after the first standardized test was used in 1914 in which average levels of performance were established, they became immensely popular. This happened according to Pearson and Stallman (1993) because of World War I (what recruits could cope), compulsory education (who should be advanced), the civil service system (it needed to be objective) and the respect for scientific objectivity (free of bias). In the 1960's, the Title I, now Chapter 1, programs asked states to be accountable for the federal funds they received; and the mastery learning philosophy was applied to the subskills of reading. In the 1980's, the outcomes-based education concept asked schools to be responsible for a certain level of performance as regarding particular objectives.

Testing directly changes instruction when the teacher takes time out from regular lesson plans to prepare for a standardized test. According to Heithaus (1989) the higher scores achieved might allow the school to gain more credibility, more financial support, and more college opportunities for the students.

Indirectly, the testing affects the instruction in that a certain basal reading program might be chosen based on its promise to increase scores and the curriculum might be altered to coordinate it with a particular test such as Heithaus found (1989). This is because school boards, administrators, and teachers have the attitude that test makers have a special understanding of what important things need to be taught and that their tests test for them.

There may be positive or negative effects from the standardized assessment depending on the intended purpose of the test and the testing situation. Pearson and Stallman (1993) reported that advocates of standardized tests see them as relatively inexpensive, efficient, taking little time from instruction, fair because they are objective, free from biases of teachers, and culturally neutral.

Levande (1993) and Sternberg (1991) pointed out the negative aspects of standardized tests. According to evidence cited the tests are culturally biased, ignore the influence of social context on learning, give minimal usable feedback concerning instructional needs, take the joy and dynamic quality away from reading, and are not helpful with classroom diagnosis. In addition, they acknowledged that the tests do not reflect new understandings of cognitive development, demand the teachers limit their curriculum by teaching to the tests' concepts, distort reality, are overemphasized and not a valid measure of real learning. Webb (1983) concluded that test results are meaningless unless used purposefully for advisement or for developing more effective teaching strategies. Heithaus (1989) believed that there is a "negative correlation between the economic feasibility of a test and the amount of rich data it provides" (p.63).

#### How are Test Data Used?

Ruddell (1985) reported that when asked to discuss how the educators used the test data and what difference the test scores had on instructional decisions, 47% of the educators said that the impact was minimal. To identify reading problems teachers made the greatest use of individual diagnostic tests and informal observation.

Some 67% of the teachers and 50% of the principals used tests to determine the progress of the school reading programs.

After extensive evaluation of three widely used standardized reading tests, Webb (1983) cautioned the educator not to use test results as infallible indicators of ability but rather he suggested using the scores along with "knowledge of the student's background, current situation and future aspirations" (p. 429). Bieger (1989) concurred as she pointed out after testing that the Degrees of Reading Power was not a valid reading assessment and showed very little relationship between the scores on it and the Metropolitan Achievement Test. However, Weller, Schnittjer and Tutem (1992) found that the Metropolitan Readiness Test given to first graders and achievement test scores in mathematics and reading for grades three, six, nine, and ten did help in making readiness decisions in the primary grades and had positive correlations with subsequent academic performance through grade ten.

Muehl and DiNello (1976) reported on a seven year follow-up study of students to determine contributions of first grade skills to their subsequent

reading test scores. They acknowledged that it is not clear on how the various skills of a young student exactly relate to the reading process. Their research indicated that the "letter-naming skill, in reflecting both sensory and language skills, is a kind of microcosm of the whole reading process, early and late" (Muehl & DiNello, 1976, p. 78). As such, it is an indicator of intellectual ability allowing for the predicting of reading performance after the first-grade. Their research showed that the low letter-name group continued in being comparatively worse off over the years. In grades four through seven the low group dropped below average.

The purposes of testing are many. They include student performance assessment, placement of students, the determination of instructional decisions in classrooms, and the distribution of finances. However, if testing data are so utilized, those using the tests need to be knowledgeable about the significance of the scores and aware of different types of tests and measurement concepts. If test scores are playing a significant role in decision making to the extent the test scores accurately portray performance, then the greater emphasis on accountability as evaluated through test scores is welcome. But to the extent that the

tests give faulty or incomplete information the increased test score usage is a cause for concern.

#### What Attitude is Held Toward Test Usage?

Some educators argue that teaching to tests forces educators and the educational system to be accountable. Since they have political value and the general approval of parents, they are indispensable tools for evaluating program effectiveness. Teachers indicated in Ruddell's (1985) study that they had acquired the notion that data from standardized or basal tests were more trustworthy than the data that they collected every day.

Others argue that changes need to be made if the current system is to remain credible and that improvements in test design should always be sought. "Testing must cover what is essential for students to know and be able to do" (Pearson & Stallman, 1993, p. 8).

By asking questions regarding testing philosophy and by asking for the ranking of various tests in terms of their usefulness for decision making, Ruddell (1985) reported that 83% of the teachers did want tests that could diagnose the progress made by a student; 50% of the principals were concerned about the correct matching

of test to curriculum; and 50% of the district personnel were concentrating their focus on the test trends and the state mandating of tests. Generally, teachers felt that tests should not be the only basis for placement decision. Overall, the educators felt standardized tests were somewhat important with all district staff and principals thinking they could be extremely useful in certain situations concerning placement.

Sternberg's (1991) thesis was that we need to be cautious in using test scores and that tests should be created that are more "realistic simulations of people's behavior in the kinds of situations in which they use the aptitudes and achievements that we measure" (p. 540). He described the differences between reading as it is found on tests with reading as it is done in normal living: test passages are short, emotionally neutral, unmotivating; the reasoning too good; there are few distractions; there is a single purpose; they don't control for the students' desire to read; and recall is immediate and intentional. While it is true that those who do well on reading tests also do well on other tests it is because the same skills are tapped.

Over the past 30 years testing in schools has increased. Ruddell (1985) thought this increase was

because there is currently more public interest in basic skills achievement and accountability. Educators as well as legislators want to determine instructional gains more clearly and to improve the local school district's decision making as well as gain more prestige for it.

Pearson and Stallman (1993) showed that these economic and political aspects of testing are real. They cited the fact that realtors quote test scores. As a result property values rise and fall according to the perceived quality of the school as shown by test scores. Also, they noted that state legislators decide which school districts get special funds depending on test scores; (but, how else might this be done?) For these reasons some educators make important instructional decisions for the only purpose of improving test scores.

#### What Affects the Test Scores?

Variables have an impact on students' standardized test scores. Educators concerned with fairness, validity, and reliability must acknowledge that there are factors that influence test results. Following are several examples gleaned from researchers.

Data indicate that ability-composition effects are of educational significance whether for high or low ability children. As Share, Jorm, Maclean and Matthews (1984) noted, reading acquisition is influenced by the child's own phonemic awareness skills as well as by the phonemic awareness skills of the child's peers. Therefore, ability grouping has long ranging effects on test scores.

Carver (1992) showed that standardized tests are really measuring general reading ability or, theoretically, efficiency which is a combination of accuracy and rate. Therefore, educators must realize that scores on standardized tests are highly impacted by rate, and that knowledge of word meanings accounted for the additional variance of scores on the items. The Hispanic children studied by Garcia (1991) revealed that unknown vocabulary in the question and answer choices was the major hindrance to the students.

In a similar study Hodges (1992) found that a strong correlation existed between teachers' assessments of their students and the students' performance on the standardized test. When the results differed the teachers usually ranked students in the next lower category. Teachers in seven kindergarten classrooms

evaluated their students using alternative evaluation techniques. Standardized tests were also given to the students. Follow-up interviews revealed that teachers and parents felt that teachers' evaluations provided more helpful information than the standardized tests did.

Hodges (1992) also pointed out how early childhood experts feel that young children are not good test takers because the strange format leads to anxiety and the results are influenced by the child's ability to sit still.

Educators should also be aware that test anxiety and self-concept are major variables in tests performance. After examining scores of high school students taking the ACT assessment, a test anxiety inventory, and a self-concept measure, Williams (1992) found that the results indicated that both test anxiety and self-concept contributed to academic performance.

Ima and Labovitz (1991) did a study with students in the San Diego, California School District where immigrants and newcomers make up about a third of the total student population. They reported that performance is strongly correlated with ethnicity, with

reading being the most sensitive to acculturation.

Ideally, a combination of a normed approach and a diagnostic-prescriptive approach is the best way to assess student progress in reading ability according to Birger (1989). Her study compared the performance on a silent informal reading inventory and achievement on a standardized reading test and found the results did not correlate highly. Direct observation combined with test results best enables the teacher to plan for corrective instruction and suitable reading materials.

Caution is in order since variables have a vital impact on the students' standardized test scores. Educators need to know what they are measuring, and what they should be measuring. They can then better evaluate alternative instructional methods.

## **Chapter III - Design of the Study**

### Purpose

The purpose of this study was to determine whether the third grade Pupil Evaluation Program Test or the Otis Lennon School Ability Index or both together are accurate predictors of future academic performance. If they are, then teachers need to pay close attention to the scores and to providing remediation services to any student not already receiving them.

If no high correlation is noted, then the ecological validity of either test would be in question. Educators would then need to be open to new formats that would more accurately serve as predictors of future reading performance.

### Experimental Hypothesis

Will the predictive relationship between the predictor variables and the criterion variables be sufficiently strong enough ( $r^2 = .30$ ) to warrant their use for grouping and placement?

## Methodology

### Subjects

Ninety-one 1993 graduates who had taken the third grade reading Pupil Evaluation Program Test and Otis Lennon School Ability Index were randomly chosen from a large suburban school district. All subjects have remained at the designated school continuously from kindergarten through grade twelve.

### Instruments

The third grade Pupil Evaluation Program reading raw score, tenth grade Otis Lennon School Ability Index, eleventh grade English Regents exam grade, unweighted high school grade point average, weighted high school grade point average, SAT reading and ACT reading scores were recorded.

### Variables

The predictor variables are the Pupil Evaluation Program Test and the Otis Lennon School Ability Index. The criterion variables are the eleventh grade English Regents exam grade, the unweighted grade point average, the weighted grade point average, the SAT reading comprehension score and the ACT reading score.

## Procedures

The x variable, or the predictor, is the score received on the third grade Pupil Evaluation Program reading test. The range of scaled scores that could be received is 0 to 56. A separate predictor is the score received on the Otis Lennon School Ability Index. The range of scores is approximately 68 to 132.

The y variables, or the criterion, are the students' eleventh grade English Regents exam grade (0-100), high school unweighted grade point average (0-4.00), weighted grade point average (0-4.66), SAT reading comprehension score (200-800), and ACT reading score (1-36).

## Analysis of Data

Scores were compared statistically to see where the strongest relationships lie. A current regression summary table was examined for each independent variable and the variables combined. Is the third grade Pupil Evaluation Program reading test score the best predictor of success as shown by the highest grade point averages and/or SAT and ACT reading scores? Perhaps the Otis Lennon School Ability Index better indicates who will be deemed a success in high school. Or finally, perhaps the PEP score combined with the Otis Lennon Index best predicts academic achievement.

The minimum coefficient of determination value that would impress this researcher in this study would be a .30 at the 95% confidence level. This unusually strong, positive relationship was desired because the necessary remediation placement could be then given to those struggling before their self-esteem and motivation suffer and advanced academic placement could be given to those needing to be challenged.

#### Summary

This study examined whether there was a predictive relationship between the third grade PEP score, tenth grade Otis Lennon score, or both together and the eleventh grade Regents English exam, the unweighted and weighted grade point averages, the SAT-Reading and the ACT-Reading scores. A current regression summary table was examined for each independent variable and the variables combined.

## Chapter IV - Analysis of Data

### Purpose

The purpose of this study was to determine whether the third grade Pupil Evaluation Program Test or the Otis Lennon School Ability Index or both together are valid predictors of future academic performance as demonstrated by high scores on the eleventh grade English Regents exam, SAT or ACT reading tests or a high unweighted or weighted grade point average.

### Findings and Interpretations

1. The relationship between the scores on the PEP test and the Regents exam was positive. The predictor variable, the PEP test, explained 39% of the variation in the criterion variable, Regents English exam, ( $r^2 = .39$ ). The correlation and regression phase of this study indicated an unusually strong relationship.
2. The relationship between the scores on the Otis Lennon and the Regents exam was positive.

The predictor variable, the Otis Lennon test, explained 14% of the variation in the criterion variable, the Regents English exam, ( $r^2 = .14$ ). The correlation and regression phase of this study indicated a minimal relationship.

3. The relationship between the combined scores of the PEP test and the Otis Lennon test and the Regents English exam was positive. Together the predictor variables explained 41% of the variation in the criterion variable, the Regents English exam, ( $r^2 = .41$ ). The correlation and regression phase of this study indicated an unusually strong relationship.
4. The relationship between the scores on the PEP test and the unweighted grade point average was positive. The predictor variable, the PEP test, explained 27% of the variation in the criterion variable, unweighted grade point average, ( $r^2 = .27$ ). The correlation and regression phase of this study indicated a very strong relationship.
5. The relationship between the scores on the Otis Lennon test and the unweighted grade point average was positive. The predictor variable, the Otis Lennon test, explained 18% of the variation in the

criterion variable, the unweighted grade point average, ( $r^2 = .18$ ) The correlation and regression phase of this study indicated a moderate relationship.

6. The relationship between the combined scores on the Pep test and Otis Lennon test and the unweighted grade point average was positive. The predictor variables together explained 31% of the variation in the criterion variable, the unweighted grade point average, ( $r^2 = .31$ ) The correlation and regression phase of this study indicated an unusually strong relationship.
7. The relationship between the scores on the PEP test and the weighted grade point average was positive. The predictor variable, the PEP test, explained 29% of the variation in the criterion variable, weighted grade point average, ( $r^2 = .29$ ). The correlation and regression phase of this study indicated a very strong relationship.
8. The relationship between the scores on the Otis Lennon and the weighted grade point average was positive. The predictor variable, the Otis Lennon, explained 21% of the variation in the criterion

variable, the weighted grade point average, ( $r^2 = .21$ ). The correlation and regression phase of this study indicated a strong relationship.

9. The relationship between the combined scores on the PEP test and the Otis Lennon and the weighted grade point average was positive. The predictor variables together explained 36% of the variation in the criterion variable, the weighted grade point average, ( $r^2 = .36$ ). The correlation and regression phase of this study indicated an unusually strong relationship.
10. The relationship between the scores on the PEP test and the SAT reading score was positive. The predictor variable, the PEP test, explained 30% of the variation in the criterion variable, SAT reading score, ( $r^2 = .30$ ). The correlation and regression phase of this study indicated a very strong relationship.
11. The relationship between the scores on the Otis Lennon and the SAT reading score was positive. The predictor variable, the Otis Lennon explained 18% of the variation in the criterion variable, the SAT reading score, ( $r^2 = .18$ ). The correlation and

regression phase of this study indicated a moderate relationship.

12. The relationship between the combined scores on the PEP test and the Otis Lennon and the SAT reading score was positive. The predictor variables together explained 35% of the variation in the criterion variable, the SAT reading score, ( $r^2 = .35$ ). The correlation and regression phase of this study indicated an unusually strong relationship.
13. The relationship between the scores on the PEP test and the ACT reading score was positive. The predictor variable, the PEP test, explained 33% of the variation in the criterion variable, ACT reading score, ( $r^2 = .33$ ). The correlation and regression phase of this study indicated an unusually strong relationship.
14. The relationship between the scores on the Otis Lennon and the ACT reading score was positive. The predictor variable, the Otis Lennon, explained 30% of the variation in the criterion variable, the ACT reading score, ( $r^2 = .30$ ). The correlation and regression phase of this study indicated a very strong relationship.

15. The relationship between the combined scores on the PEP test and the Otis Lennon and the ACT reading score was positive. The predictor variables together explained 44% of the variation in the criterion variable, the ACT reading score, ( $r^2 = .44$ ). The correlation and regression phase of this study indicated an unusually strong relationship.

## **Chapter V - Conclusions and Implications**

### Purpose

The purpose of this study was to determine whether the third grade Pupil Evaluation Program Test or the Otis Lennon School Ability Index or both together are valid predictors of future academic performance as demonstrated by high scores on the eleventh grade English Regents exam, SAT or ACT reading tests or a high unweighted or weighted grade point average.

### Conclusions

Based on these results this researcher concludes that the PEP test score be examined carefully as a predictor for performance by teachers and counselors and used for counseling students on whether to register for the eleventh grade English Regents course. Chances are that those who score well on this standardized reading test will do well on the English Regents exam.

It is recommended that the Otis Lennon score not be used as a predictor for performance on the English

Regents exam or the unweighted grade point average or the SAT reading test, however, it could be used to predict success concerning the weighted grade point average and the ACT reading test.

It is highly recommended that the PEP score together with the Otis Lennon score be examined carefully as a predictor for performance by teacher and counselors and used for counseling students on whether to take the English Regents exam, the SAT reading or ACT reading tests.

#### Implications of Education

Why do the schools do all the assessing that they do? Is it done for the school board or the classroom teacher? Is it to determine the effectiveness of the curriculum or to gauge student progress? Robinson (1990) states such basic questions must be dealt with so we can choose the appropriate assessment tools to meet our objectives. He thinks that reading assessment is very ingrained in the United States and as such he desires a strong effort be made to educate the test user, to tie the test more closely to instruction, to center on comprehension and to measure reading attitude as well. For him the main reason for assessment is to note the student's ability to understand the text and

the openness to see reading as integral to life.

However, reading researchers must support new theoretical claims with solid research before new ideas are adopted. This researcher would agree with Carver as he is quoted by Robinson (1992) that the traditional standardized tests continue to be "generally valid for measuring individual differences in efficiency, which means that they have an accuracy level component and a rate level component" (p.339). It alarms him that states such as Illinois and Michigan are developing new reading tests founded on fad and linked to the schema theory. He claims there is evidence that the prediction activities in the tests do not improve the accuracy of comprehension, that the personal differences in prior knowledge regarding a particular long passage has a small impact upon the amount of passage comprehension, the expository type of test also has minimal effect upon amount of material comprehended when the material is controlled for readability. He believes that longer passages show more variability in motivation and thus are less reliability. His own research also showed that he got less reliable and less valid scores using test questions with more than one correct answer. It is possible to him that the new tests could produce poorer

readers if all the focus continues to be on accuracy of comprehension without notice of the rate. The reading of light fiction and the making of predictions involves little gain in general knowledge thus lowering the gains made in the efficiency level.

This present research supports Carver's claim. It appears that the traditional standardized PEP reading test is a valid measure of general reading ability allowing for academic success based on relevant empirical research. Focusing only on the standardized test score is not advocated, but neither should it be ignored. If the test results would be used for evaluating teachers and for retaining students then the effect of the testing is seen to be negative. Attention always also needs to be given to direct teacher observation of student performance and understanding of the process of reading.

#### Implications for Research

Following are several examples of topics needing further exploration in this pursuit.

As it appears that there is a lack of understanding among field educators regarding test score interpretation, an inservice program should be developed

and researched according to Ruddell (1985). Such information should include information on measurement concepts, interpreting test data, and the value and limitation of such data for decision making regarding instruction, curriculum and resources. This would enable educators to better communicate with parents about the progress of their students and the schools.

Researchers agree that work should be done on the redundancy found in various tests so that testing time could be reduced and so that emphasis could be placed on that testing that is found to be useful for decision making.

Research examining whether or not state departments are biased in their interpreting of test scores would be helpful according to Henk and Rickelman (1992). A comparison of score results and interpretations determined by an independent agency to the state's findings would reveal whether there is a conflict of interest.

Assessment strategies that are consistent with the known components of strategic reading need to be researched. With it a new framework for testing needs to be developed that considers the types of decisions

that need to be made and the impact those decisions make. Acknowledging that while some measures must be administered to large groups it must be noted that those tests traditionally have provided educators with a minimum assessment of an individual's achievements, therefore, decisions for individuals might need the interview technique. An example of such a technique as recommended by Heithaus (1989) is the written retelling method involving the student's reading of a passage and then writing a retelling, thus indicating how the student's thinking is done.

Research is needed on the possibilities of making machine scorable formats available for both large scale testing and constructed response formats. Such research is beginning. Valencia and Pearson (1987) reported on research done in Illinois that is testing 15,000 students in grades 3,6,8, and 10 using new formats such as summary writing, metacognitive judgments, question selection, multiple acceptable responses and prior knowledge. In Michigan, Wixson, Peters, Weber and Roeber (1987) reported on the position paper put out by the Michigan Reading Association (MRA) in which they point out that the theory of reading has changed over the last 15 years from being regarded as a series of sequential skills to a "process of constructing meaning

through the dynamic interaction among the reader, the text, and the context of the reading situation" (p.750). The new premise required new objectives, such as constructing meaning, knowledge about reading, attitudes and self perceptions. A blueprint for a test was developed in which reading was evaluated in a more holistic manner. Also, topic familiarity was determined and the reading selections were representative of the materials students used regularly in the classroom.

What are the results of the new reading test of the 1992 National Assessment of Educational Progress? According to Robinson (1990) this test will increase the number and importance of open-ended questions while decreasing the number of multiple-choice questions. In evaluating high-level thinking skills the students will be asked to think about what they've read and then react to it in 40% of the questions. Robinson asks if we can then better assess how students interpret what they read, not just how they decode which is only one aspect of reading that happens to be easily tested.

Pearson and Stallman (1993) ask how we can make new types of assessment credible in the eyes of the public. How can we show that alternative assessment can be fair to all? They also ask that while alternative

assessments can be useful at the classroom level, can they really be useful on a larger district or national level. Finally, they ask what standards will be used to evaluate alternative assessments. "It is not clear, for example, that the conventional standards of reliability, validity, utility, efficiency, and objectivity are applicable to alternative measures" (p. 9).

#### Summary

This chapter has considered ideas for education and for further research. There was acknowledgment made for the interpretation of the data cited in Chapter IV. There was a strong relationship between the third grade PEP reading score and high school performance.

## **References**

## References

- Bieger, E. (1989). The Degrees of Reading Power--a testing and teaching tool? Reading Improvement, 26(3), 203-207. (ERIC Document Reproduction Service No. EJ 403 654)
- Birger, B. (1989). A comparison of performance on a silent informal reading inventory and achievement on a standardized reading test. Unpublished master's thesis, Kean College, CA. (ERIC Document Reproduction Service No. ED 312 628)
- Carver, R. (1992). What do standardized tests of reading comprehension measure in terms of efficiency, accuracy, and rate? Reading Research Quarterly, 27(4), 347-359.
- Davis, S. (1989). Using multiple criteria to place middle school students. Journal of Reading, 32(8), 720-728.
- Fein, S. & Solomon, A. (1990, April). The relationship between student reading achievement as measured by standardized test scores and reading attitudes as measured by students' and teachers' responses to selected national assessment of educational progress attitudinal items. Paper presented at the meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 320 958)
- Ferguson, R. (1990). Racial patterns in how school and teacher quality affect achievement and earnings. Cambridge, MA: Harvard University, John F. Kennedy School of Government. (ERIC Document Reproduction Service No. ED 347 217)

- Gaines, M. & Davis, M. (1990, April). Accuracy of Teacher Prediction of Elementary Student Achievement. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 320 942)
- Garcia, G. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. Reading Research Quarterly, 26(4), 371-391.
- Heithaus, J. (1989). College reading assessment: Dinosaurs and new theory. Reading Research and Instruction, 28(4), 60-63. (ERIC Reproduction Service No. 395 000)
- Henk, W. & Rickelman, R. (1992). Trends in statewide reading assessment: A closer look. Reading Horizons, 33(1), 67-82. (ERIC Reproduction Service No. EJ 452 780)
- Hodges, C. (1992, December). Literacy assessment in kindergarten: A longitudinal study of teachers' use of alternative forms of assessment. Paper presented at the meeting of the National Reading Conference, San Antonio, TX. (ERIC Reproduction Service No. 353 560)
- Ima, K. & Labovitz, E. (1991). Language proficiency, ethnicity and standardized test performance of elementary school students. San Diego, CA: San Diego Unified School District. (ERIC Reproduction Service No. 341 700)
- Levande, D. (1993). Standardized reading tests: Concerns, limitations, and alternatives. New Haven: Southern Connecticut State University, Department of Education. (ERIC Reproduction Service No. EJ 463 785)

- Muehl, S. & DiNello, M. (1976). Early first-grade skills related to subsequent reading performance: A seven year followup study. Journal of Reading Behavior, 8(1), 67-81.
- Pearson, P. D. & Stallman, A. (1993). Approaches to the future of reading assessment: Resistance, complacency, reform (Tech. Rep. No. 575). Champaign: University of Illinois at Urbana. College of Education. (ERIC Reproduction Service No. ED 356 457)
- Robinson, M. (1990). Reading, K-12; language arts K-5. program evaluation, 1989-90 (Rep. No. 90-115). Des Moines, IA: Des Moines Public Schools. (ERIC Reproduction Service No. EJ 451 345)
- Robinson, R. (1990). Leaders in Reading Research and Instruction. Reading Psychology: An International Quarterly, 11(4), 335-346. (ERIC Reproduction Service No. EJ 418 077)
- Robinson, R. (1992). Leaders in Reading Research and Instruction. Reading Psychology: An International Quarterly, 13 (Oct/Dec), 337-43. (ERIC Reproduction Service No. EJ 451 345)
- Ruddell, R. B. (1985). Knowledge and attitudes toward testing: Field educators and legislators. The Reading Teacher, 38(6), 538-543.
- Share, D., Jorm, A., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. Journal of Educational Psychology, 76(6), 1309-1324.
- Sternberg, R. (1991). Are we reading too much into reading comprehension tests? Journal of Reading, 34(7), 540-545.

- Valencia, S. & Pearson, P. (1987). Reading assessment: Time for a change. The Reading Teacher, 40(8), 726-732.
- Walberg, H. & Tsai, S. (1985) Correlates of reading achievement and attitude: A national assessment study. Journal of Education Research, 78(3), 159-167.
- Webb, M. (1983). A scale for evaluating standardized reading tests, with results for Nelson-Denny, Iowa, and Stanford. Journal of Reading, 26(5), 424-429.
- Weller, L., Schnittjer, C. & Tutem, B. (1992) Predicting achievement in grades three through ten using the Metropolitan Readiness Test. Journal of Research in Childhood Education, 6(2), 121-130.  
(ERIC Document Reproduction Service No. EJ 460 132)
- Williams, J. E. (1992). Effects of test anxiety and self-concept on performance across curricular areas. Oklahoma State University. (ERIC Document Reproduction Service No. ED 344 903)
- Wixson, K., Peters, C., Weber, E., Roeber, E. (1987). New directions in statewide reading assessment. The Reading Teacher, 40(8), 749-754.

## **Appendix**

Scores from 1993 graduates on PEP, Otis Lennon, Regents  
English, Unweighted and Weighted Grade Point  
Averages, SAT-Reading and ACT Reading

Appendix

Scores from 1993 graduates on PEP, Otis Lennon, Regents English. Unweighted and Weighted Grade Point Averages, SAT-Reading and ACT Reading

#	M/F	PEP	O.L.	Regents	U.G.P.A.	W.G.P.A.	SAT-RACT-R
1	M	55	125	87	2.78	2.80	590 32
2	F	40	109	72	2.15	2.15	370
3	M	61	132	91	3.54	3.77	650
4	M	50	112	90	3.39	3.46	600 24
5	F	42	96	80	2.90	2.92	470 17
6	M	55	110	91	3.92	4.13	500 30
7	F	52	132	92	3.87	4.10	540 31
8	F	44	114	72	2.08	2.05	
9	F	63	123	87	3.43	3.52	580 29
10	F	51	97	80	3.18	3.20	540 25
11	F	46	120	86	2.78	2.79	540 26
12	F	42	98	70	2.74	2.73	420
13	F	56	103	88	3.59	3.70	600 29
14	F	48	115	84	3.38	3.38	470
15	M	53	111	88	2.79	2.79	510 23
16	M	46	110	88	3.88	4.08	570
17	F	50	100	75	2.58	2.54	330
18	F	53	132	92	3.25	3.38	590 31
19	M	42	118	87	2.26	2.69	530
20	M	38	108	71	2.52	2.38	390 25
21	F	61	136	88	3.44	3.66	460 30
22	M	50	110	88	3.04	3.04	520 24
23	M	39	95	71	2.65	2.62	19
24	M	44	125	70	2.61	2.61	370 21
25	M	39	118	73	3.23	3.31	410 17
26	M	52	136	86	3.43	3.59	
27	M	50	98	92	3.43	3.44	460 22
28	F	54	114	88	3.02	3.11	630 31
29	M	42	126	73	3.31	3.35	430 14
30	F	52	110	90	3.59	3.83	510
31	M	52	101	80	2.88	2.89	470 20
32	F	45	102	83	2.94	3.01	480
33	F	40	104	75	2.62	2.67	380
34	M	54	125	88	3.16	3.19	650 21
35	M	51	132	83	3.43	3.51	430 31
36	F	59	149	74	3.00	3.03	420
37	F	49	103	82	2.68	2.68	410 21
38	F	53	107	87	3.75	3.87	570 27
39	M	36	109	68	2.33	2.27	260
40	M	39	121	86	3.00	3.00	520 26
41	F	52	107	75	2.64	2.64	370 18
42	F	50	109	90	3.06	3.08	570 25
43	M	56	150	91	3.52	3.66	690 36
44	M	41	106	81	3.14	3.14	460 21
45	M	46	123	86	3.29	3.32	410 22

46	M	45	118	75	2.56	2.56	460	19
47	M	45	109	61	2.67	2.66		16
48	M	41	130	90	4.08	4.26	600	27
49	F	52	112	89	3.51	3.58	580	28
50	M	46	102	84	2.88	2.91	680	32
51	M	51	103	78	2.87	2.91	420	19
52	M	61	128	90	3.27	3.42	710	35
53	M	54	112	84	2.99	3.01	490	28
54	M	43	110	85	2.98	2.99	560	26
55	M	44	113	72	3.40	3.41	370	21
56	M	50	110	88	3.25	3.27	530	26
57	F	54	131	89	3.75	3.96	470	31
58	F	46	126	84	3.35	3.50	630	31
59	M	40	121	75	2.81	2.81	560	25
60	M	59	117	77	2.62	2.62	490	21
61	F	46	124	76	3.44	3.50	440	30
62	F	45	113	85	3.14	3.15	470	21
63	M	51	114	75	2.77	2.77	420	24
64	M	42	114	73	2.69	2.69	470	23
65	M	45	89	84	2.95	2.97	550	27
66	F	34	113	79	2.56	2.56	370	20
67	M	55	120	90	3.17	3.23	570	30
68	F	39	104	65	2.14	2.02	390	
69	M	53	118	80	3.17	3.25	580	
70	F	50	119	80	3.22	3.24	560	27
71	F	50	107	77	2.39	2.38	430	18
72	M	54	136	94	2.81	2.90	660	
73	F	37	106	69	2.63	2.63	420	22
74	F	48	109	91	3.34	3.37	520	27
75	F	41	103	76	3.05	3.04	350	19
76	F	43	106	73	2.81	2.81	380	20
77	F	53	97	86	3.70	3.71	360	24
78	M	52	146	94	3.68	3.88	650	35
79	M	53	124	90	3.44	3.63	720	35
80	M	62	145	88	3.74	3.91	570	
81	F	22	97	65	2.27	2.17	200	15
82	M	43	107	75	2.40	2.37	360	23
83	F	51	120	86	2.88	2.88	640	29
84	M	50	117	81	3.02	3.13	420	20
85	M	52	118	94	3.41	3.59	670	35
86	M	34	97	60	2.15	2.03	420	14
87	F	54	127	85	3.46	3.60	520	35
88	F	52	131	86	3.77	3.98	570	27
89	M	51	121	77	2.75	2.78	470	22
90	M	48	110	75	2.64	2.64	400	21
91	M	50	126	77	3.08	3.14	510	27