

Running Head: TEACHER LEADERS AT WORK

Teacher Leaders at Work:  
Analyzing Standardized Achievement Data to Improve Instruction

John E. Henning  
University of Northern Iowa

John E. Henning  
545 Schindler Educational Center  
University of Northern Iowa  
Cedar Falls, IA 50614-0607  
(319) 273-7488  
Fax (319) 273-77732  
John.Henning@uni.edu

1001 Barrington Drive  
Cedar Falls, IA 50613  
(319) 277-7626

Submitted to  
*EDUCATION*  
October 25, 2004

### Abstract

The purpose of this study is to describe how 24 elementary and middle school teacher leaders analyzed standardized achievement test scores, utilizing four different approaches: comparing to the norm, analyzing trends, correlating data, and disaggregating data. The analyses were conducted in order to effect positive change in student learning and as part of graduate coursework in a Teacher Leadership Program at the University of \_\_\_\_\_ . Findings from this study have implications for the storage and organization of data, for methods of teaching teachers to analyze school data, and for the discovery of new approaches to analyzing school data. Descriptions of how educators are using student achievement data to improve instruction could provide insights for teachers, principals, and professors who are interested in practical approaches to analyzing data in school settings.

Teacher Leaders at Work:

The Analysis of Standardized Achievement Data to Improve Instruction

Introduction and Literature Review

A half century ago, standardized achievement test scores were primarily used for (a) informing teachers and parents about students' achievement relative to their peers, (b) helping place students in appropriate programs, and (c) justifying the allocation of supplemental resources. However, advances in the technology of standardized test taking, combined with the popular belief that testing improves student achievement, has led to uses of the standardized test results in ways not originally intended. Concerned educators have warned that some of these uses, e.g. evaluating schools, teachers, and as a requirement for grade promotion, are invalid and can have a negative impact on student learning (Popham, 2001a, 2001b).

Consonant with the increasing public pressure on schools to increase student achievement has been the increasing use of standardized achievement test scores to inform instruction and curriculum. For example, test makers (e.g. Hoover et al., 2003) suggest that by comparing the student, classroom, or building scores with local and national norms, teachers can identify individual or group strengths and weaknesses for the purpose of adjusting the curriculum. The efficacy of this approach has been disputed by Popham (2001a, 2001b) who argues (a) that the descriptions of knowledge and skills on standardized tests are not clear enough to provide a focus for improving instruction and (b) that classroom assessments are the best source of data for informing instruction.

Despite these caveats, there has been a proliferation of articles and texts addressing the analysis of student achievement data for the purpose of improving learning

(e.g., Anderson et al., 2004; Killion, 2002; Johnson, 2002; Streifer, 2002; Bernhardt, 1998). All of these authors offer alternatives to the approaches historically associated with educational research, e.g. random samples and control groups. The first and most common approach is analyzing trends, i.e. determining whether an instructional intervention has positively influenced student achievement over time (Streifer, 2002; Johnson, 2002; Bernhardt, 1998). A second approach is to disaggregate the data, i.e. group achievement scores by the students' ethnicity, gender, SES, or performance and then make comparisons. For example, a quartile analysis can compare subgroups by examining the percentages of students falling into the bottom, the two intermediate, and the top quartiles. This analysis can compare the progress of different student subgroups and could indicate whether instructional practices favor a particular group or groups of students (Johnson, 2002; Streifer, 2002). A third approach is to examine the relationship between student achievement scores and other indicators of student performance, e.g., grades, attendance, or discipline interventions (Johnson, 2002; Streifer, 2002).

These school-based approaches to data analysis have not been widely taught in university-based research courses, in part because faculty have been trained in research methods more commonly employed in a university context. Unfortunately, this can mean the data analysis skills needed to lead a district through the school improvement process may not be included in graduate programs. The emphasis is often on research designs and analytical procedures that are not relevant to the environment of the practitioner:

... In all too many instances, statistics are taught in a theoretically rarefied atmosphere replete with hard-to-understand formulas and too few examples relevant to the daily life of education practitioners (Bracey, 1997, p. 2).

Shifting to a more practice-based instruction has been increasingly recommended by national organizations, advisory groups, task forces, and accreditation agencies (Shakeshaft, 1999; Murphy & Forsyth, 1999).

The purpose of this study is to describe how 24 teacher leaders analyzed standardized achievement test scores from their school building for the purpose of improving instruction and hence student achievement. Descriptions like these may benefit teachers, principals, and professors who are interested in relevant approaches to data analysis in school settings.

## Methods

### Participants

Seventeen elementary teachers and seven middle school teachers participated in the study. All of the participants in the study were enrolled in a graduate program designed for teacher leaders. All were veteran teachers, many serving as leaders on building teams, as grade level leaders, or as participants on district wide initiatives. All were recommended for the teacher leader program by their principal.

### Description of the Teacher Leader Program

The teacher leader program originated at the request of a local urban school district for the purpose of increasing student achievement through school improvement and faculty development. As school districts across the nation face an increasing need for new leadership (Usdan, McCloud & Podmostko, 2000, p.5) and increased public pressure for accountability, cultivating teacher leadership offers an opportunity to take advantage of "...the best and most abundant source of leadership available to schools" (Pellicer and Anderson, 1995, p. 21). The design of the teacher leader program is based

on a job-embedded, inquiry-based philosophy of learning (Kaatzenmeyer and Moller, 2001; Darling-Hammond, 1999). Therefore, coursework is equally split between classroom instruction on campus and school-based projects that engage program participants in building learning communities, leading data assessment teams, and carrying out action research projects in their schools. University faculty visit the schools an average of three times per semester to ensure that connections are made between course work and school needs. For a fuller description of the program, see Author.

### Data Collection

Data for this study came from the program participants' analyses of the Iowa Test of Basic Skill (ITBS) scores for their school building. The teacher leaders formed teams by building to perform the analyses. Each team was required to do three analyses, one that examined building wide scores, one comparing student subgroups in the building and one analyzing a particular subject area. The 24 analyses were completed in the form of a short write up resembling a research report. In each report there were sections titled "Purpose," "Measures and Procedures," "Analysis," and "Future Steps." Below is a description of each section of the report, accompanied by an example. The examples for each section come from the same report; together they constitute a single report.

### Purpose

Each of the program participants selected the purpose of their inquiry based on perceived building needs. In this brief section, the purpose is expressed in the form of a statement or a research question:

The purpose of this analysis to examine the ITBS word analysis skill test of the "Reading Profile Summary." The question to be asked is: Do the low achieving

students share any similar characteristics in terms of their needs or preparation for learning these skills? What specific areas can be targeted to initiate growth?

#### Measures and Procedures:

In this section, the specific source of student achievement data is identified, e.g., the name of the ITBS report, and the specific procedures regarding data manipulations are described:

The "Reading Profile Summary Item Analysis" was examined to determine the specific skill areas to be targeted. The word analysis and reading comprehension scores were disaggregated to identify the success of different groups of students according to the subgroups identified by the ITBS. The actual number of students representing a subgroup was identified. Next, the test performance of specific subgroups were compared with the whole group scores to determine the specific range of need.

#### Analysis

Program participants were free to choose and apply the particular analyses which best served the purpose of their inquiries. They were given instruction on four different approaches to analyzing standardized achievement scores, as described below.

*Comparing to the Norm* - The purpose of comparing scores to the norm is to analyze the strengths and weaknesses of individual students and groups of students, e.g. a classroom, an entire building, or an entire school district. The results of the analysis can be utilized for differentiating instruction, grouping students, or determining gaps in the curriculum by identifying skill deficiencies.

*Analyzing Trends* - The purpose of analyzing trends is to determine whether standardized achievement scores have improved over time. The results of the analysis can be used to determine the effectiveness of program or school wide interventions on composite, subject area, or skill area scores.

*Correlating Data* - The purpose of correlating data is to examine the relationship of standardized achievement scores to other school measures, e.g. grades, attendance, or discipline interventions.

*Disaggregating Data* - The purpose of disaggregating data is to determine the proportion of high and low performing students within either the total population or a particular group of students as indicated by standardized achievement test. The results of the analysis can be used (a) to identify how well both low and high performing students are being served; (b) to compare the relative performance of different subgroups based on gender, ethnicity, or socioeconomic status; or (c) to discover the proportion of high and low performing students related to any skill or concept. The following excerpt is an example of analyzing student achievement by disaggregating data:

Examination of the fall 2003 ITBS "Profile Summary" and "Building Item Analysis" identified scores lower than 10 points below the national average in the skill area of phonological awareness and decoding (word analysis). Analyzing the lowest percentile scores showed that 62.7% of African American second graders scored below the 40<sup>th</sup> percentile. This compares to 44.6% of all ethnic groups. Actual student count shows that there are a total of 15 students in that subgroup out of the 60 total second graders. 85.7% of African American third graders scored below the 40<sup>th</sup> percentile as compared to 56.9% of all ethnic groups. The



African American student count shows this subgroup has 21 members out of a total of 74 third graders. Another significant group is that of students who qualify for free and reduced lunches. 70.8% of the third grade students and 70.6% of second graders scored in the low percentiles. Over 53% of our second and third graders make up the last subgroup. The specific needs of this group, as well as the African Americans will be examined.

Future Steps:

In the final section, which resembles the “Discussion” section of a research report, the participants describe future actions or future questions for study, as illustrated in the example below.

Although each student arrives at school with his or her own set of experiences, studies have identified certain common characteristics of members of a subgroup which have the potential to inhibit a person’s rate of learning. These common concerns must be researched and analyzed to identify patterns of learning. First grade teachers will benefit from specific information and strategies to implement in the area of word analysis to enhance and clarify the basic reading skills for students who are now low achievers in this area. Research methods of direct instruction and strategies must be implemented. Questions must continue to be asked. Who will implement these strategies? How will their effectiveness be assessed? How will students continue to be identified so that interventions can be implemented?

## Visual Aids

Teacher participants were required to illustrate their data through the use of tables, graphs, and charts for the dual purpose of (a) communicating their findings and (b) to provide further insight while conducting the analysis.

## Findings

### Data Analysis

Each of the 24 reports was identified as belonging to one of five categories: 1) correlating data, 2) disaggregating data, 3) analyzing trends, 4) analyzing the trends of data compared against the norm, and 5) analyzing trends of disaggregated data. The reports were separated into categories based on the research question, whether the data encompassed one or more years, whether or not the participants were divided into groups based on their scores, and whether student scores were compared against the test norms.

### Correlating Data

In three research reports, standardized test scores were correlated with each other or other measures of student performance. The first analysis utilized math and reading composite scores to determine the percentage of the students who fell below 40<sup>th</sup> percentile in both subject areas. The second analysis determined the percentage of the students who fell below 40<sup>th</sup> percentile in both problem solving and reading comprehension. In the third analysis, the ITBS national percentile rank (NPR) composite score and GPA of every fifth student was correlated to compare the relationship between the ITBS scores and grades of white and African American students. Following are the three statements of purpose associated with these analyses.

1. The purpose of the large group analysis is to examine second grade reading and math scores...to determine if the students who scored below the 40<sup>th</sup> percentile in reading also scored the same in math.
2. The purpose of the subject analysis is to determine if problem solving skills affect reading comprehension based on the ITBS scores.
3. The purpose of the analysis is to determine the correlation between ITBS national percentile rank composite scores and a student's GPA and to see if the correlation is the same between different groups.

### Disaggregating Data

In three research reports, standardized test scores representing a single year were disaggregated for the purpose of comparing the performance of different student groups. In the first analysis, the math computation and reading comprehension scores of African American, Hispanic, and Caucasian students were compared to determine the percentages of each group below the 40<sup>th</sup> percentile (defined as proficient to meet the requirements of No Child left Behind in the state of Iowa), between the 40<sup>th</sup> to 89<sup>th</sup> percentile, and above the 90<sup>th</sup> percentile. In the second, the "Reading Profile Summary Analysis" and the "Building Item Analysis" reports were utilized to compare the percentage of students above and below the 40<sup>th</sup> percentile according to the following subgroups: Caucasian students, African American students, and students receiving free and reduced lunch. The third utilized the grade equivalent scores in reading from the "Building Test Ranked by Test" report to compare achievement differences between males and females from a cohort of students who had attended the school for 3 consecutive years with students who

had moved into the district. Following are the research question and two statements of purpose associated with these analyses.

1. Do the low achieving students share any similar characteristics in terms of their needs or preparations for learning these skills? What specific areas can be targeted to initiate growth?
2. The purpose of the subgroup analysis is to compare the ethnic groups (African American, Hispanic, and Caucasian) in math computation and reading.
3. The purpose of the analysis is to identify any difference on the ITBS reading scores between males and females in the cohort group and another group of fourth grade students.

#### Analyzing Trends

In five research reports, standardized test scores were utilized to analyze trends. In four of the reports, data was examined over a 3 year period; in one report, data was examined over a 2 year period. The purpose of all five reports was to determine if there had been growth in either reading comprehension, math composite scores, vocabulary, problem solving, or math concepts and estimation. The following are a sample research question and a sample statement of purpose from the reports categorized as analyzing trends.

1. Have low SES seventh graders shown growth on the ITBS from the fall of 2002 to the fall of 2003 in the area of math concepts and estimation?
2. The purpose of this data analysis by subject area is to identify an increase in the area of comprehension in the Iowa Test of Basic Skills.

### Analyzing Trends of Data Compared against the Norm

In three research reports, trends were analyzed for data compared against the norm. All three reports tried to identify areas for improvement in student scores as compared against national norms over a 2 year period. Skill areas that could be improved included word analysis, problem solving, data and computation, and concepts/estimation. The following are two sample research questions from the reports categorized as analyzing trends of data compared against the norm.

1. What area of reading instruction could be targeted to impact the most significant growth in the ITBS reading scores?
2. When we examine the item analysis data for math on the ITBS, are the weaknesses discovered in problem solving consistent across all grades?

### Analyzing Trends in Disaggregated Data

In ten research reports, standardized achievement test scores were used to analyze trends in disaggregated data. Of these ten reports, three reports used 2 years of data, two reports used 3 years of data, and five reports over 4 years of data. In nine of these reports, the students were divided into three levels (e.g. from 0 – 39<sup>th</sup> percentile, the 40<sup>th</sup>–79<sup>th</sup> percentile, and above the 80<sup>th</sup> percentile). In one report, the students were divided into two levels: one above and one below the 40<sup>th</sup> percentile. In seven reports, participants addressed the change in percentage of low, medium, and high performers from year to year. Of these seven reports, three compared the trends of sub groups according to low, medium, and high performance; two compared Caucasian, African American, and Hispanic students as subgroups and the other compared differences in performance according to gender. The following are two sample research questions and

two sample statements of purpose from the reports categorized as analyzing trends in disaggregated data.

1. How many of our students are not proficient or score below the 40<sup>th</sup> percentile as measured by the "Primary Reading Profile"?
2. Is there an improvement in reading comprehension scores from second through fourth grade for male students?
3. The purpose of our subject analysis is to determine if there are improvements in reading comprehension scores of second to fifth grade students over a 4 year period as indicated by ITBS reading comprehension scores.
4. The purpose of this subgroup analysis is to compare the ethnic groups (African American, Hispanic, and Caucasian) in the area of math problem solving.

#### Discussion

The purpose of this study was to provide a description of how a group of teacher leaders analyzed standardized achievement test scores in order to improve instruction. To summarize briefly, 6 of the 24 research reports utilized data from a single year only. Of those 6, 3 correlated achievement scores with other data, and 3 disaggregated data. The other 18 research reports utilized multiple years of data in some form of trend analyses. Of those 18, 5 analyzed trends only, 3 analyzed trends of data compared against the norm, and 10 analyzed trends for disaggregated data. All 18 of the trend analyses were completed with 4 years or less of data: 7 analyses utilized 2 years of data, 6 utilized 3 years of data, and 5 utilized 4 years of data. None of the analyses used 5 years or more of data. The limited number of years in these trend lines may indicate that student achievement data is not yet accessible enough for teachers, an explanation

informally confirmed by the teacher leaders. Improving the storage and organization of school data may contribute to the development of historical baselines against which current data can be compared more effectively.

The lack of baseline data could also limit the usefulness of trend analysis in university courses. However, professors should take note that participants in this study made effective use of only a single year of data by disaggregating and correlating data. Professors and staff developers should also note two other considerations related to providing instruction on school data analysis to teachers. First, the insights provided by school achievement data are often revealed through visual display of data; therefore, instruction should include practice with techniques related to the manipulation and transformation of data e.g., creating frequency distribution charts, histograms, charts and graphs. Second, the conclusions drawn from an analysis must be consistent with the statistical relationships governing the comparison of percentile scores; therefore, engaging with statistical concepts relevant to comparing percentile rankings should be an important part of coursework e.g., recognizing that percentile scores should not be averaged unless they are converted to Normal Curve Equivalents (Bernhardt; 1998).

The description of research reports from this study has indicated that school data can be analyzed in a wide variety of ways. Two applications not seen in this study include mapping of high and low scores across grade levels for the purpose of discovering instructional or curricular gaps (comparing against the norm) and overlaying trend lines in order to compare the performance of two different classes or grade levels (analyzing trends). Likewise, other approaches are possible. Sharing them through the educational literature is an important part of enriching the culture for school data

analysis. Each new description adds another model of practical application for the benefit of teachers, principals, and professors who are interested in making principled decisions based on standardized achievement data.



References

- Anderson, B., MacDonald, D.S. & Sinnemann, C. (2004). Can measurement of results help improve the performance of schools? *Phi Delta Kappa*, 85 (10), 735-739.
- Bernhardt, V. L. (1998). *Data analysis for comprehensive school improvement*. Larchmont, NY: Eye on Education.
- Bracey, G. (1997). *Understanding education statistics: It's easier (and more important) than you think*. Arlington, VA: Educational Research Service.
- Darling-Hammond, L. (1999). Teacher learning that supports student learning. *Educational Leadership*, 55 (5) 6-11.
- Grogan, M. (2001). Positive and not-so-positive effects of preparing students to pass standards of learning tests in virginia's high school classrooms. *NASSP Bulletin*. 85(6), 24.
- Hoover, H.D., Dunbar, S.B., Frisbee, D.A., Oberley, K.R., Bray, G.B., Naylor, R.J., Lewis, J.C., Ordman, V.L., & Qualls, A.L. (2003). *The iowa tests: Interpretive guide for teachers and counselors*. Ithaca, IL: Riverside Publishing.
- Johnson, R.S. (2002). *Using data to close the achievement gap: How to measure equity in our schools*. Thousand Oaks, CA.: Corwin Press.
- Murphy, J. & Forsyth, P.B. (1999). A decade of changes: Analysis and comment. In Patrick B. Forsyth and Joseph Murphy (Eds.) *Educational administration: A decade of reform*. pp. 253-272 Thousand Oaks, CA: Corwin Press, Inc.
- Katzenmeyer, M., & Moller, G. (2001). *Awakening the sleeping giant: Helping teachers develop as leaders*. (2<sup>nd</sup> ed.). Thousand Oaks, CA: Corwin Press.

- Killion, J. (2002). *Assessing impact: Evaluating staff development*. Oxford, OH: National Staff Development Council.
- Pellicer, L.O., & Anderson, L.W. (1995). *A handbook for teacher leaders*. Thousand Oaks, CA: Corwin Press.
- Phelps, R. P. (1998). The demand for standardized student testing. *Educational Measurement: Issues and Practice*, 17 (3), 5-23.
- Popham, J.W. (2001a). Uses and misuses of standardized tests. *NASSP Bulletin*. 85(6), 24.
- Popham, J.W. (2001b). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Shakeshaft, C (1999). A decade half full or a decade half empty: Thoughts from a tired reformer. In Patrick B. Forsyth and Joseph Murphy (Eds.) *Educational administration: A decade of reform*. Thousand Oaks, CA: Corwin Press, Inc.
- Streifer, P.A. (2002). *Using data to make better educational decisions*. Lanham, MA: The Scarecrow Press.
- Usdan, M., McCloud, B., & Podmostko, M. (2000, October). Tomorrow's principal. *Institute for Educational Leadership*, 4-6. Retrieved from [www.iel.org/programs/21st/report/principal.pdf](http://www.iel.org/programs/21st/report/principal.pdf)