Critical Attributes of Teachers Who Have Become Practitioners of Authentic Assessment

Bernadette Meisenheimer
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Recommended Citation

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 1996 Bernadette Meisenheimer
LOYOLA UNIVERSITY OF CHICAGO

CRITICAL ATTRIBUTES OF TEACHERS
WHO HAVE BECOME PRACTITIONERS OF AUTHENTIC ASSESSMENT

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF CURRICULUM, INSTRUCTION,
AND EDUCATIONAL PSYCHOLOGY

BY

BERNADETTE MEISENHEIMER

CHICAGO, ILLINOIS
JANUARY 1996
Copyright by Bernadette Meisenheimer, 1996
All rights reserved
ACKNOWLEDGMENTS

Many people have supported my educational goals, and to them I owe a debt of thanks for their patience, assistance, and expertise.

Loyola University of Chicago faculty: Dr. Robert Cienkus, committee chair, Dr. Barney Berlin and Dr. Dan Burke, committee members, for their guidance and advice;

Mundelein High School former and present administrators and staff: Dr. Linda Hanson, Mr. John Graham, Mrs. Kathy Conway, Mrs. Linda Dzwonkiewicz, for their support and encouragement;

Professional colleagues: Dr. Robert J. Wilson and Dr. Richard Stiggins for their permission to use portions of their survey instruments in this study;

My husband: Jim for his love and understanding;

My parents: Helen, my mother, and Edward, my late father, for teaching me the values of life-long learning.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Our Testing Roots</td>
<td>3</td>
</tr>
<tr>
<td>The Need to Quantify - Polygeny and Craniometry</td>
<td>5</td>
</tr>
<tr>
<td>Intelligence Testing</td>
<td>7</td>
</tr>
<tr>
<td>Army Alpha and Beta Testing</td>
<td>10</td>
</tr>
<tr>
<td>Influences on Classroom Testing</td>
<td>11</td>
</tr>
<tr>
<td>The Problem and Its Significance</td>
<td>15</td>
</tr>
<tr>
<td>Problem and Purpose</td>
<td>19</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>20</td>
</tr>
<tr>
<td>Procedures for Analysis of the Data</td>
<td>21</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td>23</td>
</tr>
<tr>
<td>Organization of the Study</td>
<td>23</td>
</tr>
<tr>
<td>2. REVIEW OF THE LITERATURE</td>
<td>24</td>
</tr>
<tr>
<td>A New Wave of Research</td>
<td>24</td>
</tr>
<tr>
<td>The Fallibility of Testing and its Effects on the Curriculum</td>
<td>29</td>
</tr>
<tr>
<td>The Reign of ETS</td>
<td>30</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>&quot;The Lake Woebegone Report&quot;</td>
<td>31</td>
</tr>
<tr>
<td>The Report of the National Commission on Testing and Public Policy</td>
<td>32</td>
</tr>
<tr>
<td>Effects on Curriculum</td>
<td>33</td>
</tr>
<tr>
<td>Reform Movements</td>
<td>34</td>
</tr>
<tr>
<td>National Reforms</td>
<td>35</td>
</tr>
<tr>
<td>State Reforms</td>
<td>38</td>
</tr>
<tr>
<td>Professional Organization Reforms</td>
<td>39</td>
</tr>
<tr>
<td>Authentic Assessment</td>
<td>40</td>
</tr>
<tr>
<td>Definition and Characteristics of Authentic Assessment</td>
<td>40</td>
</tr>
<tr>
<td>Pilot Projects</td>
<td>45</td>
</tr>
<tr>
<td>Research Findings</td>
<td>48</td>
</tr>
<tr>
<td>Implications for School Administrators</td>
<td>50</td>
</tr>
<tr>
<td>3. METHODOLOGY</td>
<td>57</td>
</tr>
<tr>
<td>Sampling Procedure</td>
<td>57</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>60</td>
</tr>
<tr>
<td>Research Questions and Treatment of Data</td>
<td>62</td>
</tr>
<tr>
<td>4. FINDINGS AND IMPLICATIONS</td>
<td>65</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>65</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>80</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>95</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Correlation for All Teachers Between Performance Assessments and Testing Practices</td>
<td>71</td>
</tr>
<tr>
<td>2. Authentic Assessors' Mean Scores On Assessment Practices</td>
<td>77</td>
</tr>
<tr>
<td>3. Authentic Assessors' Mean Scores on Level of Use</td>
<td>79</td>
</tr>
<tr>
<td>4. Authentic and Traditional Assessors' Mean Scores on Assessment Practices</td>
<td>87</td>
</tr>
<tr>
<td>5. Authentic, Traditional, and All Assessors' Mean Scores on Assessment Practices</td>
<td>89</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Time on Different Tests, as a Percentage of the Total Student Time Devoted to Test-Taking</td>
<td>17</td>
</tr>
<tr>
<td>2. Schools' Revenue Sources</td>
<td>58</td>
</tr>
<tr>
<td>3. Student Enrollment, Ethnicity, Limited English Proficient, and Low Income Figures</td>
<td>59</td>
</tr>
<tr>
<td>4. Total Teachers by Subject Area and Sex</td>
<td>60</td>
</tr>
<tr>
<td>5. Participating Teachers by Subject Area and Sex</td>
<td>60</td>
</tr>
<tr>
<td>6. Total and Participating Teachers by Subject Area and Sex</td>
<td>66</td>
</tr>
<tr>
<td>7. Percent of Time Spent on Authentic Assessment</td>
<td>68</td>
</tr>
<tr>
<td>8. Teachers' Scores on Assessment Practices</td>
<td>69</td>
</tr>
<tr>
<td>9. Teachers' Scores on Types of Assessments Used and Assessment Practices</td>
<td>70</td>
</tr>
<tr>
<td>10. Authentic Assessors as Identified by Subject Taught</td>
<td>72</td>
</tr>
<tr>
<td>11. Authentic Assessors as Identified by Years of Teaching, Sex, Age, and Education</td>
<td>72</td>
</tr>
<tr>
<td>12. Authentic Assessors as Identified by Last Graduate College Course and Last Workshop / Conference in Assessment Procedures</td>
<td>73</td>
</tr>
<tr>
<td>13. Authentic Assessors' Ranked Response to Most Important Contribution to Testing Knowledge</td>
<td>74</td>
</tr>
</tbody>
</table>
15. Authentic Assessors' Ranked Response to Reasons for Testing as Compared to Wilson's Data
16. Authentic Assessors' Mean Percent of Time Allotted to Different Types of Assessment
17. Authentic and Traditional Assessors as Identified by Subject Taught
18. Mean Scores for All Teachers of Time Allotted to Performance Assessments
19. Authentic and Traditional Assessors as Identified by Years of Teaching, Sex, Age, and Education
20. Authentic and Traditional Assessors as Identified by Last Graduate College Course and Last Workshop/Conference in Assessment Procedures
21. Authentic and Traditional Assessors' Ranked Response to Contributions to Testing Knowledge
22. Authentic and Traditional Assessors' Ranked Response to Reasons for Testing
23. Authentic, Traditional, and All Assessors' Mean Percent of Time Allotted to Different Types of Assessment
24. Comparative Data on Reliance on Different Types of Assessments
25. Comparative Data on Level of Use on Different Types of Assessments
CHAPTER 1
INTRODUCTION

Paper and pencil, one-right answer tests have been an integral part of our schools for as long as most people can remember. By the time students graduate from high school, they have answered thousands of test questions in which they must select the correct response from multiple possibilities, determine if a statement is true or false, and match items with their descriptors, synonyms or antonyms. Students take standardized tests, criterion-referenced tests, and teacher-made tests by the dozens; yet, the format is usually the same. There is one correct answer, and students select their response by filling in a blank or a bubble on a computerized form. Rarely are students asked to provide a rationale for the answers they selected. The process students use to arrive at the correct or incorrect response is hardly ever questioned, and all too infrequently are they asked to demonstrate their knowledge in any type of original format. Judgments about a student's competence begins early - often in kindergarten or before. The judgments can frequently be based on test results which presume to measure a student's ability to perform a certain set of tasks.

Roger Farr (1994) illustrated this point in his February 16, 1994 workshop "Solving The Assessment Puzzle" by conveying the reasoning behind the selections of kindergartners on a pre-reading test. Students were faced with the following test item:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>E</td>
<td>W</td>
<td>m</td>
</tr>
</tbody>
</table>
The test item was designed to assess visual discrimination and the student's ability to match an item with a duplicate when presented with a set of distractors. The directions, given by the teacher, were to match the first item marked "A" with the one that looked just like it from choices one through five. Dr. Farr questioned the first young boy who selected response number five. Dr. Farr asked why he had selected that item as the correct response. "Oh," replied the five year old, "I picked that one because it was hard to find. Number three was easy to find. But the teacher said to look real hard, and not mark the first answer you come to. I had to look hard to find this one."

Another student also selected number five as the correct response but used a different rationale. "I picked number five because the teacher said we shouldn't worry about this test. We should just go ahead and have fun, and color our papers when we're finished. I'm going to make a horse out of number five by putting a head on the front, and a tail on the back, and coloring it brown. I wouldn't be able to make a horse out of number three."

These responses show that as early as kindergarten, judgments are being made about the skills of students based on test results. The results may qualify students for remedial or gifted classes, place students in an academic track or ability group, or require that a student be retained in a developmental or second year kindergarten before progressing to the first grade.

However, once students leave school they rarely encounter the one-right answer testing format. Certain professions with licensing requirements may require candidates to attain a designated score on a paper and pencil test; yet, even these assessments are changing. In many states, beauticians were once required to pass a multiple-choice exam to secure a license. However, new exams require students to demonstrate the skills of cutting and styling before the expert eyes of a panel of examiners. Similarly, the Bar Exam in most states was once a multiple-choice examination and a series of essay questions on
theoretical issues, but is being replaced by performance examinations. In California, for example, only one day is devoted to multiple-choice exams, and two days are written exams testing "lawyering skills." Prospective attorneys receive a "file," setting forth the problem, and a "library" which consists of back-up research. The test asks students to use the facts and the law persuasively by writing a closing argument, a grievance letter, or a claim against a company (S. Leonard, personal communication, March 10, 1994).

The paper and pencil tests that are part of a licensing program quickly become a thing of the past as students enter a profession. The beautician may reach the rank of senior stylist by continuing her education, maintaining a prescribed weekly revenue, and serving as a technical leader in the salon. And the attorney advances to partner by his win-loss record in court, his ability to attract new clients, the amount of revenue he generates from his billings, and other similar demonstrations of skill and competence.

If the purpose of school is to prepare students for the world of work, why then is there such an emphasis on paper and pencil testing which is rarely used as a measure of worth outside of the classroom walls? How did testing come to be a measure of instruction, student achievement, an instrument for placement, and a doorway or roadblock to opportunity? The testing legacy has an interesting if not educationally sound history in our education system. This chapter will provide a brief look at our testing legacy. It will also examine the research questions in this study on the practitioners of a new science being labeled "authentic" assessment.

**Our Testing Roots**

Stephen Jay Gould (1981) in *The Mismeasure of Man* traces the history of testing which has preceded our over-reliance on objective standardized and classroom tests as today's measures of worth and intelligence. Gould's research
provides a look at the faulty framework upon which our testing methods are based. Gould begins by looking at the prevailing zeitgeist on intelligence in the 1800s.

During the 1800s, polygeny (the belief that human races were different species with some races genetically superior to others) was a commonly held belief. Louis Agassiz, a Swiss naturalist, is one of the first names associated with intelligence theories in the U.S. Agassiz came to America in the 1840s, and accepted an appointment as a professor teaching zoology at Harvard University. Agassiz was familiar with, but was not a proponent of polygeny. However, it is documented that Agassiz converted to polygeny upon coming into contact with blacks in America when he encountered them as servers in a Philadelphia restaurant. His correspondence reveals that the appearance of blacks was repugnant to him and that he feared intermarriage between blacks and whites would dilute the white race (Gould, pp. 44-49). In a letter to his mother in 1846, he wrote the following:

Nevertheless, I experienced pity at the sight of this degraded and degenerate race, and their lot inspired compassion in me in thinking that they are really men. Nonetheless, it is impossible for me to repress the feeling that they are not of the same blood as us. In seeing their black faces with their thick lips and grimacing teeth, the wool on their head, their bent knees, their elongated hands, their large curved nails, and especially the livid color of the palm of their hands, I could not take my eyes off their face in order to tell them to stay far away. And when they advanced that hideous hand towards my plate in order to serve me, I wished I were able to depart in order to eat a piece of bread elsewhere, rather than dine with such service. What unhappiness for the white race - to have tied their existence so closely with that of negroes in certain countries! God preserve us from such a contact! (Gould, p. 45)

Although Agassiz did not involve himself in any scientific studies of polygeny, he is known as a theorist of the belief. Agassiz became acquainted
with Samuel George Morton who is credited with being the first scientist to "prove" the theory of polygeny (Gould, pp. 44-49).

The Need to Quantify - Polygeny and Craniometry

Morton was a physician from Philadelphia who set out to prove the biological superiority and inferiority of different human races by gathering skulls, filling them with buckshot, and then weighing them. By the time of his death in 1851, Morton had gathered and measured the cranial cavity of over 1,000 skulls. He filled the skulls with mustard seeds, and later converted to using buckshot, emptied the seed or buckshot into a container, and then recorded the weight. Although he studied and classified many groups, his findings in the U.S. "proved" that whites were superior in intelligence, followed by Indians, and lastly were the blacks. He published reports of his findings between 1839 and 1849 (Gould, pp. 50-53). In working toward his "mismeasurement" theory, Gould reanalyzed Morton's work and found "... Morton's summaries are a patchwork of fudging and finagling in the clear interest of controlling a priori convictions" (Gould, p. 54). While Gould does not accuse Morton of conscious manipulation of data, he feels Morton's belief in polygeny caused him to "finagle" to prove his belief.

Some of the points Gould reviews in his analysis of Morton's original research include the following:

1. Favorable inconsistencies and shifting criteria: Morton often chose to include or delete large subsamples in order to match group averages with prior expectations ... He made calculations for Caucasians to demonstrate the superiority of Teutons and Anglo-Saxons, but never presented data for Indian subsamples with equally high averages.

2. Subjectivity directed toward prior prejudice: Morton's measures with seed were sufficiently imprecise to permit a wide range of influence by subjective bias; ... In other words, blacks fared poorest and whites best when the results could be biased toward an expected result.
3. Procedural omissions that seem obvious to us: Morton was convinced that variation in skull size recorded differential, innate mental ability. He never considered alternate hypotheses, though his own data almost cried out for a different interpretation . . . . Morton used an all-female sample of three Hottentots to support the stupidity of blacks, and an all-male sample of Englishmen to assert the superiority of whites.

4. Miscalculations and convenient omissions: All miscalculations and omissions that I have detected are in Morton's favor. He rounded the negroid Egyptian average down to 79 rather than up to 80. He cited averages of 90 for Germans and Anglo-Saxons, but the correct values are 88 and 89. He excluded a large Chinese skull and an Eskimo subsample from his final tabulation for mongoloids, thus depressing their average below the Caucasian value. (Gould, pp. 68-69)

Gould's reanalyzation of Morton's data revealed " . . . no significant differences among races for Morton's own data" (p.67).

In the 1860s Paul Broca, the founder of the Anthropological Society of Paris, weighed the brain after autopsies and added more controls to the experiments that Morton had conducted with buckshot. "He spent months refining the technique, taking into account such factors as the form and height of the cylinder used to receive the shot after it is poured from the skull, the speed of pouring shot into the skull, and the mode of shaking and tapping the skull to pack the shot and to determine whether or not more will fit in" (Gould, p. 85). Broca also measured different parts of the skull to establish charts showing the intelligence of the different races. When Broca began measuring the length of the lower arm and upper arm and correlating these measurements with size of the brain, he began finding that blacks were surpassing whites in intelligence. Rather than question his own research results, he quickly abandoned this type of correlation as a measure of intelligence because it was not proving his belief about the relationship of intelligence and race (Gould, pp. 82-88).
When Darwin's theory of evolution poked holes in the polygenists' belief that man had descended from separate species, a new science was invented to prove racial inequality. In the late 1800s, Darwin's cousin Francis Galton began measuring various parts of the skull to test for intelligence. This science was known as crainometry (Gould, p. 75). Galton believed that almost everything could be measured, and he is considered the pioneer of modern statistics. "In the 1880s Galton began a program of measuring the intellects of English school children by testing their hearing, coordination, and reaction time. He pioneered the statistical technique of correlation analysis which the Educational Testing Service (ETS) uses today to determine the validity of its tests" (Nairn, 1980, p. 163). Galton also worked at categorizing the "civic worth" in man by developing a table with eight categories ranging from a low of criminals and loafers to a high of the intelligent class. With his categories in place, Galton then moved to measuring and weighing babies and assigning them to a category.

Galton had a theory that measurement of personal qualities could be used to change the character of human life itself. Through the practice of eugenics (the identification of the degree of desirable or undesirable qualities in individuals and the control of their reproduction through planned marriages and sterilization to maximize the desirable), Galton and his successors sought to improve the quality of the world's gene-pool. (Nairn, p. 164)

The seeds of the testing movement had begun. People were being categorized and labeled, and judgments of worth were being issued by those who were considered experts in their field. These early efforts in testing and quantifying intelligence would eventually lead to wide-scale testing efforts.

Intelligence Testing

Prior to 1900, the majority of the research on intelligence was related to measuring parts of the body and determining the weight of the cranial cavity.
Alfred Binet changed the way testing was conducted and laid the path for I.Q. testing. Binet, who was director of the psychology laboratory at the Sorbonne, first studied intelligence by using some of Paul Broca's techniques, but then began to question his methods. "I feared," Binet wrote in 1900, "that in making measurements on heads with the intention of finding a difference in volume between an intelligent and a less intelligent head, I would be led to increase, unconsciously and in good faith, the cephalic volume of intelligent heads and to decrease that of unintelligent heads" (Gould, p. 147).

When the minister of public education commissioned Binet to develop a test to identify students for special education, he switched from what he termed "medical" approaches to "psychological" methods. Binet came up with a test of 54 tasks that ranged from nursery level to mid-teens. Most of the items were ". . . short tasks related to everyday problems of life (counting coins, or assessing which face is 'prettier,' for example), but supposedly involving such basic processes of reasoning as 'direction' (ordering), comprehension, invention and censure (correction)" (Gould, p. 149). Binet was careful to avoid tasks which could be attributed to instruction or rote learning, and he cautioned that the scale not be used to measure intelligence "as linear surfaces are measured" nor to rank "all pupils according to mental worth" (Gould, pp. 151-152).

Intelligence testing as we know it in the United States can be tracked to Alfred Binet. His work made its way to the U.S. by way of H. H. Goddard who translated and popularized it here. "Goddard and his colleagues emphasized that such crucial matters as directing the educations and careers (and sex lives) of other people was a task for trained professionals only" (Nairm, p. 167). Gould identified some of the idiocies and atrocities committed in the name of research by Goddard. Gould tells of how Goddard sent two researchers to Ellis Island in 1913 for two months for the express purpose of testing immigrants. Thirty-five Jews, twenty-two Hungarians, fifty Italians, and forty-five Russians were tested.
"Binet tests on the four groups led to an astounding result: 83 percent of the Jews, 80 percent of the Hungarians, 79 percent of the Italians, and 87 percent of the Russians were feeble minded - that is, below age twelve on a Binet scale" (Gould, p. 166). The fact that these people had just endured an ocean voyage, were probably poor and hungry, and did not speak English did not play a part in the research design. Goddard was disturbed by the results and "Eventually . . . monkeyed about with the tests, tossed several out, and got his figures down to 40 to 50 percent, but still he was disturbed" (Gould, p. 166). However, Binet's work in intelligence testing has become immortalized through the Stanford-Binet Intelligence Test which bears his name and the name of Stanford University where Lewis Terman worked with expanding the test.

Terman revised the Binet test from the original 54 tasks to 90 tasks that ranged all the way up to superior adults. Not heeding Binet's earlier warning, Terman put the top end on the IQ measurement and set the standard for all IQ testing that would follow (Gould, pp. 174-176). Allan Nairn (1980) in The Reign Of ETS quotes from Terman's 1916 book of instructions for the Stanford-Binet test regarding scores falling in the seventy to eighty range:

... [these scores are] very, very common among Spanish children and Mexican families of the Southwest and also among negroes . . . [sic] Children of this group should be segregated in special classes . . . [sic] They cannot master abstractions but they can often be made efficient workers . . . [sic] There is no possibility at present of convincing society that they should not be allowed to reproduce, although from a eugenic point of view they constitute a grave problem because of their unusually prolific breeding. (Nairn, p. 170)

According to Harold Berlak (1992) , Goddard and Terman became

... prime movers of the American eugenics movement and leading advocates for developing mental measurements which could be used to identify and classify mental incompetents and
defectives. Mental tests, they argued, were a scientific means of identifying and controlling social deviants (read: troublemakers and nonconformists). (Berlak, p. 182)

Goddard and Terman were soon to join forces with their colleague Robert Yerkes who is credited with beginning the mass testing movement in the United States.

Army Alpha and Beta Testing

The first mass produced intelligence tests were developed for the Army by R. M. Yerkes, a faculty member at Harvard. With the advent of World War I in 1917, the Army was faced with over a million recruits and used Yerkes' test to classify these men and women into general categories as enlisted men and officers. During the course of the war over 1.75 million recruits were tested on Yerkes' product. Gould (1981) in The Mismeasure of Man describes the test.

The Alpha test included eight parts, the Beta seven; each took less than an hour and could be given to large groups. Most of the Alpha parts presented items that have become familiar to generations of test-takers ever since: analogies, filling in the next number in a sequence, unscrambling sentences, and so forth. This similarity is no accident; the Army Alpha was the granddaddy, literally as well as figuratively; of all written mental tests. One of Yerkes's disciples, C. C. Brigham, later became secretary of the College Entrance Examination Board and developed the Scholastic Aptitude Test on Army models. If people get a peculiar feeling of deja-vu in perusing Yerkes's monograph, I suggest that they think back to their own College Boards, with all its attendant anxiety. (Gould, p. 199)

Testing conditions were not optimal, and test items had no such indicators as validity or reliability. Sometimes, the groups Yerkes used for his testing research had to be those who were most convenient. Yerkes published a statistical monograph in 1921 that detailed his work and testing results. Gould in quoting from Yerkes's monograph tells of his efforts to round up the local
prostitutes from areas surrounding the Army base where he was working.

The results of the Army examining of prostitutes corroborate the conclusion, attained by civilian examinations of prostitutes in various parts of the country, that from 30 to 60 percent of prostitutes are deficient and are for the most part high-grade morons; and that 15 to 25 percent of all prostitutes are so low-grade mentally that it is wise (as well as possible under the existing laws in most states) permanently to segregate them in institutions for the feeble-minded. (Gould, p. 198)

Gould continues by saying, "One must be thankful for small bits of humor to lighten the reading of an eight-hundred page statistical monograph. The thought of army personnel rounding up the local prostitutes and sitting them down to take the Binet tests amused me no end, and must have bemused the ladies even more" (p. 198).

Influences on Classroom Testing

A number of events and forces came together in the waning part of the 19th Century and the beginning of the 20th Century which would impact all future testing in the United States. One of the prime forces was Carl Brigham who had been commissioned to complete a study of the World War I Army test results. Brigham had worked with a sample of 81,000 native-born whites, 12,000 foreign-born individuals, and 23,000 blacks, and in 1923 he published his results in A Study of American Intelligence. Allan Nairn (1980) in The Reign of ETS states "The army test scores showed a clear pattern: the foreign-born people who had lived longest in the U.S. scored highest and recent arrivals scored lowest" (p. 178). Since the army tests had been previously criticized, Brigham (1923) was prepared for the critics when he published his study. On concerns with the sample of his population he stated:
The tea taster samples the tea to be graded. He does not need to brew a whole bale of tea to find its worth . . . . In the same way, no one could seriously question the reliability of our sampling . . . . (Brigham, p. 31)

On the fact that the language factor may have distorted the results:

. . . if one wishes to deny, in the teeth of the facts, the superiority of the Nordic race on the ground that the language factor mysteriously aids this group when tested, he may cut out of the Nordic distribution the English speaking Nordics, and still find a marked superiority of the non-English speaking Nordics over the Alpine and Mediterranean groups. (Brigham, p. 171)

On the inferiority of races:

Our results showing the marked intellectual inferiority of the negro are corroborated [sic] by practically all of the investigators who have used psychological tests on white and negro groups . . . . Our figures [also] tend to disprove the popular belief that the Jew is highly intelligent. (Brigham, p. 190)

On intermarriage and immigration laws:

We may consider that the population of the United states is made up of four racial elements, the Nordic, Alpine, and Mediterranean races of Europe, and the negro. If these four types blend in the future into one general American type, then it is a foregone conclusion that this future blended American will be less intelligent than the present native born American, for the general results of the admixture of higher and lower orders of intelligence must inevitably be a mean between the two . . . . (Brigham, p. 205)

It was these beliefs that Brigham brought to a company not yet founded that would become known as the Educational Testing Service (ETS).
A second force at work in the late 1800s was the National Education Association. Concerned with the growing school population and the lack of uniformity in the schools, three committees were formed: the Committee of Fifteen on Elementary Education, the Committee of Ten on Secondary School Studies, and the Committee on College Entrance Requirements. According to Ornstein and Hunkins (1993), when the Committee on College Entrance Requirements "... met in 1895, it reaffirmed college dominance over the high school, in terms of admission requirements and classical subjects for mental training at the high school and college levels" (p. 87). The Committee was also charged with setting standards for admission to university study, and in 1900 was chartered as the College Entrance Examination Board. The first College Board exam was an essay test and was administered in 1901 to 973 applicants. According to Nairm (1980) two key events helped to establish its place in testing history.

The first was in 1916, when Harvard, Yale and Princeton simultaneously gave up their own examination and turned all their candidates for admission over to the Board. The second was in 1919, when Vassar, Smith, Mount Holyoke and Wellesley, by mutual agreement, ... sent their hundreds of candidates to join the already swollen ranks. (p. 185)

A third factor converging on this same time period was the work of Edward Thorndike at Teachers College Columbia University in New York. Thorndike authored hundreds of articles and books on testing and also developed scales to measure student achievement in arithmetic, handwriting, spelling, drawing, reading and language ability (Levin, 1991, p.73). Referred to as "the godfather of standardized testing," Thorndike is quoted as saying, "Whatever exists at all exists in some amount" (Levin, p. 73). His work in testing
was highly respected and found its way into teacher education programs across the country.

These three factors resulted in many significant events that impacted the already growing testing movement including the first publication of major tests which are still used today. In less than 50 years, the following events impacted testing:

1900 College Entrance Examination Board chartered
1904 Thorndike published first textbook on educational measurement
1908 Stone published arithmetic reasoning ability test
1909 Courtis published arithmetic computation test
1910 Thorndike published handwriting scale
1913 Buckingham published spelling power test
1914 Thorndike published vocabulary test
1923 Stanford Achievement tests published
1942 College Board dropped essay exams in favor of multiple choice
1937 Testing organization became ETS
1947 ETS chartered

Thorndike found his testing philosophy readily adopted by school districts who were struggling with an explosion of students. Levin (1991) offers an explanation for the mass adoption of Thorndike's testing philosophy by writing about the exploding school population during the Progressive period.

Creating mass education during the Progressive period was an immense task. From 1890 to 1918, for example, secondary school attendance increased 711 percent, from just over 200,000 to more than 1.5 million students, while the general population increased only 68 percent (Tyack, 1974). One can imagine great pressure on educators of that time, as today, to test and grade, sort and categorize, separate and distinguish pupils - in short to find slots
for them in school and ultimately society and, from educators' standpoints, to make bureaucratic and organizational sense out of a chaotic period of growth and change. (Levin, p. 75)

Thorndike's testing philosophy enabled educators to "grade" and "sort," and publishing companies followed suit with graded textbooks and workbooks; and teaching became a means to the testing end.

The measurement texts used in teacher education programs in the 1960s and the 1970s still bore marks of Thorndike's influence. Texts by Marshall and Hales (1971), Tenbrink (1974), and Dizeney (1971) provided instructions on how to build multiple-choice tests, emphasizing such items as supplying appealing detractors, writing grammatically correct items, and providing an equal number of responses to all test questions.

The Problem and Its Significance

The legacy of classroom testing has come to the 20th Century classrooms through a rather incongruous route. From measuring skulls, determining the weight of cranial cavities, devising questionnaires to identify the feeble-minded, and selecting officers for the military, educational testing has evolved to an almost universal system of assessing students through the use of objective questions and multiple-choice answers.

The research on classroom testing indicates that teachers feel ill-prepared to design tests, and predominantly use multiple-choice tests which do not adequately measure intended outcomes or higher level thinking.

Stiggins and Conklin (1992) in *In Teachers' Hands* summarized the results of the research on classroom testing. Stiggins and Conklin's analysis of the study by Fleming and Chambers of 400 teacher-made tests revealed the following:
1. Teachers use short-answer questions most frequently in their test making.

2. Teachers, even English teachers, generally avoid essay questions, that represent slightly more than 1% of all test items reviewed.

3. Teachers use more matching items than multiple-choice or true false items.

4. Teachers devise more test questions to sample knowledge of facts than any of the other behavioral categories studied.

5. When categories related to knowledge of terms, knowledge of facts, and knowledge of rules and principles are combined, almost 80% of the test questions reviewed focus on these areas.

6. Teachers develop few questions to test behaviors that can be classified as ability to make applications.

7. Comparison across school levels shows that junior high school teachers use more questions to tap knowledge of terms, knowledge of facts, and knowledge of rules and principles than do elementary or senior high school teachers. Almost 94% of their questions address knowledge categories, versus 69% of senior high school teachers' questions and 69% of the elementary school teachers' questions. (Stiggins and Conklin, pp. 13-14)

Stiggins and Conklin (1992) also reported on a study by Carter.

In another study, Carter (1984) studied the test development skills of 310 high school teachers and reported that teachers had great difficulty recognizing items written to measure specific skills, especially higher order thinking skills. She also reported that teachers learned to write original items at higher skills levels very slowly and felt insecure about their test making capabilities. (p. 14)

While teachers may feel ill-prepared to write test items, research shows that they tend to rely more on their own tests rather than on publishers' tests or standardized tests. Stiggins and Conklin reported that Dorr-Bremme and Herman found than almost 75 percent of tests used at the high school level are
teacher-made. Table 1 is from Stiggins and Conklin where it was reprinted by permission of Dorr-Bremme and Herman.

**TABLE 1**

TIME ON DIFFERENT TESTS, AS A PERCENTAGE OF THE TOTAL STUDENT TIME DEVOTED TO TEST-TAKING

<table>
<thead>
<tr>
<th>Tests</th>
<th>Tenth Grade English Teachers</th>
<th>Tenth Grade Mathematics Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tests which form part of a statewide assessment program</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>Required minimum-competency tests</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Tests included with curriculum materials</td>
<td>8%</td>
<td>17%</td>
</tr>
<tr>
<td>Other commercially published tests</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>Locally developed- and district-adopted tests</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>School- or teacher-developed tests</td>
<td>74%</td>
<td>76%</td>
</tr>
</tbody>
</table>


In reviewing research by Gullickson and Ellwein, Stiggins and Conklin found more evidence which indicates that classroom testing is based on poor methods and little strategy.
Gullickson's (1982; Gullickson and Ellwein, 1985) studies of midwestern teachers' testing strategies provide further evidence of a lack of quality control strategies. For example, few of the teachers he surveyed computed summary statistics needed to evaluate test performance. Most limited test questions to short-answer and matching, testing lower cognitive levels. Few teachers took time to improve their tests, and usually reused items without careful item analysis. Overall, Gullickson concluded that teachers have not been taught how to evaluate their test items, take necessary steps to improve quality, or accurately set criterion levels for student performance. Further, they do not value statistical analysis of test items as a helpful strategy in the classroom. (Stiggins and Conklin, p. 16)

Studies on students' perceptions of testing have been limited, but research conducted by Haertel et al (1984) surveyed over 600 high school students and found that:

While students consider tests important and are willing to work to earn high scores, they see tests as requiring mostly memorization, perhaps to the detriment of other types of learning Students understand that there should be more to schooling outcomes than answering multiple-choice questions; over half recognize that many important ideas are not tested at all. (p. 29)

Stiggins and Conklin (1992) in their own research with sixth grade and high school teachers used a variety of research techniques including questionnaires, journal entries by teachers, and observations by trained personnel. For use at the high school level, Stiggins and Conklin devised a framework for analyzing classroom testing environments. The framework covers eight dimensions and details over 400 specific items. The framework used at the high school level attempts to record (1) assessment purposes, (2) assessment methodologies, (3) criteria used in selecting the assessment method, (4) quality of assessments, (5) feedback, (6) the teacher as assessor: background time expenditure and personal/professional characteristics, (7) the teacher's
perception of the students, and (8) the assessment-policy environment. In a synthesis of their research, they found the following:

1. Assessment is largely used for the narrow purpose of grading; opportunities for such critical activities as instructional improvement including student diagnosis and evaluation of instruction are missed.

2. Higher-order thinking skills are not understood and/or not assessed and there is commonly a mismatch in thinking skills level of instruction, assignments, and tests.

3. Individual student grades are not always based on valid or reliable data and criterion- and norm-referenced systems are confounded.

4. Teachers are unfamiliar with appropriate methods for assessing performance.

5. The meaning of objective assessment and valid assessment are confused, leading teachers to regard objective tests as necessarily valid and performance assessments as necessarily less valid.

6. Teachers view instruction and assessment as entirely distinct functions and do not know how to integrate instruction and assessment in planning class time.

7. Although they wish to base their assessments of students on achievement, teachers often mix affective factors into grading equations as they strive to motivate students. (Stiggins and Conklin, pp. 151-152)

Problem and Purpose

The problems inherent in classroom testing will become more complicated in the years to come. If teachers feel they have had inadequate training to prepare them as assessors of student progress and lack the time to develop adequate tests, the reform movements concerning testing will only add to their burden. Chapter 2 reviews the problems with present day testing and the call for reform from all arenas of the educational community. What is being called for is an "authentic" type of assessment. A measure that examines
students' processes as well as products, one that looks at growth over time, and one that reflects the skills that are considered important in the world of work.

The purpose of this study, using quantitative and qualitative research designs, was to examine attributes of teachers who have shifted their paradigm from objective testing to become practitioners of the new art deemed authentic assessment. The research questions included the following:

1. Using an operational definition of authentic assessment, based on the literature, which teachers in three suburban high schools in Lake County, Illinois can be identified as practitioners of authentic assessment practices?

2. Of the identified teachers, what are the attributes and characteristics which separate them from teachers using traditional assessment techniques (attributes included personal factors, assessment practices, reasons for testing, and professional development)?

3. What conditions of the school were contributing factors in enabling teachers to become practitioners of authentic assessment (factors included financial indicators, administrative/supervisory structure, school characteristics, and school climate)?

Through analysis of the qualitative data based on questionnaires, school characteristics, and teacher interviews, a profile of teachers who have become authentic assessors was developed.

Significance of the Study

To respond to the reform movements that will be called for in testing procedures, superintendents must be cognizant of the characteristics possessed by teachers who are capable of changing paradigms. Superintendents and other supervisory personnel must also be aware of and be willing to provide the kind of environment and support systems that will enable teachers to learn the
strategies and experiment with pilot projects related to using authentic assessment designs.

A lack of clarity on the part of key personnel charged with providing for and managing the change in assessment methods will only create a greater burden for teachers struggling to institute the assessment reforms. "The frustration of small failures and unserviceable plans for improvement will lead to an inevitable cycle of business as usual in our schools" (Louis and Miles, 1990).

The reform literature must include research identifying current practices in our schools and the environments and characteristics of effective practices and practitioners. The implications of knowing which teachers are and why they are effective classroom assessors will affect hiring practices, inservice decisions, and teacher preparation programs.

Procedures for Analysis of the Data

Data was gathered using three questionnaires. A Classroom Testing Questionnaire (Appendix A) was used as a screening device to identify teachers as those who were practitioners of authentic assessment and those who were using traditional assessment practices. An In-Depth Interview Questionnaire (Appendix B) was used with teachers who fell at the two extremes on the screening device. And a Principal Interview Questionnaire was used to gather supporting data on the schools (Appendix C).

The Classroom Testing Questionnaire was developed based on a review of the literature, and on the previous research of Robert J. Wilson and Richard Stiggins. Permission was requested and granted to use portions of their survey designs in this study (Appendices D and E).

The Classroom Testing Questionnaire was divided into five sub-sections: (a) Demographic and Professional Information, (b) Reasons for Testing, (c) Assessment Practices, (d) Types of Assessments Used, and (e) Level of Use.
Numerical ratings were assigned to the responses in sub-sections B through E of the Classroom Testing Questionnaire. A formula was developed which yielded a number indicative of those respondents who could be identified as practitioners of authentic assessment. Cut-off ranges were established to identify respondents who fell at either of the extremes on the screening device. These respondents were selected for further study, and took part in an in-depth interview.

An In-Depth Interview Questionnaire was developed based on a review of the literature. The questionnaire was divided into four sub-sections identified as: (a) Personal Factors, (b) Professional Development and Involvement, (c) Administrative/Supervisory Structure, and (d) Assessment Practices.

Respondents who fell above the "high" cut-off score and below the "low" cut-off score on the screening instrument were identified as subjects for an in-depth interview. A scoring chart was developed to categorize the respondents' answers based on the presence or absence of identified characteristics and factors in some questions, and on their frequency of use or knowledge base on other questions.

Additionally, an interview was conducted with the Principal of each high school to gather background information to assist in the interpretation of the data. The Principal interview covered the topics of district organization, population and description of the school, recent leadership history, relations with the Board of Education, vision and mission of the school, recent and anticipated changes, academic organization and average teaching assignment, focus on curriculum and inservice, climate and teacher empowerment, and the school's strengths and weaknesses or challenges.

Analysis of the data was conducted using the computer software package Microsoft ® Excel (Version 2.2, copyright 1989).
Limitations of the Study

The study was limited to the three high schools, in Lake County, Illinois. The study was also limited to the characteristics of authentic assessment identified in a review of the literature, and those characteristics and factors identified in the Classroom Testing Questionnaire and the In-Depth Interview Questionnaire. Other factors and characteristics were not included. Responses of the teachers were based on their perceptions and may reflect individual interpretations of actual practice.

Organization of the Study

Chapter 2 reviews the literature on testing and authentic assessment, and presents research findings on current assessment practices.

Chapter 3 presents the methodology used in the study. Sampling procedures, instrumentation, and research questions and treatment of the data are detailed.

Chapter 4 covers the findings and implications of the study. The data collected from the questionnaires is reviewed and described.

Chapter 5 presents a summary of procedures, the conclusions derived from the results of the study, recommendations drawn from the study, and recommendations for further research.
CHAPTER 2
REVIEW OF THE LITERATURE

The literature on testing beginning in the late 1970s and early 1980s took a new direction. Key research studies began to examine areas that were interwoven in terms of intelligence, learning, and classroom testing. Studies on the nature of intelligence and how the brain functions and what we do in terms of classroom learning and testing were investigated. Exposes on the practices of standardized testing and its fallibility were written, and studies on how testing affects curriculum and instruction were conducted. The end result has been a call for reform. Without exception, all the reform movements cite a need for authentic assessment. However, a significant section - how teachers make the transition from using multiple-choice, one right answer tests, to using authentic measurements - is missing from the literature.

A New Wave of Research

Beginning in the 1980s brain based research started gaining popularity. Educators began looking into learning and its relationship to intelligence and brain processing. A variety of theories were developed based on social interaction and the way people learn outside the confines of the conventional classroom. The October 1990 issue of Educational Leadership was devoted to learning styles and research on the brain. Scientists also began studying brain function by isolating through EEG tests which areas of the brain responded to different learning stimuli, how the brain has evolved over time, and the way the brain operates when faced with solving new problems.
Generally, aside from the purely scientific research, the brain or learning research has focused on three areas. The first area encompassed attempts to understand the processing involved in learning by the "expert" or competent performer. Studies of "novices" and "experts" were performed in many fields including chess players, bridge players, computer programmers, radiologists, social scientists, athletes, and teachers. Studies found that the "experts" were always involved in patterning or organizing knowledge in ways different than the "novices." The "experts" were also highly attuned to what is termed metacognition or thinking about their own thinking as they worked (Bransford and Vye, 1989, p. 178).

The second area of research looked at the "initial" states of learners. What information did they bring to learning, and how did they resolve any misconceptions they had (Bransford and Bye, 1989, p. 183). Caine and Caine (1991) in Teaching And The Human Brain provide an example of how the state of the learner is often ignored in everyday classroom learning.

Children live with parallel lines long before they ever encounter school. By the time parallel lines are discussed in geometry, the average student has seen thousands of examples in fences, windows, mechanical toys, pictures, and so on. Instead of referring to the parallel lines students and teachers have already experienced, most teachers will draw parallel lines on the blackboard and supply a definition. Students will dutifully copy this "new" information into a notebook to be studied and remembered. Parallel lines suddenly become a new abstract piece of information stored in the brain as a separate fact. No effort has been made to access the rich connections already in the brain that can provide the learner with an instant "Aha!" sense of what the parallel lines they have already encountered mean in real life, what can be done with them, and how they exist other than as a mathematical abstraction. (p. 4)
The third area of research involved the transition of a student's initial state to a goal state: the process by which new knowledge is actively constructed by the learner. Cognitive scientists believe that if students don't have the opportunity to use new knowledge in meaningful ways it remains inert and can only be called forth in certain contexts such as testing situations (Bransford and Vye, p. 188).

This brain based research has evolved into what is termed learning theory as contrasted with earlier research efforts on teaching theory. Learning theorists believe in the social and authentic aspects of learning. They point to the "real world" where most learning is done cooperatively and where demonstrations of learning are projects, presentations, and portfolios, not multiple choice tests. Hence they believe that cooperative learning is the route to effective instruction, and that assessment should focus on students' strengths and on what they will need to accomplish in the real world. Leslie Hart (1983) says:

The ability to make plans and carry them out is the key aspect of human intelligence - a truth that becomes strikingly evident when we look at our history as humans. Yet as teachers or instructors we commonly do the planning ourselves (or follow those laid down for us by authorities), and the students, told what to do at every turn, get little chance to use their brains in this basic, human way. (p. 49)

Hart goes on to point out the fallacy of teaching and testing in small segments.

Once we begin to look critically at this notion of teaching in a logical sequence, we can see that usually a further giant - and utterly wrong - assumption has been made: that if a subject is fragmented into little bits, and the student is then presented with the bits in some order that seems logical to somebody, the student will be quite able to assemble the parts and emerge with the whole - even though never given an inkling of the whole! . . . We would hardly expect that if we show a young boy all of the parts of a
television, he would then be able to assemble the receiver, and also grasp how the interrelated, interdependent components work as a system. (pp. 52-54)

Howard Gardner (1983) in *Frames of Mind* began to look at the nature of intelligence and introduced his theory of Multiple Intelligences (MI). In his work, Gardner studied people of many cultures, educational levels, and achievement. His quest was to determine a definition of intelligence that was not unique to Western civilization. Gardner studied students who had mastered skills and displayed intelligence (navigating by the stars, composing music by using a computer and synthesizer, etc.), but felt that "... current methods of assessing the intellect are not sufficiently well honed to allow assessment of an individual's potentials or achievements . . . ." (p. xix). The paper and pencil tests that currently assess intelligence often ask the test taker to perform tasks that are often not valued by all cultures or even valued by the culture in which intelligence is being measured. What, for example, is the value in repeating in reverse order a series of numbers, or in identifying which picture in a set of four matches a "control" picture.

Gardner (1991) in his book *The Unschooled Mind* focused on how students who scored well on tests and received good grades in courses, had little understanding of the concepts taught. Gardner reviewed studies (Clement, 1982 and 1983; Arons, 1973; Carmazza, McCloskey, and Green, 1980 and 1981; and Clement, 1982) at MIT, Johns Hopkins, and other universities which looked at physics students who failed to give correct explanations or answers to questions on basic concepts, involving the principles of gravity and trajectory, when the questions were rephrased in terms other than those encountered in traditional testing situations.

In a typical example, college students were asked to indicate the forces acting on a coin that has been tossed straight up in the air
and has reached the midway point of its upward trajectory. The correct answer is that once the coin is airborne, only gravitational pull toward the earth is present. Yet 70 percent of college students who had completed a course in mechanics gave the same naive answer as untrained students: they cited two forces, a downward one representing gravity and an upward one from "the original upward force of the hand." This response reflects the intuitive or common-sense but erroneous view that an object cannot move unless an active force has somehow been transmitted to it from an original impelling source (in this instance, the hand or arm of the coin tosser) and that such a force must gradually be spent.

Students with science training do not display a blind spot for coin tossing alone. When questioned about the phases of the moon, the reasons for the seasons, the trajectories of objects hurtling through space, or the motions of their own bodies, students fail to evince the understandings that science teaching is supposed to produce. Indeed, in dozens of studies of this sort, young adults trained in science continue to exhibit the very same misconceptions and misunderstandings that one encounters in primary school children - the same children whose intuitive facility in language or music or navigating a bicycle produces such awe. (Gardner, 1991, pp. 3-4)

Gardner (1991) discusses the misconceptions that abound in schools that link correct answers on multiple-choice tests with true understandings.

These investigations document that even students who have been well trained and who exhibit all the overt signs of success - faithful attendance at good schools, high grades and high test scores, accolades from their teachers - typically do not display an adequate understanding of the materials and concepts with which they have been working. (p. 3)

If you answer questions on a multiple-choice test in a certain way, or carry out a problem set in a specified manner, you will be credited with understanding. No one ever asks the further question "But do you really understand?" because that would violate an unwritten agreement: A certain kind of performance shall be accepted as adequate for this particular instructional context. The gap between what passes for understanding and genuine
understanding remains great; it is noticed only sometimes . . . , and even then, what to do about it remains far from clear. (p. 6)

Gardner along with Yale's Robert J. Sternberg initiated a six-year research project in 1992 entitled "Practical Intelligence for School." The project combines Gardner's MI theory with Sternberg's thinking theory. Sternberg has identified twelve different thinking styles and feels classroom instruction can be improved if teachers and students can vary their thinking styles. Currently the researchers are working with teachers to develop curriculum units for use in schools in Connecticut and Massachusetts. The project also deals with what Sternberg calls 'tacit knowledge' a common sense approach to solving problems which is usually not part of a core curriculum. According to Sternberg 'tacit knowledge' would help a student who excels in mathematics, but can't calculate his own bowling score (Jacobson, 1992, pp. A9-A15).

It is evident that current tests, both standardized and teacher made tests, are not compatible with what researchers now know about brain functioning and the learning process, yet little has changed in the general practice of assessing students' learning.

The Fallibility of Testing and its Effects on the Curriculum

From its early beginnings, testing has grown to be big business in the United States. "A recent study by FairTest, a Boston-based advocacy group, found that U.S. public schools administered 105 million standardized tests to 39.8 million students in the 1986-87 school year alone - an average of more than 2.5 standardized tests per student per year" (Leslie and Wingert, 1990, p. 56). A special edition of Newsweek in September of 1990 boosted that number to 127 million tests administered each year. Add to this number the teacher-made quizzes and tests, and the number of tests students take each year rises significantly. Tests are designed to measure, rate, and evaluate intelligence,
ability, performance, and future performance. Tests are used to identify preschoolers fit for gifted kindergarten programs, and as a guide post for determining admission to law and medical schools across the country. The jargon of testing is quickly identified by those in academia: CAT, SAT, ACT, PSAT, PACT, GMAT, LSAT, MMPI, ITBS, etc.


**The Reign of ETS**

The Ralph Nader Report (1980) on the Educational Testing Service (ETS) entitled *The Reign of ETS*, looked closely at "... a process of evaluation where the educational and career opportunities of millions of people are significantly determined by multiple-choice examinations, which do not even purport to test their judgment, wisdom, experience, creativity, idealism, determination or stamina" (p. xiv). *The Reign of ETS* attacked the $94 million dollar annual income company, which is classified as a tax exempt and non-profit organization, as controlling the lives of millions of adolescents and adults. This mega-empire begins testing before a student is five with its Cooperative Preschool Inventory, and extends its services with a myriad of tests through adulthood.
In its attack, the report found that the roll of a pair of dice could be as accurate as some of ETS's predictions for success in school. From a chart on page 65, a role of the dice was shown to be between 87 percent to 92 percent as accurate in predicative validity as ETS tests.

**Percentage of Predictions in which Random Prediction With a Pair of Dice is as Accurate as an ETS Test**

- SAT (college) 88%
- LSAT (law school) 87%
- GRE (graduate school) 89%
- GMAT (business school) 92%

The report estimates that over 100 million people were tested by ETS from 1948 through 1979 and recounts some of the horror stories of lives that were ruined and opportunities blocked by the results of one of the tests.

"The Lake Wobegone Report"

Two major exposes on testing were written in 1987 and 1989 by a West Virginia physician. Dr. John Jacobs Cannell's *Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above the National Average*, looked at the phenomenon of how students in all 50 states could score above average on nationally-normed achievement tests in which only 50 percent of the students should score above average and 50 percent below average. His study was later called "The Lake Wobegone Report" by the Associated Press after Garrison Keillor's fictitious Lake Wobegone area where "all the women are strong, all the men are good looking, and all the children are above average." Cannell's study was later duplicated and confirmed

Cannell and others have looked at the effect of "high stakes" testing, a phrase coined by Professor Jim Popham at UCLA (Cannell, 1989, p. 9). "High stakes" refers to significant consequences of testing which may include promotion from grade to grade, accountability to the public, college admissions, public comparison of schools and districts, and merit pay or salary raises for teachers. George Madaus, Director of The Center for the Study of Testing, Evaluation, and Educational Policy, stated the following in an interview in *Educational Leadership*.

> When the stakes are high, people are going to find ways to have test scores go up. That's true of any social indicator; a good example is airplane schedules. When the Federal Aviation Commission started to publish arrival times of airlines - who was late and who wasn't - within six weeks the airlines added a half hour to each schedule. I used to fly into Washington in 60 minutes, now it's scheduled to be an hour-and-a-half flight. It's the same with test scores. If it's important enough, people are going to find ways to get kids over the hurdle of the tests. The school will look better, but the skill levels will not necessarily be going up. (Brandt, 1989, p. 27)

The Report of the National Commission on Testing and Public Policy

The National Commission on Testing and Public Policy was founded in 1987 to take at the role of testing in the United States, and to investigate and research not only standardized testing but testing in the classroom, workplace, and armed forces. The Commission's 1990 report *From Gatekeeper to*
Gateway: *transforming testing in America* published five indictments of the testing program in the United States.

1. Tests are imperfect and therefore potentially misleading as measures of individual performance in education and employment.

2. Some test uses result in unfair treatment of individuals and groups.

3. Students are subjected to too much testing in the nation's schools.

4. Some testing practices in both education and employment undermine important social policies and institutions intended to develop or utilize human talent.

5. Tests have become instruments of public policy without sufficient public accountability. (p. 6)

**Effects on Curriculum**

When the stakes are high, the consequences of using objective, one-right-answer testing have been found to adversely affect the curriculum. David Moody (1991) in a Policy Brief for the Far West Laboratory for Educational Research and Development stated:

> In particular, a multiple-choice, "fill in the bubble" type of examination may lead to Trivial Pursuit-type instruction that produces students who can memorize well but are rarely challenged to exercise "higher-order" thinking skills: to think critically and deeply; to apply knowledge in novel situations; to integrate many discrete pieces of information; and to collaborate with others in the solution of complex problems. Combine these two factors - high-stakes evaluative assessments that end up driving instruction and a testing instrument that reflects a narrow subset of legitimate learning objectives - and instructional quality is likely to suffer seriously. (pp. 1-2)
The National Commission on Testing and Public Policy (1990) found that testing takes up about 20 million school days that could be spent on instruction and costs between $700 and $900 million annually in expenditures (p. 14).

Educators addressing the curriculum in our schools have pointed to the fact that much of the content is fragmented, memorization-type material that will assist students in doing well on standardized tests, but does not really provide any type of in-depth understanding of concepts studied (Boyer, 1983; Goodlad, 1984; Sizer, 1984; and Gardner, 1991). Linda Darling-Hammond (1990) writing in *Phi Delta Kappan* summed up the testing and curriculum issue by stating:

Another crucial factor in the level of performance of U.S. students is the role of standardized achievement tests in American schools. In recent years, officials of the National Science Foundation, the National Council of Teachers of English, the National Council of Teachers of Mathematics, and the National Assessment of Educational Progress have attributed the steady decline in students' analytical and problem-solving abilities to the tight coupling of basic skills testing with teaching in American schools. They charge that the back-to-basics movement, with its emphasis on teaching what is tested on standardized achievement tests, has brought about the neglect of higher-order skills and performance abilities. (p. 289)

**Reform Movements**

The high public profile concerning the reform of education began in earnest in 1983 with the publication of the National Commission on Excellence in Education's "A Nation At Risk." Since that time, everyone has jumped on the reform bandwagon. ACT and SAT scores have been criticized as declining; Diane Ravitch published the much talked about *What Do Our 17 Year-Olds Know?*, and President Bush and the National Governors' Association offered grants to schools and projects ready to translate the "America 2000" goals into realities. *Educational Leadership*’s October issues in 1991 and 1993 were devoted to the topic of standards, and other journals as well as professional
organizations' conference meetings have addressed restructuring, redesigning, and revamping our schools.

One of the leading themes in the reform movements has been the need to examine and restructure testing practices. Testing, in the reform literature, has been renamed "authentic assessment." Papers and reports setting standards and addressing assessment issues have been published by national, state, and professional organizations.

National Reforms

The National Center on Education and the Economy along with the University of Pittsburgh's Learning Research and Development Center are working under a $2.5 million grant to develop assessments. Unlike traditional paper and pencil tests, these assessments would include performance-based activities, portfolios and projects (O'Neil, 1991, p. 5).

The New Standards Project (NSP) directed by Lauren Resnick and Warren Simmons, plans to develop over a three year period, assessments in math, English, and science for grades 4, 8, and 10. NSP involves 17 states and six large school districts that serve fifty percent of the school aged children in the United States. The assessments promise "... portfolios that will contain a combination of on-demand and curriculum-embedded assessments; performance-based matrix exam tasks, projects, exhibitions; and work selected by districts, schools, teachers, and students (O'Neil, 1991, p. 18). Differing from standardized tests, the NSP envisions including teachers in developing assessment strategies, designing scoring rubrics, and improving curriculum and instruction from assessment results.

The report of the National Commission on Testing and Public Policy issued in 1990 offered eight specific recommendations for improving testing in our schools.
1. Testing policies and practices must be reoriented to promote the development of all human talent.

2. Testing programs should be redirected from over reliance on multiple-choice tests toward alternative forms of assessment.

3. Test scores should be used only when they differentiate on the basis of characteristics relevant to the opportunities being allocated.

4. The more test scores disproportionately deny opportunities to minorities, the greater the need to show that the tests measure characteristics relevant to the opportunities being allocated.

5. Test scores are imperfect measures and should not be used alone to make important decisions about individuals, groups, or institutions; in the allocation of opportunities, individuals' past performance and relevant experience must be considered.

6. More efficient and effective assessment strategies are needed to hold institutions accountable.

7. The enterprise of testing must be subjected to greater public accountability.

8. Research and development programs must be expanded to create assessments that promote the development of the talents of all our peoples. (pp. x-xi)

One of the latest documents, Learner-centered psychological principles: Guidelines for school redesign and reform, published in January 1993, was produced by the Presidential Task Force on Psychology in Education and the American Psychological Association and supported by the Mid-continent Regional Educational Laboratory (McREL). Twelve learner-centered principles are set forth with the hopes that these principles will guide schools in restructuring for the 21st Century. Based on the principles, were eighteen points regarding assessment. Prominent among the eighteen points was the need that assessments "... should be based on authentic and meaningful tasks that are aligned with the regular curriculum... should include exhibits, portfolios, and
performances to demonstrate achievement, . . . and [should] provide for multiple plausible responses and growth in understanding through errors" (p. 15-16).

In addition to these national efforts, other centers for research and study have been engaged in looking at testing reform. These include Research for Better Schools (Philadelphia), Council for Basic Education (Washington, D.C.), Center for Research, Evaluation, Standards and Student Testing (University of California at Los Angeles), the Urban District Leadership Consortium (Washington, D.C.), and the National Center for Fair and Open Testing (FairTest, Cambridge, Massachusetts). These groups advocate performance-based and authentic testing because they believe it:

1. leads to improved curriculum and teaching since its use will spur instruction in higher order skills;

2. eliminates the narrowing of the curriculum found with standardized test use;

3. responds positively to "teaching to the test" since teaching to the performance-based test will result in better instruction and curriculum;

4. results in a greater understanding of the student's abilities than is provided by the standardized type of test;

5. alleviates the boredom of classroom instruction keyed to standardized test driven instruction and replaces "multiple-choice teaching" with a "thinking curriculum";

6. provides teachers with professional growth and gives them the opportunity to see the effects of their instruction in the performances of their students;

7. contributes to the advancement of teacher empowerment (greater decision-making authority) by expanding the participation of teachers in the development of performance assessment programs and by providing teachers with an active role in the scoring process for such alternative assessments; and
8. holds promise for use in the restructuring process of urban center education systems since alternative testing results in a greater understanding of the students' abilities than is provided by the standardized test. (New York State Education Department, 1991, pp. 12-13)

State Reforms

State testing practices were addressed in the 1990 Gateway study which found the following:

The Commission finds that testing primary school children for entry to or exit from a grade is poor education practice. Nonetheless, prekindergarten tests are mandated in more than 16 states, widely used in seven states, and known to be used at the district level in more than 37 states. Kindergarten exit/first grade entrance tests are used in at least five states and known to exist at the district level in an additional 37. In some school districts as many as 60 percent of the kindergartners are judged to be "unready" for first grade because of their scores on 'readiness' tests. Achievement testing is required for first graders in nine states; for second graders in nine states; and for third graders in 27 states. (National Commission on Testing and Public Policy, pp. 14-15)

The New York State Education Department (1991) in "Student Assessment: A Review of Current Practices and Trends in the United States and Selected Countries" indicated that two major surveys on performance testing have been compiled to give an indication of the states' involvement in reforming testing. The studies were conducted by Pamela Aschbacher of UCLA's Center for Research, Evaluation, Standard, and Student Testing (CRESST) in the spring of 1990, and the Council of Chief State School Officers (CCSSO) in December of 1990. Their findings indicated that:

1. 34 states have some type of performance-based testing implemented;
2. 25 of the 34 states that have some form of performance-based assessment practices are planning/developing or discussing expansion of alternative assessment;

3. 9 of the 16 states without existing performance-based assessment practices are planning/developing programs, and 2 other states are discussing alternative assessment; and

4. 5 states have expressed no interest in moving toward alternative assessment activity. (New York State Department of Education, p. 16)

Professional Organization Reforms

Professional teacher organizations have also joined the call for reform in the testing movement. Some have completed standards documents which emphasize the development of hands-on types of assessment (science), portfolio assessments (English), and projects and performances, while other organizations are in the process of developing their documents. John O'Neil (1991) in "Can National Standards Make A Difference" in Educational Leadership delineated the efforts of professional organizations involved in reform. They are as follows:

- **Social Studies**
  - National Center for History in the Schools at the University of California-Los Angeles
  - The National Council of Geographic Education
  - The Center for Civic Education
  - National Council for the Social Studies
  - National Task Force for Social Studies Standards

- **English**
  - The National Council of Teachers of English
  - The International Reading Association
  - Center for the Study of Reading at the University of Illinois at Urbana-Champaign

- **Mathematics**
  - National Council of Teachers of Mathematics

- **Science**
  - National Science Education Standards
The term "authentic assessment" has been used in many reports and research projects. Used almost interchangeably have been other phrases such as "performance-based assessment" and "alternative assessment." While some researchers use the terms interchangeably to mean anything other than multiple-choice testing, they are not truly synonymous. "Alternative assessment" is generally used to indicate testing situations other than multiple-choice, one-right answer questions; this may include "authentic" and "performance-based" assessment. However, "authentic" and "performance-based" assessment may differ radically even though they are often used synonymously.

Definition and Characteristics of Authentic Assessment

"Performance-based assessment" is thought to differ from "authentic assessment" in a major way. While both types can include a performance or the production of a product, "authentic assessment" has its basis or philosophy grounded in what is expected in the "real" world. For example, a student in home economics may be required to hem a length of material to demonstrate proper hem stitching (performance), but she is not producing something that is valued in the real world. To sew a skirt, which would also require that hem stitching be demonstrated would be more authentic. Similarly, a student in English may write to an author asking for clarification on something he read in a
novel. If the letter is graded only for its style and grammar, it qualifies as a performance-based assessment, but not as an authentic assessment. The letter would actually need to be sent to the author, and the reply used as a learning experience relating to reading comprehension to constitute an authentic learning experience.

Grant Wiggins (1990) who has written extensively on authentic assessment contrasts, in the following chart, authentic assessment with traditional assessments by enumerating student behaviors, task orientation, and reliability and validity measures.

<table>
<thead>
<tr>
<th>Traditional Assessment</th>
<th>Authentic Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>reveals whether students can recognize, recall or &quot;plug in&quot; what was learned out of content</td>
<td>requires students to be effective performers with acquired knowledge</td>
</tr>
<tr>
<td>are limited to paper-and-pencil, one answer questions</td>
<td>presents the student with the full array of tasks that mirror the priorities and challenges found in the best instructional activities: conducting research; writing, revising, and discussing papers; providing an engaging oral analysis of a political event; collaborating with others on a debate, etc.</td>
</tr>
<tr>
<td>asks students to select or write correct responses - irrespective of reasons</td>
<td>attends to whether the student can craft polished, thorough and justifiable answers, performances or products</td>
</tr>
<tr>
<td>standardizes objective items to one right answer for each question</td>
<td>achieves validity and reliability by emphasizing and standardizing the appropriate criteria for scoring products</td>
</tr>
<tr>
<td>are like drills, assessing static and arbitrarily discrete or simplistic elements</td>
<td>involves ill-structured challenges and roles that help students rehearse for the complex ambiguities of the game of adult and professional life (p. 3)</td>
</tr>
</tbody>
</table>
Wiggins (1991), along with Arthur Costa (1993), Joe M. Steele (1992), and Fred M. Newmann and Gary C. Wehlage (1993) have characterized authentic assessment in terms of its philosophy, scope of tasks, and grading practices. The following summarizes their work:

On philosophy:

From Wiggins:
- involves tasks we value, and at which we want students to excel - tasks worth learning and "teaching to."
- simulates the challenges facing adults or workers in a field of study, or the real-life "tests" of civic and personal life in which our educational knowledge is required.
- is composed of "ill-structured" challenges that require (a) problem clarification and knowledge in use, (b) effective use of a repertoire of knowledge, (c) good judgment in solving the problem, and (d) overcoming realistic constraints to fashion an effective and appropriate response in context.
- involves de-mystified and non-secret tasks, criteria and standards; allows for thorough preparation and accurate self-assessment by the student.

From Steele:
- utilizes real life settings where possible or simulates them as realistically as possible.
- requires an application of skills to the kinds of problems found outside the classroom.
- addresses complex, ill-defined tasks that demand application of higher order thinking skills.

From Newmann and Wehlage:
- involves constructing meaning and producing knowledge.
- uses disciplined inquiry to construct meaning.
- has value and meaning beyond the instructional context.

On scope of tasks:

From Wiggins:
- focuses on the students' ability to produce a quality product and/or performance.
- calls upon different forms of communicating and means of
displaying mastery - in an integrative "performance" or set of products, e.g., an oral report, supported by a paper.

From Costa:
- involves collecting logs, journals, and portfolios of selected artifacts of learning excellence.
- involves maintaining checklists recording indicators of growth toward desirable habits of mind.
- employs media and advanced technology to assist in collecting and recording information.

From Newmann and Wehlage:
- requires students to manipulate information and ideas in ways that transform their meaning and implications, such as when students combine facts and ideas in order to synthesize, generalize, explain, hypothesize, or arrive at some conclusion or interpretation.

On grading:

From Wiggins:
- relies on trained assessor judgment, in reference to clear and appropriate criteria (as opposed to those most easily observed or scored).
- is typically composed of interactions between assessor and student. Focuses on the student's ability to justify answers and respond to follow-up or probing questions.
- involves patterns of response and behavior, consistency of performance: emphasis is on consistency of quality, habits of mind.

From Steele:
- assures measures have been validated and standards identified.
- plans for the measures themselves to be learning experiences.

From Costa:
- directly observes performance in collaborative problem-solving situations.
- observes performances while conducting extended cooperative projects.
- conducts interviews to discover students' self perceptions as problem solvers.
- assesses displays, exhibitions and performances according to both internal and external criteria.

From Newmann and Wehlage:
- assesses students' depth of knowledge and understanding. (Wiggins, 1991, Toward One System of Education, p. 25; Costa, 1993, pp. 50-51; Steele, 1992, p. 1; and Newmann and Wehlage, 1993, pp. 8-12)

Archbald and Newman (1988) have also characterized authentic assessment as having "aesthetic" or "utilitarian" meaning outside of the school context. They believe:

Authentic demonstrations of mastery often share three features uncommon in most school testing situations: the production of discourse, things, or performances; flexible use of time; and collaboration with others.

1. **Production of discourse, things, performances.** Beyond school we demonstrate knowledge by providing original conversation and writing, by repairing and building physical objects, and by producing artistic, musical, and athletic performances.

   In contrast, assessment in school usually asks students to identify the discourse, things, and performances that others have produced (for example, by recognizing the difference between verbs and nouns, between socialism and capitalism; by matching authors with their works; by correctly labeling rocks and body parts).

2. **Flexible use of time.** The significant achievements of disciplined inquiry often cannot be produced within rigidly specified time periods. Adults working to solve complicated problems, to compose effective discourse, or to design products are rarely forced to work within the rigid time constrains imposed on students such as the 50-minute class or the two-hour examination period.
Standard, predetermined time schedules based on bureaucratic procedures for managing masses of students and diverse course offerings, rather than on the time requirements of disciplined inquiry, can reduce the authenticity of student achievement.

3. **Collaboration.** Achievements outside school often depend on the opportunity to ask questions of, to receive feedback from, and to count on the help of others, including peers and authorities. In contrast, typical assessment of school achievement focuses primarily on what the student can accomplish while working alone. Assessment tasks that deny opportunities to cooperate can thereby diminish the authenticity of the achievement. (pp. 3-4)

**Pilot Projects**

While "authentic assessment" may be a relatively new term, the practice is very old. Socrates engaged in discourse with his students, Dewey's philosophy was based on a learner-centered curriculum which involved authentic assessment, and Eliot Wigginton enabled his students to learn grammar and writing skills through the publication of *Foxfire*. In addition teachers in fine, performing, and applied arts have for years based students' grades on authentic measures such as producing an oil painting, playing a musical composition, tuning-up an engine, or baking a cake. The ground-breaking use of authentic assessment is gaining entry into the core fields such as English, science, math, social studies, and foreign language where multiple-choice testing has prevailed.

Efforts to institute authentic assessment have cropped up in schools where teachers have learned about the concept in graduate courses or workshops, but the most ambitious, cross-curricular efforts have been in pilot projects usually in cooperation with universities and/or funded through research grants.

One of the first pilot projects was Project Spectrum, part of Harvard University's Project Zero, which was instituted in 1984 and is founded on
Gardner's (Harvard) work in Multiple Intelligences theory and David Feldman's (Tufts) theory of development in non-universal domains. The project, located in schools in Massachusetts, is funded by grants from the William T. Grant Foundation, the Rockefeller Brothers Fund, and the Spencer Foundation.

"Spectrum is based on the assumption that every child has the potential to develop strength in one or several content areas and that it is the responsibility of the educational system to discover and nurture these proclivities. Rather than building around a test, the Spectrum approach is centered on a wide range of rich activities; assessment comes about as part-and-parcel of the child's involvement over time in these activities" (Kreschevsky, 1991, p. 42). Features of the program include the following.

1. Blurring the line between curriculum and assessment.
2. Embedding assessment in meaningful, real-world activities.
3. Using measures that are "intelligence-fair."
5. Attending to the stylistic dimensions of performance.
   (Kreschevsky, 1991, pp. 45-46)

Gardner is also one of the partners in the ATLAS Communities Project which involves schools in Gorham, Maine; Norfolk, Virginia; and Prince George's County, Maryland. ATLAS which stands for Authentic Teaching, Learning, and Assessment for All Students blends "... the reform philosophies of four of education's gurus: ... Theodore Sizer of the Coalition of Essential Schools, the Education Development Center's Janet Whitla, Howard Gardner from Harvard University's Project Zero, and James P. Comer from the School Development Project" (Viadero, 1995, p. 26). ATLAS was funded by a $2.5 million grant from NASDC and is based on five design principles.

Authentic teaching and learning is driven by questions; focuses on habits and understanding; and involves challenging, purposeful, and sustained work.
Ongoing cycles of planning, action, and reflection characterize effective teaching, learning, assessment, and organizational change.

Relationships matter because learning is a social activity.

Shared leadership, commitment, and communication build a collaborative culture of learning.

Members of ATLAS schools and pathways see themselves as part of broader, more integrated learning communities. (Viadero, 1995, p. 29)

Another pilot project is the University of Chicago School Mathematics Project which was founded in 1983, and is partially funded by the National Science Foundation. One of its main goals is to use real-world applications in the study of mathematics, and to determine if the National Council of Teachers of Mathematics Standards can be implemented in classrooms. The project is concerned with developing curriculum units and teacher materials for grades kindergarten through twelve (Usiskin, 1993, p. 14).

The Center for Problem-Based Learning (PBL) is housed at the Illinois Math and Science Academy (IMSA) in Aurora, Illinois, and helps teachers there develop units based on PBL philosophy. Teachers and researchers work on developing entire courses along the PBL theory or on developing "post-holes" - short problems can be used for a few lessons or a few weeks. "Through problem-based learning, students learn how to use an iterative process of assessing what they know, identifying what they need to know, gathering information, and collaborating on the evaluation of hypotheses in light of the data they have collected" (Stepien and Gallaghery, 1993, p. 25).

Other sites across the country are also experimenting with authentic learning and assessment. Some of these unique programs include the Passages Program at Jefferson County Open High School (Evergreen,
Colorado), the Rite of Passage at Walden III High School (Racine, Wisconsin), Learning Unlimited at North Central High School (Indianapolis, Indiana) and the Self Assessment System at Alverno College (Milwaukee, Wisconsin).

Research Findings

While there is much experimentation regarding authentic assessment, much of the literature is devoted to descriptions of the projects and ways of implementing the particular programs. Data on the effectiveness of authentic assessment has shown mixed results.

The ATLAS Project has RAND Corporation anthropologist Donna Muncey chronicling the events at the five project schools in Prince George's County, Maryland. According to Muncey, changes have been made, but "it's been slow going" (Viadero, 1995, p. 31). Muncey and Theodore Sizer estimate it will take up to ten years for a school to become a true ATLAS school. Meanwhile, results in the Maryland schools have been promising:

[Adelphi Elementary School] last month learned that the test scores of its Title I students had improved dramatically for the first time in years, making it likely that the school will get off the state's endangered list. . . . The P.T.A., once practically nonexistent, now has more than 100 members.

At High Point High School . . . not a single teacher at the school asked to transfer this year. Even more impressive, more than 100 asked to transfer into the school this spring.

On ATLAS's 'school climate' surveys, which reflect everything from teacher morale to whether students feel cared for, all three pathway schools improved in some areas this school year. (Viadero, 1995, pp. 29-31)

Positive results have also been noted in the "Work Sampling System" which is being used in 3,000 classrooms in the United States in pre-school
through fifth grade. Generally, report cards and traditional tests are eliminated and "... grade-level guidelines, checklists, portfolios, and summary reports [are used] to measure children's progress" (Miller, 1995, p. 8). Research conducted on the Work Sampling System has indicated a high reliability in terms of students' achievement.

In a test involving 100 kindergartners, the system proved to be an accurate predictor of performance on norm-referenced tests, even when researchers controlled for the potential effects of gender, age, and initial ability. (Miller, 1995, p. 9)

However, authentic assessment has also come under fire by expert panels examining the use of authentic assessments at the state testing level. Kentucky has been using KIRIS (Kentucky Instructional Results Information System) "... a battery of essays, physical tasks, and collected classwork intended to better represent student accomplishment and improvement" (Harp, p. 12). The KIRIS battery does not contain any multiple-choice questions, and is considered to be "... at the heart of Kentucky's 1990 education-reform act, which replaced the state's entire education system" (Harp, p. 12). In 1994 the legislature commissioned a panel to examine the KIRIS battery, and the results were not positive.

... members of the review panel concluded that performance standards used to gauge test scores are too narrow and unreliable, scoring of portfolios is too subjective and inconsistent, efforts to equate assessments from one year to the next are problematic, and student gains on the tests do not match changes in performance on other standardized tests. (Harp, p. 13)

Lynn Olson (1995) in Education Week analyzed some of the state programs and found additional problems.
Arizona, for instance, has a set of 'essential skills' that school districts are supposed to teach. Districts administer a test known as Form A that demonstrates whether students have mastered the essential skills in reading, writing, and math. In grades 3, 8, and 12, a sample of the state's students takes Form D, a statewide test that was presumed to measure the same skills in a more integrated fashion. But a recent study found almost no correlation between the two assessments, suggesting that they are measuring different things. (Olson, p. 11).

In Vermont, the lack of consistency in implementing and scoring portfolios has resulted in not issuing results at the school level, and in California similar problems with scoring "... led to inaccurate results for a number of schools" (Olson, p. 11). According to Edward H. Haertel, a professor of education at Stanford University, "I think we're already seeing some signs of a retreat from large-scale performance assessments on the part of states" (Olson, p. 11).

Although results on the effectiveness of authentic assessment has started to come into the literature, there is a lack of research on how authentic assessment has influenced teachers. Most studies describe the types of assessments used and the grade levels at which they are implemented. How teachers make the shift to using authentic assessments has not been examined.

Implications for School Administrators

Where does all this leave the teacher? Over fifty percent of the teaching work force in elementary and secondary schools are over the age of forty. Roughly 1,169,000 teachers received their basic training in the late 1960s and early 1970s when assessment courses were still heavily influenced by Thorndike's measurement work (U.S. Bureau of the Census, 1992, p. 150). Although the majority have probably returned to school for advanced degrees and kept abreast of new innovations through inservice programs, the research
indicates that not much has changed in terms of assessment practices in the classroom.

Robert J. Wilson of Queen's University in Canada is involved in a study of over 100 teachers and their testing practices. In other research with colleagues he is also looking at policies and procedures that govern teachers in assessing students' progress. In asking student teachers and practicing teachers to rank reasons for evaluating students, he found a disparity among the two groups. While student teachers were more concerned with students' progress, practicing teachers were more concerned with generating marks (administrative and external aims) which were marked lowest by the student teachers. Wilson feels that the student teachers will change their views once they are placed in the actual teaching situation. He hypothesizes "... that the policies and procedures concerning student achievement devolved upon teachers from levels 'above' them in the administrative hierarchy will force their evaluation activities into relatively narrow areas" (Wilson, 1990, p. 7). The ranking for different reasons for assessing students that Wilson found for secondary practicing teachers and intermediate senior (grade 9 through Ontario Academic Course or grade 13) student teachers are summarized below.

<table>
<thead>
<tr>
<th>Reasons</th>
<th>Practicing Teachers</th>
<th>Student Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>To check students' progress against course objectives</td>
<td>2</td>
<td>1 (tie)</td>
</tr>
<tr>
<td>To compare students' achievement to others</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>To generate marks for reporting purposes</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>To ensure that students do assigned work</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>
To prepare students for this kind of evaluation in the future 5 8
To have students practice or apply what has been learned 3 5
To diagnose students' weaknesses with the material 8 3
To enable students to monitor their own progress 7 1 (tie)
To help me decide what to teach next 10 6
To allow me to see how well I taught the material (Wilson, p. 4) 9 4

Wilson is also engaged in research concerning the types of assessments that teachers use. He feels that the timelines for reporting students' grades and the tightly woven symbol system (letter grades) leads teachers to use assessment items which suit these administrative and bureaucratic needs rather than to improve their teaching and students' learning. A natural consequence of this practice is that the curriculum is affected. In collecting teachers' tests, Wilson says:

... it is clear that the cognitive demand of single-word completions and short-answer items (questions which appeared on 44% of the instruments we collected) is not likely to be high if for no other reason than that the format does not allow any higher level than recall of specific bits. (Wilson, p. 8)

Bikkar S. Randhawa (1990) in a paper presented at the 2nd Conference on Classroom Testing in Canada reviewed some of the major research of the 1980s involving teachers' extent of testing, knowledge in testing, and problems with classroom testing. From the studies of Carlberg (1981), Newman and
Stallings (1982), Gullickson (1982), Fennessey (1982), Yeh (1980), Green and Stager (1886-87), and Ebel (1980), the following points are summarized from Randhawa’s paper:

On extent of classroom testing:
- 10 to 25% of instructional time is spent on the assessment of student progress and on diagnostic information gathering;
- 95% of teachers test at least biweekly;
- 40 to 50% of course grades of students are dependent on test scores.

On knowledge of testing:
- teachers' knowledge of testing techniques and their skills in classroom testing practices was less than adequate;
- conflicting research exists on the relationship between teachers' knowledge of measurement and evaluation and the purposes for which they test and the number of test item types they employ in testing.

On problems in classroom testing:
- teachers tend to rely primarily on their own subjective, but presumably absolute, standards in evaluating achievement;
- teachers tend to put off test preparation to the last minute, then they do it on a "catch-as-catch-can" basis;
- many teachers administer tests that are too poorly planned, too short, or too inefficient in form to sample adequately the intended content and abilities in the subject;
- teachers often put too much emphasis on trivial or unnecessary details in their tests but neglect to include basic principles, understandings, and applications of the subject;
- teachers often write test questions, both essay and objective, whose effectiveness is reduced by ambiguity or by irrelevant clues to the correct answer;
- many teachers underestimate or overlook the influence of sampling errors on test scores;
- most, if not all, teachers fail to examine the effectiveness of their tests by even a simple statistical analysis of the items or the results of their tests. (pp. 39-53)
The emphasis on the need for teachers to shift their testing practices has been addressed by three prominent groups - the National Education Association (NEA), the American Federation of Teachers (AFT), and the National Council on Measurement in Education (NCME) - who jointly published their 1990 Standards for Teacher Competence in Educational Assessment of Students. Ronald K. Hambleton (1990) of the University of Massachusetts at Amherst outlined the purposes of the document and the competencies addressed by these groups

Purposes:

1. a guide for teacher educators who design and/or approve teacher education programs,
2. a basis for teachers conducting a self-evaluation of their educational testing skills,
3. a guide for the design of testing workshops for teachers,
4. a directive to educational measurement specialists and teacher trainers to broaden their conception of student assessment and convey this broader conception in their research, writing, and teaching.

Competencies:

1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2. Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
3. Teachers should be skilled in administering, scoring and interpreting the results of both externally produced and teacher-produced assessment methods.
4. Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, and developing curriculum and school improvement.
5. Teachers should be skilled in developing valued pupil grading procedures which use pupil assessments.
6. Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
7. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information. (Hambleton, pp. 94-95)
Richard J. Stiggins (1990) of the Northwest Regional Educational Laboratory (NWREL) conducted a study on teacher training in the Pacific Northwest states including Alaska, Hawaii, Idaho, Montana, Oregon, and Washington and:

... found that less than half of the largest undergraduate and graduate teacher training programs in our six-state region offers the option of assessment training to their students. Further, less than a quarter of these programs require the successful completion of this course by their students. (p.97)

Stiggins feels classroom testing is inadequate because of the lack of training teachers receive.

We have known for decades that teachers and administrators alike are inadequately trained in assessment. Yet despite research-based reminders of this fact about once every ten years for the last 50 years, nothing has changed. . . . [this] suggests that there are purposeful forces at work within and outside of the education community to prevent assessment training from becoming part of the professional preparation of educators . . . one of the primary causes of the absence of assessment training in the teacher training curriculum has been the chronic and deep-seated mismatch between what teachers need to know about assessment and the content of assessment courses when they are offered. (Stiggins, 1990, p. 97)

Generally, it appears that most teachers view testing as a means to assigning letter grades, and have a limited amount of knowledge in designing good assessment measures. How then can they be expected to shift paradigms and embrace the authentic assessment movement, especially if they view this as an unwelcome and irrelevant burden.

While those teachers at pilot projects have resources both in terms of top-name educators and researchers guiding their efforts and in terms of monetary support, the majority of classroom teachers are not that fortunate. Most teachers
are not in touch with research unless they are involved in graduate studies or involved in one of the university-connected pilot projects. As Ornstein (1989) pointed out, teachers "... have little motivation for reading the research, lack research knowledge and are unable to understand the data, or feel that research is not relevant to the practice of teaching" (p. 95).

There are teachers, however, using authentic approaches in the classroom. For them the shift to the new paradigm has been described as the "Aha!" experience that all learners share when insight is achieved; others have described their philosophical shift as a "born again" experience. The characteristics and experiences these teachers share in terms of philosophy, education, and support for their endeavors can have important implications for school administrators who are endeavoring to enhance teachers' knowledge and encourage teachers' use of authentic assessment. This study investigated the identification of teachers as authentic assessors using an operational definition based on the literature, the attributes and characteristics of authentic assessors as compared to traditional assessors, and the conditions of the schools as contributing factors in enabling teachers to become authentic assessors.
CHAPTER 3
METHODOLOGY

The purpose of this study was to identify characteristics of teachers who had become practitioners of authentic assessment and delineate support systems or conditions which enabled them to shift their paradigm from the more traditional role of classroom assessor. The first step in the procedure was to review the literature and develop a definition of authentic assessment. From the review of the literature in Chapter 2, a questionnaire was developed which was used to identify teachers as authentic assessors.

This chapter is organized into three sections: the sampling procedure, the instrumentation, and the research questions and the treatment of data.

Sampling Procedure

Three suburban high schools in Lake County, Illinois were selected based on similar enrollment figures and on varying financial indicators. The three schools were located in different socio-economic communities within the county and were supported by local revenue by varying degrees. The schools also had student populations that ranged from largely white to predominately minority. School one was located in a blue collar community and served a student population that was largely composed of minorities with 68.2% African Americans and 11.4% Hispanics. Local revenues supported only 32% of the budget, and 30.7% of the students were classified as coming from low income families. School two was located in a community that was a combination of blue and white collar. The school population was 75.5% white, but also had a
significant Hispanic population of 19.6%. Low income students comprised 12.1% of the student population, and local revenue supported 90.6% of the budget. School three was located in a predominately white collar or professional community. This school had the largest percentage of white students at 93.1%, and there were a small percentage of minority students. The school had no low income students, and secured 97.5% of its revenue from local sources. The descriptive information was obtained from the 1992 Ed.dat Databook - Volume II: Education Finance and the 1993 Ed.dat Databook - Volume 1: School Report Card Data. Tables 2 and 3 provide information on revenue sources and student population.

**TABLE 2**

SCHOOLS' REVENUE SOURCES

<table>
<thead>
<tr>
<th>School</th>
<th>Total Revenue Per Pupil</th>
<th>Local (%)</th>
<th>State (%)</th>
<th>Federal (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Revenue Per Pupil</td>
<td>Revenue Per Pupil</td>
<td>Revenue Per Pupil</td>
<td>Revenue Per Pupil</td>
</tr>
<tr>
<td>School 1</td>
<td>6,453</td>
<td>2,062 (32.0%)</td>
<td>2,749 (42.6%)</td>
<td>1,642 (25.4%)</td>
</tr>
<tr>
<td>School 2</td>
<td>10,349</td>
<td>9,381 (90.6%)</td>
<td>731 (07.1%)</td>
<td>237 (02.3%)</td>
</tr>
<tr>
<td>School 3</td>
<td>18,532</td>
<td>18,061 (97.5%)</td>
<td>422 (02.3%)</td>
<td>49 (00.3%)</td>
</tr>
</tbody>
</table>
TABLE 3
STUDENT ENROLLMENT, ETHNICITY, LIMITED ENGLISH PROFICIENT, AND LOW INCOME FIGURES

<table>
<thead>
<tr>
<th>Enroll ment</th>
<th>% White</th>
<th>% Afr. Amer.</th>
<th>% Hisp.</th>
<th>% Asian Amer.</th>
<th>% Nat'v Amer.</th>
<th>% LEP</th>
<th>% Low Inc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>School 1</td>
<td>849</td>
<td>17.9</td>
<td>68.2</td>
<td>11.4</td>
<td>1.4</td>
<td>1.1</td>
<td>2.7</td>
</tr>
<tr>
<td>School 2</td>
<td>1,133</td>
<td>75.5</td>
<td>1.3</td>
<td>19.6</td>
<td>3.5</td>
<td>0.1</td>
<td>5.3</td>
</tr>
<tr>
<td>School 3</td>
<td>1,093</td>
<td>93.1</td>
<td>1.2</td>
<td>1.2</td>
<td>4.5</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

A letter was sent to the superintendents of the three schools requesting the participation of their teachers in the study (Appendix G). Teachers in the core subject areas of English, social studies, foreign language, math, and science were asked to participate.

All three superintendents agreed to participation. The study was explained to the principals at the three schools during phone conversations during the time period between October 1 and December 1, 1994. It was explained that teachers' participation was voluntary, and that their responses would be reported anonymously with neither the teachers' identities nor the schools' names reported in the research findings. The questionnaire asked teachers to supply their names, addresses, and phone numbers for possible later contact for an in-depth interview. The number of teachers in the core subject areas at the three schools is reported in Table 4.
The teachers agreeing to take part in the study and completing the questionnaires were representative of the four content fields, and were distributed among those fields as outlined in Table 5.

**Instrumentation**

From a review of the literature and drawing upon previous studies by Robert J. Wilson and Richard Stiggins a five-part Classroom Testing Questionnaire was developed. The questionnaire identified (a) demographic and professional information, (b) reasons for testing, (c) assessment practices, (d) types of assessments used, and (e) level of use of different types of assessment (Appendix A).

### TABLE 4

TOTAL TEACHERS BY SUBJECT AREA AND SEX

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>41</td>
<td>25</td>
<td>27</td>
<td>26</td>
<td>22</td>
<td>75</td>
<td>66</td>
<td>141</td>
</tr>
</tbody>
</table>

### TABLE 5

PARTICIPATING TEACHERS BY SUBJECT AREA AND SEX

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>31</td>
<td>18</td>
<td>15</td>
<td>20</td>
<td>11</td>
<td>45</td>
<td>50</td>
<td>95</td>
</tr>
</tbody>
</table>
Permission was received from Robert J. Wilson and Richard Stiggins to use parts of questionnaires developed by them and used in previous studies (Appendices D and E).

An In-Depth Interview Questionnaire (Appendix B) was used with teachers who were identified as authentic assessors and those identified as traditional assessors. The questions were culled from a review of the literature on authentic assessment and from the document Standards for Teacher Competence in Educational Assessment of Students by the American Federation of Teachers, the National Council in Measurement in Education, and the National Education Association (Stiggins, 1994, pp. 457-460). The questionnaire focused on illuminating contrasting factors between the two groups of teachers including their knowledge of assessment practices and their background or preparation in assessment. The questionnaire included four sections which addressed: (a) personal factors, (b) professional development and involvement, (c) administrative/supervisory structure, and (d) student characteristics.

A Principal Interview Questionnaire (Appendix C) was used to determine if conditions at the schools or contributing factors enabled teachers to become practitioners of authentic assessment. Questions covered support systems common to schools such as staff development and inservice programs, financial assistance for professional development, and competencies in assessment from Standards for Principals in Educational Assessment (Stiggins, 1994, pp. 461-466). The Principal Interview Questionnaire was composed of ten sub-sections which covered the topics of: (a) district organization, (b) population and description of the school, (c) recent leadership history, (d) relations with the Board of Education, (e) vision and mission of the school, (f) recent and anticipated changes, (g) academic organization and average teaching
assignment, (h) focus on curriculum and inservice, (i) climate and teacher empowerment, and (j) the school's strengths and weaknesses or challenges.

Research Questions and Treatment of Data

Survey data was analyzed using the statistical package Microfsoft Excel (Version 2.2, copyright 1989).

The research questions and methods of analysis were as follows:

Research Question 1: Using an operational definition of authentic assessment, based on the literature, which teachers in three suburban high schools in Lake County, Illinois can be identified as practitioners of authentic assessment practices?

Method of Analysis: Two sections of the Classroom Testing Questionnaire were used to identify teachers as practitioners of authentic assessment. Sub-section C - Assessment Practices - used descriptors that represented a continuum from authentic to traditional on such testing characteristics as students' backgrounds, administration, scoring, and content of assessments. A score of "one" on each item was considered representative of authentic assessment. The individual scores on each item were transposed so a "one" became a "five," etc. A maximum score of "fifty" was considered the most representative of authentic assessment.

In sub-section D - Types of Assessments Used - teachers were asked to indicate the percent of use of six types of assessments. Two types of performance assessments were listed - structured and spontaneous - and these were considered most representative of authentic assessment.

Teachers were identified as practitioners of authentic assessment based on the sum of their scores on sub-section C and the sum of percentages allotted to the two types of performance assessments on sub-section D of the
questionnaire. Teachers identified as authentic assessors were those who scored 90 percent or more of the 150 points allotted to sub-sections C and D.

**Research Question 2:** Of the identified teachers, what are the attributes and characteristics which separate them from teachers using traditional assessment techniques (attributes included demographic and professional information, reasons for testing, assessment practices, types of assessments used, and level of use)?

**Method of Analysis:** The Classroom Testing Questionnaire was used to classify respondents into two categories based on their scores in the various sub-sections: authentic assessors and traditional assessors. Cut-off ranges of ten percent based on the total possible number of points in sub-sections C and D were used to classify teachers as authentic and traditional assessors. Using the information from sub-sections A through E of the questionnaire, the characteristics of teachers who were practitioners of authentic assessment was then contrasted with teachers who scored at the opposite end of the continuum and were identified as traditional assessors.

**Research Question 3:** What conditions of the school were contributing factors in enabling teachers to become practitioners of authentic assessment (factors included personal, professional development and involvement, administrative/supervisory structure, and student characteristics)?

**Method of Analysis:** An In-Depth Interview Questionnaire (Appendix B) was developed to verify categorization of the teachers as authentic and traditional assessors, and to determine if any factors in the school environment contributed to teachers' testing philosophies. Those teachers scoring at the high and low ends based on the initial questionnaire were contacted for an in-depth interview. Teachers' responses were recorded and later categorized on a tally sheet. Factors contributing to teachers' being able to shift their paradigm from traditional assessors to authentic practitioners were delineated.
A Principal Interview Questionnaire (Appendix C) was developed to gather background information to assist in the interpretation of the data. The Principal interviews covered the topics of district organization, population and description of the school, recent leadership history, relations with the Board of Education, vision and mission of the school, recent and anticipated changes, academic organization and average teaching assignment, focus on curriculum and inservice, climate and teacher empowerment, and the school's strengths and weaknesses or challenges.
CHAPTER 4
FINDINGS AND IMPLICATIONS

This chapter reviews the data obtained through the Classroom Testing Questionnaire, the In-Depth Interview Questionnaire, and the Principal Interview Questionnaire. The research questions addressed in this study were:

Research Question 1: Using an operational definition of authentic assessment, based on the literature, which teachers in three suburban high schools in Lake County, Illinois can be identified as practitioners of authentic assessment practices?

Research Question 2: Of the identified teachers, what are the attributes and characteristics which separate them from teachers using traditional assessment techniques (attributes included personal factors, assessment practices, reasons for testing, and professional development)?

Research Question 3: What conditions of the school were contributing factors in enabling teachers to become practitioners of authentic assessment (factors included financial indicators, administrative/supervisory structure, school characteristics, and school climate)?

Research Question 1

The purpose of research question 1 was to use an operational definition of authentic assessment to identify teachers who were practitioners of authentic assessment. Teachers in three high schools in Lake County, Illinois took part in the study. Table 6 outlines the total number of teachers in the selected subject areas and those participating in the survey by subjects taught and sex.
TABLE 6
TOTAL AND PARTICIPATING TEACHERS BY SUBJECT AREA AND SEX

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>41</td>
<td>25</td>
<td>27</td>
<td>26</td>
<td>22</td>
<td>75</td>
<td>66</td>
<td>141</td>
</tr>
<tr>
<td>Partici.</td>
<td>31</td>
<td>18</td>
<td>15</td>
<td>20</td>
<td>11</td>
<td>45</td>
<td>50</td>
<td>95</td>
</tr>
<tr>
<td>Percent</td>
<td>75.6</td>
<td>72.0</td>
<td>55.6</td>
<td>76.9</td>
<td>50.0</td>
<td>60.0</td>
<td>75.8</td>
<td>67.4</td>
</tr>
</tbody>
</table>

Ninety-five teachers completed the questionnaire. Some sub-sections of the questionnaire were completed by less than the total number of respondents, and these differences are noted when the results of those sub-sections are discussed.

Two sections of the questionnaire were used to identify teachers as practitioners of authentic assessment: sub-section D - Types of Assessments Used and sub-section C - Assessment Practices.

In sub-section D - Types of Assessments Used, teachers were asked to indicate the percent of time they allotted to six different types of assessments based on a total of 100% of the time that they allotted for classroom assessment. Definitions of each of the six types of assessments were given in the questionnaire. The six types of assessments and their definitions are outlined below.

\[
\begin{align*}
\text{____} \% & \quad \text{Paper and pencil tests} \\
\text{____} \% & \quad \text{Curriculum-embedded tests} \\
\text{____} \% & \quad \text{Standardized tests} \\
\text{____} \% & \quad \text{Oral questioning in the classroom} \\
\text{____} \% & \quad \text{Performance assessment - structured} \\
\text{____} \% & \quad \text{Performance assessment - spontaneous} \\
\text{100} \% & \\
\end{align*}
\]
Paper and pencil tests are those you develop for your own use in the classroom. This category includes all true-false, multiple-choice, matching, fill-in, and short-answer tests and quizzes which YOU DEVELOP to determine if students have mastered the material taught.

Curriculum-embedded tests or included in the instructional materials are those that may be found in the textbooks or workbooks you use. They may also be found in an instructor's guide or may take the form of questions at the end of chapters in the materials themselves.

Standardized tests are offered by test publishers, such as the Stanford Achievement Test, Comprehensive Test of Basic Skills, Metropolitan Achievement Test, and Iowa Test of Basic Skills. This category also covers state-wide or district-wide tests, including norm and criterion referenced tests.

Oral questioning in the classroom is the daily question-and-answer process used on a day-to-day basis during instruction to track whether individual students or the class as a group are learning the material.

Performance assessments are those assessments in which you, the teacher, observe students in the process of doing things (e.g., speaking or oral reading) or examine products created by students (e.g., writing sample or art project). Then, on the basis of your professional judgment, you judge or rate student performance.

Performance assessments take one of two forms. Some are STRUCTURED tests and include: (1) a clearly defined reason for assessment; (2) pre-planned exercises to elicit student responses; (3) a pre-specific response to be evaluated; and (4) carefully spelled out scoring procedures. SPONTANEOUS assessments can be much less structured. A spontaneous classroom event may provide a teacher with an informal opportunity to observe and evaluate a student's performance and to judge the student's proficiency.

Based on the definitions, the last two types labeled "performance assessments" were considered representative of authentic assessment. Scores
for structured and spontaneous assessments were totaled with a resulting range of scores between 0% and 100%. Table 7 indicates the amount of testing time in quartiles that teachers spent on authentic assessment.

TABLE 7

<table>
<thead>
<tr>
<th>PERCENT OF TIME SPENT ON AUTHENTIC ASSESSMENT</th>
<th>0-25%</th>
<th>26-50%</th>
<th>51-75%</th>
<th>76-100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers</td>
<td>31</td>
<td>36</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>N=95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Seven teachers indicated that they used performance assessments less than 10% of the time, and eleven teachers indicated that they used performance assessments more than 90% of the time. These teachers' scores were compared to their responses on sub-section C - Testing Practices in the questionnaire. Sub-section C listed ten statements as attributes of authentic assessment based on a review of the literature. The statements were listed with contrasting statements representative of traditional assessment. Respondents marked each statement with a number from one through five. A score of one was representative of the statement at the left side of the continuum which described authentic assessment, while a score of five was representative of the statement at the right side of the continuum which described traditional assessment. In recording the data, these scores were transposed so that a one became a five, etc., with the higher scores representing authentic assessment. The lowest possible score on this sub-section was a ten, and the highest possible score was a fifty. Teachers' results represented in quartiles are presented in Table 8.
### TABLE 8

**TEACHERS' SCORES ON ASSESSMENT PRACTICES**

<table>
<thead>
<tr>
<th></th>
<th>10-20</th>
<th>21-30</th>
<th>31-40</th>
<th>41-50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers</td>
<td>8</td>
<td>28</td>
<td>35</td>
<td>23</td>
</tr>
<tr>
<td>N=94</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In comparing the teachers' scores from both extremes from sub-section D - Types of Assessments Used with their scores from sub-section C - Assessment Practices, they were found to be consistent. Teachers who indicated that they did not use performance assessments also indicated that their assessments were traditional in nature. Similarly, teachers who used performance assessments more than 90% of the time described their assessments as authentic. The scores of the seven teachers identified as "traditional" and the eleven identified as "authentic" based on sub-sections C and D of the questionnaire are presented in Table 9.
TABLE 9

TEACHERS' SCORES ON TYPES OF ASSESSMENTS USED AND ASSESSMENT PRACTICES

<table>
<thead>
<tr>
<th></th>
<th>% of Time on Performance Assessments</th>
<th>Score on Assessment Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Assessors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher 1</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Teacher 5</td>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td>Teacher 6</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Teacher 7</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Authentic Assessors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher 1</td>
<td>90</td>
<td>42</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>90</td>
<td>48</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>90</td>
<td>38</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>90</td>
<td>46</td>
</tr>
<tr>
<td>Teacher 5</td>
<td>90</td>
<td>42</td>
</tr>
<tr>
<td>Teacher 6</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td>Teacher 7</td>
<td>90</td>
<td>44</td>
</tr>
<tr>
<td>Teacher 8</td>
<td>90</td>
<td>33</td>
</tr>
<tr>
<td>Teacher 9</td>
<td>95</td>
<td>44</td>
</tr>
<tr>
<td>Teacher 10</td>
<td>95</td>
<td>36</td>
</tr>
<tr>
<td>Teacher 11</td>
<td>100</td>
<td>50</td>
</tr>
</tbody>
</table>

The correlation coefficient, for all ninety-five teachers in the study, between scores on sub-section C - Assessment Practices and the total amount of time allotted to performance assessment from sub-section D - Types of Assessments Used was .68 using the Pearson r formula which was significant at the .001 level. The scattergram depicting that correlation is shown in Figure 1.
The eighteen teachers listed in Table 9 were identified as "traditional" assessors and "authentic" assessors based on their scores, and were used as the basis for the profiles in further discussions.

The profile of the authentic assessors was created using the responses of the eleven teachers listed in Table 9 and the information from sub-sections A through E on the Classroom Testing Questionnaire. Tables 11 through 14 and Figures 2 and 3 profile the teachers who were identified as authentic assessors using the data from the questionnaire from sub-section A- Demographic and Professional Information.
As depicted in Table 10, the majority of the teachers identified as authentic assessors, or nine of the eleven, taught English. Two teachers identified as authentic assessors taught social studies, and no teachers were identified as authentic assessors in the fields of foreign language, science, or math.

Table 11 shows that the authentic assessors were fairly evenly distributed in terms of teaching experience. Two teachers had less than five years of experience, one had six to ten years, two had eleven to fifteen years, three had
sixteen to twenty years, and three had twenty years or more. Four of the teachers were male, and seven female. The majority of the teachers, eight of the eleven, were in their forties. The majority of teachers, nine of the eleven, also held advanced degrees. Six teachers indicated they had a master's degree plus thirty hours of college work, and one teacher held a doctor's degree.

Table 12 indicates that the majority of teachers identified as authentic assessors had received formal training in testing in the past five years. Eight teachers indicated that they had taken a graduate course in testing in the past five years, and nine teachers indicated they had attended a workshop or conference in the past year that dealt with testing.

**TABLE 12**

**AUTHENTIC ASSESSORS AS IDENTIFIED BY LAST GRADUATE COLLEGE COURSE AND LAST WORKSHOP/CONFERENCE IN ASSESSMENT PROCEDURES**

<table>
<thead>
<tr>
<th>N</th>
<th>Last Formal Graduate College Course</th>
<th>N</th>
<th>Last Workshop or Conference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>never had one</td>
<td>1</td>
<td>never had one</td>
</tr>
<tr>
<td>5</td>
<td>in past year</td>
<td>9</td>
<td>in past year</td>
</tr>
<tr>
<td>3</td>
<td>1 to 5 years ago</td>
<td>0</td>
<td>1 to 5 years ago</td>
</tr>
<tr>
<td>0</td>
<td>6 to 10 years ago</td>
<td>1</td>
<td>6 to 10 years ago</td>
</tr>
<tr>
<td>1</td>
<td>10+ years ago</td>
<td>0</td>
<td>10+ years ago</td>
</tr>
</tbody>
</table>

Table 13 shows that professional collaboration was rated as the most important contribution to their testing knowledge by the teachers identified as authentic assessors. Trial and error in the classroom and professional reading were ranked as numbers two and three. Even though many of the teachers had advanced degrees and had taken graduate level courses in assessment during
the past few years, graduate testing courses ranked last as a contribution to their testing knowledge. None of the authentic assessors indicated that graduate testing courses had contributed significantly to their testing knowledge. Other college level courses, such as undergraduate testing and methods courses were also ranked low as contributing to testing knowledge. Only one teacher indicated that an undergraduate testing course contributed significantly to testing knowledge, and two teachers indicated that undergraduate methods courses had made significant contributions to their testing knowledge.

**TABLE 13**

**AUTHENTIC ASSESSORS' RANKED RESPONSE TO MOST IMPORTANT CONTRIBUTION TO TESTING KNOWLEDGE**

<table>
<thead>
<tr>
<th>Most Important Contribution to Testing Knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 Undergraduate testing course</td>
</tr>
<tr>
<td>8 Graduate testing course</td>
</tr>
<tr>
<td>6 Undergraduate methods course</td>
</tr>
<tr>
<td>5 Experience as a student</td>
</tr>
<tr>
<td>3 Information and ideas / professional reading</td>
</tr>
<tr>
<td>4 Inservice training</td>
</tr>
<tr>
<td>2 Trial and error in the classroom</td>
</tr>
<tr>
<td>1 Professional collaboration / team teaching / peer coaching</td>
</tr>
</tbody>
</table>

In sub-section B - Reasons for Testing, teachers were asked to rank order a list of ten purposes for testing. Robert J. Wilson (1990) used this scale in previous studies, and identified items A, F, G, and H as "informing the teaching-learning process" (p. 9). Wilson studied the attitudes of "...51 practicing teachers and 101 student teachers in grades nine through the Ontario Academic Course or grade thirteen" (p.4). Wilson felt that "... the policies and procedures
concerning student achievement developed upon teachers from levels 'above' them in the administrative hierarchy will force their evaluation activities into relatively narrow areas" (p. 7). Table 14 indicates the ranking for these reasons for testing for the teachers who were identified in this study as authentic assessors. The four items which Wilson identified as "informing the teaching-learning process" - items A, F, G, and H - were ranked as the top four items by the teachers identified as authentic assessors.

**TABLE 14**

AUTHENTIC ASSESSORS' RANKED RESPONSE TO REASONS FOR TESTING

<table>
<thead>
<tr>
<th>Most Important Contribution to Testing Knowledge</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. To check students' progress against course objectives</td>
<td>3</td>
</tr>
<tr>
<td>B. To compare students' achievement to others</td>
<td>8</td>
</tr>
<tr>
<td>C. To generate marks for reporting purposes</td>
<td>7</td>
</tr>
<tr>
<td>D. To insure students do assigned work</td>
<td>9</td>
</tr>
<tr>
<td>E. To prepare students for this kind of evaluation in the future</td>
<td>10</td>
</tr>
<tr>
<td>F. To have students practice or apply what has been learned</td>
<td>1</td>
</tr>
<tr>
<td>G. To diagnose students' weaknesses with the material</td>
<td>4</td>
</tr>
<tr>
<td>H. To enable students to monitor their own progress</td>
<td>2</td>
</tr>
<tr>
<td>I. To help me decide what to teach next</td>
<td>6</td>
</tr>
<tr>
<td>J. To allow me to see how well I taught the material</td>
<td>5</td>
</tr>
</tbody>
</table>

When comparing Wilson's results to the authentic assessors identified in this study, it is evident that the authentic assessors were more closely aligned with the student teachers. Wilson felt that the student teachers would change their views once they were placed in the actual teaching situation. Table 15 compares the authentic assessors' rankings to those Wilson found in his study (p. 4).
TABLE 15

AUTHENTIC ASSESSORS' RANKED RESPONSE TO REASONS FOR TESTING AS COMPARED TO WILSON'S DATA

<table>
<thead>
<tr>
<th>AA N=11</th>
<th>ST N=16</th>
<th>PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
</tbody>
</table>

A. To check students' progress against course objectives
B. To compare students' achievement to others
C. To generate marks for reporting purposes
D. To insure students do assigned work
E. To prepare students for this kind of evaluation in the future
F. To have students practice or apply what has been learned
G. To diagnose students' weaknesses with the material
H. To enable students to monitor their own progress
I. To help me decide what to teach next
J. To allow me to see how well I taught the material

AA = Authentic assessors as identified in this study
ST = Student teachers as identified in Wilson's study
PT = Practicing teachers as identified in Wilson's study, (Wilson, 1990, p.4)

In sub-section C - Assessment Practices, teachers were asked to assign a ranking of one through five to ten pairs of statements that described testing practices. These ten statements were synthesized from a review of the literature. Each pair of statements was designed to represent a continuum with a descriptor of authentic assessment in the left hand column, and a descriptor of traditional assessment in the right hand column. Teachers marked each item from one to five using a one to represent the statement in the left column, and a five to represent the statement in the right hand column. The data was then transposed so that a "one" became a "five," etc. Thus, the higher scores became representative of authentic assessment. Mean scores were calculated for the eleven teachers who were identified as authentic assessors. Those mean scores are represented in Figure 2.
Fig. 2. Authentic Assessors' Mean Scores on Assessment Practices

Statements Representing the Authentic End of the Continuum

1. Students' backgrounds, native skills, prior training are taken into account in assessment designs.
2. Assessments are not administered during designated time periods. There is really a blend of assessment and instruction.
3. Assessment involves working with others.
4. Assessment involves a judgment on the process a student used.
5. Students know how they will be assessed by clearly defined criteria.
6. Assessments are subjective and based on growth of individuals.
7. Assessments require higher order thinking skills.
8. Assessments allow for growth over time and may be made at varying times for individual students.
9. Assessments involve interactions with the teacher and justifications for students answers.
10. Assessments include presentations and demonstrations of knowledge.

Statements representing both ends of the continuum are included in Appendix A. N=11
The mean scores in Figure 2 indicated that the authentic assessors perceived that their testing practices were consistent with the attributes of authentic assessment. The lowest possible score was a one, and the highest possible a five. If divided into quartiles, the mean scores on all ten statements for the authentic assessors fell into the top two quartiles or the range of scores from three to five.

Sub-section D - Types of Assessments Used, asked teachers to assign a percent to the six types of assessments listed. Definitions were presented for the six types of assessments. Table 16 indicates that the authentic assessors relied on the two types of performance assessments more than 90% of the time.

<table>
<thead>
<tr>
<th>N=11</th>
<th>Paper and pencil tests</th>
<th>Curriculum-embedded tests</th>
<th>Standardized tests</th>
<th>Oral questioning in the classroom</th>
<th>Performance assessment - structured</th>
<th>Performance assessment - spontaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71.4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sub-section E - Level of Use, asked teachers to describe the level of use from non-use to comfortable use for the six types of assessments in sub-section D of the questionnaire. The descriptor statements were as follows:

A. I do not currently use them and do not plan to use them in the future. (non-use)
B. I have decided to start using them in the future, but have not started to do so yet. (anticipated use)

C. I currently use them, but I find them difficult to use and it takes great effort. (effort in use)

D. I use these tests on my own as a regular part of my instruction and do so comfortably. (comfortable use)

The descriptors were transferred to numeric data with an "A" equalling a "one," etc. A score of "one" represented non-use, a "two" anticipated use, a "three" effort in use, and a "four" comfortable use. Figure 3 represents the mean level of use scores for the authentic assessors.

**Fig. 3. Authentic Assessors' Mean Scores on Level of Use**

Level of Use: 1 = Non-use, 2 = Anticipated use, 3 = Effort in use, 4 = Comfortable use.

Types of Assessments: 1 = Paper and pencil tests, 2 = Curriculum-embedded tests, 3 = Standardized tests, 4 = Oral questioning in the classroom, 5 = Performance assessment - structured, 6 = Performance assessment - spontaneous

N=11
Level of use scores for paper and pencil tests, curriculum-embedded tests, and oral questioning were at the 1.0 or non-use level. Even though standardized test scores received a 0% score on time allotted to use, the corresponding level of use score was 3.9 or effort in use. In reviewing the data, it was noted that all eleven teachers identified as authentic assessors scored standardized test scores at 0% in time allotted for use. However, these same teachers scored the level of use for standardized tests at threes and fours. A few teachers made notes on the questionnaire and indicated that they were referring to the Illinois Goal Assessment Program (IGAP) tests or tests they were required to give as part of the district's testing program. The authentic assessors indicated the highest level of use for the performance assessments; ranking structured assessments at 3.9 and spontaneous assessments at 3.3.

Research Question 2

The purpose of research question 2 was to examine the attributes and characteristics of authentic assessors which separated them from traditional assessors. The results from the questionnaire were used to contrast the responses of those teachers identified as authentic assessors with those teachers identified as traditional assessors. The ways in which the two groups of teachers differed are represented in Tables 18 through 24 and in Figures 4 and 5.

Table 17 compares the number of authentic and traditional assessors as identified by subject taught.
### TABLE 17

AUTHENTIC AND TRADITIONAL ASSESSORS
AS IDENTIFIED BY SUBJECT TAUGHT

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>A</th>
<th>I</th>
<th>A</th>
<th>I</th>
<th>A</th>
<th>I</th>
<th>A</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Social Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A=Authentic Assessors  
N=11  
T=Traditional Assessors  
N=7

The majority of the authentic assessors taught English with two from social studies; while the traditional assessors were evenly divided in the fields of foreign language, science, and math.

The mean scores for the sum of time allotted to performance assessments were computed by subject area for all teachers and are shown in Table 18.

### TABLE 18

MEAN SCORES FOR ALL TEACHERS OF TIME ALLOCATED TO PERFORMANCE ASSESSMENTS

<table>
<thead>
<tr>
<th>%</th>
<th>Subject Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>63%</td>
<td>English</td>
</tr>
<tr>
<td>28%</td>
<td>Foreign Language</td>
</tr>
<tr>
<td>25%</td>
<td>Math</td>
</tr>
<tr>
<td>30%</td>
<td>Science</td>
</tr>
<tr>
<td>43%</td>
<td>Social Studies</td>
</tr>
</tbody>
</table>

N=95
Based on the data in Table 18, it appeared that the authentic assessors were more likely to be found in the fields of English and social studies.

Table 19 identified the authentic assessors and the traditional assessors by years of teaching experience, sex, age, and education level. Both the authentic assessors and traditional assessors seemed to be evenly distributed in terms of years of teaching experience. There were more females - seven of the eleven or 64% - in the authentic assessors category; and more males - five of the seven or 71% - in the traditional assessors category. In the age category, none of the authentic assessors fell into the age ranges of the fifties and sixties, while two of the seven, or 28%, of the traditional assessors were in their fifties or sixties. Both groups seemed to be evenly distributed in terms of educational level. Six of the eleven authentic assessors or 54% reported at least a Master's degree plus thirty graduate hours, and three of the seven traditional assessors or 43% fell in the same educational category.

<p>| TABLE 19 |
| AUTHENTIC AND TRADITIONAL ASSESSORS AS IDENTIFIED BY YEARS OF TEACHING, SEX, AGE, AND EDUCATION |</p>
<table>
<thead>
<tr>
<th>Years</th>
<th>Sex</th>
<th>Age</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 1</td>
<td>I 1</td>
<td>4 5 Male</td>
<td>2 2 20s</td>
</tr>
<tr>
<td>2 1</td>
<td>Less 5</td>
<td>7 2 Female</td>
<td>1 1 30s</td>
</tr>
<tr>
<td>1 1</td>
<td>6-10</td>
<td>8 2 40s</td>
<td>1 2 MA</td>
</tr>
<tr>
<td>2 1</td>
<td>11-15</td>
<td>0 1 50s</td>
<td>2 1 MA+15</td>
</tr>
<tr>
<td>3 2</td>
<td>16-20</td>
<td>0 1 60s</td>
<td>5 3 MA+30</td>
</tr>
<tr>
<td>3 2</td>
<td>20+</td>
<td>0 Ph.D.</td>
<td>1 0</td>
</tr>
</tbody>
</table>

A=Authentic Assessors
N=11
T=Traditional Assessors
N=7
Table 20 identified the authentic and traditional assessors by last graduate college course and last workshop or conference in assessment procedures. While five of the eleven, or 45%, of the authentic assessors had completed a graduate course in assessment procedures during the last year, five of the seven traditional assessors, or 71%, had not had a course in assessment in the past six years. A similar difference was noted in workshops and conferences attended. Nine of the eleven authentic assessors, or 82%, had been involved in a workshop or conference in assessment procedures during the last year, while three of the seven traditional assessors, or 75%, had not attended a workshop or conference in assessment procedures in the last ten years.

**TABLE 20**

**AUTHENTIC AND TRADITIONAL ASSESSORS AS IDENTIFIED BY LAST GRADUATE COLLEGE COURSE AND LAST WORKSHOP/CONFERENCE IN ASSESSMENT PROCEDURES**

<table>
<thead>
<tr>
<th>Last Formal Graduate College Course</th>
<th>Last Workshop or Conference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td><strong>T</strong></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

A=Authentic Assessors, N=11  
T=Traditional Assessors, N=7
Table 21 presents the authentic assessors' and traditional assessors' ranked responses to their contributions to testing knowledge. While the authentic assessors credited professional collaboration as their most important contribution to testing knowledge, the traditional assessors ranked professional collaboration third in a tie with undergraduate methods courses and professional readings. Both groups ranked trial and error in the classroom high, and ranked undergraduate and graduate testing courses low.

TABLE 21

AUTHENTIC AND TRADITIONAL ASSESSORS' RANKED RESPONSE TO CONTRIBUTIONS TO TESTING KNOWLEDGE

<table>
<thead>
<tr>
<th>Most Important Contribution to Testing Knowledge</th>
<th>A</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate testing course</td>
<td>7</td>
<td>6 (tie)</td>
</tr>
<tr>
<td>Graduate testing course</td>
<td>8</td>
<td>6 (tie)</td>
</tr>
<tr>
<td>Undergraduate methods course</td>
<td>6</td>
<td>3 (tie)</td>
</tr>
<tr>
<td>Experience as a student</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Information and ideas / professional reading</td>
<td>3</td>
<td>3 (tie)</td>
</tr>
<tr>
<td>Inservice training</td>
<td>4</td>
<td>6 (tie)</td>
</tr>
<tr>
<td>Trial and error in the classroom</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Professional collaboration / team teaching / peer coaching</td>
<td>1</td>
<td>3 (tie)</td>
</tr>
</tbody>
</table>

A=Authentic Assessors, N=11
T=Traditional Assessors, N=7

Table 22 presents the authentic assessors' and traditional assessors' ranked response to reasons for testing. Both groups ranked as their top reason "to have students practice or apply what has been learned." However, while the authentic assessors ranked "to enable students to monitor their own progress" second, the traditional assessors ranked this item sixth. "To
enable students to monitor their own progress," was identified by Grant Wiggins (1991) as a characteristic of authentic assessment. Wiggins states that authentic assessment "Involves de-mystified and non-secret tasks, criteria and standards; allows for thorough preparation and accurate self-assessment by the student" (Toward One System of Education, p. 25). In this regard, the authentic assessors ranked response as second in priority for reasons for testing is consistent with characteristics revealed in a review of the literature.

**TABLE 22**

AUTHENTIC AND TRADITIONAL ASSESSORS' RANKED RESPONSE TO REASONS FOR TESTING

<table>
<thead>
<tr>
<th>Rank</th>
<th>A</th>
<th>T</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>F. To have students practice or apply what has been learned</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6</td>
<td>H. To enable students to monitor their own progress</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>8 (tie)</td>
<td>A. To check students' progress against course objectives</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>7 (tie)</td>
<td>C. To generate marks for reporting purposes</td>
</tr>
<tr>
<td>7</td>
<td>4 (tie)</td>
<td>8</td>
<td>B. To compare students' achievement to others</td>
</tr>
<tr>
<td>8</td>
<td>8 (tie)</td>
<td>7</td>
<td>D. To insure students do assigned work</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>9</td>
<td>E. To prepare students for this kind of evaluation in the future</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>1</td>
<td>G. To diagnose students' weaknesses with the material</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>1</td>
<td>I. To help me decide what to teach next</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>10</td>
<td>J. To allow me to see how well I taught the material</td>
</tr>
</tbody>
</table>

A=Authentic Assessors, N=11  
T=Traditional Assessors, N=7

Figure 4 depicts the authentic and traditional assessors' mean scores on assessment practices. The data was tabulated based on teacher's responses to sub-section C of the Classroom Testing Questionnaire (Appendix A). In this section, teachers were asked to respond to statements that represented two
ends of a continuum regarding their assessment practices. Teachers marked each of ten items with a number between one and five, with a one representing the authentic end of the continuum and a five representing the traditional end of the continuum. These scores were later transposed so a one became a five, etc. Thus, the higher numbers represented the authentic assessment practices.

Figure 4 indicates that the authentic and traditional assessors' scores were consistent with the types of assessments used. In each of the ten items, the authentic assessors' mean scores were at 3.9 or above, while the traditional assessors' mean scores were at or below 3.3.
Fig. 4. Authentic and Traditional Assessors' Mean Scores on Assessment Practices

Statements Representing the Authentic End of the Continuum

1. Students' backgrounds, native skills, prior training are taken into account in assessment designs.
2. Assessments are not administered during designated time periods. There is really a blend of assessment and instruction.
3. Assessment involves working with others.
4. Assessment involves a judgment on the process a student used.
5. Students know how they will be assessed by clearly defined criteria.
6. Assessments are subjective and based on growth of individuals.
7. Assessments require higher order thinking skills.
8. Assessments allow for growth over time and may be made at varying times for individual students.
9. Assessments involve interactions with the teacher and justifications for students answers.
10. Assessments include presentations and demonstrations of knowledge.

Statements representing both ends of the continuum are included in Appendix A.
Figure 5 depicts the mean scores of the authentic and traditional assessors on assessment practices as compared to the mean scores of all teachers. On all ten of the statements, the authentic assessors' scores were highest and most representative of authentic assessment. On nine of the ten statements, the traditional assessors had the lowest scores or those most representative of traditional assessments. In item number five, the traditional assessors' mean score was a 3.3; while the mean score for all assessors was slightly lower at 3.1.
Fig. 5. Authentic, Traditional, and All Assessors' Mean Scores on Assessment Practices

Assessment Practices Statements

1. Students' backgrounds taken into account vs. all students receive same assessment.
2. Blend of assessment and instruction vs. administered at specific times.
3. Involves working with others vs. requires students to work alone.
4. Involves judgment on process vs. depends on product (correct score).
5. Students judged on clearly defined criteria vs. students can expect varied examples.
6. Requires subjective appraisal based on growth vs. objective and based on correct responses.
7. Requires higher order thinking skills vs. requires knowing correct answer.
8. Allow for growth over time and may be given at different times for individual students vs. are specific measurements given at designated intervals over course of semester/year.
9. Involves interactions with teacher vs. assessments are paper and pencil tests.
10. Includes presentations and demonstrations vs. involves answering essay or multiple-choice questions.
Table 23 outlines the mean percent of time allotted to different types of assessments by authentic, traditional, and all assessors. Of the total time allotted to testing, the authentic assessors spent 1% on the first three categories which included teacher-made paper and pencil tests, curriculum-embedded tests, and standardized tests. In comparison the traditional assessors allotted 90% of their time to these same categories. Both authentic assessors and traditional assessors allotted 7% of assessment time to oral questioning; while the mean percentage for all teachers was 18%. Authentic assessors allotted 92% of assessment time to structured and spontaneous performance assessment; while traditional assessors allotted 3% to performance assessment, and all teachers allotted 39% of their assessment time to performance assessment.

<table>
<thead>
<tr>
<th></th>
<th>Authentic</th>
<th>Traditional</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=11</td>
<td>N=7</td>
<td>N=77</td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>60%</td>
<td>29%</td>
<td>Paper and pencil tests</td>
</tr>
<tr>
<td>0%</td>
<td>26%</td>
<td>12%</td>
<td>Curriculum-embedded tests</td>
</tr>
<tr>
<td>0%</td>
<td>4%</td>
<td>2%</td>
<td>Standardized tests</td>
</tr>
<tr>
<td>7%</td>
<td>7%</td>
<td>18%</td>
<td>Oral questioning in the classroom</td>
</tr>
<tr>
<td>71%</td>
<td>1%</td>
<td>27%</td>
<td>Performance assessment - structured</td>
</tr>
<tr>
<td>21%</td>
<td>2%</td>
<td>12%</td>
<td>Performance assessment - spontaneous</td>
</tr>
</tbody>
</table>

Richard J. Stiggins and Nancy J. Bridgeford (1985) conducted a study "... to probe assessment practices in a stratified sample of teachers selected from eight districts across the country, varying in size and geographic location"
In one area of their study, they sought to determine ". . . the relative importance teachers assigned to the various test types for diagnosing the strengths and weaknesses of individual students, grouping for instruction, assigning grades, evaluating the effectiveness of an instructional treatment, and reporting results to parents" (p. 275). Table 24 compares Stiggins and Bridgeford's work to the results found in this study. Stiggins and Bridgeford's data is reported for only two categories: "assigning grades" and "reporting results to parents." In Stiggins and Bridgeford's study, teachers were asked to assign percentages to six types of assessments based on their reasons for testing. This study asked teachers to assign percentages to the same six types of assessment based on the total amount of time allotted for assessment. Stiggins and Bridgeford's data is reported in what they termed reliance percentages, and is described as follows:

Because teachers assigned higher percentages to the methods that contribute most to each decision, these data are hereafter called reliance percentages in describing and interpreting the results. The higher the reliance percentage, the more weight given to a type of test for that purpose. (p. 275)

Their data is also reported in terms of percent of respondents, while the comparative data from this study is reported in percent of time allotted to the type of test. Although Stiggins and Bridgeford studied teachers in grades two, five, eight, and eleven, only the results of the eleventh grade teachers are reported in Table 24. The categories of curriculum-embedded tests and standardized tests were combined under the category published tests. Stiggins and Bridgeford did not report results on oral questioning in the classroom, even though it was included in the survey question. Therefore, in comparing results oral questioning was omitted from Table 24.
### Table 24

**Comparative Data on Reliance on Different Types of Assessments**

<table>
<thead>
<tr>
<th></th>
<th>Stiggins and Bridgeford Results in % of respondents</th>
<th>Results in % of time devoted to assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grading N=55</td>
<td>Reporting N=55</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>ST PA</td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>SP PA</td>
<td>10</td>
<td>14</td>
</tr>
</tbody>
</table>

OBJ stands for teacher-made objective tests, PUB for published tests, ST PA for structured performance assessment, SP PA for spontaneous performance assessment.


Stiggins and Bridgeford summarized their results as follows:

In analyzing the role of each test type for different test purposes, certain consistent patterns emerge. For example teachers indicate they use their own objective tests more frequently than other assessments for all purposes. However, teachers also report heavy reliance on both types of performance assessments. Published tests consistently play a secondary role. Clearly, teacher-made tests dominate. (pp. 275-277)

Although the results in Table 24 are reported in different terms, it is possible to make some comparisons. As in Stiggins and Bridgeford's study, all assessors reported that they relied on their own tests more than any other type. However, the authentic assessors indicated that they did not rely on objective tests at all. All assessors reported that they relied on performance assessments a little less.
than did the teachers in Stiggins and Bridgeford's study; while authentic assessors relied on this method more, and traditional assessors relied on performance assessment far less of the time when compared to Stiggins and Bridgeford's data. Stiggins and Bridgeford found that published tests "played a secondary role," and this data was consistent with all the teachers in this study, but not to the same extent. The traditional assessors relied on published tests thirty percent of the time, and while the authentic assessors did not rely on published tests at all.

Table 25 compares the results on level of use on four types of assessments. Teachers were asked to respond to the ways in which they used teacher-made objective tests, standardized tests, curriculum-embedded tests, oral questioning in the classroom, structured performance assessments, and spontaneous performance assessments (Classroom Testing Questionnaire, Appendix A - sub-section E). Teachers were asked to indicate their level of use of each assessment by marking them with statements A through D.

A. I do not currently use them and do not plan to use them in the future. (non-use)

B. I have decided to start using them in the future, but have not started to do so yet. (anticipated use)

C. I currently use them, but I find them difficult to use and it takes great effort. (effort in use)

D. I use these tests on my own as a regular part of my instruction and do so comfortably. (comfortable use)

In Table 25 standardized tests and curriculum-embedded tests were collapsed into a category entitled published tests, and the category of oral questioning was omitted.
TABLE 25
COMPARATIVE DATA ON LEVEL OF USE ON DIFFERENT TYPES OF ASSESSMENTS

<table>
<thead>
<tr>
<th>Assessment Type</th>
<th>A (N=11)</th>
<th>T (N=7)</th>
<th>All (N=75)</th>
<th>S &amp; B (N=55)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Teacher Made Tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-use</td>
<td>91</td>
<td>0</td>
<td>29</td>
<td>11</td>
</tr>
<tr>
<td>Use</td>
<td>9</td>
<td>100</td>
<td>67</td>
<td>47</td>
</tr>
<tr>
<td>Refinement</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>15</td>
</tr>
</tbody>
</table>

| **Published Tests**                    |          |         |            |               |
| Non-use                                | 100      | 29      | 45         | 53            |
| Use                                    | 0        | 71      | 43         | 39            |
| Refinement                             | 0        | 0       | 12         | 8             |

| **Structured Performance Assessment** |          |         |            |               |
| Non-use                                | 0        | 100     | 10         | 8             |
| Use                                    | 91       | 0       | 65         | 57            |
| Refinement                             | 9        | 0       | 25         | 36            |

| **Spontaneous Performance Assessment** |          |         |            |               |
| Non-use                                | 20       | 83      | 19         | 11            |
| Use                                    | 70       | 17      | 56         | 70            |
| Refinement                             | 10       | 0       | 25         | 19            |

A=Authentic Assessors
T=Traditional Assessors
All=All Assessors
S&B=the results of the study by Stiggins and Bridgeford (1985), p. 277.
Statements indicating non-use and anticipated use (statements A and B) were collapsed into the non-use category, with statement D representing comfortable use, and statement C representing refined use. For teacher-made objective tests, 91% of the authentic assessors reported non-use, and only 9% reported comfortable use; while 0% of the traditional assessors reported non-use, and 100% reported comfortable use. This data is compared to the results of all teachers in this study and to the results of Stiggins and Bridgeford (1985).

Across all four categories, the results of all teachers are fairly consistent with the data from Stiggins and Bridgeford. While the results for the authentic and traditional assessors are almost diametrically opposed. Traditional assessors showed a comfortable level of use with teacher-made objective tests and published tests, while authentic assessors showed a comfortable level of use with structured performance assessments and spontaneous performance assessments. Conversely, authentic assessors showed a high level of non-use for teacher-made objective tests and published tests, while traditional assessors indicated a high level of use for these same two types.

Research Question 3

The purpose of research question 3 was to find contributing factors which enabled teachers to become practitioners of authentic assessment. These factors included personal factors, professional development and involvement, administrative/supervisory structure, and student characteristics, along with demographic data on students and teachers.

An In-Depth Interview Questionnaire (Appendix B) was developed to verify categorization of the teachers as authentic and traditional assessors. Those teachers scoring at the high and low ends based on the initial questionnaire were contacted for an in-depth interview. The teachers' responses were recorded and later categorized on a tally sheet. Factors contributing to
teachers being able to shift their paradigms from traditional assessors to authentic practitioners were delineated.

In-depth interviews were conducted with sixteen of the teachers who were identified as authentic and traditional assessors. Of the seven teachers identified as traditional assessors, five were interviewed. Two teachers were not interviewed because one teacher was not identified by name on the initial survey, and the other teacher declined the request for an interview. All eleven of the teachers identified as authentic assessors were interviewed. The traditional teachers interviewed included a foreign language and a social studies teacher, two math teachers, and one science teacher. The authentic assessors interviewed included ten English teachers and two social studies teachers.

The interviews were composed of twenty-four questions which covered personal factors, professional development and involvement, administrative/supervisory structure, and student characteristics. Teachers were also asked to talk about the assessments they used in class, explain their philosophy underlying assessment, and supply a typical assessment that they might use.

Additionally, an interview was conducted with the principal of each high school to gather background information to assist in the interpretation of the data. The principal interview covered the topics of academic organization, description of the school, recent leadership history, relations with the Board of Education, vision and mission of the school, average teaching assignment, focus on curriculum and inservice, and climate and teacher empowerment.

Based on the interviews, four areas were identified as enabling or supporting teachers' efforts to become authentic assessors: professionalism, collegiality, philosophy, and support systems.
Professionalism

**Teacher Interviews:** There were marked differences between the interviews of the authentic and traditional teachers based on questions 7 through 14 on sub-section B of the In-Depth Interview Questionnaire used for this study (Appendix B). Professionalism covered two general areas: ongoing education as identified by college courses, workshops, conferences, and professional readings; and leadership roles such as presentations at conferences, workshops conducted, and publications.

All of the authentic assessors were very involved in professional activities, and all eleven of the teachers had attended one or more off-site workshops or conferences during the past year. The authentic assessors seemed to fall into two groups labeled as neophytes and leaders. Three of the eleven teachers were labeled neophytes. The neophytes had been at their school for one to two years, and had taught less than five years. One teacher was in his first year of teaching, one in his second year, and one in her fourth year. These teachers were viewed as emulating or modeling the philosophy and instructional and assessment practices of their peers who were labeled leaders. Although the neophytes were excited about their work, it was obvious that they were being "brought into the fold." They did not have the rich background in terms of teaching experience that the other teachers identified as authentic assessors possessed. It appeared that the neophytes were hurrying to catch up with their colleagues in terms of professional activities. One had a Master's degree, and the other two were enrolled in Master's programs.

In contrast, the eight teachers identified as leaders had all taught over fourteen years with an average of seventeen years of teaching experience. Of the eight teachers described as leaders, one had a Ph.D., three were enrolled in doctoral programs, three had thirty hours beyond the Master's degree, and one had fifteen graduate hours beyond the Master's degree.
The leaders were also involved in other professional activities. Among their activities were the following:

1. Four were department or division chairs at their schools - two English chairs, and two social studies chairs. One was a part-time curriculum coordinator - teaching two classes in the morning and assisting other teachers during the rest of the day.

2. Nine teachers had made presentations at professional conferences which included the:
   - Association for Supervision and Curriculum Development (ASCD)
   - Constitutional Rights Foundation
   - Illinois Association for Supervision and Curriculum Development (IASCD)
   - Illinois Council for the Social Studies (ICSS)
   - Illinois Council of Teachers of English (IATE)
   - Illinois Education Association (IEA)
   - Illinois Whole Language Association
   - International Reading Association (IRA)
   - National Association of School Administrators (NASA)
   - National Council for the Social Studies (NCSS)
   - National Council of Teachers of English (NCTE)
   - National Forum on Education (Phoenix, Arizona)
   - Whole Language Umbrella Association

3. Five of the teachers had experience in presenting workshops to teachers in other districts. Five of the teachers had actually served as consultants, charging for their services at other schools. Two of the teachers estimated that they had easily presented over 100 workshops each in districts in Illinois and surrounding states. Three teachers had made presentations for or worked as consultants for the Illinois State Board of Education, the Constitutional Rights Foundation, and for Educational Service Centers in Illinois.

4. Two teachers had written a trade book on using portfolio assessment at the secondary school level which was published by Heinemann and Sons, and three teachers had written articles or chapters in professional journals and books.
5. Three of the teachers served on state or national committees. One teacher had been appointed to two boards: the National Board for Professional Teaching Standards sponsored by the Education Development Center at the Massachusetts Institute of Technology, and to the Commission on Reading a sub-committee of the National Council of Teachers of English. Another teacher was part of the Social Science Education Consortium's grant project sponsored by the U.S. Office of Education which involved and supported fifteen teachers in five states in writing and designing authentic assessments for civics, law, and government classes. And the third teacher served as the Chairperson for the Illinois Advisory Committee for the Social Studies IGAP, and served on the panel for the Longitudinal Study of Assessment sponsored by the National Assessment of Educational Progress (NAEP).

6. At one of the schools, the English teachers had been featured in a video on student-centered education produced by the Mid-Continent Educational Laboratory (McREL).

7. Many of the teachers had initiated or were the sponsors of ongoing programs such as a freshman orientation program, parents clubs, the drama club, the forensics team, and others.

All of the eleven teachers were members of at least two professional organizations, and said that they regularly read professional publications. All of the teachers were familiar with a list of topics and names asked in question nine of the In-Depth Interview Questionnaire which included: learning styles, brain-based research, student-centered instruction, Bloom's Taxonomy, Howard Gardner, Richard Stiggins, reform movements, and professional standards. The teachers labeled neophytes could give a brief description of the topics listed, and seemed to have a general understanding of the theories espoused by the people on the list. While the teachers labeled leaders, could articulate on which points they agreed with the people on the list, and where their personal philosophies differed. The teachers identified as leaders also added names and topics to the list of influences on their instruction and assessment philosophies. The leaders
could also give specific examples of how topics and people on the list had influenced their instruction and assessment styles. Teachers cited the Mid-Continent Regional Educational Laboratory's (McREL) publication Learner-centered psychological principles, and added names such as Nancy Atwell, Harvey Daniels, Smokey Daniels, Toby Fulwiler, Bob Gundlach, Jerome Harste, Fred Newmann, Janet Rico, and Regie Routman as being influential in their development of assessment strategies. The ways in which the teachers were influenced will be addressed in the section on philosophy.

The five traditional assessors matched the authentic assessors on educational level - all had at least a Master's degree, two had a Master's plus thirty graduate hours, and one has a Master's plus fifteen graduate hours. But the similarities between the two groups ended there.

Only one traditional assessor had attended a workshop or conference in the last year. The traditional assessors tended to have a few more years in teaching experience than the authentic assessors. Three had over 20 years of experience, and one teacher had 40 years of experience. The average number of years of teaching experience was 20 years.

The traditional assessors did not participate in leadership roles in the school. One teacher had been a former chair of the math department, but had been relieved of that position five years ago. While the traditional assessors were not involved in sponsoring extra-curricular activities, they were involved in coaching. One was the head football coach, one had been the head basketball coach and now served as an assistant coach, and one had coached a variety of sports including track and basketball. The traditional assessors also had no experience in presenting workshops, serving as consultants, or writing for professional publications.

Two of the five teachers belonged to professional organizations, but only one had attended a national conference during the past year. None of the
teachers said they read professional journals. When presented with the list of topics and names in question number nine, only "learning styles" and "professional standards" were cited as being familiar. Their responses to these topics and names will be discussed in the section on philosophy.

**Principal Interviews:** The interviews with the principals supported the level of professional activity that was noted in the teacher interviews with the authentic assessors. School two had the largest number of authentic assessors, and a variety of programs that encouraged professionalism. The school had an active building-wide inservice program organized and managed by a committee of teachers. The year's theme was "What Constitutes Quality Assessment?" All teachers were reading Kay Burke's book *How To Assess Authentic Learning*, and were meeting in teams to discuss issues during the course of the year. The first semester was devoted to developing a knowledge base on quality assessment, and the second semester's focus was on experimenting with assessments that had been investigated the previous semester. Teams of teachers and/or the entire faculty had met with Dr. Kay Burke of IRI/Skylight in Palatine, Illinois, Dr. Fred Newmann of the University of Wisconsin, and other consultants to assist them in the curriculum and assessment revisions. A combination of institute days and late arrival and early dismissal days were scheduled to provide inservice time.

The Principal of school two also stated that the travel and substitute budgets allowed teachers to travel to a variety of conferences as attendees and presenters. The travel enabled teachers to meet nationally recognized experts in the field of assessment, and opened up new doors for professional activity. Two teachers had written a book entitled *The Portfolio As A Learning Strategy* with a forward by Yetta Goodman, and two other teachers had attended an assessment workshop with Grant Wiggins at CLASS (the Center on Learning, Assessment and School Structure) in Geneseo, New York. The CLASS
workshop had resulted in Wiggins sending a team of his researchers to the school to video tape a semester exam that was based on authentic learning principles. The videotape was to be used in one of CLASS's teacher training videos. Professional travel had also resulted in teachers becoming acquainted with representatives from McREL (Mid-Continent Regional Educational Laboratory), and the selection of the school's English department as one of three schools featured in their teacher training video on student-centered learning. The Principal also stated that the school was very liberal in allowing teachers to serve as active consultants to other districts. Honorariums received by the teachers were placed in a departmental account which could then be used for additional supplies or as a supplement to the travel budget.

At school three, with the second largest number of authentic assessors, the inservice program was described as serving departmental needs with an absence of a building-wide focus. One day of inservice was spent on articulation with the feeder schools. Often a consultant spoke on a current educational issue such as gender bias or gifted education with the majority of the day devoted to teachers sharing information on the curricula in the different schools. Other inservice time was spent on specific projects such as math teachers developing instructional units using graphing calculators, all teachers working on writing learner objectives for the IGAP, or all teachers developing required syllabi for all courses that stated learning outcomes and grading practices.

Professional development also included teachers on committees such as the Professional Development Committee which awarded teachers "minisabbaticals" or released time to develop projects, and the Teacher Institute Committee which was developing a professional growth plan for teachers to earn credit on the salary schedule or stipends by involvement in professional development other than the standard college graduate course.
School one had no identified authentic assessors. The Principal felt that the inservice program at the building level focused on ways of enhancing instruction. Most of this was accomplished, he said, by sending teachers to off-site locations to participate in workshops or training. The teachers also took part in a county-wide inservice program each fall sponsored by the Regional Superintendent of Schools Office. The Principal did not feel that a building-wide focus on inservice had been established, because "we are not at that hurdle yet - we still have to identify our needs."

**Synthesis:** Authentic assessors were identified at schools two and three which seemed to have programs in place to support teachers in terms of professional development. At school two, which had a district-wide focus on inservice education, all of the teachers had been provided with a copy of Kay Burke’s book which was being used as a foundation for the inservice program. The inservice program was also structured throughout the year with teachers knowing their responsibilities in terms of reading chapters in the book and meeting in discussion groups. This school seemed to practice what it preached in that the inservice program was teacher-centered: teachers planned the program, were expected to participate in discussions, were asked to experiment with authentic assessments, and then contribute an outline of an authentic assessment to a handbook that was being developed for staff use. Inservice at school three did not appear to be as teacher-centered, but a sizeable budget allowed for expert speakers on a variety of topics that were currently being discussed in the educational literature.

While schools two and three had designated funds for inservice and professional travel, it seemed that the authentic assessors took advantage of the professional activities while the traditional assessors did not. The same environments had teachers at both ends of the assessment spectrum: authentic and traditional.
School one, in which only traditional assessors were identified, seemed to lack support for professional development. The Principal indicated that teachers participated in a yearly county-wide program that presented speakers on a diversity of topics; but other than this yearly program, no other professional growth activities or programs were in place. The Principal stated that teachers could be sent to off-site locations for professional development, but neither of the traditional assessors at the school had participated in professional travel during the last year.

**Summation of Professionalism:** The following summarizes the authentic assessors professional involvement:

1. eleven had attended off-site workshops or conferences in the past year,
2. one had a Ph.D. from Stanford University, and three were enrolled in doctoral programs,
3. four were department or division chairs,
4. nine had presented at professional conferences,
5. five had served as consultants,
6. two were trade book authors,
7. three served on national committees or boards,
8. eleven were members of at least two professional organizations,
9. eleven were familiar with names from professional readings.

The traditional assessors professional activities are summarized as follows:

1. one had attended an off-site workshop or conference in the past year,
2. two belong to a professional organization,
3. one had attended a national conference in the past two years,
4. none read professional journals.
Teacher Interviews: The authentic and traditional assessors also differed markedly in the area described as collegiality. Their responses were from questions 2-4, 13, 20, and 21 in the questionnaire. The authentic assessors seemed to enjoy a great deal of collegiality which centered on curriculum writing, planning, sharing and debriefing lessons, and making presentations. Each teacher was also able to describe a relationship with a mentor who had influenced his/her professional development.

Nine of the eleven teachers identified as authentic assessors were English teachers at schools two and three. At both schools, the English teachers had a departmental office space with an open area for working or meeting, and the department chair had an office attached to the departmental area.

At both schools, curriculum writing was done by teams of teachers. English teachers at one of the schools had presented a series of curriculum workshops for the English teachers at the other school during the past year, so both schools had virtually the same English curriculum at the freshman level. The team approach to curriculum writing also included the designing of assessments. One school's model for curriculum design differed from a traditional model. Rather than 1) setting objectives, 2) selecting content, and then 3) determining assessment; determining assessment was the second step. The teachers at school two were entering their fourth year of completely redesigning the English program to what they termed a "student-centered approach." While the English teachers at school three were in their first year of the new curriculum, and had plans to redesign the sophomore, junior, and senior programs in succeeding years.

The English curricula at both schools two and three were portfolio driven with both written and video portfolios used to assess progress in reading, writing, and speaking. A "core" curriculum insured that all students covered and
completed the same basic material and assessments. An "enrichment" portion enabled teachers to add topics and/or assignments to the curriculum.

The English teachers at both schools two and three spoke about the daily interactions with their peers as they planned instructional strategies, shared ideas, and debriefed lessons. Most of the English teachers ate lunch together in their departmental offices and spent their planning periods there. A few teachers remarked that they felt other teachers in the school thought the English teachers were too "cliquey" because they never set foot in the teachers' cafeteria or teachers' lounge.

The English teachers also extended their collegiality to their professional activities. Some of the teachers described as leaders discussed how they would work on conference proposals in the departmental office and draw the younger teachers into their presentations. Both English department chairs described a concerted effort to draw these younger teachers into presentations by designing segments they could successfully present, and then assisting them with developing their presentations.

The two social studies teachers who were included in the authentic assessors category provided a unique situation. They were both department chairs, they were the only teachers in their departments to be identified as authentic assessors, and they were married to each other. These two teachers did not describe the same level of collegiality in their departments that was evidenced by the English teachers, but were working to bring a higher level of collegiality to their departments. One teacher had served in the role of chair for four years, and the other teacher had been at the school and in the role of chair for three years.

Both of these social studies teachers had initiated the team approach to curriculum writing, and had involved teachers in their department in conference presentations. These department chairs had also worked with teachers in their
department to design authentic assessments, had team taught units with their teachers, modeled how the assessments could be integrated into the curriculum, and assisted teachers in designing scoring rubrics for the assessments.

All of the eleven teachers identified as authentic assessors could identify one or more mentor relationships that influenced their assessment methods. Most frequently, the younger or neophyte teachers identified the department chair or another teacher in their department as their mentor. They described how their mentor or mentors would assist them which included:

1. "talking them through" a lesson before they tried it in front of their class;
2. assisting them in debriefing a lesson;
3. "opening them up" to new learning by suggesting they read a particular article or book, and then discussing it with them;
4. observing a lesson and then assisting them in analyzing strengths and weaknesses;
5. encouraging them to feel that their ideas were as valued as other department members during meetings or curriculum writing sessions.

The teachers labeled as leaders identified their mentors as the department chair or as university professors. These teachers described relationships that were both personal and professional. Some of the teachers had become friends with their mentors, and had turned conference presentations with them into family vacations, with spouses and children, in locations such as San Diego, Orlando, and New Orleans. Other teachers labeled as leaders described a more professional relationship with their mentors which included co-authoring professional articles or book chapters, guest lecturing in their college courses, or assisting them with their research by providing material from their teaching situations.
The traditional assessors who were interviewed were at a loss to describe collegiality in their departments. They cited that they met at department meetings, and everyone was free to "say what they wanted." All of these teachers followed a textbook-driven curriculum, and their view of collegiality was delivering a final exam that had been culled from test items from the teachers in the department.

The mentors the traditional assessors identified were most frequently their own high school teachers. Rather than a true mentor who offered assistance and guidance during the course of professional development, these mentors were teachers who were admired and respected. Many of the traditional assessors stated that they had modeled their teaching styles and assessment styles on their former high school teachers because that particular style had worked for them as students, and they had been successful in their mentors' classes. The traditional assessors, for the most part, did not think their mentors were aware that they were so identified, and the majority of the traditional assessors had not had contact with these teachers since their own high school days.

Principal Interviews: At schools two and three, with authentic assessors, the aspect of collegiality was supported primarily by an organizational structure, and with programs and funds that adequately supported curriculum writing. School two was academically organized into six divisions with each division chaired by an Academic Director who taught two classes. The Academic Directors were classified as administrators, and were charged with overseeing curriculum development. Each division had an office area which accommodated most of the teachers and enabled curriculum collaboration. The typical teaching assignment was five classes on an eight period day, with an average class size of 18.8. The school had an active Curriculum Council and a part-time Curriculum Coordinator whose job was to assist teachers in designing curriculum
and instruction. A five-year curriculum review plan insured that each course was reviewed on a cyclical calendar. The Curriculum Coordinator held Writing Workshops for teachers who were to write curricula over the summer to insure that the curricula would be written according to the adopted guide which stressed student involvement and authentic assessments. Teachers on the Curriculum Committee along with administrators and the Curriculum Coordinator reviewed and approved each curriculum before payment was made to the teachers.

School two also had an active Mentor Program. All new teachers signed an agreement promising that they would actively participate in the Mentor Program for a period of two years. The Mentor Program was organized and facilitated by two teachers. In addition to teaming the new teachers with a mentor, the program offered workshops during the year which were designed to introduce the new teachers to the instructional philosophy which was valued at the school.

School three had a similar organizational design, but approached curriculum a little differently. The organization of the school was based on twelve academic departments, with an average teaching load of five classes on an eight period day, and with a Department Chair released from one or two classes and not categorized as an administrator. The Department Chairs reported to the Principal who was charged with all curricular and instructional demands, supervised all teachers, and was responsible for staff development.

The Principal stated that curriculum development was based on a "grass roots philosophy." Ideas were generated by teachers in the departments, and then proposals were reviewed at a Department Chairs' meeting and again by a Curriculum Council. The Curriculum Council was founded the previous year and was composed of teacher representatives and the Principal. If a course were endorsed by the Department Chairs and Curriculum Council, it was sent to the Superintendent and eventually to the Board of Education for approval.
At school one, with no identified authentic assessors, the academic organization of the school was based on five divisions: math and science; music and art; health and physical education; English, social studies, and foreign language; and electives in business, home economics, and industrial arts. The Principal said that Division Heads had been eliminated as a "cost cutting measure," but that each department had a Lead Teacher. The Lead Teacher positions were not considered administrative, and each Lead Teacher was relieved of a supervisory duty. The Lead Teachers reported to the Coordinator of Curriculum and Instruction who was charged with "overseeing all departments." Teachers taught five classes on an eight period day, with an average class size of twenty-eight students, and had one supervisory period.

When asked about curriculum, the Principal stated that there was a Curriculum Committee with representatives from all the content areas, and that a formal process for proposing changes to the curriculum was being developed.

**Synthesis:** In terms of collegiality, schools two and three had formal processes for encouraging teachers to work collaboratively on curriculum. The majority of the authentic assessors were English teachers, and had previously worked together with one group of teachers providing training for the other group in portfolio assessment. Both schools also had departmental offices or workrooms which provided an environment for collaboration on instruction and assessment. Both schools also had Department Heads or Academic Directors who taught in the department, and took active roles in curriculum development. The English Department Chairs at both schools were respected by their teachers and regarded as instructional leaders.

While schools two and three had the same organizational plan for all departments, the only other authentic assessors at the schools were the Department Heads of the Social Studies Departments. These two teachers seemed to be struggling with bringing about a change in their departments. Both
teachers were extremely active in professional organizations, and had reported that they had made concerted efforts to involve teachers in their departments in these activities. In reviewing the raw data from the Classroom Testing Questionnaire, there were teachers in both Social Studies Departments who had high scores in the areas used to identify authentic assessors. However, their scores were not at the cut-off range used to designate authentic assessors.

The Social Studies Department Head at school two was also the head of the Foreign Language Department. She spoke about how she was trying to engage teachers in working together and in experimenting with authentic assessment. Three foreign language teachers were specifically mentioned. The department head indicated that a French teacher had been working on a final exam in French 1 which required students to engage in authentic tasks at a simulated family reunion/Christmas party. Students were required to order gifts by phone from a mail-order catalogue, issue invitations, order food, and introduce family members at the party. All of the activities that constituted the final exam had to be conducted in the target language. Phone calls were conducted from class with native-speakers on the receiving end of the call who graded the students using a teacher-designed rubric. This French teacher had begun work on the final exam while attending a workshop at Grant Wiggins' Center on Learning, Assessment, and School Structure in Geneseo, New York during the summer of 1994. The department also spoke of Grant Wiggins' plans to send a video technician to school two in the spring of 1995 to film this assessment activity for possible inclusion in a teacher-training video.

Spanish teachers, at school two, were also involved in authentic assessment activities. One teacher had served as an author for McDougall-Littell's Spanish series Bravo! Another teacher was in the process of developing authentic exercises for the student workbook which accompanied the Bravo! texts.
According to the department head, the foreign language teachers were in a state of transition. They were moving toward developing and using authentic assessments, but still relied on objective-type tests for much of their grading purposes. In review the raw data, these three teachers had scored far below the cut-off range used for identifying teachers as authentic assessors.

Curriculum writing at both schools two and three followed a formal procedure and involved the principals in the processes that were used to bring about instructional and assessment changes. While the principals were members of the school-wide Curriculum Council, they were not active in the actual writing of the curriculum, and felt the leadership in this area should be left to the Department Chairs.

School one did not yet have in place any formal procedures for reviewing curriculum. The Principal was not aware of any curriculum changes in the past few years, and did not have a budget for curriculum revision. Although plans for procedures were being considered, these plans were under the supervision of the Assistant Superintendent of the district and removed from the building level.

School two also had a Mentor Program in place, but this did not appear to have a significant bearing on teachers becoming authentic assessors. Most of the authentic assessors were seasoned veterans with sixteen or more years of teaching experience, and their mentors were the department chairs or university professors.

**Summation of Collegiality:** The authentic assessors:

1. shared department / office space,
2. wrote curriculum in a team approach,
3. shared lesson planning,
4. had established times to confer and share ideas when teaching similar courses,
5. worked in teams to write conference proposals and present at professional meetings,

6. could identify mentors who had influenced their professional growth.

The traditional assessors:

1. were not involved in collegial instructional activities,

2. identified former high school teachers as role models, but were not engaged in formal mentor relationships.

Philosophy

Teacher Interviews: Perhaps the greatest chasm between the authentic and traditional assessors was brought to light in the category labeled philosophy. The teachers' answers were from interview questions 1, 6, 9-12, 16, and 24; and from being asked to talk about their philosophies relating to instruction and assessment, and from being asked to supply a typical classroom assessment and explain how it was developed and how it was used.

The authentic assessors were able to clearly articulate a personal philosophy on instruction and assessment. The authentic assessors could also attribute their philosophies to ways they had been influenced from their readings of professional books and journals. The authentic assessors also discussed how students were involved and played a part in designing assessment and in evaluating their own progress.

In contrast the traditional assessors gave a general philosophy statement with references to the fact that they were doing the best job they could in preparing students for college, the work world, or the 21st Century. Some traditional assessors were familiar with authentic assessment in general terms, but none had experimented with it in their classrooms. The traditional assessors who were familiar with authentic assessment felt it was not appropriate nor could it be integrated in their content fields. The traditional assessors did not involve
students in planning instruction and assessment, nor did they involve students as self-assessors.

The differences between the two groups are perhaps best illuminated by their statements regarding their personal philosophies on teaching, the ways in which they developed or changed their paradigms on assessment, and the role of students in instruction and assessment.

When asked about their personal philosophies on instruction and assessment, the authentic assessors were not at a loss for words. They could clearly articulate how and why instruction and assessment should be integrated, could cite influences on their personal philosophies, and could give examples of assessments they used that had personal meaning for students and provided a link to the real world. The following represent some of their comments.

Brain-based research has helped me approach instruction differently. I want to help kids transfer their learning strategies across content areas. Instead of integrating the curriculum we need to start integrating the process of how kids learn.

I read Gardner's stuff, and I like his ideas on how kids can demonstrate their understanding other than by writing. I'm now trying to have some of the kids who are really artistic, draw a picture of the story, talk about it, and then write. I have these students use their art work as sort of an advanced organizer or map. It helps them to get started writing because they usually can't get going. I've also had kids make videos and perform plays to demonstrate their understanding.

Harste believes curriculum should be inquiry - like Gardner he thinks there are multiple ways of knowing. I'm trying to get the kids to see how they need to interpret from different perspectives. For example, if there was a car accident a biologist might be looking at the body parts, a writer at the emotions of the people, a psychologist at the trauma and how people would deal with it, and a P.E. person at how long rehabilitation might take for those who were injured. Kids need to see that there is not always one right way or one right answer.
I've been influenced by Newmann and Wiggins. I remember two main ideas when I design instruction and assessment. Newman says that assessment should have disciplined inquiry, integration of knowledge, and value beyond evaluation. Wiggins asks: "What counts as evidence in what you're going to assess?" and "So what?" If I can't outline or describe what the learner will know now that he didn't before, and what evidence there will be of this learning, than I go back to square zero in designing the assessment.

In my law class, the content is based on real life. For example, kids that serve as jurors in a trial make decisions based on the evidence that is presented. This is linked to the real world because there is a high probability that they will serve on a jury some day. We also discuss issues that are present in our society. We watched the movie Philadelphia where a lawyer is fired because he has AIDS. The movie models for them what they will have to do in a mock trial. The cases they try are based on legitimate issues that people wrestle with in our system such as custody cases, teen abortions, and drunken driving.

For my students I believe the first step in real learning and assessment involves the community in which they live. In Political Science class, the kids researched local issues based on the candidates running for election. They [students] contacted the candidates, created a survey to give them, analyzed the data, drew conclusions, and then wrote position papers on each candidate based on the issues they researched.

The traditional assessors seemed to be uncertain of what response they were supposed to give when asked about their personal philosophies. None of the teachers cited readings nor nationally known educators as being influential in establishing a personal philosophy. Many of the traditional assessors indicated that they never had to answer this types of question, and consequently had given it little thought. Their responses were vague and general.

I do the best job I can in presenting the material in a way that they [students] can understand. I lay the foundation for later growth.
I try to mold the students' minds and give them support. Many of them think they can't do it, and they need someone there who will tell them that they can.

I test the same way that I took tests when I was a high school and college student. I will also give a kid credit for homework and for his attitude. I won't fail a kid if I think he's really trying.

I cover the material in the textbook. If you read the way I teach on paper, you'd probably hate me. It sounds boring, but I try to develop relationships with the kids and have them learn to trust and respect me. I get in few - really zero - blatant disagreements with kids and have no complaints from parents. I don't attribute that to good teaching. It's just because I care about the kids. I could fail a kid, but the kid knows I care.

I just do the best job I can for the students. I'm retiring at the end of this year. I really want to keep working so I hope I can find a job. I just can't stand teaching anymore. It is so boring going into the classroom and doing the same thing day after day.

The authentic assessors differed from the traditional assessors in being able to describe how and when they changed their paradigms regarding instruction and assessment. Their answers indicated that they had spent time in self-reflection and evaluation.

I used to have beautiful lesson plans that I developed in the first five years of teaching. They were all in plastic sheet protectors, and in three-ring binders organized by the courses I taught. Everything was laid out for the kids. I thought I was a good teacher, but I wasn't especially good - I was entertaining and because of that, well liked. I can't believe how much more satisfying and exciting teaching is this way. I feel like I've had one of those experiences Born Again Christians describe. It was like I had a funeral for all my old binders. I hated to see them go because they represented so much time and work, but I knew it was time to burn them.
Before I became certified in English, I taught P.E. and home economics. Both of those are hands-on demonstration classes. Students had to demonstrate their learning by doing authentic tasks valued in the real world - by playing tennis, sewing a skirt, or cooking a meal. Since that time, I've been trying to duplicate that learning and testing in English class. The Whole Language movement helped me to see the link and began my conversion.

I always knew a piece of the puzzle was missing. I tried new things in the classroom all the time. I was moving toward involving students, but I was still lecturing a good deal of the time. When I came here [school two], it was like finding some soul mates. By talking to some of the other English teachers I could see what they were doing and how I could make this same thing work in social studies. The missing link was letting the students know what was going on and what they were going to do. Before, I just assumed that they saw the link between my objectives and assessments. When you let them [students] in on the secret, teaching becomes much easier.

Most of the traditional assessors did not see the need for change. One of the math teachers was aware that the professional standards in his field was calling for the use of authentic assessment in the teaching of mathematics, but he felt that the teaching of math would always involve paper and pencil testing and that using authentic assessment in introductory math courses was not possible.

The NCTM [National Council of Teachers of Mathematics] standards are trying to get away from rote memorization, and move to application - I totally agree. That's why I like the fact that our department is pushing graphing calculators. But real life examples always involve numbers. You can have them [students] build pyramids, but it will still come down to paper and pencil.

My program [the Saxon math program] has generated results. Theirs [NCTM standards] is based on who they are and what they hope will happen. They have no proof this [call for authentic assessment] will produce results. They can't say look at this increase in ACT scores like we can.
The professional standards [NCTM] say that we should use graphing calculators and every day math situations which represents applied math. I don't do it because in Algebra 1 and 2 it is difficult to do. The basic courses don't have applications to every day life. It's probably easier to do if you're teaching advanced math.

The traditional assessors also relied heavily on multiple choice, short answer, and matching formats for their assessments. Their rationales for this type of testing was based on the models that they had been provided with as students. Most had not considered testing any other way, and did not seem to be aware of ways in which students could demonstrate their learning other than by this type of format. One teacher described changing his scoring format, but it was still based on his traditional style of testing.

I use multiple choice and short answer - one to two words - one-hundred percent of the time. It is easiest for me and for the students. I can grade the papers quickly with the scan-tron, and the students know what to expect. I do this because I have always taken tests of this type - from high school through college.

There really isn't time to get into long, involved projects. There is so much content to cover in the science courses and so little time, that I use objective type tests. This moves the curriculum along.

I started thinking about what that speaker [Kay Burke] said in our workshop on rubrics. If there is a score range of 1 to 5, most people give a 3 because it is the middle of the road. I don't use rubrics, but I give kids partial credit for showing their work in math. I don't give 1 or 3 points anymore. I give 2 or 4 points. If they had the right approach, I'll give 4 points; if they had the wrong approach, I'll give 2 points.

The authentic assessors were very clear regarding how and why students need to be involved in instruction and curriculum. Students' involvement was a major part of their planning. Additionally, the authentic assessors felt that it was important for their students to know them as human beings not as just teachers.
The authentic assessors provided examples of how they were partners in the learning process, and shared not only their personal lives with students, but their own learning processes.

We need to listen to students more. Their voices need to be heard when we're planning curriculum, instruction, and assessment.

Students need to be involved in their own learning. They need to be valued as individuals and be willing to voice their opinions. I have to create a comfortable environment for that to happen. Choice is crucial. Classrooms need to be a model for democracy.

Students are involved in the creation of all assessments. We create them inductively before the kids begin to work on a project. I try to immerse them in experiences that relate to what they will do. Then we determine effective characteristics, design a grade sheet, and assign points to different categories on the grade sheet. I have three different classes, and they all use different assessments and grading sheets - it's no problem at all. The kids are determining all the grading for their poetry unit. They have to select a poet, tell about him and his time period, give a dramatic interpretation of a poem, summarize it, interpret it, and use a visual aid. They established criteria for every one of these points. Each group will have their presentation video taped, and then they will grade their own project. We use the grading criteria to start our discussion on what grade they earned. This is the first time kids will grade their own work. They have designed and used rubrics before, and typically their grades are higher and the quality of their work is better because they fully understand the expectations.

When kids design a rubric it isn't without my guidance. We review a lot of models I provide. For instance, if they are going to write an opinion essay, I give them some good ones and some bad ones. Kids use the language familiar to them to verbalize a concept, and then I teach them that we may call it thesis statement or transition. Our critics [of authentic assessment] say we are not covering these basics - but, it's just that we aren't covering them up front. Kids can come up with the right ideas, and then we just connect the terms.
I don't do anything formal with learning styles, but we talk about different styles as we use different strategies. I tell the kids that I really struggle when we do authors' circle because I am a visual learner, and I have trouble paying attention when the information is auditory.

Curriculum, instruction, and assessment must be interwoven. Ultimately the learner must see the link between content, process, and how success will be measured. If they [learners] don't see this link, what they do in the classroom is just an exercise and it has no meaning. The learning has to have real content and not be devoid of what is happening in the real world.

I share everything with my students - within reason of course. If I expect them to take risks in my classroom and be vulnerable as learners, they have to know that I have my warts too. They know about my family, my husband, etc. I don't mean to sound like I take up all the class time rattling on about my personal life, I use examples when it is appropriate so that they get to know me as a person.

The traditional assessors, however, did not involve students in instruction and assessment. They provided some examples of what they shared with students about their personal lives, but indicated that this information was kept to a minimum. Most of the traditional assessors felt it was the responsibility of the student to adapt to his or her environment. By teaching students to adapt to their teaching and assessment styles, the traditional assessors felt that they were preparing students for the real world of work.

We have to teach kids that different disciplines have different characteristics. It's not fair to teach kids that math is always fun - it isn't - it's lots of drill and practice.

Yes, I'm familiar with all that learning styles stuff. But my methods don't account for different learning styles, and I don't feel bad about it. We're letting kids off the hook when we teach them that they don't have to respond to the situation. I can do the touchy-feely, but it's not what kids need, and it's not what life is like.
If a kid can't succeed in my classroom, that should indicate to the kid where he's going in life. When I was in high school, I had trouble with artsy-craftsy things, so I didn't go into acting.

I do lots of volunteer work with the Hispanic community. I tell the students about this to show them that they will be able to use Spanish in the community in which they live.

I don't really tell the kids anything about myself. Sometimes a kid will ask about something because he is curious. If I feel it's something I want to share, I'll answer his question. But I don't regularly volunteer information.

I enjoy yard work, painting, and gardening but for short periods of time. No, I don't tell the kids any of this. It's none of their damn business what I do.

Principal Interviews: The principals of the school were asked to describe their schools and to explain the vision and mission that drove the instructional and assessment program. At school two, the Principal described the school as "evolving," and attributed the many changes the school had undergone in the last few years to the past Superintendent. He described that Superintendent as a curriculum leader with a "hands-on" style of leadership. Since there was only one school in the district, the Principal and Superintendent had shared many of the leadership roles, and had worked as a team to energize the faculty and bring about needed changes in the instructional and assessment programs. He explained that the school had been rather stagnant under the previous leadership of a Superintendent who had served thirteen years and been quite ill during his last few years until his death during the mid-term of the 1989-90 school year.

The Principal was able to clearly articulate the vision and mission statement of the school, and felt both served as filters through which all proposed changes were processed. The vision was to build a "student-centered" instructional environment that best served the diverse needs of the growing
population. The mission statement was "To provide interaction that fosters commitment to learning, self, and others."

The Principal felt the changes under the immediate past Superintendent had been very positive and had resulted in many honors for the school including the U.S. Department of Education Blue Ribbon Award and the selection as one of three schools featured in the Mid-Continent Regional Laboratory's (McREL) video entitled "For Our Students, For Ourselves: Learner-Centered Principles In Practice."

The Principal also noted that he was challenged by the reluctance of some faculty members to accept the new "vision" of the school. The Principal stated that some of the older faculty members felt threatened by the emphasis on curricular and instructional revision which focused on the students as active participants in the instructional process. The Principal explained that these faculty members felt the revisions were an indictment of their current instructional styles which were dominated by the lecture approach followed by multiple-choice tests which were easily scored by a Scan-tron machine.

At school three, the Principal described the school as a very academically oriented school and close to the ideal cycle of education. She believed that the community members and parents valued and respected education, and students were sent to school for a definite purpose. She felt this was evidenced by the fact that 93% of the senior class went on to college. She said that the school was not bothered by gangs nor violence which enabled faculty to spend time on curriculum and approaches to instruction that worked. The Principal expressed the mission of the school as "providing a rigorous academic program within a nurturing environment." She felt the vision was to be able to provide individually for all the students in ways that would best suit their special needs, and felt that with a population of just over 1,000 this should be possible.
The interview with the Principal school one started with his explanation of the district's recent reorganization. He explained that the high school was part of a K-12 district that had previously been three separate school districts - with the high school as one of those districts. A 1989 referendum consolidated the districts for the purpose of combining the financial resources to "brighten the future" for two of the financially troubled districts. The high school was one of the financially troubled districts. The Principal stated that even though the rationale behind the districts' consolidation was considered sound, the anticipated outcome did not occur. The Principal stated that the consolidation required an equalization of salaries for all certified and classified staff in the three districts which "ate up all the money" that was originally viewed as available for areas other than salary. The Principal also spoke about the problem of impact aid which had contributed to the district's financial problems. He stated that roughly twenty-seven percent of the high school students were government dependents, and that the school received only roughly fifty percent of the actual cost of educating these students. The Principal stated that the high school used to be one of the best schools in the county, but that the lack of federal aid over a number of years had significantly impacted the financial picture.

The high school had a current enrollment of 811 students, and was described as having the foundation and potential to become an excellent school. The Principal felt that with adequate financial and human resources and his leadership, he could make an impact on the educational process at the high school. His personal goals for the school included raising test scores, improving academic performance, improving behavior management, involving parents and community members in the educational process, accentuating the positive aspects of the school, and down playing the negatives and negative perceptions which he felt were greatly exaggerated by the press and media.
When asked about the mission and vision of the school, the Principal replied that it was "to prepare students for the 21st Century." He stated that this was not a district or building vision that had been formally adopted, but rather his own vision. He felt the current school year was being devoted to "assessing and evaluating," and that it would be followed next year by "charting the course" for the high school and district. The Principal felt that the formal goals and objectives of the high school were ones he "had inherited," and that he had not yet had time to develop his own.

**Synthesis:** There appeared to be a positive correlation between the number of authentic assessors identified at the schools and the clarity of the school's vision or mission. At school two, the vision and mission were reinforced in a variety of ways. The Principal stated that the "student centered" philosophy was a guiding principle in reviewing and approving curriculum and extended into decisions regarding extra-curricular activities. The teachers at this school, whether they agreed or disagreed with the vision, were aware of the district's goals and direction.

At school three, the vision seemed less pronounced and less relevant to decision-making. Many of the teachers interviewed, both authentic and traditional assessors, referred to a document that stated the school's official philosophy. Some even admitted to working on a committee that wrote the document, but none could remember exactly what it said. These teachers called it some "educationaleze" or "jargon" that was on a shelf somewhere and was probably written for a North Central Evaluation. At school three the authentic assessors, who were mostly English teachers, seemed to have developed their own departmental philosophy. The traditional assessors appeared to be following their own agendas with little to no concern regarding a school-wide vision.
A positive relationship between the number of authentic assessors and the involvement of the administration in curricular matters was also noted. At school two, the Principal was keenly aware of the curriculum - what was being taught, how assessment was conducted, what changes were made, etc. The Principal also explained that the past Superintendent had also been a "hands-on" administrator. Since the district had only one school, the Superintendent had worked directly with department heads, teachers, and the curriculum council to reinforce the philosophy and vision expected in curriculum development and student assessment. The Principal felt he was following the course she had set in working closely with these same groups in curriculum matters.

School three evidenced a lower level of administrative involvement. The Principal felt the Superintendent was fairly removed from curriculum input and decision-making. The Principal served on the Curriculum Committee, but did not appear to be as knowledgeable about assessments used and curricular developments. This perception was reinforced during the teacher interviews by teachers who stated that they thought the Principal could probably describe the curriculum in general terms, but could not describe the types of assessments used in daily instruction.

Both schools two and three had administrators involved in working with teachers on curriculum whereas in school one - with no identified authentic assessors - the administration seemed to be removed from the curriculum process. At school one the "mission/vision" questions were interpreted by the Principal as personal goals. This Principal seemed to be overwhelmed by financial problems and daily operation of the school. He spoke about managing student behavior and improving the image of the school. He felt he had not yet had time to develop a vision or mission and was not aware of a district vision. Similarly, the traditional assessors interviewed at school one had a relatively narrow interpretation of vision and mission. Some teachers did not seem to
understand the question in terms of a "school's or district's" direction, and answered in terms of why they thought students attended school. One teacher responded by saying she thought the students had different missions. Some came to cause trouble, some to socialize, and some to get an education. When the question was rephrased or asked again in terms of "vision of the school" teachers still responded in a personal way by indicating they tried to help students develop a positive self-concept or see the worth of a high school diploma.

**Summation of Philosophy:** The authentic assessors:

1. could articulate how and why instruction and assessment should be integrated,
2. could cite influences on their personal philosophies,
3. could give examples of assessments used that had personal meaning for students and provided a link to the real world,
4. could describe how and when they changed their paradigms regarding instruction and assessment,
5. had spent time in self-reflection and evaluation,
6. gave reasons how and why students needed to be involved in instruction and curriculum planning,
7. felt it was important for their students to know them as human beings not as just teachers,
8. provided examples of how students and teachers were partners in the learning process, and shared not only their personal lives with students, but their own learning processes.

The traditional assessors:

1. had generalized philosophy statements such as "doing their best" for students,
2. taught the way they had been taught in high school,
3. felt their subject matter did not lend itself to real world applications,
4. relied heavily on objective testing and did not see a need for change,
5. followed their own agendas with little or no regard for a school-wide vision,
6. did not involve students in planning instruction and assessment, nor did they teach students to be self-assessors,
7. shared little about their personal lives with students.

Support Systems

Teacher Interviews: The last way in which traditional assessors and authentic assessors differed can be described in terms of the support systems that were present at their schools and their perceptions of these support systems. Teachers' responses were from questions 15, 17-19, and 21 in the In-Depth Interview Questionnaire. The items that were included under the category support systems were the involvement of the administration in classroom assessment, the climate of the department and school, the attitude of teachers and administrators toward experimentation and empowerment, and the financial support available for training and curriculum writing.

The authentic assessors felt empowered in their classrooms and schools. They felt the administration was knowledgeable about and supportive of their assessment efforts and of experimentation in terms of instruction and assessment. For the most part, the traditional assessors did not feel the administration was either supportive or non-supportive of their assessment practices, but ignorant of what actually happened in the classroom. Some felt the administration was imposing or mandating change for change sake with little or no understanding of how the change would impact content areas.

Most of the authentic assessors agreed that the knowledge and involvement of the administration regarding classroom assessment was greatest with the department chair and diminished in the cases of the principal and
superintendent. All of these teachers also felt that this was appropriate. The department chairs at schools two and three, with identified authentic assessors, were both described as very involved in what was going on in the classroom and very supportive of the teachers' efforts. The chairs were considered part of the team and not viewed as outside administrators. The authentic assessors saw the administration as supporting their efforts. The English teachers at school three reported that their teaching load had been reduced by one class so that they were available to conference with the students regarding their portfolios and writing assignments. These teachers saw this as direct administrative support for their efforts.

The authentic assessors also saw the climate at their school to be very conducive to experimentation and empowerment. They felt that they were entrusted with curriculum decisions and regarding as professionals by the administration and Board of Education. Some of the authentic assessors also felt that this empowerment was acting as a divisive factor between them and the teachers who clung to traditional modes of instruction and assessment.

We [teachers] have total professional freedom as long as there is a research base for the decisions we make. This [freedom] frightens some people in the building. There is no excuse for not experimenting. People used to say there is no administrative support, but "I" am current. Now the administration is saying "go ahead" and it terrifies some people.

Experimentation is extremely time consuming. Those [teachers] not involved leave here at 3:45 PM with nothing in hand. They don't want to devote the time.

There was also financial support for training and curriculum writing at schools two and three. School two reimbursed teachers up to $150 per approved, graduate level course. This was considered to be one of the highest reimbursement levels in the county. A year-long, inservice program was also
being conducted to provide training for teachers in authentic assessment. The culmination activity was a "best practices" session in which teachers explained what types of assessment they had experimented with and what results they had experienced. A handbook with model assessments used by the teachers in the school was printed for every staff member and was used as a reference during the inservice program. The teachers at school three also reported that money was available for them to travel to conferences and conventions to learn about authentic assessment techniques, and that they were financially rewarded for working on curriculum. Teachers were generally paid between $400 and $1,000 for summer curriculum work depending on whether the work involved revision of a course or a new course, and whether they worked in a small team of two people or a large team of ten to twelve.

Authentic assessors at school three also reported financial support for their efforts. Departments could apply for "summer workshop" monies which enabled them to be reimbursed for the time they spent revising curriculum. Teachers were paid $32 per hour for curriculum development. Teachers felt there was enough money allotted for travel, and felt encouraged to attend professional conferences and meetings.

The traditional assessors had a different view of the support system in their schools, even though some of these teachers were in the same schools as the authentic assessors. For the most part they believed the department chair was knowledgeable about and involved in what they were doing in the classroom, but felt the principal and superintendent had little knowledge of what actually happened in the classroom. One teacher described the administration's involvement as follows:

I think the principal could probably present an overview to some community group, and that's probably her job. She could describe in general terms what we are doing, but couldn't describe the types of tests that I give.
The traditional assessors also responded differently to the climate question. While the authentic assessors talked about collegiality and academic freedom, the traditional assessors interpreted the climate question in terms of department members teaching at the same pace, a relaxed administrative structure, and the presence or absence of interpersonal conflicts. A sampling of teachers' answers included the following:

The climate is better than it was a few years ago. People in my department are now working together, using the same textbook, mandating the homework examples at the end of the chapter, and trying to get to the same place in the text at the same time. The climate is better too because people aren't being reprimanded all the time if they are late - people are more relaxed now.

Our department is filled with turmoil. Some teachers are supportive, and some are the exact opposite. The problems are both personal and related to curriculum. I don't really want to talk about it, because it is a big focus at our school right now. But you can say we are definitely not a cohesive unit.

The traditional assessors also responded to the questions on experimentation and empowerment in a different light. Where the authentic assessors answered in terms of experimentation based on research and ideas garnered at conferences and workshops, the traditional assessors answered in terms of personal freedom in their classrooms. The traditional assessors also saw experimentation as isolated instances rather than ongoing efforts to improve the curriculum. One teacher also described experimentation in very narrow terms as using a variation of an objective test item in lieu of those normally used. Other teachers provided responses that suggested little appreciation for revising the curriculum and for the administration's support of experimentation. The following responses typify their answers.

We tried to change. The four pre-algebra teachers got together and planned what do do every week because the math standards
say to try cooperative learning, hands-on experiments, and things
to hit different learning styles. We had the kids in the computer lab
for a week working on statistics and doing graphs. On Friday, they
were supposed to do a presentation using a circle graph. In six
sections of the class, only 25% did it or wanted to do it. Some kids
didn't want to make the graph, others didn't want to do the
presentation. I even modeled a good and bad presentation. I've
had the most success in the last five weeks with here's the
worksheet, get it done, turn it in tomorrow. You have to look at the
maturity level and responsiveness of the kids. Kids don't want to
work in cooperative groups in our classes because they've already
had it four hours all day. They want to sit down, open the book,
and work.

Sure we experiment in my department all the time. There are only
three teachers, but we discuss what we do on tests. If a teacher
has a good matching section, we may try it on a test and not use
short answers or as many multiple-choice questions.

The ASCD [Association for Supervision and Curriculum
Development] that's a great journal for this school. It says here's
the fad of the month, try it. We have teachers who just love that
stuff.

The financial support available to the traditional assessors at the two
schools was the same as that available to the authentic assessors. For the most
part, the traditional assessors did not avail themselves of the opportunities that
existed.

I haven't worked on curriculum recently because I'm not interested.
They [teachers in the department] rewrote some stuff over the
summer, but if you just follow the textbook you cover the
curriculum.

These teachers did not participate in curriculum work nor express interest
in attending professional meetings. Only one teacher indicated attendance,
during the past year, at a national conference which was held in Chicago. At
school one, with only traditional assessors, the financial support was basically
non-existent. Monies were not freely available for curriculum revision and professional travel. An anticipated reorganization of the district-wide curriculum was to include a proposed budget for these expenses.

**Principal Interviews:** The interviews with the principals sought to identify the supporting foundation for teachers being able to become authentic assessors. Topics included relations with the Board of Education, leadership history, and climate and teacher empowerment.

The Board of Education at school two was described as extremely supportive of the teachers. The Board meet yearly with the Superintendent to establish goals for the coming year and supported curriculum revision with financial backing. The Board met twice per month, and at the second meeting of each month recognized students and teachers for their accomplishments by awarding plaques and certificates of recognition. One meeting per month also included a presentation on curriculum or the work-in-progress of one of the school committees. Teachers preparing a professional presentation were often asked to provide a "dress rehearsal" for the Board. The Principal felt the Board knew the teachers individually and were aware of their professional accomplishments both in and out of the classroom.

The Principal had been at the school for the past 35 years; the last ten as Principal and previously as a social studies teacher and dean. During the 1994-95 school year, an Interim Superintendent was on board due to the resignation of the past Superintendent who had been employed for four years. A new Superintendent had been hired and was to assume his post as of July 1, 1995.

The climate at school two was described as "positive." There had never been a teachers' strike, and relations between the school Board and the teachers' organization were characterized as "friendly." The Principal stressed that the word "union" was never used, and that the local teachers' bargaining unit was always referred to as an "organization." The principal felt the positive
climate at school two was due in part to teacher empowerment. Teachers were charged with all curriculum revisions, and were financially compensated for their work outside of the school day. Teachers were also responsible for the inservice and mentor programs, and were active participants on the curriculum, evaluation, and other committees.

The Principal felt the strength of school two was its focus on being "student-centered." "Not one decision is made in this building unless it is good for students," said the Principal. "We encourage teachers to experiment in the classroom and take risks only to bring about greater student success." Another strength was the "family" atmosphere that pervaded the environment. The Principal felt the size of the school enabled the faculty and students to know one another. The Principal was proud of the Blue Ribbon Award in 1994, the recognition by McREL, and the professional involvement of the faculty. Many faculty members had presented at local, state, and national conferences such as the American Association of School Administrators (AASA), the Association for Supervision and Curriculum Development (ASCD), the National Council of Teachers of English (NCTE), the Leadership Forum, and others.

School two also had a generous curriculum writing budget which reimbursed teachers for working in teams during the summer and writing curriculum which always included designing assessments. The average budget allotted to curriculum development was $35,000 to $40,000 for the past three years. The Principal commented that the proposed budget for curriculum work for the FY96 year was $54,000. This figure was to be offset with grant monies resulting in a district cost of $37,000. Teachers were generally paid between $500 and $1,000 for curriculum writing based on how many teachers were involved in writing a particular course of study.

The Principal at school three was in her ninth year as Principal of this school. The Superintendent was in his third year in that post, and had followed a
retiring Superintendent who had served as Superintendent for twenty-nine years. The Principal stated that the retired Superintendent had a "wonderful reign" during which the climate and culture of the school were very stable. While the new Superintendent was characterized as "energetic" he was also described as "task oriented" and contrasted with the retired Superintendent who was "people oriented." The Principal said that while people knew it was time for a change in leadership there was a "mourning period" for the style of leadership that was now gone and an anxiety in facing the unknown. The Principal believed that a stabilization was taking effect in the school climate with staff becoming more comfortable with the new Superintendent and knowing what to expect.

The Principal described relations with the Board of Education as generally good with the Board making overtures to promote the "healing process" which was necessary following negotiations three years ago. During these negotiations, the faculty developed a distrust of the Board when the Board failed to respond to negotiation talks in a timely manner. The "healing overtures" were described as dinners during which the Board members met with invited faculty to discuss concerns. The Principal also felt that the Board had started to "micro-manage" school three during the latter years of the retired Superintendent. The Principal also stated that two CEOs were recently elected to the Board and that their business experience, along with reminders by the current Superintendent, had helped the Board to see that their role was not daily management of the school. The Principal described the Board members as educated and articulate and as taking an active and energetic role as Board members.

The climate of school three was described as "settling down" based on the turnover in the superintendency. The Principal stated that twenty-five percent of the faculty was new in the past two years, due to the "5 + 5" retirement incentive, and that the new teachers were aligned with the new Superintendent. The Principal also felt that the "old guard" were getting ideas
from new faculty which was enabling them to make positive changes in their instructional practices. She also felt that the new administrative structure with six administrators reporting to the Superintendent was helping to break down the "we versus them" attitude.

School three also had a curriculum writing budget that allowed teachers to apply for a stipend based on a pay rate of $32 per hour. The previous year's budget for curriculum writing was $55,000.

At school one the Principal was in his first year as Principal of the high school, returning to the district after a two year absence. The Principal had first served in the district beginning in 1990 as an Athletic Director and Director of Intramural Sports for grades 6-12 for one year, and then for one year as Assistant Principal and Athletic Director of the high school. The Principal then left the district for an Assistant Principal position in a Chicago suburban district where he served for two years. Upon his return to the district as the high school Principal, he was allowed to bring in a "hand-picked team" of administrators from "in-house" and "out-of-house" which included the Associate Principal, the Head Dean, the Head Counselor, and the Athletic Director.

The Principal also stated that the majority of the central office staff was new this year including the Superintendent, the Associate Superintendent, the Business Manager, and the Director of Student Affairs and Special Education. He stated that the instability of leadership in the district contributed to their problems. In the past four and one-half years, six superintendents had come and gone. In reverse chronological order, the current Superintendent had just been hired as of January in the 1994-95 school year, and had been in the position less than one month. His predecessor had been an Interim Superintendent for a total of four days taking over from a previous Interim Superintendent who was a retired administrator and had served his maximum of one hundred days beginning with the start of the 1994-95 school year. Another
Interim Superintendent had completed the 1993-94 school term when a Superintendent who had served six months was released after being hired at the beginning of that school year. The Superintendent prior to the 1993-94 school year had served for two and one-half years.

Relations with the Board of Education were described as "another sticky-wicket" because the previous school year - 1993-94 - had begun with the longest strike in the history of the school lasting six weeks. The Principal stated that only two members of the Board were left from his previous employment in the district between 1990 and 1992. The Principal felt the Board had a diverse make-up with parents, educators, and business people making up the current Board. The Board also had non-voting members from the nearby government facility which accounted for twenty-seven percent of the district's students. The Principal felt the Board was typical in that members could form coalitions "depending upon the issues on the table."

The Principal said that he was trying to improve the climate of school one by empowering teachers to share in the decision making process. He wanted teachers to have the opportunity to have their voices heard in all phases concerning the operation of the school. He said that committees had been established in all areas to address concerns and issues of the school. Specifically, these committees included the Discipline Committee, the Curriculum Committee, the Social Committee, and the Principal's Advisory Committee. The Principal also met with the faculty once a month, and held "Fireside Chats" once a month. The "Chats" were open to any teachers who wanted to attend and discuss issues.

The Principal of school one stated that there was a "Special Projects" rate of pay which ranged from $12 to $19 per hour, and was used to reimburse teachers for a variety of extra assignments. The Principal did not know how much had been spent on curriculum development during the previous year, but
said it was probably "very little due to the turmoil." The "turmoil" he explained was the fact that school had been in session until the end of June due to the teachers' strike at the beginning of the school term. The Principal also stated that "no dollar amount had been budged for curriculum development" in the current school year, and that the amount was "kind of open." The Principal explained that the new Associate Superintendent was focusing on the curriculum at every building in the district. The Principal said he expected to see money budged for the curriculum alignment process along with released time for individuals to serve on a district-wide committee dedicated to bringing about the proposed curriculum alignment.

**Synthesis:** The support systems at schools two and three, with identified authentic assessors, differed dramatically from school one where only traditional assessors were identified. At schools two and three empowerment and experimentation were encouraged. Teachers were involved in committees that planned and made decisions regarding teacher inservice programs and curriculum development. There was a focus on continual learning and instructional improvement.

Schools two and three also shared a feature of stability in terms of leadership. The Principals had held their posts for over nine years, with one Principal having 35 years of experience in the school. Similarly, the positions of Superintendent had a stable history with 16 and 4 years by two consecutive Superintendents at school two, and 29 and 3 years by Superintendents at school three. Financial resources were also present at these schools. The annual budget range reported for curriculum development was between $30,000 and $55,000. Adequate budgets also existed for inservice programs, professional travel, and graduate study reimbursement.

Climate at the schools also appeared to be relatively positive. Although one traditional assessor described his department in a state of turmoil, no other
teachers remarked on this situation. The Boards of Education appeared supportive of the teachers' efforts, and one Board was making conciliatory overtures following negotiation misunderstandings three years ago. Both schools two and three reported a strike-free history.

In contrast, school one lacked many of these support systems. Teachers did not talk about empowerment in terms of school-wide issues, but felt they could experiment in their classrooms if they so chose. The leadership history at the school and district level were also marked by instability. The Principal was in his first year in that post, and the district had seen six Superintendents in the past four years. Financial resources were also lacking. Although a dollar figure existed for compensating teachers on a per hour basis for curriculum work, the Principal was not aware of the school's budget nor of any work that had been completed in the past year. Financial difficulties also negatively impacted inservice and professional travel in that no organized inservice program existed, and professional travel seemed nonexistent. The climate in this school was also considered to be "recovering" since the teachers had returned in the fall following the longest strike in the district and one of the longest in the state.

**Summation of Support Systems:** The authentic assessors:

1. felt empowered in their classrooms and schools,
2. felt the administration was knowledgeable about and supportive of their assessment efforts and of experimentation in terms of instruction and assessment,
3. described the department chairs as very involved in what was going on in the classrooms and very supportive of the teachers' efforts,
4. considered the department chairs as part of the team and not as outside administrators,
5. felt the school climate was conducive to experimentation and empowerment,
6. felt entrusted with curriculum decisions,
7. were regarded as professionals by the administration and Boards of Education,
8. had financial support for travel, training, inservice programs, and curriculum writing.

The traditional assessors:

1. felt the administration was ignorant of what actually happened in the classroom,
2. felt the administration was mandating change with little or no understanding of how the change would impact content areas,
3. viewed climate as a relaxed administrative structure and the absence of interpersonal conflicts,
4. viewed experimentation in terms of personal freedom in their classrooms,
5. regarded experimentation as isolated instances rather than ongoing efforts to improve the curriculum,
6. did not avail themselves of the financial opportunities that existed for travel and curriculum writing.

Summary

The data presented in Chapter 4 indicated that there are discernable differences in the methods teachers use for classroom assessment of students' learning. Based on definitions and descriptions found in the literature, these methods were classified as traditional and authentic. The critical attributes that distinguished authentic assessors from traditional assessors included: professionalism, collegiality, philosophy, and support systems.

Professionalism included the propensity of authentic assessors as characterized by their continued involvement in formalized education through college courses, their attendance at professional workshops and conferences,
and their professional leadership roles demonstrated by presenting at
conferences and publishing works.

Collegiality was defined as the authentic assessors' involvement with
colleagues which included cooperative curriculum writing, ongoing interaction
regarding daily lessons, collaboration on professional presentations, and the
existence of defined mentor relationships.

Philosophical differences between the authentic assessors and the
traditional assessors was attributed to the assessors' abilities to articulate their
personal philosophies of assessment based on research and professional
readings, cite ways they had been influenced by these readings, and provide
specific examples of assessments they used which enabled students to make
connections with the real world. Authentic assessors were also able to explain
how and why they had shifted their paradigms concerning assessment to
embrace more authentic measures, to reflect on ways they evaluated their
assessment measures, and to explain how they involved students in the decision
making process concerning assessment.

Support systems that separated authentic assessors from traditional
assessors included both perceptions and realities regarding the administrative
structure of the school, the school climate, the attitudes toward and opportunities
for empowerment and experimentation, and the financial support available for
curriculum and professional development.
Chapter 5 presents a summary of procedures and findings, conclusions, recommendations, and recommendations for further study.

Summary of Procedures

The purpose of this study was to examine the attributes of teachers who had shifted their paradigms from objective testing to become practitioners of the new art deemed authentic assessment. Three research questions were developed. Research question 1 sought to identify teachers in three suburban high schools in Lake County, Illinois who could be identified as practitioners of authentic assessment. Research question 2 attempted to identify the attributes and characteristics which delineated authentic assessors and traditional assessors. Research question 3 proposed to identify contributing factors of the schools which enabled teachers to become practitioners of authentic assessment.

The schools involved had similar enrollments, but different financial support bases and student populations. Tables 2 and 3 on pages 58 and 59 outlined the schools' revenue sources and profiled the student populations. The teacher sample included ninety-five teachers in the subject fields of English, social studies, math, science, and foreign language. Data was gathered using three questionnaires. A Classroom Testing Questionnaire (Appendix A) was used as a screening device to identify teachers as those who were practitioners of authentic assessment and those who were using traditional assessment.
practices. An In-Depth Interview Questionnaire (Appendix B) was used with teachers who fell at the two extremes on the screening device. And a Principal Interview Questionnaire (Appendix C) was used to gather background information to support the data gathered from the teachers.

The Classroom Testing Questionnaire was developed based on a review of the literature, and on the previous research of Robert J. Wilson and Richard Stiggins. Permission was requested and granted to use portions of their survey designs in this study (Appendices D and E).

The Classroom Testing Questionnaire was divided into five sub-sections: (a) Demographic and Professional Information, (b) Reasons for Testing, (c) Assessment Practices, (d) Types of Assessments Used, and (e) Level of Use. Numerical ratings were assigned to the responses in sub-sections B through E of the Classroom Testing Questionnaire. An In-Depth Interview Questionnaire was developed based on a review of the literature. The questionnaire was divided into four sub-sections identified as: (a) Personal Factors, (b) Professional Development and Involvement, (c) Administrative /Supervisory Structure, and (d) Assessment Practices.

The data from the questionnaires was analyzed using a combination of descriptive frequency distributions and qualitative analysis.

**Summary of Findings**

The purpose of research question 1 was to use an operational definition of authentic assessment to identify teachers who were practitioners of authentic assessment. Teachers in three high schools in Lake County, Illinois took part in the study. Tables 4 and 5 outlined the total and participating number of teachers in the study as identified by subject area. Ninety-five teachers completed the questionnaire. Seven teachers were identified as traditional assessors, and eleven teachers were identified as authentic assessors. The means of
identifying the teachers as authentic and traditional assessors was discussed on pages 65-70. Tables 10 through 16 and Figures 2 and 3 profile the teachers identified as authentic assessors using the data from the questionnaire.

The authentic assessors were found to use performance assessments more than 90% of the time in instructional testing, while traditional teachers used performance assessments less than 10% of the time. The authentic and traditional assessors were also consistent in responding to descriptors which profiled assessment practices.

Authentic assessors were generally found in the teaching fields of English and social studies, had more than sixteen years of experience, were in their forties, and had a Master's plus fifteen hours or more of graduate credit. Authentic assessors were also active in continuing education having taken a graduate course or workshop in assessment procedures during the past year.

Authentic assessors were likely to add to their knowledge of testing through professional collaboration which included team teaching and peer coaching. The authentic assessors felt the most important reasons for classroom testing were to have students practice what they learned and to enable students to monitor their own progress. In planning assessment, authentic assessors were more likely to devote most of the time to structured performance assessments (71.4%) and spontaneous performance assessments (20.5%) than any other type of testing. The authentic assessors were also comfortable using performance assessments; scoring on the average a 3.9 for structured performance assessments and a 3.3 for spontaneous performance assessments on a four point scale which ranged from non-use to comfortable use.

The purpose of research question 2 was to examine the attributes and characteristics of authentic assessors which separated them from traditional assessors. The results from the Classroom Testing Questionnaire were used to
contrast the responses of those teachers identified as authentic assessors with those teachers identified as traditional assessors. The ways in which the two groups of teachers differed are represented in Tables 17 through 25 and in Figures 4 and 5.

In comparing authentic assessors to traditional assessors, the greatest number of authentic assessors were found in English and social studies; while the traditional assessors were evenly distributed in foreign language, science, and math. Based on the results of the 95 teachers in this study, English teachers were found to use performance assessments 63% of the time followed by 43% for social studies teachers, 30% for science teachers, 28% for foreign language teachers, and 25% for math teachers.

No apparent differences were noted in terms of teaching experience or educational level between authentic assessors and traditional assessors. While authentic assessors were most likely to be in the forties, two of the seven traditional assessors were in their fifties and sixties.

In terms of keeping abreast of assessment knowledge, seven of the eleven authentic assessors had taken a graduate course on assessment during the past year, and five of the seven traditional assessors indicated that it had been six years or more since they took a graduate level course in assessment. Similar results were found for the last workshop or conference attended that dealt with assessment procedures. The majority of authentic assessors - ten of the eleven - had attended an assessment workshop or conference during the past year; while four of the seven traditional assessors had attended a workshop or conference in the last one to five year range, and three of the traditional assessors had not attended one in the past ten years.

Authentic assessors indicated that their most important contribution to testing knowledge was from professional collaboration, while traditional assessors relied most heavily on trial and error in the classroom. Both groups
indicated that to have students practice or apply what had been learned was their most important reason for testing. Having students monitor their own progress was ranked second in importance by authentic assessors, while the traditional assessors ranked this item sixth in importance.

The authentic and traditional assessors were also consistent in responding to descriptors of assessment practices. On a scale of one to five, with a five most representative of authentic assessment practices, the authentic assessors' scores ranged from 3.9 to 4.7; while the traditional assessors' score range was between 1.4 and 3.3.

When comparing the time devoted to different types of testing, authentic assessors indicated that they use a combination of teacher made multiple-choice tests, curriculum embedded tests and standardized tests only 1% of the time. The traditional assessors used this same combination of tests 90% of the time, and the average time allotted to this same combination for all the other teachers in the study was 43% of the time. At the other end of the spectrum, authentic assessors used performance assessments 92% of the time, traditional assessors used them 3% of the time, and all the other teachers in the study used them 39% of the time.

The purpose of research question 3 was to find contributing factors which enabled teachers to become practitioners of authentic assessment. Factors included financial indicators, administrative/supervisory structure, school characteristics, and school climate.

In-depth interviews were conducted with sixteen of the teachers who were identified as authentic and traditional assessors and with the three Principals of the schools. Based on the interviews, four areas were identified as enabling or supporting teachers' efforts to become authentic assessors: professionalism, collegiality, philosophy, and support systems.
At two schools both traditional and authentic assessors were identified, and at one school only traditional assessors were identified. In explaining the differences between the teachers and the schools regarding the four identified characteristics (professionalism, collegiality, philosophy, and support systems) it was apparent that a hierarchy existed. If a hierarchy similar to Maslow's could be applied to a school organization, it seemed that two of the schools had this structure in place, and that it was lacking in the third school which had no identified authentic assessors. In these two schools there were ample financial, administrative, and organizational supports to allow teachers to grow and develop professionally. The authentic assessors seemed to avail themselves of this structure, and moved up the hierarchy taking advantage of all the support that was available in the school. The traditional assessors seemed to be at the bottom rungs of this hierarchical ladder. Although opportunities for professional growth and development existed, they did not appear interested or motivated in moving up this ladder.

The school with no identified authentic assessors seemed to have only the rudimentary characteristics or initial steps of a hierarchy in place. This school was struggling with financial problems, student management, changing leadership, and organizational activities. Although some formal processes and committees were being established, they had not yet begun to meet or function on a regular basis.

Conclusions
The following conclusions are based on the limitations of this study which was confined to teachers of English, social studies, science, math, and foreign language in three suburban high schools in Lake County, Illinois.

1. There are a small number of teachers using authentic assessments to measure students' learning. However, the
number of teachers committed to this type of assessment is not consistent with the call for authentic assessment by the various reform movements of governmental and professional organizations.

2. Teachers identified as authentic assessors are characterized by a high degree of professionalism as exhibited by ongoing education, a high degree of collaboration with colleagues, an active role in professional organizations, leadership roles in either administration or curriculum development, and a well defined mentor relationship.

3. Teachers identified as authentic assessors have a strong philosophical understanding of the merits and benefits of authentic assessment as evidenced by professional readings. These teachers are able to support and defend their assessment practices with examples from the literature.

4. Teachers identified as authentic assessors were employed in schools that were supportive of their efforts. Support was demonstrated by the empowering of teachers, the encouragement of experimentation, and the financial backing for staff training and curriculum development.

Recommendations

A significant number of reform movements, as outlined in Chapter 2, call for the use of authentic assessment in the nation's schools. If this paradigm shift from objective testing to authentic assessment is to be realized, a number of factors need to be instituted in terms of administrative practices, inservice education, and support systems.
Regarding administrative practices:

1. Administrators interested in complying with the reform philosophies regarding authentic assessment must familiarize themselves with the research regarding this issue. The authentic assessors, in this study, felt empowered and supported by department chairs, principals, and superintendents who were familiar with the professional literature on authentic assessment, and believed in the benefits of authentic assessment as evidenced by student learning.

2. Employment interviews should include questions on assessment practices. All of the traditional assessors, in this study, based their assessment methods on their own testing experiences in high school and college which were objective-type tests. These teachers saw no need to change their assessment methods since they had never been exposed to any other types of assessments. The hiring of teachers committed to objective testing does not move the school toward assessment reform.

Regarding continuing education:

3. Inservice education as provided by the school needs to be an integrated, ongoing part of the school year. The greatest number of authentic assessors, identified in this study, were from the middle socio-economic school. The school was involved in a year-long program focusing on identifying and experimenting with quality assessments.

4. Continued education must be supported by tuition reimbursement or salary incentive programs. The teachers identified as authentic assessors, in this study, were either
enrolled in Master's or Doctoral level degree programs or were taking courses leading to advanced certification. Both the middle and high socio-economic schools offered tuition reimbursement and/or advancement on the salary schedule for professional development.

5. Continued education must be supported by financial support which allows travel to conferences and workshops. The authentic assessors, in this study, were highly active in professional organizations and professional activities as conference and workshop attendees and presenters, and board members of state or national organizations. Not only released time, but travel expenses were provided for these teachers to engage in these activities.

Regarding support systems:

6. Personal support systems such as mentor programs should be instituted. The authentic assessors, in this study, could all identify a mentor who influenced their assessment philosophy.

7. A framework for collegiality should be established. This framework could include departmental offices, team teaching opportunities, shared planning periods for same-course teachers, and/or opportunities for curriculum writing collaboration. The authentic assessors, in this study, all felt their efforts were supported through professional collaboration.

8. Resource assistance should be available for teachers developing authentic assessments. The largest number of authentic assessors, in this study, were at the middle socio-economic school which had a part-time curriculum coordinator. The curriculum coordinator was available to meet with teachers
individually or in groups to assist in the designing of assessments and complimentary scoring rubrics.

9. Financial reimbursement must be available for teachers for work outside of the school year. Both and middle and high socio-economic schools had ample budgets which allowed teachers to be compensated for curriculum development work during the summer months.

**Recommendations for Further Study**

This study was limited to the identification of the critical attributes which enabled teachers to shift their paradigms from objective testing to authentic assessment. It was further limited to three schools in Lake County, Illinois with similar enrollments and differing economic environments. Based on the limitations of this study, the following are recommendations for future study.

1. This study was limited to three schools within one county. Future research could include a larger geographic region and a greater number of schools.

2. Ninety-five teachers participated in this study from a population of 141 which represents 67% of the teachers. A larger number of participants would yield more reliable results.

3. Of the ninety-five participants, only eleven teachers were identified as authentic assessors and seven as traditional assessors based on the survey used and the limits for selection that were established. Future research might include additional avenues of selection such as identification by the school principal or by department heads.

4. The means for selecting teachers as authentic and traditional assessors was established by identifying teachers who used
authentic assessments over 90% of the time and less than 10% of the time. These cut-off scores identified only teachers at the very extremes of the continuum in terms of using authentic assessments. Selecting more moderate cut-off scores would result in the inclusion of a larger number of teachers for the study.

5. No controls existed for financial resources of the schools. Using the economic support system of a school as a control factor would assist in identifying whether money for programs such as inservice and professional travel was a contributing source in becoming an authentic assessor.

6. No controls existed for subject matter taught. The majority of teachers identified as authentic assessors were English teachers and two were Social Studies teachers. Establishing controls for subject matter would assist in identifying whether certain subject areas are more easily adaptable to the use of authentic assessment.

Although the concept of authentic assessment can be traced to Socrates and Dewey, its use in the recent past has been relatively restricted to classes in fine, performing, and technical arts. Authentic assessment as a way of evaluating student performance in English, social studies, foreign language, math, and science is relatively new. While many journal articles describe its use and benefits, data on whether authentic assessment contributes to student learning is sparse. Professional journal articles usually describe the assessments of teachers and students associated with a major university or research lab. While these articles describe assessments rich in student activities such as investigative practices, computer applications and simulations, and
service learning projects, these assessments are generally not found in the typical high school. Creating quality assessments and scoring procedures can be a painstaking activity requiring much in the way of time and money for training and development. Whether authentic assessment will become a fixture of schools or a passing fad is yet to be seen.
APPENDIX A

CLASSROOM TESTING QUESTIONNAIRE
CLASSROOM TESTING QUESTIONNAIRE

Name: ___________________ Phone: ___________________

School: ___________________

Subject: (circle one) English  math  science  social studies  foreign language

A. Demographic and Professional Information

1. Years of teaching
   ____ less than 5  ____ 6 to 10  ____ 11 to 15  ____ 16 to 20  ____ 20+

2. Sex
   ____ Male  ____ Female

3. Age
   ____20s  ____30s  ____40s  ____50s  ____60s

4. Education
   ____ BA  ____ BA+15  ____ MA
   ____ MA+15  ____ MA+30  ____ Ph.D./Ed.D.

5. Last formal graduate college course that covered assessment procedures
   ____ in past year  ____ 1 to 5 years ago
   ____ 6 to 10 years ago  ____ 10+ years ago

6. Last workshop or conference that covered assessment procedures
   ____ in past year  ____ 1 to 5 years ago
   ____ 6 to 10 years ago  ____ 10+ years ago

7. Place an A next to the one source that made the most important contribution to your knowledge about how to develop and use classroom assessments. Place Bs next to all others that contributed significantly to your current practice. Leave those blank that made minimal or no contribution.

   ____ Undergraduate testing course
   ____ Graduate testing course
   ____ Undergraduate methods course
   ____ Experience as a student
   ____ Information and ideas/ professional reading
   ____ Inservice training
   ____ Trial and error in the classroom
   ____ Professional collaboration/team teaching/peer coaching
B. Reasons for Testing

Directions: Rank order the following purposes of testing. Use numbers from 1 to 10. Number 1 would indicate the most important reason for classroom assessment and number 10 would indicate the least important reason for classroom assessment.

_____ A. To check students' progress against course objectives
_____ B. To compare students' achievement to others.
_____ C. To generate marks for reporting purposes.
_____ D. To insure students do assigned work.
_____ E. To prepare students for this kind of evaluation in the future.
_____ F. To have students practice or apply what has been learned.
_____ G. To diagnose students' weaknesses with the material.
_____ H. To enable students to monitor their own progress.
_____ I. To help me decide what to teach next.
_____ J. To allow me to see how well I taught the material.
C. Assessment Practices

The statements in the two columns below represent a continuum. Mark each statement on where you believe you fall in using classroom assessment measures. A "1" would represent the comment in the left column and a "5" would represent the statements in the right column. Mark each statement with a number between "1" and "5" based on your classroom practice.

Mark each statement. Do not skip any. Do not use an in-between rating scale such as a "3.5" or similar system. Do not use any numbers not between 1 and 5.

1. Students' backgrounds, native skills, prior training are taken into account in assessment designs. All students receive the same assessment measure and are judged against the number of correct answers possible.

2. Assessments are not administered during designated time periods. There is really a blend of assessment and instruction. Assessments are clearly defined and administered at specific times.

3. Assessment involves working with others. Assessments require students to work alone.

4. Assessment involves a judgment on the process a student used. Assessment depends on the product (or correct score) a student receives.

5. Students know how they will be assessed by clearly defined criteria. Assessments are designed around material covered in class, and students can expect varied examples.

6. Assessments are subjective and based on growth of individuals. Assessments are objective and based on correct responses.

7. Assessments require higher order thinking skills. Assessments require students knowing the correct answer.

8. Assessments allow for growth over time and may be made at varying times for individual students. Assessments are specific measurements given at designated intervals over a semester/year.

9. Assessments involve interactions with the teacher and justifications for students answers. Assessments are paper and pencil tests.

10. Assessments include presentations and demonstrations of knowledge. Assessments involve answering essay or multiple-choice questions.
D. Types of Assessments Used

Listed below are six types of assessments that you might use in your classroom. Of the total amount of time you allot for assessment, indicated what percent each of these represent. The numbers should equal 100%. Definitions of each type of test follows.

Teacher-developed paper and pencil tests you develop for your own use in the classroom. This category includes all true-false, multiple-choice, matching, fill-in, and short-answer tests and quizzes which YOU DEVELOP to determine if students have mastered the material taught.

Tests embedded or included in the instructional materials are those that may be found in the textbooks or workbooks you use. They may also be found in an instructor's guide or may take the form of questions at the end of chapters in the materials themselves.

Standardized achievement test batteries are offered by test publishers, such as the Stanford Achievement Test, Comprehensive Test of Basic Skills, Metropolitan Achievement Test, and Iowa Test of Basic Skills. This category also covers state-wide or district-wide tests, including norm and criterion referenced tests.

Oral questioning in the classroom is the daily question-and-answer process used on a day-to-day basis during instruction to track whether individual students or the class as a group are learning the material.

Performance assessments are those assessments in which you, the teacher, observe students in the process of doing things (e.g., speaking or oral reading) or examine products created by students (e.g., writing sample or art project). Then, on the basis of your professional judgment, you judge or rate student performance. Performance assessments take one of two forms. Some are STRUCTURED tests and include: (1) a clearly defined reason for assessment; (2) pre-planned exercises to elicit student responses; (3) a pre-specific response to be evaluated; and (4) carefully spelled out scoring procedures. SPONTANEOUS assessments can be much less structured. A spontaneous classroom event may provide a teacher with an informal opportunity to observe and evaluate a student's performance and to judge the student's proficiency.

<table>
<thead>
<tr>
<th>%</th>
<th>Type of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paper and pencil tests</td>
</tr>
<tr>
<td></td>
<td>Curriculum-embedded tests</td>
</tr>
<tr>
<td></td>
<td>Standardized tests</td>
</tr>
<tr>
<td></td>
<td>Oral questioning in the classroom</td>
</tr>
<tr>
<td></td>
<td>Performance assessment - structured</td>
</tr>
<tr>
<td>100</td>
<td>Performance assessment - spontaneous</td>
</tr>
</tbody>
</table>
E. Level of Use
The six types of tests are listed below. For each type, please indicate which statement best describes your current level of use. Use the definitions labeled A through D to identify your level of use for each type of test.

A. I do not currently use them and do not plan to use them in the future.

B. I have decided to start using them in the future, but have not started to do so yet.

C. I currently use them, but I find them difficult to use and it takes great effort.

D. I use these tests on my own as a regular part of my instruction and do so comfortably.

_____ 1. Objective Paper and Pencil Tests
_____ 2. Text Embedded Tests
_____ 3. Oral Questioning
_____ 4. Standardized Tests
_____ 5. Structured Performance Assessments
_____ 6. Spontaneous Performance Assessments
APPENDIX B

IN-DEPTH INTERVIEW QUESTIONNAIRE
IN-DEPTH INTERVIEW QUESTIONNAIRE

A. Personal Factors
   1. Are you aware of your own learning style and/or the learning styles of your students?
   2. Did you have a mentor or special relationship with someone who influenced your professional growth?
   3. Are you a parent? How many children do you have, and what ages are they?
   4. Are you single or married?
   5. What outside interests do you have? Do you share these with your students?
   6. Can you define or express your personal philosophy statement?

B. Professional Development and Involvement
   7. What kinds/types of college courses/workshops have you taken recently? When did you take them?
   8. What kinds of degrees/training do you have?
   9. Are you familiar with any of the following names or topics (that provided a basis for authentic assessment)? How have they influenced your work in classroom assessment?
      Learning styles
      Brain-based research
      Student-Centered Instruction
      Bloom's Taxonomy
      Howard Gardner's Multiple Intelligences Theory
      Richard Stiggins
      Grant Wiggins
      Reform Movements
      Professional Standards in your field
   10. Have any other readings or national figures influenced your use of assessments in the classroom?
   11. Where did you receive your information on classroom testing?
       Undergraduate courses? Which years?
       Graduate courses? Which years?
       Colleagues?
       Workshops?
       In-house staff development programs?
       Outside workshop or conferences?
   12. What types of professional books or journals do you read regularly?
   13. What professional organizations do you belong to? How active are you in these organizations?
14. Have you ever presented at a regional/state/national conference? Have you ever conducted a staff development program?

C. Administrative/Supervisory Structure
15. To what degree is your department chair/principal/superintendent involved in what happens in classroom testing?
16. Can you define or explain the mission/philosophy of the school? Is it clearly articulated by the administration?
17. How would you describe the climate of your department/school?
18. What is the attitude in your department/school toward experimentation? Is it discouraged/encouraged?
19. How empowered are teachers in your department/school? Do you have input in designing curriculum/classroom tests? Is the curriculum a lock-step procedure?
20. What opportunities exist for collegiality in your department/school? Do teachers design curriculum/classroom tests in teams, individually? Are there any peer coaching or team teaching programs?
21. How are teachers evaluated? Are the types and/or results of classroom tests or any other tests used in the evaluation process?

D. Student Characteristics
22. Are the students tracked? What tracks do you teach?
23. How would you describe the students you teach? What is the socio-economic, ethnic, LEP population?
24. What role, if any, do students play in designing classroom assessments?
APPENDIX C

PRINCIPAL INTERVIEW QUESTIONNAIRE
PRINCIPAL INTERVIEW QUESTIONNAIRE

District Organization
1. Describe the organizational structure of the district.
2. What types of leadership positions exist?
3. Who reports to whom?

Population and Description of the School
4. How would you describe this school.
5. What is the population of the school?
6. How would you describe the student population?
7. What are the parents like?

Leadership History
8. What is the leadership history?
9. How long have you been Principal?
10. How long has the Superintendent held his/her post?

Board of Education
11. Describe relations with the Board of Education.
12. How would you describe the make up of the Board?

Vision and Mission
13. Does the school have a mission statement?
14. Is there a district-wide or school-wide vision?

Changes
15. Have there been any recent changes in the school organization, processes, curriculum, etc.?
16. Do you anticipate any changes?

Academic Organization and Average Teaching Assignment
17. What kind of academic organization exists?
18. Are their departments or divisions?
19. Are their department heads?
20. Are they classified as administrators?
21. Do they conduct teacher evaluations?
22. Do they have released-time to perform their supervisory duties?
23. What is the average teaching assignment?
24. What is the average class size?
Curriculum and Inservice
25. Describe the inservice program that exists for professional development.
26. How is it organized, run, and evaluated?
27. Is there a budget for professional travel or organizational dues?
28. Is there a reimbursement system for graduate level education?
29. How is curriculum developed?
30. Is there a process for changing, adding, deleting courses to the curriculum?
31. Are teachers reimbursed for writing curriculum?
32. How much are they paid?
33. What is the yearly budget for curriculum development?

Climate and Teacher Empowerment / Experimentation
34. How would you describe the climate of this school?
35. Are teachers in this school empowered?
36. What avenues are there for teacher empowerment?
37. Is experimentation encouraged?
38. What examples can you give regarding classroom experimentation?
39. Do teachers use authentic assessment in their classrooms?
40. What examples of authentic assessment can you describe?

Strengths and Weaknesses/Challenges
41. What are the strengths of the school?
42. What are the weaknesses or challenges that this school faces?
APPENDIX D

PERMISSION LETTERS TO / FROM

DR. RICHARD STIGGINS
November 1, 1994

Dr. Richard Stiggins, Director  
Center for Classroom Assessment  
Northwest Regional Educational Laboratory  
101 S. W. Main, Suite 500  
Portland, Oregon 97204

Dear Dr. Stiggins:

I am completing a doctoral dissertation at Loyola University of Chicago entitled "Critical Attributes of Teachers Who Have Become Practitioners of Authentic Assessment." Last year I attended a workshop you presented in Gurnee, Illinois for the Lake County Educational Service Center, and spoke with you briefly about my work. You referred me to the book written by you and Nancy Faires Conklin - *In Teachers' Hands: Investigating the Practices of Classroom Assessment*.

This letter is a request to use some of the questionnaires you have developed in my research, and to reprint portions of your questionnaire in my dissertation. The questionnaires are listed in the appendices of your book *In Teachers' Hands: Investigating the Practices of Classroom Assessment*. I have enclosed a copy of the questionnaire I plan to use in my research.

The requested permission extends to any future revisions and editions of my dissertation, including non-exclusive world rights in all languages, and to the prospective publication of my dissertation by University Microfilms, Inc. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you. Your signing of this letter will also confirm that you own the copyright to the above-described material.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you very much.

Sincerely,

Bernadette Meisenheimer  
824 Paddock Lane  
Libertyville, Illinois 60060  
(708) 367-8741 phone  (708) 680-7881 fax

PERMISSION GRANTED FOR THE USE REQUESTED ABOVE

__________________________________________________________________________  
Dr. Richard Stiggins ___________________________ Date ___________________________
November 1, 1994

Dr. Richard Stiggins, Director  
Center for Classroom Assessment  
Northwest Regional Educational Laboratory  
101 S. W. Main, Suite 500  
Portland, Oregon 97204  

Dear Dr. Stiggins:

I am completing a doctoral dissertation at Loyola University of Chicago entitled "Critical Attributes of Teachers Who Have Become Practitioners of Authentic Assessment." Last year I attended a workshop you presented in Gurnee, Illinois for the Lake County Educational Service Center, and spoke with you briefly about my work. You referred me to the book written by you and Nancy Faires Conklin - *In Teachers' Hands - Investigating the Practices of Classroom Assessment*.  

This letter is a request to use some of the questionnaires you have developed in my research, and to reprint portions of your questionnaire in my dissertation. The questionnaires are listed in the appendices of your book *In Teachers' Hands - Investigating the Practices of Classroom Assessment*. I have enclosed a copy of the questionnaire I plan to use in my research.  

The requested permission extends to any future revisions and editions of my dissertation, including non-exclusive world rights in all languages, and to the prospective publication of my dissertation by University Microfilms, Inc. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you. Your signing of this letter will also confirm that you own the copyright to the above-described material.  

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you very much.

Sincerely,

Bernadette Meisenheimer  
824 Paddock Lane  
Libertyville, Illinois 60060  
(708) 367-8741 phone   (708) 680-7881 fax

PERMISSION GRANTED FOR THE USE REQUESTED ABOVE
APPENDIX E

PERMISSION LETTERS TO / FROM

DR. ROBERT J. WILSON
November 1, 1994

Dr. Robert J. Wilson
Faculty of Education
Queen's University
Kingston, Ontario, Canada K7L3N6

Dear Dr. Wilson:

This letter is a follow-up to our telephone conversation of September 6, 1994.

I am completing a doctoral dissertation at Loyola University of Chicago entitled "Critical Attributes of Teachers Who Have Become Practitioners of Authentic Assessment."

This letter is a request to use a questionnaire you have developed in my research, and to reprint the questionnaire in my dissertation. The questionnaire is from your paper "The Context of Classroom Procedures in Evaluation Students" which was reprinted in a compilation of the papers presented at the Second Canadian Conference on Classroom Testing in June of 1990. I have enclosed a copy of the questionnaire I plan to use in my research.

The requested permission extends to any future revisions and editions of my dissertation, including non-exclusive world rights in all languages, and to the prospective publication of my dissertation by University Microfilms, Inc. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you. Your signing of this letter will also confirm that you own the copyright to the above-described material.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you very much.

Sincerely,

Bernadette Meisenheimer
824 Paddock Lane
Libertyville, Illinois 60060
(708) 367-8741 phone (708) 680-7881 fax

PERMISSION GRANTED FOR THE USE REQUESTED ABOVE
November 1, 1994

Dr. Robert J. Wilson
Faculty of Education
Queen's University
Kingston, Ontario, Canada K7L3N6

Dear Dr. Wilson:

This letter is a follow-up to our telephone conversation of September 6, 1994.

I am completing a doctoral dissertation at Loyola University of Chicago entitled "Critical Attributes of Teachers Who Have Become Practitioners of Authentic Assessment."

This letter is a request to use a questionnaire you have developed in my research, and to reprint the questionnaire in my dissertation. The questionnaire is from your paper "The Context of Classroom Procedures in Evaluation Students" which was reprinted in a compilation of the papers presented at the Second Canadian Conference on Classroom Testing in June of 1990. I have enclosed a copy of the questionnaire I plan to use in my research.

The requested permission extends to any future revisions and editions of my dissertation, including non-exclusive world rights in all languages, and to the prospective publication of my dissertation by University Microfilms, Inc. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you. Your signing of this letter will also confirm that you own the copyright to the above-described material.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you very much.

Sincerely,

Bernadette Meisenheimer
824 Paddock Lane
Libertyville, Illinois 60060
(708) 367-8741 phone (708) 680-7881 fax

PERMISSION GRANTED FOR THE USE REQUESTED ABOVE

[Signature]

Dr. Robert J. Wilson

[Signature]

Nov. 10, 1994
APPENDIX F

SAMPLE LETTER TO SUPERINTENDENTS
November 15, 1994

Mr. John Graham  
Interim Superintendent  
Mundelein High School  
1350 W. Hawley Street  
Mundelein, Illinois 60060

Dear Mr. Graham:

This letter is a request for you and your district to participate in a doctoral dissertation involving the classroom testing practices of high school teachers.

I am a doctoral candidate at Loyola University of Chicago, and I would like permission to distribute a questionnaire to teachers in the English, social studies, math, science and foreign language departments. I have enclosed a copy of the questionnaire I will be using.

Teachers' participation in this study is completely voluntary. Teachers who select to participate do have do be identified by name so that I may contact some of them for follow-up in-depth interviews. However, in the dissertation the teachers and their home schools will remain anonymous. I would like permission to contact the principal of your high school, and to distribute the questionnaires during the month of November.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you very much.

Sincerely,

Bernadette Meisenheimer

__________________________________________  ____________________________
Mr. John Graham  Date
November 15, 1994

Mr. John Graham
Superintendent
Mundelein High School
Mundelein, Illinois 60060

Dear Mr. Graham:

This letter is a request for you and your district to participate in a doctoral dissertation involving the classroom testing practices of high school teachers.

I am a doctoral candidate at Loyola University of Chicago, and I would like permission to distribute a questionnaire to teachers in the English, social studies, math, science and foreign language departments. I have enclosed a copy of the questionnaire I will be using.

Teachers' participation in this study is completely voluntary. Teachers who select to participate do have to be identified by name so that I may contact some of them for follow-up in-depth interviews. However, in the dissertation the teachers and their home schools will remain anonymous. I would like permission to contact the principal of your high school, and to distribute the questionnaires during the month of November.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you very much.

Sincerely,

Bernadette Meisenheimer

Bernadette Meisenheimer

[Signature]

Mr. John Graham

[Signature]  

11-16-94

Date
November 15, 1994

Dr. Robert Kessler
Superintendent
Lake Forest High School
Lake Forest, Illinois 60045

Dear Dr. Kessler:

This letter is a request for you and your district to participate in a doctoral dissertation involving the classroom testing practices of high school teachers.

I am a doctoral candidate at Loyola University of Chicago, and I would like permission to distribute a questionnaire to teachers in the English, social studies, math, science and foreign language departments. I have enclosed a copy of the questionnaire I will be using.

Teachers' participation in this study is completely voluntary. Teachers who select to participate do have to be identified by name so that I may contact some of them for follow-up in-depth interviews. However, in the dissertation the teachers and their home schools will remain anonymous. I would like permission to contact the principal of your high school, and to distribute the questionnaires during the month of November.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you very much.

Sincerely,

Bernadette Meisenheimer

Bernadette Meisenheimer
November 15, 1994

Dr. Cuttie Bacon
Interim Superintendent
North Chicago High School
1717 - 17th Street
North Chicago, Illinois 60064

Dear Dr. Bacon:

This letter is a request for you and your district to participate in a doctoral dissertation involving the classroom testing practices of high school teachers.

I am a doctoral candidate at Loyola University of Chicago, and I would like permission to distribute a questionnaire to teachers in the English, social studies, math, science, and foreign language departments. I have enclosed a copy of the questionnaire I will be using.

Teachers' participation in this study is completely voluntary. Teachers who select to participate do have to be identified by name so that I may contact some of them for follow-up in-depth interviews. However, in the dissertation the teachers and their home schools will remain anonymous. I would like permission to contact the principal of your high school, and to distribute the questionnaires during the month of November.

If these arrangements meet with your approval, please sign this letter where indicated below and return it to me in the enclosed return envelope. Thank you very much.

Sincerely,

Bernadette Meisenheimer

Bernadette Meisenheimer

Dr. Cuttie Bacon

11-18-94

Date
BIBLIOGRAPHY


Farr, Roger. (1994, February 16). *Solving the assessment puzzle* [Lake County Educational Service Center Workshop, Gurnee, IL].


Girl, 10, gets perfect score on math SAT. (1994 March 7). *Chicago Tribune*, Section 1, p. 4.


Kiernan, L. (1993, November 14). Old or revised, SAT spells dread for many. Chicago Tribune, Lake Section, 1, 3.


VITA

Bernadette Kinsey Meisenheimer graduated from Bradley University in Peoria, Illinois, with a Bachelor of Arts degree in English / Education in 1970, and a Master of Arts degree in Guidance and Counseling in 1971. In 1977 she received a Reading Specialist Certificate from the National College of Education in Evanston, Illinois; and in 1982 received a Certificate of Advanced Study in Curriculum and Instruction from Northern Illinois University in DeKalb, Illinois.

She began her career in 1971 as a secondary school counselor in Chicago, Illinois; and then served for six years as a Chapter I counselor and reading specialist in the Round Lake Unit School District in Round Lake, Illinois.

In 1979 she joined Mundelein High School in Mundelein, Illinois as a reading specialist, and then served as the Department Chair of the Communication Arts Division. She later became the Director of Student and Staff Services, and currently serves in that position.

Since joining Mundelein High School, she has presented workshops at local and national conferences on topics in curriculum and instruction and computer technology, authored numerous grant winning proposals, and served as a visiting faculty member and assistant co-chairman on North Central Evaluation committees.

In 1994, she received a Distinguished Service Award from the Illinois Chapter of the National School Public Relations Association, and a Those Who Excel Award of Merit as a school administrator from the Illinois State Board of Education.

The author resides with her husband Jim in Libertyville, Illinois.
THESIS/DISSERTATION APPROVAL SHEET

The dissertation submitted by Bernadette Meisenheimer has been read and approved by the following committee:

Robert Cienkus, Ph.D., Director
Associate Professor
Department of Curriculum, Instruction, and Educational Psychology
Loyola University Chicago

Barney Berlin, Ph.D.
Associate Professor
Department of Curriculum, Instruction, and Educational Psychology
Loyola University Chicago

Daniel J. Burke, Ed.D.
Professor
Department of Curriculum, Instruction, and Educational Psychology
Loyola University Chicago

The final copies have been examined by the director of the dissertation study and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the committee with reference to content and form.

The dissertation is, therefore, accepted in partial fulfillment of the requirements for the degree of Ph.D.

[Signature]

Date: 11/9/95

Director's Signature