

The district wanted improved student achievement but not at the expense of low teacher morale, dissatisfied parents, or exploited students.

Criterion-Referenced Testing for Outcomes-Based Education

Richard P. Manatt and Glenn R. Holzman

Overview

K-12 schools in Kansas and nationwide are seeking methods to establish outcomes-based education. A team of researchers and trainers led by Dick Manatt and Shirley Stow has specialized in this task beginning with mathematics and reading in the landmark experiment, the School Improvement Model (SIM), in the Minneapolis area (1978-85). The model of outcomes-based education was first tested for all subjects and all grades in the Hot Springs County School District No. 1 (Thermopolis, Wyoming) in the late 1980s. In this report Glenn Holzman, who directed the project's indistrict effort, and Dick Manatt, director of the university based team, explain how it was accomplished.

Development of all training activities, curriculum planning, criterion-referenced test development, and pilot testing was the responsibility of the Iowa State University team. The district wanted achievement to improve, but equally important, teachers and students were to be treated with consideration and the entire process was to be systematized so that the curriculum materials, tests and methodology could be shared with other public and independent schools.

The district administered the SRA achievement tests annually to all grades, so these measures were selected to be the indicator of improved student learning. School climate, parental and student satisfaction were also to be measured annually.

The Process

In the fall of 1985, the administration and school board of Hot Springs School District No. 1 approached the School Improvement Model office at Iowa State University request-

Richard P. Manatt is Professor and Director of the School Improvement Model, Iowa State University, Ames, Iowa.

Glenn R. Holzman is Research Associate with the School Improvement Model, Iowa State University, Ames, Iowa.

ing help in the development of a comprehensive teacher and administrator evaluation system. The overarching goal was to improve student achievement. It was generally accepted in the district, however, if long-term meaningful improvement of student achievement was going to take place, as much or more attention would need to be given to the "what" as to the "how" of instruction. Subsequently, the school improvement effort in Hot Springs County moved beyond mere performance evaluation of personnel to take on the monumental task of curriculum development/renewal at every grade level, for every subject, with accompanying criterion-referenced measures of student achievement. It was obvious that such an effort was beyond the capabilities of a small rural school district, thus the cooperative endeavor of the district and university-based SIM was expanded to an additional three year school improvement effort.

The foundation upon which the curriculum development project was to be constructed was based on four generic questions posed by Ralph Tyler in 1949, which have since come to be known as the *Tyler Rationale*:

- (1) What educational purposes should the school seek to attain?
- (2) How can learning experiences be selected which are likely to be useful in attaining objectives?
- (3) How can learning experiences be organized for effective instruction?
- (4) How can the effectiveness of learning experiences be evaluated?

There was strong commitment within the district that the people most qualified, most appropriate, to answer these questions were classroom teachers. Whatever was developed needed to be teacher based and specific to the needs of students and staff in Hot Springs County. This involvement would later provide the commitment and ownership that would support the successful implementation of the curriculum. Teachers would not be able to say, "You didn't measure what I was actually teaching!"

In order to address those questions, subject area K-12 curriculum committees were formed. A framework for developing the curriculum was adopted that included six parts: (1) philosophy statement, (2) strands of learning, (3) program goals, (4) instructional objectives, (5) instructional activities (teachers' and students'), and (6) criterion-referenced measures. Of paramount importance in this effort was the notion of curriculum alignment: *i.e.*, the congruent relationship of the written, taught, and tested curriculum. Does each one support the other two?

It was the charge of the committees to ensure that the district's mission, subject area philosophy, major strands, program goals, instructional objectives, and test items all were aligned with a logical flow from one to another. Working directly with the consulting team from SIM, the teachers moved through this process step by step over a period of twelve to eighteen months. The teachers were viewed as the subject area experts and were given the responsibility of fitting the appropriate content within the given framework. While some of this process could be done on an individual basis, it was important to work across grade levels to ensure continuity and alignment of the objectives throughout the curriculum and from grade to grade. The importance of K-12 representation in each subject area cannot be overemphasized.

The most difficult component in terms of time and effort was the student achievement measures. Most teachers have had little training in or experience with the development of valid and reliable tests. Thus, training and support provided by the consultants played an integral role during

this process. Multiple choice test items were developed to match the instructional objectives that had previously been identified. Strict test writing guides were followed and each test at each grade level was critiqued a number of times before pilot testing. After pilot testing the results were statistically critiqued and appropriate revisions were made before the final copy of the test was printed.

Early into the test development process, teachers raised the issue of testing in those areas that were only performance based, *i.e.*, fine arts, physical education, vocational education, *etc.* Certainly, these areas were based on performance of certain activities, but there are cognitive skills and knowledge necessary to achieve the performance objective. In these areas performance checklists were developed along with limited paper and pencil tests to measure the appropriate knowledge base.

Another question that had to be constantly addressed was whether the instructional objectives being developed, and their related measures, were minimums, maximums, aimed at the average student, essential skills or those skills needed to pass on to the next grade. It was decided to identify the "essential skills" that they, the teachers as experienced and trained professionals, felt the students should master to succeed at the next level of instruction or grade. In the process of identifying those essential skills the following questions were put to the teachers:

- given an infinite amount of information that could be taught, but only a finite amount of time to teach, what do you want the students to learn in the allotted time available?
- if we accept the premise that students will generally forget the vast majority of what is presented to them, what is the essential information they must retain?
- what are those skills that build upon prior learning and what are those skills that are required prerequisites before additional learning can take place in the next unit, class, or grade?
- what have the major textbook publishers identified as essential skills across the nation for a particular subject area?
- what does the most recent research and recognized experts have to say about what is important for students to know?
- what does the community want its students to learn?
- as an experienced professional, what does the classroom teacher feel is important for the student to learn?

Certainly these questions or guidelines are not as clear-cut or objective as many faculty members would have liked. They left much room for subjectivity, professional judgment, and debate. However, the district was strong in its belief that the staff would make the appropriate decisions that would eventually serve the best interests of the district's students.

The Results

Outcomes-based testing was highly successful in reaching the prime objective of raising student achievement district-wide as measured by the composite results of the SRA achievement tests. Equally important were the positive results of the formative measures of teacher and administrator performance evaluation, student and parent feedback, and the measures of school climate.

The district wanted improved student achievement but not at the expense of low teacher morale, dissatisfied parents and exploited students. Because of continuous measures of climate, student feedback to teachers and parent feedback (in the form of a School Report Card), the district could be certain that achievement gains were an unmixed

blessing. Moreover, the improvement curve continued for five years; it was not simply a Hawthorne effect.

School Climate

Climate factors were measured in May of Years Two, Three, and Four using the School Improvement Inventory (SII). The SII is administered by the School Improvement Model Projects office to over 100 schools each year and provides a comparison with national norms. To illustrate, Table 1 contains the three years of climate data for the district's middle school. The inventory was completed by all certified personnel in the school.

Table 1
Middle School Climate Survey—School Improvement Measures as Assessed by Teachers—Hot Springs County School District No. 1

Dimension ^a	Mean Response ^b			
	Year Two (1986)	Year Three (1987)	Year Four (1988)	National Norms (1989)
Goal Orientation	5.77	6.28	6.15	5.81
Esprit	5.56	5.80	5.81	5.83
Cohesiveness	5.81	5.97	5.91	5.33
Teacher Expectations	5.95	6.36	6.25	6.01
Administrator Dedication				
Enthusiasm	6.55	6.96	6.73	5.54
Student Attitudes	5.14	5.46	5.54	6.30
Supports Teachers	4.68	5.92	5.92	6.11
Evaluates Pupil Progress	5.82	5.83	5.69	4.78
Coordinates Instructional Curriculum	4.91	5.35	5.73	4.48
Instructional/ Curriculum Emphasis	4.82	5.46	5.35	4.18
Learning Environment	5.36	5.78	5.85	5.48

^aSchool Improvement Inventory. SIM, Research Institute for Studies in Education, Ames, Iowa.

^bRange of Responses: 1 = low to 8 = high.

Inspection of the table reveals that by Year Four, the climate of the middle school equalled or exceeded the national norms in all areas except student attitudes. Generally speaking, each measure had improved over the three years. This was especially satisfying because the faculty had undertaken tremendous efforts to improve both teaching and curriculum content during that time frame. Note that esprit actually improved during this effort while goal orientation, cohesiveness, and teacher expectations rose markedly. Previously, the great emphasis placed on improved performance by the SIM Model had been accompanied by a drop in teacher morale for the first two years (Petrone, 1989).

Student Achievement

At the outset, this school improvement effort anticipated that criterion-referenced tests would show positive

results. But because the criterion-referenced tests would undergo constant refinement and revision during the five year study, the norm referenced SRA tests were also used as the criterion of success.

Student Achievement by Subject

Efforts to improve the curriculum started in the core academic areas of math and reading, including language arts, then moved to other subject areas as time and staffing patterns allowed. Changes in student achievement tended to follow the same pattern.

Mathematics and English were the most improved subjects during the years under study (see Table 2). Composite percentiles for mathematics rose 16 points, while English increased by 14 points. Reading composite percentiles changed from 60 to 71, a gain of 11 points. Social studies and science had lesser gains of 5 and 8 percentiles respectively. In four years, percentile composites had risen from a range of 55-61 to a range of 66-77. The total composite percentile changed from the 59th to the 73rd, a gain of 14 (see Table 3).

Table 2
Student Achievement by Subject—
Percentile Composite Scores, National Norms*

Subject	YEAR					Change
	Year One (1985)	Year Two (1986)	Year Three (1987)	Year Four (1988)	Year Five (1989)	
Reading	60	63	68	70	71	11
English	55	60	65	65	69	14
Mathematics	61	68	73	76	77	16
Social Studies	61	62	64	67	66	5
Science	61	64	66	69	69	8
Total	59	65	69	71	73	14

*Science Research Associates Student Achievement Tests

Table 3
Student Achievement by School—
Percentile Composite Scores, National Norms*

Level	YEAR					Change
	Year One (1985)	Year Two (1986)	Year Three (1987)	Year Four (1988)	Year Five (1989)	
1-4	61	71	78	78	81	20
5-8	59	62	67	69	70	11
9-12	59	58	59	62	67	9
District Composite	59	65	69	71	73	14

*Science Research Associates Student Achievement Tests

Student Achievement by School

Elementary school achievement (Table 3) was at the 61st percentile (composite) in 1985; it had risen to 81 by 1989 for an overall gain of 20 points. Middle school achievement had risen 11 points (59 to 70) during the same period. High school composite achievement moved from the 59th percentile to the 67th by 1989 (+8). The district composite increased 14 percentile points since 1985 as noted previously.

Conclusion

With improved student achievement as the criterion, the School Improvement Model was an unqualified success for the Hot Springs County School District No. 1. Changes on norm-referenced test results of this magnitude are rare, especially in a working class rural district with all of the usual effective schools concerns, viz., high student turnover, economic factors, family influence, and gender differences. The district made wave one (curriculum improvement) and wave two (more effective teaching) school reform a reality. Wave one became more than just adding courses for graduation; vigor was infused.

The district had several very positive characteristics. School board leadership, which stayed vitally committed to the project for the entire five years, had a major impact. The teachers and their leadership from the NEA-affiliated local association played a major role in planning, directing, and refining all of the components of the School Improvement Model. The district's administrative team was particularly skillful in operating the school improvement components and fortunately remained in the district throughout the long endeavor. But perhaps the most salient factor of the model was time duration. This was not a quick fix. Consuming five years allowed enough time to fully develop each of the improvement components. Each item was invented, field tested, reinvented, and then tried again and again with time for critiquing and improvement. The curriculum content was clearly richer at the end of the project. The "what" of learning was fully developed. Indeed, the pre- and posttesting of student learning via CRTs meant that teachers and students share a keen interest in being task-oriented and businesslike.

At the end, the district had a total systems approach to managing and improving instruction. Everyone had more usable information, not just more information. This provided the "wheaties effect"—feedback that makes a difference.

References

- Manatt, R. and Daniels, B. (1991). Relationships between principals' ratings of teacher performance and student achievement. *Journal of Personnel Evaluation in Education*.
- Manatt, R. (1987). Lessons from a comprehensive appraisal project. *Educational Leadership*, 44: 8-14.
- Petrone, J. (1989). A follow-up of selected School Improvement Model (SIM) projects 1978-88. E005 Lagomarcino Hall, Iowa State University, Ames, IA.
- Tyler, R. (1949). *Basic principles of curriculum and instruction*. University of Chicago Press.