

CHAPTER 5

SUMMARY, CONCLUSIONS, DISCUSSION, POLICY CONSIDERATIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

The Challenge

It will be urged, Some men have such weak intellects that it is not possible for them to acquire knowledge. I answer, it is scarcely possible to find a mirror so dulled that it will not reflect images of some kind, or for a tablet to have such a rough surface that nothing can be inscribed on it. (Comenius, 1632, p. 86)

Summary of Survey Data

My purpose in attempting this study was to document teacher beliefs about the outcomes that were likely to flow from the implementation of Virginia's *Regulations Establishing Standards for Accrediting Public Schools in Virginia* (2000) and from the testing requirements in the standards. After completing a review of related literature, I developed a survey to be sent to Virginia teachers. The instrument was subjected to extensive content validity studies during 1999 and 2000 and was refined several times. Lacking an effective strategy for communicating with all teachers on a statewide basis, I sought and received permission to survey a sample of the approximately 46,000 members of the Virginia Education Association (VEA). The survey was mailed to 464 VEA members in early November of 2000. Surveys were received and included in the study until mid-January of 2001. 352 surveys were returned for a return rate of 76%.

Survey responses were coded into the Statistical Program for the Social Sciences (SPSS, 1999), and descriptive statistics were computed for five independent (demographic) variables in five domains and then refined into tables. Analysis of variance (ANOVA) was computed to determine if teacher opinions were significantly influenced by three independent variables: school socio-economic level, SOL test grade status, and tenure status. Tables were also produced for these data.

When the data from the mailed survey were analyzed, mean responses were compared by domain to a favorability mid-point of 2.5 on a scale that ranged from 1.0 representing a high favorability rating for SOL testing practices and 4.0 representing a low favorability rating for SOL testing practices. Mean domain scores (the average of all teacher responses for a domain) always exceeded the mid-point of 2.5. Mean domain scores for student outcomes, school outcomes, and outcomes for public confidence were approximately 3.00, while mean domain scores for instructional practices and outcomes for teachers were approximately 3.30. Given a choice of expressing positive or negative beliefs about SOL testing outcomes, teachers chose to give appraisals that were relatively less favorable. Teachers rated the outcomes associated with teaching practices and outcomes associated with outcomes for teachers more unfavorably than they did outcomes for students, outcomes for schools, or outcomes associated with public confidence. A possible explanation for these ratings is that teachers felt the consequences of high-

stakes testing and measurement-driven instruction most keenly as applied to their own lives as teachers while consequences for students, schools, and public confidence were not as salient.

The ANOVA analyses indicated that teacher opinions did not vary significantly by any of the variables studied or by any combination of these variables. The following were findings: (1) Responses from teachers with tenure did not vary significantly from those without tenure. (2) Responses from teachers in SOL test grades did not vary significantly from those in non-SOL test grades. (3) Responses from teachers in low SES schools did not vary significantly from teachers in high SES schools. (4) Responses from teachers were not affected by the interaction between tenure status and SOL test grade status. (5) Responses from teachers did not show significant interaction between school socio-economic level and tenure status. (6) Responses from teachers did not show significant interaction between SOL test grade status and school socio-economic level. (7) Responses from teachers did not show significant interaction among tenure status, SOL test grade status, and school socio-economic level. In other words, teacher opinions were remarkably alike regardless of teachers' relative job security, the challenges teachers faced by virtue of the socio-economic level of their school, or whether teachers taught in an SOL test grade. In short, even though I surveyed a systematic sample of VEA members from throughout the state, the thinking of this group of teachers was remarkably alike. The obvious question to this circumstance is *why*.

Summary of Interview Data

Twelve teachers were interviewed from a total of 122 who indicated a willingness to be interviewed. Interviewees were randomly selected from this group according to a twelve cell matrix that included all possible combinations of SES level, SOL test grade status, and tenure status. Data were analyzed with the Constant Comparative Method described by Maykut and Morehouse (1994). Eight outcome propositions were derived from the data. Identified propositions were favorable, unfavorable, and neutral with regard to SOL testing. However, an analysis showed that four of eight categories portrayed an unfavorable view of outcomes from SOL testing. As with the survey data described above, one must ask *why*.

Conclusions

After careful reflection, I offer the following as conclusions from this study: (1) The sample of VEA members who participated in the study held views that were relatively unfavorable toward Standards of Learning Tests and related regulations. (2) Teachers were less favorable in their views about outcomes for instruction and outcomes for teachers than they were about outcomes for students, outcomes for schools, or outcomes related to public confidence. (3) Teachers' opinions did not vary significantly according to the independent variables identified in the study. Teaching experience, school socio-economic level, SOL test grade status, teaching assignment, tenure status, and combinations of these variables were not significantly associated with differences in teacher opinions about outcomes from SOL testing. (4) Teachers identified a variety of positive outcomes of SOL testing and related curriculum standards.

Discussion of Survey Data and Interview Data

An obvious answer to the question of *why* teachers responded so similarly to survey statements and to interview questions is that teachers in the sample simply viewed the statements in the survey through the same lens. That is, this group of teachers may have read similar newspaper articles; they may have attended similar staff development activities related to SOA and SOL; they may have been asked to alter their instruction in a similar manner; they may have noted similarly low levels of student performance on test scores; they may have experienced unfavorable changes in the principal-teacher relationship; they may have noted elevated levels of frustration from students, parents, and colleagues; they may have read about new plans to reward or punish teachers; or some combination of these or other factors may have led them to respond with such similarity of thought. If these suggestions are accurate, it is possible that the views expressed by these educators are representative of the entire VEA sample.

The purpose of interview questions was to extend the depth of my understanding about teacher opinions expressed in the survey, and my understanding was, indeed, extended. The interviews elucidated key conclusions derived from the survey data. There is no question that teachers' survey responses produced relatively lower favorability scores concerning Virginia's SOL testing program, and there is no question that teacher concerns about outcomes for teachers, students, and for instructional programs were confirmed in the interview data. It would be relatively easy to reach the conclusion that most members of the sample held unfavorable views of SOL testing *because* of the tests and testing practices. However, it is important for a researcher to make every effort to consider rival hypotheses and conclusions (Miles & Huberman, 1994); good researchers should be their own worst critics. With this in mind, several alternative hypotheses are offered to explain why teachers may have responded to survey statements and interview questions in such a similar manner.

First, it is possible that teacher responses about the SOL testing program are artifacts of failed relationships between teachers and principals exacerbated by principals' attempts to broker change in a high accountability instructional model. Second, it is possible that teachers have united intellectually against state and local leaders whom they may believe are requiring them to compete in a "scientific management" model in which the uniformity of production procedures and products is paramount. Third, it is possible that teachers have succumbed to the phenomenon of "group think." That is, teachers may be somewhat blindly accepting and espousing negative perceptions associated with the beginning of the standards movement in Virginia because of the reported experiences and views of others. While many other rival hypotheses might be considered, I believe that these three may be responsible individually or in combination for some of the limited variability in teacher responses.

The Principal-Teacher Relationship

The requirement for increasingly higher student pass rates on SOL tests for schools to maintain their accreditation, to avoid public media censure, to avoid reviews by state assessment teams, or to avoid being placed into receivership has thrust principals into new high-stakes leadership roles. Though many principals are trained and able to lead teachers in instructional

improvement, this work has often been collegial (Goldhammer, 1969), democratic (Lucio & McNeil, 1962), and low-stakes. That is, principals (and other supervisors) have worked with teachers in annual evaluation processes that were oriented toward the maintenance of self-esteem and processes rather than toward measurable results.

Blumberg (1974) wrote persuasively that the principal-teacher relationship is fraught with difficulties, misperceptions, conflict, and lack of trust. Blumberg made the following research-based observations: (1) Teachers rarely identify their principal as the person to whom they would go for ideas to improve their teaching practices. In fact, teachers usually do not ask for help from their principals *or* their peers. (2) Teachers feel that their supervisors are out of touch with the classroom. (3) There is a “we-they” gulf between teachers and supervisors that is exacerbated by “perceptual screens.” These screens affect the way that each party receives and interprets the ideas, attitudes, and feelings of the other, and (4) Teachers may believe that principals participate in the implementation of unwise innovations because they are easily impressed by those peddling seemingly good ideas. “Change in schools frequently means that someone has sold the administration a bill of packaged goods” (Blumberg, 1974, p. 129). This bill of goods is loaded on top of other packages that have been implemented half-heartedly. Teachers may expect, then, that the related plans and mandates will eventually join all the others in oblivion.

Though Blumberg’s analysis may seem harsh, my experience has been that his observations are close to the mark. Therefore, when Virginia’s high-stakes standards were first proposed in 1995, it is realistic to think that many principals were unprepared for their new roles as change agents to ensure high student achievement on SOL tests. It is also realistic to believe that many principals stumbled in their plans and interactions to achieve higher scores and that their stumbling led to more unfavorable perceptions about state-mandated curricula and testing.

Combinations of these factors may have led to similarities in teachers’ responses to the survey and to interview questions. Both the surveys and the interviews indicated that teachers are quite concerned about the outcomes of SOL testing on teachers and on instructional practices. Both surveys and interviews also documented teacher concerns about the outcomes of SOL testing for students themselves. And, supporting Blumberg’s views, teachers indicated that principal behavior does influence the stress teachers feel about SOL testing.

Scientific Management as Applied to Teachers

A second hypothesis may explain the similarity of teacher opinions about survey statements and interview questions. As achievement mandates began to flow from the Virginia State Board of Education to local school boards and to schools, administrators are likely to have felt pressured to take some actions to ensure the production of the desired results. These actions may have taken the form of long-range consensus-building and teacher involvement in developing solutions. However, they may also have resulted in top-down directives in which teachers were seen as workers employed to produce a standard product through the most efficient use of resources according to the carefully laid plans of educational supervisors and leaders. In 1910, similar notions came to be known as *scientific management*.

In its earliest context in the metal forming industry, proponents of scientific management held that managers and workers each had separate strengths; managers were skilled at planning and workers were skilled at implementing the plans. Therefore, managers should break production tasks into their smallest sub-components. The time for all sub-components added together would yield the anticipated time to complete defined production tasks (Drury, 1922). Rather than relying on the uncertain output or even the delaying tactics of laborers, workmen were given production targets based on a scientific assessment of the time actually required to do a job.

Critical components of scientific management include the following: (1) a belief that managers only should plan while workers should implement these plans, (2) dissatisfaction with the current ratio of cost to productivity, (3) a determination of the task or product to be completed, (4) a determination of the accuracy with which the product must be produced and the time that may be devoted to its production, (5) the employment of only exceptionally competent workers, and (6) a determination of the increased compensation that will be associated with increased productivity (Drury, 1922). Each of these principles is embodied to some degree in today's quest for higher student scores in our vast education industries. First, Virginia's educational leaders exercised high levels of state control in envisioning and producing a state-sponsored curriculum and assessment program. No referenda were held to determine educators' opinions about the concept, and teacher input into the planning phases of these initiatives was only token.

Secondly, there is no question that political and business leaders and the public in general have expressed dissatisfaction with student achievement in relationship to the monies expended to support educational systems. In Virginia it is not unusual for more than two-thirds of the budgets of county and city governmental agencies to be devoted to K-12 public education. It is logical, therefore, for those in business and governmental leadership positions to ask if they are getting their money's worth from their large investment.

Third, Virginia's Board of Education, like state boards almost everywhere else, has stipulated the standards that must be reached (numbers of items correct on state-mandated tests that assess state-mandated curricula) as well as the time available for the completion of this task—180 school days with five and one half hours a day available for instruction each day.

Fourth, virtually all public schools have for years required a strict program of professional preparation as a prelude to licensing. Teacher preparation programs in Virginia routinely require a minimum of a four-year preparation program, minimum grades in the area of major, supervised field experiences, student teaching and related evaluations, portfolio development and analysis, as well the requirement to take and pass the Pre-Professional Skills Test (PRAXIS) with required scores among the highest in the nation.

And last, concerning the issue of compensation, most Virginia school systems have chosen to eschew differential salary scales that are tied to student scores on standardized tests. Teachers have been asked to achieve increasingly higher results without compensation related to this achievement. However, the practice of paying teachers for higher scores is becoming increasingly more common. One respondent to the telephone interviews in this study indicated that “merit

pay” for higher test scores would take effect in 2001/02 in her school division in northern Virginia. And the Roanoke City school system plans to institute teacher bonuses in 2001/02 based on school-wide test score improvement (Roanoke Times, August 7, 2001). It is possible, then, that pressures and requirements from state and local leaders for teachers to produce pre-determined outputs under circumstances defined according to the principles of scientific management could have resulted in homogeneity of thought when teachers responded to the survey and interviews in this study.

The Concept of “Group Think”

Teachers are often located in a relatively confined physical environment in which they work and associate with a defined group of peers for a minimum of 190 to 200 days a year. They often travel to conferences with peers and associate with peers in university classes, degree programs, and in staff development experiences. They work on long-range projects and serve on school committees with peers, and many educators form personal relationships that extend beyond the walls of the schoolhouse. Janis (1996) postulated that in close social environments such as schools, co-workers can succumb to the danger of accepting group norms that bolster morale to the detriment of critical thinking. The author termed this phenomenon “groupthink.” He uses the term to describe “the mode of thinking that persons engage in when concurrence-seeking becomes so dominant in a cohesive group that it tends to override realistic appraisal of alternative courses of action” (Janis, 1996, p.184).

Groupthink is postulated to become most evident when group members, in an effort to concur with the dominant opinions of their group, avoid being harsh or judgmental toward the ideas and thinking of their leaders or co-workers. Groupthink, the author argued, becomes more pronounced when group cohesiveness increases. Symptoms of groupthink include the following: (1) Invulnerability. If the group feels this way, then this must be the right way. (2) Morality. What we believe is inherently good (for children in this case). (3) Stereotypes. All reformers are bad people. They do not look ahead, and they do not have the best interests of students or teachers at heart. They have political agendas. (4) Pressure. Group members who momentarily express opinions that are contrary to group norms and beliefs will be subjected to pressure to reform their thinking. (5) Self-censorship. Victims of groupthink often avoid creative thinking and tend to silence themselves before others have a chance to do so. (6) Unanimity. Victims of groupthink share the view that all members in the group think alike and have the same opinions. So, when the members of the group reach a unanimous accord on an issue, group members feel that the position must be the right one.

What more likely event would cause increased group cohesiveness among teachers than working in an era of top-down mandates to achieve a defined product according to the dictates of their managers who are often perceived as lacking understanding and competence? Therefore, if teachers hear and participate in daily conversation about just how bad SOL tests are, just how inept school-level leadership is, and just how out-of-touch the state educational leadership is with the realities of classroom teaching, one could easily imagine how teachers could be led, even unwittingly forced, to conform to the norms of the group and to see the state reform initiatives in a largely negative light.

The actual effect of all these factors, individually or as a group, may have altered the outcomes of the study and may have contributed to the error associated with interpretation of the data. Though I took great care to ensure that the theoretical model, the research design, the research instruments, data entry, and data analyses were unbiased, each of the caveats above must be considered as possible threats to the validity of the findings of the study.

Policy Considerations

Having spent more than two years working on this study, I am struck by an important and sobering question: If the conclusions of my study are accurate, and if Virginia teachers *do* hold relatively unfavorable views of the state's assessment program, what *replacement* policies should be implemented? My recommendations are: (1) the granting of more authority concerning curriculum and instruction to localities, (2) a redefining of the role of the Virginia Department of Education (VDOE), (3) the development of incentives for localities to assess and meet the needs of *all* students, and (4) allowing localities to set their own standards for school and student success.

First, I recommend that localities be granted more decision-making authority in matters related to *what* is taught in their schools in place of the overly prescriptive policies that are in effect at present. After all, Virginia began its current reform initiative only to march to the same drummer as other states in the politically charged accountability environment of the 1980's and early 1990's. Reform initiatives are often undertaken to create a public impression that *something is being done* to improve the quality of education (Airasian, 1987; 1988). However, the efficacy of state-mandated curricula and high-stakes testing has no substantial research base; on the contrary, much evidence exists that school problems are best defined and solved at the school level (Darling-Hammond, 1994b) and that external accountability systems actually inhibit this local problem-solving ability (Newmann, King, & Rigdon, 1997). And, the ill effects of coercive strategies inherent in high-stakes testing are well documented (French, 1998). I would propose, therefore, that the Virginia Department of Education fund a full range of initiatives to help localities engage in "bottom-up" planning to find and solve their own instructional problems. It is this type of problem-finding and problem-solving that is likely to engender support and enthusiasm from state educators and local citizens as well (Clune, 1993).

Secondly, the role of the Virginia Department of Education (VDOE) should be redefined from an enforcement agency to a helping agency. By any account of testing, it is clear that some Virginia localities and schools produce relatively high achievement scores and that student achievement in some localities and schools continues to languish. Also clear is that certain sub-populations of Virginia students achieve at significantly lower levels than their peers—in spite of many years of accountability mandates and assurance by the VDOE that SOL test scores are improving the quality of education in the state. For example, Virginia's grade 4 reading scores on the National Assessment of Educational Progress are essentially unchanged from 1992 to 2000 (<http://nces.ed.gov/nationsreportcard>) despite years of mandates and rhetoric attesting to improvement related to the Standards of Learning curriculum and related testing program. This

same phenomenon can also be observed in the high-stakes states, Kentucky, Minnesota, and Texas.

To help reverse this history, the VDOE should engage in partnerships with low-achieving schools to improve student achievement *as determined by localities*. Such assistance might come in the form of teacher training, innovative pilot programs in partnership with state colleges and universities, or by providing additional highly-trained staff for those who have been chronically underserved: the poor, minorities, immigrants, ESL students, and students with disabilities. By one recent account (Pipho, 1999), the state spends some \$50 million annually on its SOL testing program. This amount alone would do much to help implement these initiatives while doing relatively little to diminish student learning.

Third, I would propose that the VDOE tie a portion of its annual state funding for local school districts to the *percentage of students that local districts actually test* via their own assessments (and not to student scores on the tests). In the quest to look good on test scores, schools and school districts engage in many creative strategies to “hide” students from the state testing program to avoid the condemnation associated with low scores (Koretz, 1988). Virginia is no exception. A review of school “report cards” on the Virginia Department of Education website (www.pen.k12.va.us) will quickly illustrate that some 5 to 25 % of students are routinely excluded from SOL tests in various schools throughout the state. Relatively few schools test all of their students. Unfortunately, this practice masks the true needs of schools and school divisions to educate all the students in this population—often ignoring those who are poor, minority, or disabled (Allington & McGill-Franzen, 1992a, 1992b; McDonnell, et al., 1997; Fries, 1998; McNeil, 2000). Therefore, these students do not become a part of the funding equation for appropriate assessment strategies or compensatory instructional programs and become, essentially “educational throwaways.” If we are to truly recognize Virginia’s stated goal to “Provide an essential foundation of educational programs of high quality in all schools for all students” (Regulations Establishing Standards for Accrediting Public Schools in Virginia, 2000, p.1), then we must come to learn that *all* means all.

Lastly, I propose that the state abandon its punitive accreditation policies in which school accreditation is based solely on student test scores and that it allow localities to develop their own standards for student programs and achievement (Oakes, 1989). Some localities may value vocational programs that could lead to a high school diploma. Some may wish to ensure that students have broad exposure to the arts, while some may value student competency in some field of technology. Yes, this problem-solving should involve some manner of periodic assessments, but these should be locally developed and implemented. Inherent in this proposal would be an abandonment of the use of state-mandated tests to determine high school graduation. The power of high-stakes testing, after all, is not based on a well-reasoned, universally accepted analysis of what an educated student should know or be able to do, but on threats—threats to students that they may not graduate; threats to educators that they will lose self-esteem, salary, employment, or the right to professional judgement; and threats to school districts that they will suffer a loss of autonomy and public confidence. We can be certain that threatening relationships of any kind will eventually crumble. Virginia teachers may be telling us exactly this. Perhaps we should listen to them.

Implications for Further Research

I have developed several recommendations for further research as a result of my work in this study:

(1) The present study should be replicated with a different sample of Virginia teachers to learn more about their attitudes toward the state testing program and related standards. Though my own study indicated relatively unfavorable teacher attitudes about SOL testing, these conclusions should be verified before they are accepted.

(2) Demographic characteristics of schools that have improved their SOL test scores should be studied to help determine the characteristics of improving schools. A report presented to the Virginia House Appropriations Subcommittee on Elementary and Secondary Education (Timberlake, 2000) showed strong negative correlations between the attainment of passing SOL scores and the percent of students participating in the state's free lunch program. DOE officials have been quick to point out the increase in the percentage of schools that have reached full accreditation status as evidenced by their SOL test scores; are these schools from relatively affluent communities who were already on the cusp of accreditation? How many of the improving schools were at or near the poverty level? Put differently, what kind of schools and students are improving on state-mandated tests?

(3) A study should be undertaken to determine the effects (if any) of Virginia's high-stakes testing program on principal behavior. In an era of high accountability, what factors motivate Virginia principals to make and implement instructional decisions? Are these factors related to what is good for students, what will lead to personal rewards, what is good for teachers, or what will avoid sanctions? In a North Carolina study, Danielson (1999) determined that principals' compliance with accountability standards was primarily related to the *avoidance* of negative sanctions rather than to what was good for students.

(4) A study should be conducted to determine parental opinions about SOL testing and related standards and sanctions. Virginia parents have had little opportunity to learn about or respond to Virginia's mandated curriculum and assessment program. DOE leaders have typically allowed parent input at statewide meetings conducted at only four sites throughout the state. At the last meeting to discuss the July, 2000 revisions to the Standards for Accrediting Public Schools in Virginia, I drove for over five hours for an opportunity to speak for three minutes before VDOE representatives. Many parents have neither the time, the resources, nor the knowledge of VDOE policies to attend regional meetings of this type. One poll of registered voters in Virginia reported in the *Bristol Herald Courier* (September 12, 2000) reported that 51% of respondents said the SOL tests are not working compared to 34% who said they are. Is this an accurate portrayal of parental opinions?

(5) A study should be conducted to determine student attitudes toward the related policy decisions that shape their everyday life in school. How will able students feel about the standards?

Are they seen as valuable or as a waste of time? Do they enrich the curriculum, or reduce it to only that which is testworthy? And what of struggling students? Do they view the standards as fair? Do they feel as though they are learning more than they would have otherwise? What outcomes will Virginia's testing program have on the lives of students who simply cannot meet the standards to graduate? What will become of Virginia's new category of "forced drop-outs?"

(6) A study should be undertaken to determine the effect of testing on the estimated 25% to 35% of students estimated to be *at risk* (Frymier & Gansneder, 1999). Virginia's accreditation standards currently focus on only 70% of students. If a school reaches the 70% pass rate on SOL tests, it becomes accredited and many kudos are forthcoming. My own school was so recognized in 2000, and we received congratulations from many sources. However, no one asks about the welfare or the needs of the remaining 30% or so of students who do not pass. What are the needs of students with disabilities who are routinely excluded from state testing programs? What are the needs of children who live in poverty? What are the needs of students who are victims of violent environments? What are the needs of transient students? What are the needs of students who simply learn school curricula more slowly?

(7) A study should be undertaken to determine differences in opinions between Virginia teachers and policymakers. Do policymakers believe that SOL testing is causing students to learn more? What do teachers believe about the same topic? Young (1996) and Noll (1999) found substantial differences in the views of teachers and policy-makers about state-mandated testing and related outcomes. A wide gulf in opinion between Virginia teachers and state policymakers fully six years after the implementation of Virginia's curriculum and testing program would argue for substantial efforts toward policy revision.