

# Is There a Better Way? Applying Rules of Science to the Process of Improving Schools

R. Wade Smith

Models of school reform centered around high stakes tests for students and schools are sweeping across the educational landscape of America. All students in the third through eighth grades are now mandated by the No Child Left Behind Act (2001) to take annual tests that will likely provide impetus for a radical reorganization of many schools and school systems. A key theme in the legislation is the elimination of the achievement gap that currently exists between students from affluent and disadvantaged environments. Interestingly, the language suggests that educational improvement initiatives should be data-driven and grounded in sound principles of scientific research. In what may be the ultimate irony, it is possible that the entire theoretical framework of the act, and indeed all accountability programs that use standardized tests as the sole criteria for measuring student achievement, are in violation of the very principles of scientific research that they profess to uphold.

High stakes tests serve several purposes: (a) They are relatively inexpensive to administer; (b) They can be externally mandated; (c) They can be rapidly implemented; and (d) Results are visible (Linn, 2000). The last purpose may be the most attractive one to policymakers because of the likelihood of increased scores over the first few years of a program (Linn, Graue, & Sanders 1990). Whether real, sustainable improvement in student learning has occurred is debatable. Regardless of the reason(s), high stakes tests have become a major emphasis in school accountability models. However, it is clear that for any school improvement model to be effective it must be consistently based on a conceptual model and must measure the relationship(s) between the variables to be studied.

## The Issue

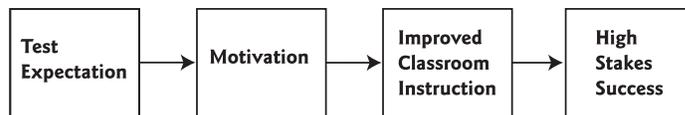
An elementary tenet of scientific research is the identification of variables. This activity drives the entire process of inquiry that ensues. Without a clear understanding of the variables to be studied and their relationship, research becomes a hit-and-miss proposition where serendipity and happenstance are just as likely to produce results as deductive reasoning. If a model of research design is fundamentally flawed, then conclusions drawn from the study are fundamentally flawed as well. This brings us back to two essential questions to be

answered regarding any model for high stakes accountability. Namely, does it accurately portray the relationship of the variables, and what is the strength of the relationships between the variables and the expected outcome?

There are two kinds of variables in a research design: independent and dependent. In an experimental design, the independent variable is manipulated to determine its relationship to the dependent variable. To work backwards from the dependent variable to the independent variable is untenable because one cannot be sure at all that the results are in fact due to the particular independent variable included in the study. For example, one might have an experiment where the relationship between stress and sleep deprivation is explored. In this experiment, stress level would be an independent variable that would be expected to influence sleep duration and quality. If one increased stress levels, it would be likely that a pattern of sleep deprivation would occur. If this pattern were replicable, then a generalized theory for the relationship between stress and sleep might be developed. If, on the other hand, one starts with lack of sleep and tries to conclude that it must be from stress, one is met with a litany of problems. Many other extraneous variables might account equally well for inability to sleep – drug use, pulled back muscle, headache, and loud noises might account for the exhibited sleep patterns. Only by creating a model where the independent variable (stress) can be manipulated and the dependent variable (sleep deprivation) can be measured can relationship be established that might lead to theory development.

Unfortunately, research into student learning typically does not allow for such clean identification of variables as the example given above. Humans are complex, and human behavior typically is influenced by variables that mediate for the effect of other variables. These intermediary variables may exert considerable influence upon the courses of action that are considered and undertaken. Consider the flow chart in Figure 1 that illustrates a hypothetical outcome expectancy for high stakes tests. It is hypothesized that the high stakes test will create a heightened sense of urgency in students and teachers alike. This in turn increases motivation for teaching and learning and improved classroom instruction. If these hypotheses are supported, success on the end-of-year high stakes test is an expected outcome.

**Figure 1**  
**High Stakes Model for School Accountability (HSMSA)**



The High Stakes Model for School Accountability (HSMSA) offered above includes no mediating variables although a direct cause and effect relationship is posited to exist between test expectation, motivation, improved classroom instruction, and success. If the ultimate outcome of accountability models is the improvement of student learning and achievement, particularly for disadvantaged groups, then it is critical to determine if the model in Figure 1 and its hypotheses are correct, both from a practical as well as a moral perspective. Practically speaking, billions of dollars are being pumped into school accountability programs across the nation and if the "medicine" of high stakes accountability is an incorrect prescription for obviating systemic poverty (a keystone of No Child Left Behind), then it is somewhat analogous

---

**R. Wade Smith is Assistant Professor of Educational Leadership in the College of Education at Auburn University.**

to a doctor knowingly treating a patient with an improper drug. From this perspective, the high stakes outcomes of the law extend far beyond the scope of classrooms into the fabric of American society itself.

### **Variables That Mediate for Teaching and Learning**

Methodologically, an age old question arises when considering the variables in any high stakes accountability model. Namely, are there variables outside the control of the school? Elmore, Abelman, and Furhman (1996) note:

One side of this issue... argues that schools can fairly be held accountable only for factors that they control, and therefore that performance accountability systems should control for or equalize student socioeconomic status before they dispense rewards and penalties... The other side of the issue argues that controlling for student background or prior achievement institutionalizes low expectations for poor, minority, low achieving students (pp 93-94).

The authors succinctly summarize the debate. For what can we hold schools accountable? Current high stakes models hope, and No Child Left Behind mandates, that schools improve student learning across all social and demographic strata. Data will be disaggregated by race, gender, socioeconomic level, and special learning needs with all subgroups expected to show long-term continuous academic growth. Further, the argument continues, by becoming aware of the achievement gaps, educators will apply appropriate research-based methods to eliminate these inequities.

On the other hand, critics of the high stakes model argue there are variables outside the school's control. Traub (2000) notes that reforming schools in America has been a stated goal since the 1960s, and yet four decades later little has been done to make a significant dent in educational inequality between affluent and disadvantaged students. Payne and Biddle (1999) reported on this phenomenon and document the acute nature of the problem. According to the authors, when looking at data from the Second International Mathematics Study (SIMS), North American students attending well-funded schools with low child poverty would have ranked higher than every country except Japan. Alternatively, North American students in poorly funded schools with high child poverty scored approximately the same as students from Nigeria and Swaziland.

Payne and Biddle observe that well-funded American schools with low levels of student poverty tend to perform much higher on average than disadvantaged American schools consistent with previous research on this issue (Berliner and Biddle 1995). Their observation was not lost on the lawmakers crafting No Child Left Behind, resulting in the call for greater disaggregation of student data. An intent of data disaggregation is to prohibit more affluent schools from masking the lack of progress being made by their disadvantaged populations within the rosier picture provided by the scores of more advantaged students. However, there remain many questions about the degree to which schools actually can influence the academic progress of privileged and disadvantaged students.

Traub (2000) notes that schools themselves may not be a powerful enough social engine to overcome the kinds of systemic inequalities noted by Payne and Biddle (1999), as follows:

School, at least as we understand it now, is not as powerful an institution as it seems. Most children do not encounter school

until age 5 unless they happen to be in an unusually rigorous preschool program. Anyone who has ever reared a child knows how immense and lasting, are the effects of those first five years. Nor is school quite as all-encompassing as it seems: academic work typically takes up only about half the time that children spend in school. And whom you hang out with, both during and after school, can matter more than what happens in the classroom (p. 6).

Although they may not agree in total, Traub and Payne and Biddle both have noted mediating variables in the relationship between schools and student learning. Traub argues that the collective effect of human and social capital over the first five years can mediate for even the most effective instructional strategies while Payne and Biddle note the strong relationship between poverty levels, school funding, and student achievement. Both perspectives offer compelling evidence that a simple, linear model for high stakes testing is suspect. Further support for this conclusion can be found within a social-cognitive view of learning and motivation.

### **Relation of Self-Beliefs to Learning**

From a social-cognitive perspective, self-efficacy is an important variable expected to act as a mediating variable in Figure 1. According to Bandura (1997), self-efficacy refers to one's ability to organize and execute courses of action required to produce given attainments. These beliefs have a broad influence upon courses of action people choose to pursue, how long they will persevere, amount of effort expended upon a task, resiliency to adversity, the role one's thoughts play in hindering or aiding goal attainment, levels of stress, and, ultimately, levels of accomplishment. Clearly one's personal self-efficacy for academic achievement would be expected to play a powerful mediating role in the ultimate level of academic success experienced.

From a social-cognitive perspective, motivation can be understood as a function of one's general beliefs about his or her competence for a task. If a person believes s/he has adequate ability to perform a task, failure is likely to be ascribed to lack of organization, effort, or the like. Typically, people with these beliefs are likely to make adjustments in their original area of deficiency and retry the task. On the other hand, if a person believes failure at a task is due to insufficient ability, there is a high likelihood they will shut down more quickly, expend less energy, and become resigned to failure more easily (Bandura 1997).

Development of cognitive competencies is most likely through sustained involvement in appropriate activities. These activities are most effective when they integrate mastery experiences into an environment that fosters creation and implementation of challenging self-set goals (Bandura, 1997). The vast differences in social and human capital that exist in students from varying backgrounds seems to require a contextualized curriculum grounded in social constructivist principles or these teaching methods (Brooks & Brooks, 1993). High stakes tests are generally not sensitive to this issue. If the tests act as a screening agent for deficiencies in social and human capital among students, then it is likely that a long-term result will be the further Balkanization of students. If this Balkanization occurs, it would be ironic that the reform program designed to eliminate the achievement gap perpetuated it instead.

Human and social capital, school spending, student self-efficacy, and personal motivation are just four of many variables that have a mediating effect upon teaching and learning. From the perspective of

a school-related variable, quality of instruction must be considered central to any efforts to improve student learning. Hallinger and Heck (1996) report that teachers contribute the greatest variance in student achievement. In another twist of irony, it is possible that a technical rational reform model such as the HSMSA may lower the quality of instruction within many classrooms. Popham (2001) and Darling-Hammond (1991) have both argued elegantly that high stakes tests tend to narrow the curriculum and stifle the enriched learning activities that are most likely to provide meaningful opportunities for enactive mastery experiences and self-directed learning. If this is the case, then it is reasonable to conclude that teaching skills may be supplanted by "teacher proof" curricula that de-emphasize teacher input.

Any discussion of a high stakes testing model would be incomplete without an analysis of the validity of the whole process. High stakes proponents argue that the tests serve as a tool for parents and teachers, offering information regarding what their students know and can do (No Child Left Behind, para. 3). Critics of high stakes testing note that this may be the stated purpose of the tests, but that the reality might be quite different. For example, Freeman et al. (1984) reported that every standardized test used at that time included material that was not covered by any appropriate textbook 50% of the time. In some cases more than 80% of the information was not covered in any meaningful fashion. Admittedly, this study is close to twenty years old. However, few, if any, states with high stakes testing have undergone the rigorous process of validating items and item content with actual textbook information. Even if state leaders were to align the tests with the curriculum, one is still faced with the conundrum of reducing a year's worth of instruction in a content area to a test that typically lasts approximately half a day.

The questions raised to this point can be argued to be speculative. We have just entered the national phase of high stakes testing, and data are only beginning to emerge; but there is historical evidence that can be used to inform us of possible consequences. Linn (2000) addresses the historical evidence about high stakes testing:

As someone who has spent his entire career doing research, writing, and thinking about educational testing and assessment issues, I would like to conclude by summarizing a compelling case showing that the major uses of tests for student and school accountability during the past 50 years have improved education and student learning in a dramatic way. Unfortunately, I cannot. Instead, I am led to conclude that in most cases the instruments and technology have not been up to the demands that have been placed on them by high stakes accountability. Assessment systems that are useful monitors lose much of their dependability and credibility for that purpose when high stakes are attached to them. The unintended effects of high stakes accountability uses often outweigh the intended positive effects (p.14).

## **Conclusions**

Current conceptualizations of high stakes models for student accountability appear to overlook several factors that are critical to creating an effective teaching and learning environment for all students. Variables defined in the model do not account for powerful factors known to mediate for student achievement. Omission of these variables renders the theoretical model overly simplistic and inadequate to understand the relationship between school-related outcomes and student learning. The model also places too much emphasis upon a single high stakes test as an accurate barometer of how much learning has occurred. Such a practice is not best suited to gauging

learning over time, is not particularly sensitive to gaps in human and social capital, and is known to create a "teach to the test" mentality among teachers. Using a student's results on a high stakes test as a proxy measure for teacher effectiveness may further exacerbate critical shortages of teachers in low performing schools. Even when gains are demonstrated via high stakes tests, history demonstrates these gains are transient and regression towards the mean typically occurs within a few years after initial testing occurs. Finally, the social and economic forces behind high stakes testing may Balkanize America's educational systems and widen the divide between advantaged and disadvantaged school systems and students within these systems.

Current efforts at school accountability, with an emphasis upon high stakes testing, appear to be grounded in a questionable theoretical model that is insensitive to many important variables that affect student success. Such a theoretically impaired model should not be allowed to hold sway, particularly given the potential impact to be felt in schools dealing with high levels of student poverty. Rules of science and the moral implications of implementing a well-intentioned, but ill-conceived, high stakes testing program demand more than what is accounted for in this simplistic model.

## References

- Bandura, A. (1997.) *Self efficacy: The exercise of control*. New York: W.H. Freeman and Company.
- Berliner, D. C. & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Perseus Books.
- Brooks, G. B. & Brooks, M. G. (1993). *The case for constructivist classrooms*. ASCD: Alexandria, VA.
- Darling-Hammond, L. (1991, June). *The implications of testing policy for educational quality and equality*. Paper prepared for the American Educational Research Association Forum Washington, DC.
- Elmore, R. R., Abelman, C. H. & Fuhrman, S. H. (1996). The new accountability in state education reform: From process to performance. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education*. (pp. 65-98). Washington, DC: The Brookings Institute.
- Freeman, D. J., Kuhs, T. M., Porter, A. C., Floden, R. E., Schmidt, W. H. & Schwille, J. R. (1984). Do textbooks and tests define a national curriculum in elementary school mathematics? *Elementary School Journal*, 83(5), 501-513.
- Hallinger, P. & Heck, R. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly*, 32(1), 5-44.
- Linn, R. L. (2000). Assessments and accountability, *Educational Researcher*, 29(2), 4-16.
- Linn, R. L., Graue, M. E. & Sanders, N. M. (1990). Comparing state and district results to national norms: The validity of the claims that "Everyone is above average." *Educational Measurement: Issues and Practice*, 9(3), 5-14.
- No Child Left Behind*. 2001. [Electronic version]. Retrieved July 30, 2002 from <http://www.nochildleftbehind.gov/start/facts/testing.html>.
- Payne, K. J. & Biddle, B. J. (1999). Poor school funding, child poverty, and mathematics achievement. *Educational Researcher*, 28(6), 4-13.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Traub, J. What no school can do. (2000, January 16). [Electronic version]. *New York Times Magazine*. Retrieved July 30, 2002 from <http://www.augsburg.edu/education/dc480/noschool.html>.