

Running Head: MAKING SENSE OF NUMBERS

**Making Sense of Numbers – Leveraging
Data From Standardized Tests to Improve Student Performance**

Bruce I. Matsui, Gwen Garrison, Jennifer Rachford, and Nicole Wickler

Institute at Indian Hill

Affiliated with

Claremont Graduate University

American Educational Research Association (AERA)

Seattle 2001 Convention

April 10, 2001

Abstract

Several states now embark on a new approach to focusing the attention of schools on student performance that links 'high stakes incentives and sanctions' to standardized tests. Since 1999, California's public schools have been ranked using a fairly complex formula called the Academic Performance Index (API). Monetary incentives and a host of consequences have been employed to focus the performance of schools on state standards. The Stanford Achievement Test version 9 (SAT 9) represents the primary source of data upon which the API formula is based. Although a heated debate swirls around the use of norm-referenced data as primary measures of student achievement, the potential use of sensemaking practices around the SAT 9, and other data, by members of a given school community can serve as a catalyst for higher achievement (Ornstein & Levine, 1993; Weick, 1995). This paper tells the story of an applied educational research Institute's approach to effectively engage teachers, parents, and students in data derived from the SAT 9 exam for use in improving classroom practices for targeted underperforming students in an urban school district.

In the context of this paper, we document specific practices that were employed utilizing standardized test results through guided interactions that facilitate the interpretive skills of school personnel and community members. These actions include: organizing data into logical subgroups, presentation of data in "user friendly formats," and enhancing the interpretation of data through the use of other quantitative and qualitative sources. This process led to a more robust portrait of the school community and the development of sensemaking skills through the construction of continuous feedback loops (Argyris, 1990; Senge, 1990; Weick, 1995) that served to guide the process of inquiry.

Making Sense of Numbers – Leveraging

Data From Standardized Tests to Improve Student Performance

After nearly three decades of educational reform, we have failed to dissociate socio-economic variables from student performance. Student achievement remains inexorably linked to income, and other variables such as race, language, parental educational background and neighborhoods. For educational reformers who reside in institutions of higher learning nothing is more important than dissociating socio-economic indices from student performances. Unless we can uncouple race, language, and economic backgrounds from student achievement – our efforts will remain largely unsuccessful and episodic. Information garnered from tests, including standardized tests, forms and informs decisions that can aid the process of generating improved student performances. Unfortunately our attempt to use data derived from tests has degenerated into positions that do not serve the best interests of our children. The pejorative statements lobbed by political pundits, policy makers, and educators have not moved us closer to a sustainable solution.

The apparent, but probably momentary, détente among the combatants of the “reading wars”, have shifted to a new front - a battle between those who favor standardized tests as a means of accountability and those who would choose to use other measures such as portfolios and performance standards. The rhetoric that surrounds their respective positions threatens to drown out the many uses of standardized tests that can actually benefit students. This paper argues that effective reform benefits can be gained from the careful analysis of test data. We developed an inquiry-based approach to standardized test score data for members and stakeholders of underperforming schools. This reflective process engaged school-communities members with a much-needed understanding of the patterns about current levels of student achievement. This technique of sensemaking served to shape pedagogy for our work with schools.

Collaboration between institutions of higher education (IHE) and school personnel who work at K – 12 public schools is often touted as an ideal solution to the barriers that effect

student performances. Recent policy and legislative mandates enacted by the state of California have given this partnership a new focus. The state passed legislation that asks external evaluators to work closely with school communities whose students are underperforming. Although California's legislation is embedded in its unique policy structures, other states too, have begun to consider the enactment of similar legislation. Scholars from institutions of higher education are now being asked to move beyond their theories of reform into the arena of action. Said somewhat differently, scholars are being asked to move their theories from the plane of applicability to one of actionability. For most of us at colleges and universities, this shift calls for a metanoia in the way we interact with local public schools. We are no longer able to rely on or simply identify the flaws that link most reform initiatives and school practices. Instead, we are being asked to successfully move schools through and beyond such obstacles.

These high stake testing programs often accompany policies that test the efficacy of older and more traditional collaborative models between IHE's and schools. Simply collecting and analyzing data for the purposes of decreeing recommendations for organizational change are no longer adequate. School communities are asking professors from IHE's to design causalities that engage stakeholders in the actual task of improving rates of student achievement. Designing causalities, based on Argyis's work (1993), is a process of identifying underlying forces that produce unintended consequences and new patterns of belief and action that cause desired results. School communities no longer allow academics the luxury of espousing theory without engaging the actual task of its implementation. Designing causalities puts theory into action by developing "if...then" statements with stakeholders to be used in guiding the implementation of strategies to increase student achievement. The accountability measures attached to the task of creating equitable schools force colleges and universities to serve clients directly or risk sitting

on the sidelines while other agencies engage the work. The change harkens back to Kurt Lewin's belief that the task of social scientists is one of making the world a better place. Quoting from Argyris (1993), "Lewin changed the role of those being studied from subjects to clients. He made it explicit that he was there to be of help because the help, if effective, would both improve the clients' quality of life and produce more valid actionable knowledge" (Argyris, 1993, p10).

This paper defines a pedagogical approach developed by researchers from the Institute at Indian Hill, a non-profit applied research institute affiliated with the Claremont Graduate University that successfully works in Californian and Hawaiian schools. The model employed is analogous to the design studio that architects and engineers employ when they construct buildings. We find this architectural metaphor is particularly useful and appropriate for educational change for it connotes the use of human, social, and ecological capital (Bourdieu, 1983; Coleman, 1988; Putnam, 2000). Within the design studio model, we design causal loops that depend on competence and expertise in research and facilitation skills to guide the change process. This approach engages the semi-permeable organizing structures that govern the daily routines of schools. These structures allow for the use of feedback loops that conduct information continuously while maintaining a focal point for reform.

This paper introduces our particular iteration of this process, identifies the specific skills associated with each stage of designed causalities, and documents a case study that has successfully utilized the approach. As in most cases of innovating new approaches, the development of this particular pedagogy presented numerous challenges to all involved. The discussion and debate around the data continues to swirl in the midst of conflicting values by those seeking to operationalize theory into action. For many of our clients, their first encounter with the process of designing causalities can prove intimidating; for others, it serves as a

springboard for heated conversations about what must be done to increase student achievement. Indeed, developing an environment that takes advantage of social capital through inclusive processes is a delicate and complex act that calls upon skills that are lacking among most researchers. However, it is our thesis that without such capacity on the part of scholars from colleges and universities, most theories will languish in the developmental phase of strategic plans and fail to move into action.

Despite the exigency associated with most reforms over the past three decades, rarely have unintended consequences of policies and mandates been documented. Most accounts of reform focus on processes and largely describe the problems. This falls far short of a comprehensive investigation based on the experience of taking what is applicable and making it actionable. In other words, many of the key issues that occur during the actionable phase of reform are lost because most researchers are absent during the critical phase of implementation and only return to see if their designs have worked. Thus, the broader actions and decisions that lead to results are often lost and remain undocumented. Unfortunately, what is constant over the past three decades of reform are the links between levels of socio-economics and student achievement, and our most pressing problem that remains unchallenged is the creation of excellence and equity for every student.

Our action hypothesis suggests that university researchers must collaborate with public schools as active partners that engage in the task of bringing about desired changes. This view of change requires that scholars from universities develop skills that enable them to “coach” their clients into taking sound actions based on a careful analysis and synthesis of data. We believe that this coaching role helps to facilitate change and enhances the work of the scholar. Our direct work with schools rises to the challenge of collaboration.

The following section explains the Institute at Indian Hill and the design studio process. Then, to demonstrate the inquiry and collaborative processes, we present two brief case studies. The first focuses on a California legislative initiative and the second on an elementary school in Hawaii.

Institute at Indian Hill

Over the past five years, our collaboration with public school communities in the task of improving achievement rates for all students exemplifies an effective partnership between university scholars and school practitioners. Scholars often accuse schools of being overly reticent and resistant to much needed change—often moving at a snail’s pace—while practitioners look upon university faculty as ivory tower theorists who know little if anything about what *really* goes on in schools. Consequently, scholars from higher education and field practitioners regale audiences with ideas to reinvent education for those most in need. However, in practice, the conditions of student learning, especially for those with the least resources in public education, fail to improve. In fact, many indices and measures suggest that circumstances and outcomes for student failure in schools have actually worsened.

During the course of our work, we developed a series of hypotheses about the work of school change that begins with collaborative inquiry and results in actions based on the practice of sensemaking. Our experiences shed light upon the challenge of producing equity and excellence for students enrolled in public schools. Stories told by those who participate in the public-school change processes are too often punctuated by an inability to reverse persistent cycles of predictable failure for students from poorer neighborhoods. This is a point of consensus among scholars and practitioners, although disagreement on particular causes continues to exist.

As the Institute at Indian Hill considered the context of school reform in California and elsewhere, we reminded ourselves that people's perceptions are built through their experiences, feelings, imaginings, hopes, uncertainties, fears, regrets, and triumphs. In California, and we suspect in the rest of the nation, these stories remain central to the way in which decisions and actions are viewed and taken. Moreover, we learned that the combined histories of the members of a workplace constitute the organizational context of a school. The Institute's work in this context drives the following question: "Can we as an educational institute engage schools in a way that causes students who have been unsuccessful to improve significantly?" At present, after making sense by reflecting on what we have seen, we believe the following sentence captures our theory of action:

By creating focus on critical leverage points, and informing action with data and stories taken from classrooms, student achievement can be significantly improved and the belief that "all children can learn" will be operationalized rather than simply espoused.

Sensemaking and Institute's Design Studio Model

When we work with schools we ask permission to examine the actions that produce the current levels of student achievement. To facilitate this process, faculty and graduate students from Claremont Graduate University's School of Educational Studies develop their capacity to organize test results in consumable formats. Data from the SAT 9 standardized tests are packaged in ways to be understood by parents, teachers and students. This reverses the current cycle in which most districts prepare data for use by those in charge of programs--not those who instruct students. During their examination of the data from SAT 9 results, Institute staff ask community members to make sense of the results by sharing their interpretations to two questions:

- How did the data come to be?
- What additional information do we require?

The Institute's design studio model helps to facilitate this sensemaking approach that enhances educational change. It uses methodology, equipment and software to support both quantitative and qualitative research. The Institute's model takes advantage of, and facilitates graduate students' proficiency in, basic research methodologies and the latest software tools. Four points illustrate our design features.

First, our design studio uses state of the art computer equipment with the latest versions of statistical and qualitative software. The Statistical Package for the Social Sciences (SPSS) is the primary statistical software for quantitative analysis. But given the needs, skills, and interests of the graduate students we have access to other statistical packages such as SAS, STATA, and AMOS. The two qualitative software programs in the lab are QSR NUD*IST and HyperRESEARCH. Having access to and use of the equipment and software allows students to explore and gain mastery with their skills and interests.

Second, the Institute at Indian Hill supports the design studio by pursuing a variety of small research contracts that allows students to go full circle with a research or evaluation project. Typically, these contracts are one to three years that begin with the research design, follow up with data collection and data analysis, and conclude with report writing. Therefore, a student can see a project through from beginning to end and delve deeper into areas that relate to personal research interests.

Third, this approach complements the university course work to fulfill graduate degree requirements for students. Students at Claremont can take up to four courses in statistics and four courses in qualitative research. The Institute design studio provides real life settings for

students to practice and deepen their theoretical understanding of quantitative and qualitative research. Two examples will help illustrate this complementary approach.

Quantitatively, the Institute often receives data in its most raw format. One of the specialties of the Institute is the ability to handle SAT9 data from the publisher itself, made available through school district central assessment and evaluation offices. A typical school district data set can have 30,000 student test scores with 750 variables. Graduate students learn how to develop the syntax code and transfer data into SPSS for analysis. Additionally, districts provide second measure test score data such as a local reading or math assessment. After building the data sets, the analysis uses traditional methods to disaggregate the data to the point that teachers and school officials can make effective curriculum decisions. Therefore, working with large data sets, second measure local assessments, and analytical methods that facilitate curriculum discussions between teachers and researchers helps develop the confidence of graduate students and gives them necessary experience in working with varieties of assessment data.

Qualitatively, the Institute provides graduate students with the experiences of doing case study research. Graduate students conduct interviews, observations, focus groups and archival analysis using qualitative research training and software to develop and substantiate hypotheses. This provides important information beyond just the analysis of standardized tests.

Finally, the greatest benefit derived from the design studio approach is the collaboration and collegiality that develops among the graduate students, university faculty, and school practitioners. The Institute develops an induction process that calls upon the veteran graduate students to train the newcomers under the supervision of the Institute directors and professorial staff. This produces reflective practices that facilitate insights and deeper sensemaking.

Because this idea may be applauded or frowned upon as a sound bite typical of the reform community, we owe the reader an honest picture of this theory of action as it takes place in real time and in real settings. The following case studies illustrate our attempt to design effective causalities that move ideas from applicability to action.

Lessons learned from a legislative initiative

In an attempt to reform schools, California's legislators introduced and passed Senate Bill 1X that calls for 'carrots and sticks' for schools identified as underperformers by using a single measure derived primarily from the Stanford Achievement Test 9 (SAT 9). Although scholars and practitioners criticize the use of a single measure to summarily describe student academic progress, this is the mandated context. Much like economic laws that leave out human and environmental capital, the SAT 9 instrument leaves out important indicators such as efforts, quality of teaching, linguistic appropriateness and conditions for learning. Nevertheless, schools are expected to improve student achievement or face consequences that are part of a popular high stakes strategy for school reform.

Unlike other agencies that serve underperforming schools, the Institute builds its change pedagogy by scaffolding additional information upon a careful analysis of the results from the SAT9. This formative stance allows the Institute to capitalize on human and material assets that improve student achievement. Our approach involves gathering comprehensive data and creating feedback loops that generate a constant flow of information around the school community's work. Specific opportunities for school community members to share stories and engage in productive dialogue around the information, allows principals, teachers, and other school community stakeholders to develop and operationalize strategies as they emerge from the

process. In doing so, we believe that school communities are in a better position to make decisions that gain desired results.

Traditionally, scholars distance themselves from sensemaking processes by using language that actually complicates their ideas for practitioners in the field. Our goal is to frame school problems in comprehensible language that is accessible to all stakeholders. Our past experiences indicate that the typical language used to frame school problems makes it impossible for parents and students to comprehend. By using visuals and avoiding terms that only educators understand, we discovered that parents and students comprehend data at remarkable levels. Therefore, this repackaging of information has become an important task of the Institute members. If we are to become an active partner in the process of creating change, we must learn to communicate with those who are engaged in the work of improving student achievement.

The Institute views itself as a partner in educational change and is vested in measurable student achievement outcomes. From experience, we know that creating environments that embrace change is a complex and intensive process. As such, members of the Institute are available to serve a host of functions—as coach, critical friend, as sounding board—and stand ready to engage in all efforts at improvement, including attendance at events, offering professional development strategies to stakeholders on-site, assisting in the determination of resource needs, and assessing the impacts of interventions and programs on student achievement.

Lessons learned from a specific school community

Salt Lake Elementary (SLE) is located west of the city of Honolulu and is part of Hawaii's Central School District. The school is of average size, servicing approximately 600 mostly middle class students (determined by Title I non-status). The teaching force is primarily

Japanese American in ethnicity and numbers 30-35. This case study will illustrate how quantitative data coupled with teacher stories created new insights about student achievement.

To begin, we initially engaged SLE in a review of their SAT 9 test results. Utilizing the design studio model, we specifically analyzed and disaggregated the reading, math, and the open-ended writing tests into logical subgroups: gender, ethnicity, socio-economic status, teachers, and grade levels. We packaged the data in a “user friendly format” that matched the state’s use of reporting SAT 9 test data in stanines. We created and formatted the graphs so that teachers could easily compare one graph to another.

During the presentation of SAT 9 results we asked members of the school community to look for patterns or anomalies in the test data that provide information regarding student performance. The ensuing dialogue among the stakeholders’ jumpstarted multiple interpretations of the test scores. Teachers’ professional stories, and additional data such as video/audio tapes, observation notes, portfolios of student projects, and writing samples augmented the examination. This sensemaking dialogue surfaced patterns of teacher assumptions that influence their own schools’ student achievement. The following is an excerpt from the session.

Teachers: *The boys seem to be doing much worse in reading than the girls.*

Institute Coach: And how did you come to that conclusion?

Teachers: *If you look at the boys’ graph versus the girls’ graph you’ll notice the difference.*

Institute Coach: Let’s see. The boys have four students in the high range and 13 in the low range, while the girls have six in the high and six in the low. That is, the girls have more students in the high stanine range while at the same time having fewer students in the low stanine

range. Is there anything here that could inform us about teaching and learning?

Teachers: *Isn't there research that describes the difference between boys and girls?... Research that says girls are better readers at an earlier age. Could this be showing us that phenomenon?*

Institute Coach: Yes there is research that looks at the difference between boys and girls language and math development. A decade ago data demonstrated that boys did better in math and would catch up in reading later on in 4th or 5th grade. But the research now concludes that on a national scale girls are beginning to outdo boys in all categories.

This dialogue illustrates the powerful capacity to engage in responsive dialogue around test data. Consequently, teachers, in a facilitated conversation, examine their pedagogical practice, which is traditionally left to district and state personnel. For the first time, teachers actively use state data to reflect on their own instruction, a practice previously thought to be burdensome. As they participate in the process of sensemaking, they begin to show an interest in, and appetite for, relevant research and its implications for their teaching and learning.

Another story further captures the power of shared interpretations of test data. The fifth grade teachers initiated an interesting dialogue around the graphs that depicted Total Reading and Open-ended Reading (which is a writing sample about a short story the children read). The letters “W”, “X”, “Y”, and “Z” are used to protect the confidentiality of teacher’s names. Teacher “X” was singled out for discussion because she had every student performing at mid and high levels on the open-ended portion of the SAT9, and her students’ average reading scores were higher than students scores in the other classes. This was important for two reasons. One, it confirmed research by Douglas Reeves (2000) suggesting that a child’s ability to read is directly correlated with his or her ability to write. And two, it revealed a counter-intuitive finding

about classroom conditions for learning. SLE has an issue with the layout of its classrooms in the 5th grade. Three teachers share a large open space partitioned by soft-sided half walls, and one teacher is in a self-contained portable. Teachers located in the shared open space were sure teacher “X” was in the self-contained portable. But as the inquiry process continued with the principal, it was discovered that teacher “X” was teaching in the large open space! Though she had to contend with less than ideal conditions, she was able to meet the learning needs of her students and offered a powerful example to her colleagues. Furthermore, the Institute coach pointed out that these kinds of insights help to promote promising practice that invite professional sharing within the context of the school site.

During the dialogue of data depicted in the ethnicity graphs, teachers showed interest in knowing how each ethnic group preformed on the SAT9, regardless of the population size. Specifically, they were interested in the Hawaiian and Part-Hawaiian students, even though they number less than two percent of the total population of the school. We were able to reconfigure the graphs and share them with the stakeholders during a presentation two months later. This emphasized their concern for every child. We responded to the request and provided the information for further dialogue, thus constructing a double loop feedback cycle (Argyris, 1990).

In addition to SAT 9, Salt Lake Elementary administered a second measure test. The test was an end-of-the-unit exam from the school’s core McGraw Hill reading curriculum. The teachers graded this assessment as proficient or not proficient for each student. Subgroup graphs displayed the number of proficient students by gender, ethnicity, family income, and teacher. To enhance the understanding of the SAT 9 data and second measure test, we performed a correlational analysis between the two tests to demonstrate the relationship between the two test scores. We found a very high correlation ($r=.96$) between a third grade student’s SAT 9 total

reading score and his or her proficiency score on the McGraw Hill reading test. For teachers, this increased their confidence that the use of the core curricular content embedded tests gave a reasonable idea of how the student will do on the SAT 9 total reading test. This finding is important for two reasons. First, since teachers and schools are judged by how well their students do on the SAT 9 test, having another test to measure and monitor each child's performance significantly increases the ability of a teacher to directly impact academic achievement for individual students. Second, it reassured teachers that they did not have to teach to the test in the traditional sense. Instead, they could rely on the core curriculum of the school, full bodied and rich in literature, to adequately prepare students for the high-stakes standardized test.

The next step asks teachers to develop hypotheses for future actions by developing "if...then" statements that address problems surfaced through the sensemaking phase (Argyris, 1993; Senge, 1990). Through developing these hypotheses, teachers are able to merge test scores results with good teaching practices. These conversations employ a whole range of practical, classroom-based strategies that can be used to meet children's individual needs. Therefore, the strategies run the gamut of pedagogies and challenge teachers' beliefs about teaching and learning. It is through these dialogues that sensemaking takes what is applicable and makes it actionable.

SLE decided that the data indicated they should concentrate on reading comprehension and vocabulary. The school picked vocabulary and comprehension because these disaggregated scores were poor, especially for Filipino boys. Therefore, the hypothesis stated that "if" the teachers concentrate on specific language arts and reading strategies, "then" Filipino boys will do better on the SAT 9.

The Institute then designed a week long training which incorporated the collaborative lesson-planning model described by Jim Stigler and James Hiebert in The Teaching Gap (1999). The initial staff development provided promising practices in vocabulary development and the use of non-fiction science materials to increase reading comprehension. To improve reading comprehension for more boys, the Institute specifically employed strategies in the use of non-fiction science materials. Teachers then planned lessons together in grade level groups by incorporating many of the promising practices and “tested” them with small groups of students. During delivery of the lesson, fellow teachers and Institute members observed and took notes. After the lesson, the teachers examined the observation notes and the results of student work. Lessons were altered to better match the objectives/standards desired and then delivered again the following day to another small group of students. By the end of the week each grade level perfected a series of lessons that could be fully implemented in the curriculum.

Several months after the beginning of the new school year, we engaged with SLE’s entire staff in another phase of our cyclic process. We provided a feedback loop that incorporated learnings and a synoptic version from the first professional development group. The staff revisited the “if...then” causal statements and requested data analysis which compared the Filipino boys’ growth in reading from 1999 to 2000. The data analysis indicated that SLE was successful in significantly ($p < .05$) raising the reading comprehension scores of the male Filipino boys. Being able to track sub-groups longitudinally will now add an additional source of data to be incorporated into the sensemaking/feedback loop process during the next professional development opportunity.

This case study illustrates that test scores can be used to facilitate change effectively. Through dialogue and reflection, facilitated by faculty members and graduate students, members

of the school community at SLE were able to make sense of their students' current level of performance. Charts and graphs help to convey the students' level of achievement on a number of measures. In many cases, the data provided an epiphany for those engaged in the process. In the words of Weick (1995), people made "sense of things by seeing the world on which they already impose what they believe." Dialogue (Senge, 1990) around specific data allowed teachers to talk openly and intelligently about the barriers and opportunities to increase student achievement. Identification of specific barriers allowed the group to develop specific lessons aimed at overcoming such barriers.

Summary

This paper posits that the act of sensemaking and cycles of inquiry effectively and meaningfully create a body of knowledge that informs members of a school community about "What they Know and How they Know it." This awareness augments and improves the ability to make decisions that will ultimately improve the performance of diverse groups of students. The role of the university as coach and facilitator using data derived from quantitative and qualitative methods targets classrooms instead of the district or state level. This is a markedly different method of collaboration between a local school, university faculty members, and graduate students. By unpacking our experiences, we hope to shed new light on the roles of the university and their counterparts in K-12 as we try to make sense of what data mean and find avenues to address and improve student achievement in an increasingly 'high stakes' testing environment.

References

Argyris, Chris. (1990). Overcoming Organizational Defenses: Facilitating Organizational Learning. Upper Saddle River, NJ: Prentice Hall, Inc.

Argyris, C. (1993). Knowledge for action: A guide to overcoming barriers to organizational change. San Francisco: Jossey-Bass

Bourdieu, P. (1983). Forms of capital, In J.G. Richardson, (Ed.) Handbook of theory and research for the sociology of education. New York: Greenwood.

Coleman, J.S. (1988). Social capital in the creation of human capital, American Journal of Sociology 94, S95-S120.

Ornstein, A. and Levine, D. (1993). Foundations of education (5th edition). Boston: Houghton Mifflin Company.

Putnam, R. B. (2000). Bowling alone: The collapse and revival of American community. New York: Simon & Schuster.

Reeves, D. (2000) Accountability in action: A blueprint for learning. Denver, CO: Center for Performance Assessment.

Senge, P. (1990). The fifth discipline : the art and practice of the learning organization. New York: Doubleday

Stigler, J. & Hiebert, J. (1999). The teaching gap: best ideas from the world's teachers for improving education in the classroom. New York: Free Press.

Weick, K.E. (1995). Sensemaking in Organizations. Thousand Oaks, CA: Sage Publications, Inc.