

OPERATIONALIZING HIGH EXPECTATIONS THROUGH TEACHER APPRAISALS: A NEW ERA IN SCHOOL REFORM AND THE PURSUIT OF EQUAL OPPORTUNITY

General Session Address at *Beyond Compliance: Embracing Teacher Appraisal* Conference Delivered Friday, April 26th, 2013 by Hardy Murphy, Ph.D.

We are in the beginning stages of our latest reform movement: Evidence based teacher appraisals that include research based instructional practices rubrics and the incorporation of student growth measured in a variety of ways including the use of standardized tests. It comes with the requisite manifestos of other movements. Just as we have the MET studies, other movements have also been accompanied by the use of public research based documents with a national scope to frame the question and call to action, e.g., The Coleman Report of the 1960's, "Nation At Risk" Crisis, 1980s, 1991, Secretary's Commission on Achieving Necessary Skills: What Work Requires of Schools...For America 2000 (SCANS). All emanate from concerns regarding the ability of our schools to educate all children and to prepare them and our country to compete successfully in the new global society.

Each movement in the past as--is the current movement--was accompanied by federal legislation to address social justice and to improve a public school system subject to the criticism of pundits and politicians. To wit: Brown vs. Board of Education and the elimination of separate but equal, The Elementary and Secondary Education Act of 1965 (ESEA) and the infusion of resources into schools to better educate students below the poverty line, Title IX of the education amendments of 1972 and the pursuit of gender equity 1974, and the *Lau* remedies to ensure educational resources for students with language backgrounds other than English.

In the past decade Title 1 was reauthorized in the middle of a standards movement under the requirements of NCLB that included school choice, Vouchers, and accountability consequences for non-performing schools. This latest movement also has federal and state legislative initiatives addressing curriculum standards and teacher appraisal with Race To The Top, The Common Core State Standards, and College and Career Readiness.

Each of these movements, the legislation that accompanied them and the documents that defined the school reform rhetoric have all used a variety of statistics—predominantly standardized test scores to justify the measures that defined them.

With the development of the new approaches to teacher appraisal that define this latest era in school reform it is important to understand that the relationship of these new approaches with standardized tests have spawned an intense debate with their specifications, including reliability and validity being scrutinized with an unprecedented intensity. The debate over their appropriateness has given birth to new statistical models used to determine whether or not adequate student progress happens during a year of instruction. Where the results of these assessments once was limited to percentiles, proficiency, and mastery, we now are looking at various way of measuring growth, including value added approaches to determine the impact of a teacher upon a student's growth, scaled scores, propensities and other exotic metrics in an attempt to capture the elusive growth measure on which the new appraisal systems are based. Indeed, a valid and reliable measure of student growth has become the holy grail of education.

Of course the use of assessments is not new in a profession that in some ways is defined by the process of passing the test. Over the course of a half-century of educational reform, changes in legislation and instructional methods have been accompanied by the emergence of various approaches to determining student learning. In addition to the use of grades, standardized testing has been around in one form or another as gauge for student learning. Tests have been around since the one room school-house.

Standardized testing and educational reform have been dancing partners since the early years of the twentieth century. This relationship can be viewed by the types of tests associated with the period of reform and their purpose. In general, from the early years of the twentieth century until the later decades prior to and entering the twenty first century, standardized testing was defined by the use of norm referenced tests developed from national samples that distributed student performance into percentiles used to sort students in comparisons against a national

norming sample. In the early 1920's, SATs, a norm referenced measure emerged as a way to determine college readiness. It and the ACT remain today the two major gate keeping mechanisms into the nation's colleges and how a student performs on them has an impact upon their life experiences.

In the 1950s and 60's these tests and other norm referenced measures that emerged out of the large testing houses, e.g., the ITBS, CTBS, and the Stanford Achievement tests, were used for tracking into course sequences and selection for vocational and or college oriented subjects in the public schools. In the 1960's the use of these norm referenced measures for program evaluation, e.g., ESEA, and a variety of research efforts took center stage in the educational equity discourse.

In the later years of the twentieth century and the last two decades of this new century criterion based tests have emerged as instruments to be used in the evaluation of educational effectiveness. It is my belief that their emergence in part can be attributed to the stubbornness exhibited by an achievement gap that continues even to this day to demonstrate that public schools are differentially effective for those who differ by race, ethnicity, income, and other demographic indicators including disability and language.

We have not been able to "close the achievement gap" that defines the student comparisons of the norm referenced standardized tests, e.g., the scores of African American, Hispanic students, those on free and reduced lunch, with disabilities, and language differences. These students continue to be represented by scores at lower percentiles than White, Asian and income advantaged students.

Since the comparison of students against each other through norm referenced testing only illustrates that public education has not been able to address the egalitarian ideal of equity in achievement outcomes, a new measurement indices was needed. Thus we redefined education's purpose through the measurement of student performance against the criteria of curriculum standards necessary that ostensibly defined mastery of skills and concepts. These standards allowed education to veer away from the comparisons of students against others and the intractable achievement gap to a comparison of students against standards.

The movement to standards was also accompanied by the slogan that has defined motives and goals in the educational reform movement: high expectations. This term has become ingrained in the national educational lexicon. And, it became the slogan that defined the rhetoric of the NCLB reform era. Hardly a discussion goes by concerning educational reform that the term is not used. However, the commitment of action and resources to make the term a reality is often missing.

The norm reference standardized testing that dominated the pre No Child Left Behind era educational landscape now is accompanied by the use of criterion referenced tests mainly developed at the state level to measure student progress against state standards in the high stakes accountability era ushered in by NCLB. This new approach of using criterion referenced tests to determine student achievement and to evaluate program effectiveness began in the 1970s with minimum competency programs and continued through the 1980s with district accountability through program evaluations at the state and federal level. It lasted through the 1990s to the early late 2000's with standards based accountability systems most notably elevated to prominence through the Title 1 legislation known as No Child Left Behind.

Although criterion referenced testing has been a part of the assessment field since its birth, and was an integral part of the outcomes based instruction movement, their recent use in such a prolific way in the world of accountability and legislative initiatives emanated from documents such as the SCANS report that advocated higher standards in order to improve the preparation of students for success in a work force.

This in turn was a reflection of our political times and the social and political imperative to ensure that America would continue to be competitive in the global economy. Indeed educational reform demands often coincide with economic or political crises of one form or another. The Coleman report of the 1960's was published on the heels of an economic recession that lasted ten months during which unemployment reached 7.1%. A Nation At Risk was a contemporary of the recession that was concomitant with the Iranian oil embargo and an unemployment rate that rose to 10.8%. The SCANS report was issued during the

savings and loans crisis and a negative national growth index. NCLB arrived on the national scene in the early 2000's and a recession that began in 2001 and during which unemployment reached 6%. Of course, these unemployment rates pale in comparison to those that we are experiencing in the current recession and which accompany the release of the MET studies. And, while it would be a stretch to say that the economic strength or lack of such precipitates reform movements, it would not be as much of a stretch to conclude that the economic stress of a country looking for answers contributes to the rhetoric that accompanies reform movements.

This movement to the use of criterion referenced testing comparing student performance against specified standards rather than the norm referenced comparisons of students and their percentile ranks against others in the norming sample was accompanied by the search for criteria determining the acquisition of skills and concepts necessary for success in school and later in the workforce. However, these tests proved not to be the panacea for the disparate outcomes demonstrated in demographic differences in norm referenced test scores. Student performance on the criterion was and is still predictable by demographic characteristics. In fact, the use of these tests had the unintended consequence of states engaging in the manipulation of proficiency standards to avoid the consequences of failing to meet the federally mandated proficiency standards associated with NCLB.

The movement to the inclusion of criterion referenced tests to assess mastery and proficiency on standards was characterized by a recurring array of problems. The use of state developed tests became troublesome because it became apparent early on that the tests and their pass/fail criteria, and the curricula from state to state differed greatly.

Agreement upon which skills and what knowledge was necessary for student success and how to define and express them in the context of learning standards was and still is problematic. Further, how to measure these standards and determining performance that indicated mastery and proficiency proved to be extremely difficult. Lastly, it became apparent that the

problem with using proficiency, mastery, and percentile ranks in the appraisal of teachers was confounded by the fact that where students end up in attainment is related to where they start.

Just as perplexing is the research demonstrating that the use of these tests and the federal initiatives associated with them have not made any more of a difference in eliminating disparity and achieving equity in student outcomes than the norm referenced tests that predated them in the field.

One other theme in the present day assessment discussion is the concept of authentic assessment and the use of student portfolios. Opinions regarding this assessment process have become a significant part of the equity and outcomes discussion. Additionally, the search for valid assessments of student learning is now heightened by the role of teacher evaluation in high stakes accountability growth models, including value added, to determine the impact of a teacher and year's worth of instruction upon student's progress.

However, the use of authentic assessment to determine learning, student portfolio analyses, and the like suffers from the same flaws as the "old" teacher appraisal systems. Subjectivity leads to questions of reliability and validity in the judgment of student work similar to the criticisms of the "old" teacher appraisal systems. A sole reliance upon appraiser judgments of portfolio contents is little different from appraiser judgments of the teaching process. Judgments of portfolio ratings are as vulnerable as the ratings of observers to criticisms regarding validity and reliability. Accuracy is always an issue in the absence of objective measures.

Recent developments to these problems involve the use of student growth measured in a variety of ways to ascertain the impact of a teacher's instruction upon student learning and the use of rather sophisticated statistical analyses including scaled scores to address concerns regarding validity and reliability of varying assessment instruments. The research done in Tennessee regarding student learning is generally seen as the genesis of value added approaches. However, in spite of these developments, the question remains concerning

whether one can, with confidence for appraisal purposes, use an assessment of student learning to determine the impact of a teacher's instruction upon student's academic growth.

In order to address this concern, the literature is urging the use of multiple measures given weights in an appraisal "formula" along with the clinical judgment of effectively trained personnel to ensure fairness and validity in appraisal ratings. As this new era unfolds, any system, in order to be considered effective will have to address these issues in its design, and include extensive professional development to ensure that teachers and appraisers are able to have a meaningful dialogue about what is happening in the teaching and learning process.

School reform to date has not had the broad effect on America's classrooms that is needed to ensure that every child is prepared to take advantage of the American and World opportunity horizon. Even the use of these criterion referenced assessments has not been able to disguise this lamentable reality. And, because of the other issues that have accompanied the assessment dimension of the reform movement defined by NCLB, we are seeing the emergence of new, exotic statistics in the student learning dialogue that have their origin in norm referenced testing. Growth determined through percentiles, propensities for individual and group performance, scaled scores, stoniness, etc. are experiencing an unprecedented importance in education and in particular teacher appraisals.

The current movement seeks to deal with two problems—creating evaluation systems for current teachers so that ineffective teachers can be removed from the classroom. And, in some ways it also has extended the NCLB standards based approach to educational reform with the concepts of Common Core and College and Career Readiness. However, a much more ambitious and comprehensive agenda to improve the quality of teaching ought to be pursued.

These two prongs of the new educational reform movement are rooted in a body of research that shows that the teacher quality makes a significant difference in a student's academic outcomes. Although there is still an intense debate about how much of a difference, the quality of teaching makes in academic performance when compared to among other issues such as parenting and school and family resources, it appears that good teaching does make a definable difference.

And, in order for the goals of Brown vs. Board of Education to be realized, the vision of a better prepared work force in the SCANS report, and the better citizenry to continue the American way of A Nation at Risk, the new agenda of the common core, college and career readiness, and evidenced based teacher appraisals that include valid measures of student growth must be addressed, regardless of how difficult it may be for some to accept if we are to ensure that our schools serve the needs of all children.

In this new era the imperative must be for high expectations to be operationalized in a way that is more than the rhetoric of sound bytes and political posturing. Simply exhorting and affirming more rigorous standards is not enough. High expectations operationalized through effecting teaching of challenging standards are an imperative upon which the future of public schools rests.

If we think that by simply raising standards we will improve our schools, we are mistaken. Just raising standards will not raise student achievement, or close the achievement gap, because teaching will not improve just through the raising of standards. Teaching will improve by professional development, collegial reflection and problem solving, better prepared teachers in revamped schools of education, and appraisal systems that provide valid and reliable information regarding the teaching and learning process, and the use of this information in thoughtful appraisal that creates professional discussion and support for improvement. In this way the effective implementation of the higher standards with fidelity by schools and teachers becomes a probability that can significantly raise student achievement.

The rhetoric of the just passed era where every speech and document was seasoned with the well-worn term “high expectations” were really the expressions of high hopes that interventions based mainly on creating additions to the system rather than fundamentally changing the classroom experiences of students would yield the intended results. They have not. They created the political illusion of commitment through rhetoric. The new reform

movement has to move beyond the politics of rhetoric and into the commitment to change public education so that the potential of all children, regardless of their demographic profile, are realized through high levels of academic achievement.

The high hopes of the previous reform movement must be replaced with clear and definitive policy including the allocation of resources where they are most needed and just as importantly in the setting of appraisal standards for instruction that just as rigorous as the learning standards we are setting in this new era of the Common Core and College and Career Readiness for our students.

We have had high hopes in all of the past efforts of school reform. We passed them off as high expectations. High expectations start with a clearly understood and stated policy position articulating rigorous requirements for the teacher appraisal process involving both sides of the instructional equation: The process or inputs for teaching and the outputs or measures for learning. High expectations require that we prepare the school for the teacher in its supports and its operations so that the teacher and the school can be prepared for the students. This in and of itself will require a shift from the present thinking of preparing students for the school.

And, it has to include the imperative of developing the human capital that both teachers and students represent. For teachers this amounts to meaningful professional development and their affirmation in a collegial process in their work environment.

When this happens our schools and teachers will be able to meet their responsibility for teaching any child that walks through the doors of the public school and its classrooms. When this happens in our public schools there will be no need for us to fear charter schools or vouchers. Parents will choose us because we are the pathway to the American Dream. In this public school environment we will acknowledge and address the needs of each child and not use their differences as explanations for failure. When we manage to make this a reality in our classrooms, every child's growth will count in an highly effective appraisal process where distinguished teachers and high achieving students demonstrate growth in a vibrant, challenging, rigorous, meaningful teaching and learning process.

So what are the practical and necessary actions for us to take? We can treat teachers as true professionals and involve them in the decision-making that impacts the educational experiences of the students in their classrooms. We can prepare them for effective classroom instruction through ongoing professional development and continuing education. We can elevate the image of the profession by changing the rhetoric concerning teaching and teachers. We can develop the right incentives for drawing teaching candidates from the top tiers of college students. We can develop pay scales commensurate with the responsibility that our country has invested in them. They are custodians of our nation's future and their compensations should reflect this. And, we should initiate a call to action for colleges of education to develop standards that would result in better prepared teachers. We should develop mentor relationships with teachers who have demonstrated mastery to support new teachers. We should encourage ineffective teachers to choose another profession. By and large we are a profession of competence. However, our inability to deal in a professionally effective fashion with incompetence contributes to the rhetoric that is destroying the image of the profession. We have to incent the most effective teachers to the schools with the most challenging student populations. It is in this way that education will truly realize its mission of being the route to equal opportunity and productive citizenship for all. And we must develop true instructional leadership in principals. Without this instructional leadership the respect that results in collegial discourse and effective problem solving will not occur at the building and classroom levels.

Policy decisions at the level of governance and management are also a critical component of the culture change. We have to operationalize high expectations with clearly defined student outcomes that are reasonable and measured with validity and reliability. These expectations for student outcomes must accompany clearly articulated expectations for the design and delivery of instruction with a thoughtful process for giving feedback through appraisal. And, public education has to become less obsessed with the threat of charter schools and vouchers and more concerned about being competitive in an environment in which choice exists.

Making this happen would be truly effective educational reform. Perhaps the vision of Brown vs. Board would be realized, America's public schools would optimize the development of the human capital that sits in the classrooms of our schools and equal opportunity for all would become the greatest achievement of our time.