NCME national council on measurement in education

Volume 14, Number 4 NEWSLETTER December 2006

FROM THE PRESIDENT

Dan Eignor, Educational Testing Service

As I write this column, I'm thinking about the very productive two and half day Board meeting that took place in Washington, DC on November 9 – 11. We have been conducting our fall Board meetings in Washington because it allows us to meet directly with Robert Smith and Jerry Sroufe from AERA Headquarters to talk about details of the upcoming annual meeting. While the Rees Group has taken over almost all NCME management activities, we are still holding our annual meeting in conjunction with AERA's meeting and hence need to work out a number of details. We had a productive discussion, and I think the membership will be happy with the arrangements.

2007 Annual Meeting News

All NCME sessions will be held in the Intercontinental Hotel next April. We will have rooms for five sessions per time slot. AERA is planning on having all AERA Division D sessions in the Marriott Chicago Downtown. Both properties are on Michigan Avenue, almost directly across from each other, so switching hotels between sessions should not be a problem. The main AERA hotels are the Sheraton Chicago Hotel and Towers (for registration) and the Hyatt Regency Chicago (for exhibits). Both are about a ten minute walk from the Intercontinental, so NCME members should be able to switch to these hotels during the time period between sessions. Given the problems with distance between NCME and non-Division D AERA sessions encountered in both Montreal and San Francisco, this represents a welcome change.

Because of publication dates for *EM:IP*, we will be sending conference and hotel registration materials, information on training sessions, and upcoming program highlights to members via a separate brochure, as was done last year. This information will also be available on the NCME Website. We are planning to have this information available to you in early December.

2007 Program co-chairs Mary Pitoniak and Mike Jodoin are making some final decisions about the program for next April. I have seen the almost final version of this program, and I think that NCME members will find it to be very informative. Cheryl Cardell has also put together a fine set of training sessions for this year. With respect to our program, NCME is dependent on AERA having finalized their program first, so we can check that there is no double scheduling of individuals. AERA has let us know that they are ahead of schedule compared to the last few years, so you should receive your paper meeting program earlier than usual. This material will also be on the NCME Website, again earlier than it has been the last few years.

This year's program will follow the format followed prior to last year, with one exception—the annual run/walk will occur on Thursday rather than Wednesday morning. The No-Host Reception will still take place Tuesday night with the annual Graduate Student Issues Committee session just before it. The Breakfast will again be held Wednesday morning. Rees Group staff will be on-site at the Intercontinental prior to the beginning of training sessions, so the whole training session on-site registration and room assignment process should go much more smoothly than in past years. Also, if you've pre-registered, look for a table in the Intercontinental where you can pick up materials (badge holders, supplements to programs, dining guides). If you have not pre-registered, you will need to go to the AERA/NCME registration area in the Sheraton.

AERA/APA/NCME Standards

The Standards Management Committee and representatives from the three organizations met in Washington, DC on October 30 – 31. I am happy to inform you that work on the next revision to the *Standards* has begun. Members of the Management Committee have started the process of selecting Joint Committee co-chairs and the individual members of the Committee. Doug Becker, present chair of the NCME Standards and Test Use Committee, will be serving as the NCME liaison to the Joint Committee and will be attending all upcoming Joint Committee meetings.

On-Line Manuscript Submission System

Terry Ackerman, chair of the Publications Committee, *EM:IP* editor Susan Brookhart, incoming *JEM* editor Jim Carlson, and incoming Publications chair Steve Sireci have been hard at work getting the details of an on-line manuscript submission system in place for use in 2007. Based in part on a recommendation from Blackwell staff, we have chosen the ScholarOne online submission system. Certain details need to be worked out, but after this and a trial period, the new system will be ready for operational use. I hope you will agree that this represents an exciting new development for our organization.

Handbook of Teacher Evaluation

After some discussion by the Board, we have decided to go ahead with a new edition of the *Handbook of Teacher Evaluation.* Terry Ackerman and Board member Judy Koenig have been working with Advisory Group members Drew Gitomer, Karen Mitchell, Andy Porter, and Marnie Thompson on a plan for this activity, possible content of the new edition, and the identification of names of individuals who would be good choices for co-editors. Terry, Judy, and the Advisory Group have made amazing progress on this project and deserve our congratulations.

NCME Website

While John Hofman from the Rees Group has been able to make systems-related changes for the updated NCME Website, the revision process has been slowed because we haven't had people and a process in place for dealing with new Website content. With the help of the Board, Terry Ackerman is putting together a special Task Force that is charged with dealing with content-related issues for the updated Website. Thus far, George Engelhard has agreed to chair this Task Force, and past NCME President Jim Impara and current website coordinator Dave Miller have agreed to serve as members of the Task Force. I believe we soon will have people in place to facilitate what, I think you will all agree, are needed changes to the Website.

Long Term Planning for NCME

Given the rapidity with which new developments and changes are taking place for NCME, Board members have thought for some time that some formal attention should be paid to long-term planning for the organization. The last time the Board addressed issues of this sort was ten years ago. With the help of a planning consultant, the Board spent one day of its recent two and a half day meeting working on strategic planning issues. Developing a strong long-term plan for NCME is key to the health and growth of our organization. We will keep you informed as the long-term plan develops.

International Standards

I wrote about the International Organization for Standardization (ISO) and upcoming work on International Standards for psychological assessment in my last column. Representatives of the three organizations (AERA, APA, and NCME) met in September with the American National Standards Institute (ANSI), the American member of ISO, to investigate ways in which organizations could influence the content of the new International Standards. We learned that such input could be provided via a Technical Advisory Group (TAG), but in order to have such a group, an organization willing to serve as a TAG administrator needed to be identified and join ANSI. There are fairly substantial monetary implications for joining ANSI and for becoming a TAG administrator. None of the three organizations (AERA, APA, NCME), separately or collectively, were willing to pay the up-front costs necessary to form a TAG. It appears that the Association of Test Publishers (ATP) may pay for and take on TAG administrator responsibilities. Details as to how NCME would provide input to the International Standards preparation process through ATP as the TAG administrator, should they definitely take on this role, need to be worked out. The bottom line is that while it appears a means for NCME to influence the content of the International Standards will exist, the mechanism for doing so is going to be different from what might have been initially envisioned.

CAREER OPPORTUNITIES IN CERTIFICATION AND LICENSURE

Mark Raymond, Director of Psychometric Services, The American Registry of Radiologic Technologists (ARRT)

During my third year of graduate studies at Penn State, I decided to arrange an internship. My colleagues in clinical and school psychology were headed down the internship path, and it seemed like a good idea for me given my interests in psychological assessment. I wrote to a Penn State graduate who had recently accepted employment with a certification board, and he generously invited me to work with him for a semester. That was fall 1983. I mention this bit of my past because had it not

been for the internship, I would not have been aware of certification boards as a potential employer until much later in my career. Since that time, I have worked almost exclusively in certification and licensure.

Who Works for Certification Boards?

I've recently returned from a conference sponsored by the National Organization for Competency Assurance (NOCA). Conference participants included members of certification boards (e.g., crane operators, dieticians, lactation consultants), staff employed by those boards, and representatives from various testing companies. Of those employed by a certification board as a measurement specialist, most had job titles such as vice president for research and development, director of psychometrics, certification manager, psychometrician, or test specialist. A few had become executive directors of a board and, as such, are leaders in the profession that they serve.

The educational backgrounds of measurement specialists in certification are probably more varied than in other areas of testing. Although most have a degree in educational psychology, I/O psychology, or quantitative methods, some have backgrounds in other social sciences (e.g., educational administration, social psychology). Others have advanced degrees in their own field (e.g., nursing), sometimes coupled with a master's degree or PhD in education.

What Do Board Psychometricians Do?

Responsibilities vary considerably for those employed by certification boards. Job title has an obvious influence on work activities, but even more important is the size of the organization and whether that organization contracts with a testing company for services. For example, the ARRT has a staff of 50, with seven positions in psychometric services (4 professional; 3 support). The four professional staff are measurement generalists. Each is involved in almost all aspects of testing, from job analyses to test assembly, from item calibration to equating, from explaining eligibility requirements to educators to explaining score reports to unsuccessful candidates. In contrast, some boards may employ large numbers of measurement specialists. Psychometric staff who work for larger boards will find more opportunity to specialize, and will have responsibility for just a few measurement-related activities.

Reasons to Choose this Career Path

Some of the most significant recent innovations in testing have occurred largely through efforts of licensure and certification boards. Organizations such as the American Institute of Certified Public Accountants (AICPA), the National Board of Medical Examiners (NBME), the National Council of Architectural Registration Boards (NCARB), and the National Council of State Boards of Nursing (NCSBN) have been leaders in areas such as computer-adaptive testing, performance testing, and computerized simulations. Employment with one of these boards can be every bit as exciting as working for a major testing company, if not more so. Most large boards have resources sufficient for funding worthwhile projects. In addition, the larger boards may employ a large enough group of psychometricians so that one can benefit from collegial support.

However, most boards employ smaller numbers of psychometricians. One highlight of working for a small- or mid-sized board is autonomy. Another is impact—you *will* make a difference. As one of a few measurement specialists in an organization (or perhaps the only one), your opinion will weigh heavily on matters related to test development, equating, score reporting, and so on. Your executive director and board will ask for your advice and, more often than not, act on it.

I had dinner a few weeks ago with a colleague who works for a major testing company. Our jobs are very similar in almost every way but one: She was five layers down from the top; I am one. Of course the flip side of autonomy is accountability. A good night's sleep can be a challenge when you don't have a panel of psychometricians to call on if something does not go according to plan. Decisions are on your shoulders, and you learn to deal with it.

Other Things to Think About

Certification and licensure is a specialty in search of a home; it is part educational measurement, part personnel testing. We have a minor presence at NCME, a nagging one at SIOP, and none at all within most other societies. Although the 1985 version of the *Standards for Educational and Psychological Testing* finally included a chapter on certification and licensure, the 1999 version integrated that chapter with the one on employment testing. None-the-less, this subspecialty is experiencing some deserved recognition. The recently published *Handbook of Test Development* includes at least one chapter specific to licensure and certification, while the forthcoming edition of *Educational Measurement* promises a comprehensive chapter on certification and licensure. But, the bottom line is still this: If you prefer the social mainstream of educational testing, don't work for a certification board.

Although a little social isolation is tolerable, any amount of intellectual isolation is not. Working for a small or mid-sized board requires additional effort to stay current. What to do? Read journals. Attend conferences. Establish relationships and

stay connected. I was fortunate enough to have been employed by ACT for five years in the late 1980s, where I had the privilege of knowing some of the country's prominent psychometricians. Still today, I am quick to contact former ACT colleagues to seek advice on technical issues…and so far they have been more than charitable with their responses. So long as one is willing to maintain such contacts, working for a certification board can provide a truly rewarding environment in which to apply the science of educational and psychological measurement.

NCLB GROWTH: WHAT ARE WE LEARNING AS REAUTHORIZATION APPROACHES?

Jennifer L. Dunn and Scott F. Marion, Center for Assessment

Introduction and Background

In November 2005, U.S. Secretary of Education, Margaret Spellings released a letter inviting states to submit proposals to the U.S. Department of Education (USED) for developing growth models to meet the principles of *No Child Left Behind* (NCLB). As of November 2006, five growth models (TN, NC, DE, AR, and FL) had been approved for use in 2006-07 NCLB accountability decisions. USED plans to approve no more than five of nine additional proposals, submitted November 2006, by March 2007 (U.S. Department of Education 2006, November 9). While the requirements for incorporating growth models into NCLB accountability systems prohibit the use of some of the more familiar longitudinal growth models, they open the door for innovative amendments of these models. The goal of this article is to outline some of the recent policy and technical developments that have arisen because of the increased flexibility, namely the use of growth models, in NCLB accountability.

In theory, incorporating growth into accountability should shift what it means to be a well-performing school. Under status models, a "good" school is determined using student achievement levels from a single time point. An "improving" school is one where the school's achievement level, based on different student cohorts, increases from one year to the next (Carlson, 2006). In contrast, growth models evaluate schools according to how much students learn or develop beyond their achievement when they entered. Whether or not a school is "improving" is based on comparing the change in student achievement from the current year to the change that occurred the previous year (Carlson, 2006). Many argue that growth-based models are more valid than status or improvement approaches for holding schools accountable for student achievement because they focus on the same students over time (Carlson, 2006; Hill, et al., 2005, Goldschmidt, et al., 2005).

A range of measurement models have been developed to evaluate how well schools (or other units) elicit student growth. The simplest approaches calculate the difference between measures of a student's achievement across two or more time points. Such models, often called change score models, are dependent on the test score distributions from grade to grade, the normative group, or if used, the vertical scale. Because these models involve the direct comparison of two observed scores, they have been criticized for their unreliability. To correct for unreliability by including multiple time points, multilevel models that rely on sophisticated statistical techniques have been developed (e.g., Sanders, Raudenbush, Bryk, McCaffery, Choi, Goldschmidt). These models are able to explain the variation in student test scores that can be attributed to the status and growth of the student and the status and growth of the school. They have been criticized for measuring growth using normative as opposed to criterion-based techniques and for being difficult to understand. Consequently, a third class of models designed to bridge policy goals with technical considerations has emerged. While these models vary considerably, they fall under the general classification of hybrid models. Some of the more common approaches are Value Tables (Hill et al., 2005), rate of expected academic change (Doran & Izumi, 2004), and the Hybrid-Success (Kingsbury, Olson, McCahon & McCall, 2004) models. While these newer approaches provide a way to move growth models into a standards-based context, they still fall short of the requirements for growth models under NCLB.

Growth and NCLB

Seven Core Principles, outlined by USED, were used to ensure the growth proposals were technically sound, could be validly incorporated into school accountability systems, and were consistent with the spirit of NCLB (U.S. Department of Education, 2006, January 25). The growth models must: (1) ensure all students are proficient by 2013 – 14 and set annual goals to ensure that the achievement gap is closing for all groups of students, (2) establish high expectations for low-achieving students regardless of student demographic characteristics and school characteristics, (3) produce separate accountability decisions about student achievement in reading/language arts and in mathematics, (4) include all students and hold schools and districts accountable for subgroup performance, (5) be based on annual assessments, comparable from year to year, in each of grades three through eight and high school in both reading/language arts and mathematics that have been operational for more than

one year and approved through the NCLB peer review process, (6) track student progress across years, and (7) include student participation rates and student performance on an additional academic indicator.

In order to meet the Core Principles, states were forced to align their definitions of growth so that the target for all low performing students was proficiency. As a result, in developing growth proposals, states adjusted existing measurement models to incorporate proficiency. The hybrid models were adjusted to incorporate proficiency or progress towards proficiency as the target for non-proficient students while students starting at or above proficiency were held to goals of proficiency or higher. The multilevel models were adjusted to calculate projections to proficiency. All available assessment data were used to predict whether each student would score at or above proficiency at some specified time point in the future. The projections are not based on the demographic characteristics of the student but do assume that each student attends the average school in the state. The change score models were adjusted so that instead of calculating the difference between two scores, the difference between each student's score and proficiency was determined. This difference was used to establish student growth targets, based on making up a proportion of the "distance" to proficiency. Regardless of the model implemented, if students meet or exceed their target they are deemed as making adequate growth.

The growth models adopted by the states varied substantially. North Carolina and Florida use a change score model based on attaining proficiency in three years. Arkansas also makes use of a change score model, but the student growth targets are based on attaining proficiency by eighth grade. Tennessee uses a projection model where the growth decision is based on the number of students projected to score at or above proficient three years into the future. Delaware uses a value-table based system.

Once the measurement of growth had been defined, the manner in which growth was to be incorporated into NCLB accountability needed to be tackled. The most obvious solution was to use growth as a replacement for status. Instead of holding a school accountable for the percentage of proficient students, a school would be held accountable for the percentage of students meeting their growth targets. Given the high-stakes nature of NCLB and the suspicion that growth is occurring at a low rate, states were justifiably reluctant to use growth as a replacement for status. Alternatively, growth could be used in conjunction with status. More specifically, a school would be deemed successful if it meets or exceeds the status requirements or the growth requirements. Because the guidance indicated that all students (not just those who score below proficient) must be included in the growth model, states were essentially left with two options: adding the number of non-proficient students meeting their growth targets to the counts of proficient students, or a separate growth-based decision that includes growth targets for proficient students. Tennessee, Florida, and Arkansas made use of the latter approach. In their growth models, proficient students count positively towards a school's score only if they meet their growth targets. Proficient students who do not meet their growth targets count against the school. Although North Carolina made use of the former approach (simply adding the number of non-proficient students meeting their growth targets to the count of proficient students), growth targets are reported for all students. Delaware's model represents a combination of the two systems. All students are assigned a value-table based score, but while the score for non-proficient students is based on moving up at least one performance category, the score for all proficient students is constant. This means that proficient students cannot penalize a school unless their performance falls below proficiency.

The world of NCLB accountability may be expanding to incorporate status, improvement, and growth, but has it broadened the definition of what it means to be a successful school? The NCLB pilot criteria have essentially shifted the definition of a "good" school to one where the students are proficient or on track to proficient. A school is improving if it increases the number of proficient or on track to proficient students. Although the NCLB pilot criteria have constrained the more traditional view of growth, this definition of what it means to be an effective school represents an improvement over a pure status model. Time will tell how the allowable growth models will impact the validity of NCLB accountability decisions.

References

Carlson, D. (2006). *Focusing state educational accountability systems: 4 methods of judging quality and progress.* Retrieved November 21, 2006, from http://www.nciea.org/cgi-bin/pubspage.cgi.

Doran, H. C., and Izumi, L. T. (2004). *Putting education to the test: A value-added model for California*. San Francisco, Pacific Research Institute.

Goldschmidt, P., Roschewski, P., Choi, K. C., Auty, W., Hebbler, S., Blank, & Williams, A. (2005). *Policymakers' guide to growth models for school accountability: How do accountability models differ?* Washington, DC: CCSSO.

Hill, R., Marion, S., DePascale, C., Dunn, J., & Simpson, M. A. (2006). Using value tables to explicitly value student growth. In R.W. Lissitz (Ed.), *Longitudinal and value added models of student performance* (pp. 255-282). Maple Grove, MN: JAM Press.

Kingsbury, G. G., Olson, A., McCahon, D., and McCall, M. S. (2004). *Adequate yearly progress using the hybrid success model: A suggested improvement to No Child Left Behind*. Retrieved November 22, 2005, from http://www.nwea.org/research/grd.asp.

U.S. Department of Education (2006, January 25). *Peer review guidance for the NCLB Growth model Pilot Applications.* Retrieved November 21, 2006 from http://www.ed.gov/admins/lead/account/growthmodel/index.html.

U.S. Department of Education (2006, November 9). *Secretary Spellings approves additional growth model pilots for 2006-2007*. Retrieved November 21, 2006 from http://www.ed.gov/news/pressreleases/2006/11/11092006a.html.

U.S. Department of Education (2005, November 21). *Key policy letters signed by the Education Secretary or Deputy Secretary*. Retrieved November 21, 2006 from http://www.ed.gov/policy/elsec/guid/secletter/051121.html.

MEASURING ALIGNMENT

Andrew Porter, Vanderbilt University

The No Child Left Behind Act of 2001 (NCLB) requires that each state align assessments to content standards. If the content assessed is exactly the same as the content represented in the standards, alignment is perfect. There are two ways in which alignment can be less than perfect: Content in the standards may not be assessed, and content assessed may not be in the standards. Even more important is the alignment of instruction to assessments and standards (Porter, 2006).

Measuring alignment has been of interest for decades (Cohen, 1995; Freeman et al., 1983), but recently interest has heightened. Several reviews of different approaches to measuring alignment have appeared (Ananda, 2003; Bhola, Impara, & Buckendahl, 2003; CCSSO, 2002; Olson, 2003; Rothman, 2003).

The currently most popular method for measuring alignment of assessments to content standards was developed by Norman Webb (1997, 2002). The procedure involves experts' judgments on four criteria related to content agreement between assessments and standards. According to Webb, categorical congruence is met if there are at least six items measuring the topics represented by a standard. Depth of knowledge consistency asks experts to judge whether items in the assessment are as demanding cognitively as what the students are expected to know and do as stated in the standards. At least half the items corresponding to an objective (within a standard) have to be at or above the level of knowledge of the objective. The four ordered levels of depth of knowledge assume that when a student demonstrates achievement on a higher "depth of knowledge" they necessarily have achieved on all lower depths of knowledge. Range of knowledge correspondence is met if at least half the objectives for a standard are measured by at least one assessment item. Balance of representation indicates the degree to which one objective within a standard is given more emphasis on the assessment than another. Webb's index of balance of representation is a function of only those objectives for a standard that had one or more items assessing the objective; the index ranges from 0 to 1.0.

There are several aspects of Webb's measure of alignment that are important. First, a set of content standards is a given, and Webb measures how a particular assessment aligns to those content standards. Second, for each of the four dimensions of alignment, Webb sets a criterion for how much alignment is enough. Third, alignment is a function of content, and content is defined by topics and cognitive demand. In all cases, content standards dictate the level of detail at which topics and cognitive demands are defined. The more general the standards, the less precise the distinction among types of content.

Another frequently used measure of the alignment between assessments and content standards was developed by Porter and his colleagues (Porter & Smithson, 2001; see http://www.ccsso.org/projects/surveys_of_enacted_curriculum). Content languages are used to complete content analyses of a state's content standards and (separately) a state's assessment of student achievement. Content languages have been developed to describe the content of mathematics, science, and English language arts (see http://www.ncrel.org/sec/). The languages are two-dimensional and can be presented in a rectangular matrix with *topics* as rows and *cognitive demands* as columns. A language might have 100 topics (rows) and five levels of cognitive demand (columns). Content is defined at the intersection of a particular topic (row) and a particular cognitive demand (column)—the cells of the matrix. For example, content might be to use a linear equation (the topic) to solve a novel problem (the cognitive demand).

The basic data are proportions of content emphasis in each cell of the topics-by-cognitive-demand matrix. Alignment is perfect if the proportions for the assessment match cell by cell the proportions for the standards. A conceptually straightforward index of alignment is defined as:

$$\text{Alignment} = 1.0 - \frac{\sum |x - y|}{2}$$

where x = assessment cell proportions and y = standards cell proportions. The index ranges from 0 (no alignment at all) to 1.0 (perfect alignment). The alignment measure has been used to describe alignment in many states in the subjects of mathematic, science, and reading/language arts.

This approach to measuring alignment has several important properties. First, the basic data are a function of a common content language. Thus, all measures of alignment are a function of the same levels of detail in distinctions among topics and cognitive demand. Second, because measuring alignment begins with content analyses using one common content language, once a state's content standards have been content analyzed, data can be used not only to measure the alignment of one state's standards with that state's test, but with other tests and standards as well. Further, if a state should change its test but not its content standards, then the standards do not need to be reanalyzed—only the new test needs to be analyzed.

Third and perhaps most importantly, the procedure is not limited to measuring the alignment of tests to standards. Anything having content that can be described using the common content language can be aligned with anything else having content that can be described using the common content language (e.g. instruction, curriculum materials, tests, standards). The alignment among content standards across states can be measured. Similarly, alignment across states can be measured for instruction or tests. The alignment of instruction to content standards can be measured to assess opportunity to learn. Alignment between instruction and standards can be measured at the individual teacher level, and variance across teachers can be studied. Fourth, the alignment index is symmetric. For example, degree of alignment is a function of both content that is tested but not in the standards, and content that is in the standards but not tested.

What the index does not indicate is how much alignment is enough. A reasonable but somewhat arbitrary criterion for how much alignment is enough could be set as Webb has done for each of his four dimensions of alignment. Alternatively, alignment can be judged comparatively. For example, is alignment of a state's test to that state's standards higher than the alignment of the state's test to other states' standards? At the same time, alignment of a test with content standards for only one form of the test should not be perfect. One form of a test is a sample of the content in the standards. If the state uses a different form of the test each year, as it should, then content analyses could be conducted across multiple forms of the test. At a point when sufficient numbers of items are present so that the sample becomes close to being the domain, the test should be perfectly aligned to the content standards. But the specificity of the content standards also limits the degree of alignment. The more vague and general the content standards, the less perfect alignment can be.

Use of a common content language allows for some powerful data displays. For example, if one has content-analyzed standards and assessments from each of several states and perhaps national professional standards, the results can be displayed in a standards-by-assessment matrix of alignment values (Porter, 2002). The main diagonal of the matrix reports alignment of each state's test with that same state's content standards. Presumably, a state's assessment would be more highly aligned with its own content standards than with other states' content standards. Another display is in terms of topographical maps that can be created using a variety of charting software, including Excel (Porter, 2002). On the maps, what ordinarily is thought of as north and south represents topics and east and west represents cognitive demand. Shading represents relative content emphasis and is analogous to attitude on a topographical map. The maps clearly show not only what content is emphasized in the content standards, but what content is not.

There are other approaches to measuring alignment including Achieve's (Rothman, Slattery, Vranek and Resnick, 2002) and Project 2061 (Kesidou and Roseman, 2002), though these appear to be less often used. For each method of measuring alignment, one should ask about quality and utility. A key issue for all approaches is the reliability of content analyses and for Porter's approach, the reliability and validity of teacher reports of instructional practice. Results available are promising, but more work is needed (Porter, 2006). Another key issue is how much alignment is enough. The answer undoubtedly depends on the use, but more thinking here is needed as well.

References

Ananda, S. (2003). *Rethinking issues of alignment under No Child Left Behind.* San Francisco: WestEd.

Bhola, D.S., Impara, J.C., & Buckendahl, C.W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice, 22*(3), 21 – 29.

Cohen, S.A. (1995). Instructional alignment. In Lorin W. Anderson (Ed.), *International encyclopedia of teaching and teacher education* (2nd ed.). New York: Pergamon.

Council of Chief State School Officers (2002, September). *Models for alignment analysis and assistance to states.* Washington, DC: Author.

Freeman, D.J., Kuhs, T.M., Porter, A.C., Floden, R.E., Schmidt, W.H., & Schwille, J.R. (1983). Do textbooks and tests define a national curriculum in elementary school mathematics? *Elementary School Journal, 83,* 501-513. (Reprinted in *The Education Digest,* 1984, March, 47 – 49.)

Kesidou, S., & Roseman, J.E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review [Electronic version.]. *Journal of Research in Science Teaching, 39*(6), 522 – 549.

Olson, L. (2003, Spring). Standards and tests: Keeping them aligned. *Research Points: Essential Information for Education Policy, 1*(1).

Porter, A. C. (2006). Curriculum assessment. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of Complementary Methods in Education Research*, pp. 141 – 159. Washington, DC: American Educational Research Association.

Porter, A.C. (2002, October). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3 – 14.

Porter, A.C., & Smithson, J.L. (2001). Are content standards being implemented in the classroom? A methodology and some tentative answers. In S.H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states—One hundredth yearbook of the National Society for the Study of Education, Part II* (pp. 60 – 80). Chicago: University of Chicago Press.

Rothman, R. (2003). *Imperfect matches: The alignment of standards and tests.* Manuscript in preparation, National Research Council.

Rothman, R., Slattery, J.B., Vranek, J.L., & Resnick, L.B. (2002). *Benchmarking and alignment of standards and testing.* (CSE technical report No. 566.) Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Webb, N.L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education.* (Research Monograph No. 6). Madison: University of Wisconsin-Madison, National Institute for Science Education.

Webb, N.L. (2002, December). *Alignment study in language arts, mathematics, science, and social studies of state standards and assessments for four states.* Washington, DC: Council of Chief State School Officers.

LEGAL CORNER: CALIFORNIA EL CASE

S.E. Phillips, Consultant

The California High School Exit Examination (CAHSEE) was challenged in a California court by a group of English Learners (ELs) just prior to their scheduled graduation last spring. In *Valenzuela v. O'Connell*, the ELs sought class status and a preliminary injunction barring implementation of the graduation test requirement for students in the Class of 2006.

Description of the CAHSEE

The CAHSEE consists of two untimed sections, English language arts (ELA) and mathematics, administered initially in tenth grade. The ELA section consists of 72 multiple-choice items and one essay writing prompt. It measures skills through tenth grade including word analysis, reading comprehension, literary responses & analysis, and writing strategies, conventions & applications. Reading and writing are weighted approximately equally and the passing standard was set at 60% correct on the initial form. The math section consists of 80 multiple-choice items covering number sense, algebra & functions, measurement & geometry, probability & statistics, and mathematical reasoning through seventh grade plus Algebra I with a passing standard of 55% correct on the initial form.

Passing Rates

For the initial administration of the CAHSEE to the Class of 2006 in tenth grade, 69% of students passed both sections. For subsequent administrations, cumulative passing rates for both CAHSEE sections by subgroup were as follows:

CLASS OF 2006	PASSING RATES		
	SEPT 2005	JAN 2006	JULY 2006
AFR-AMER	63%	80%	83%
ECON DISADV	66%	82%	86%
ELS	51%	69%	77%
HISPANIC	68%	82%	85%
TOTAL	**78%**	**89%**	**91%**

Plaintiffs' Claims

The *Valenzuela* plaintiffs argued that their equal protection rights had been violated because they had not received an adequate opportunity to learn the tested material due to attending schools lacking fully aligned curricula and fully credentialed teachers. They asserted that this lack of opportunity to learn was a denial of their (state) fundamental right of equal access to public school education. In addition, they alleged that ELs had been disproportionately affected by the scarcity of resources in poor districts. The *Valenzuela* plaintiffs further argued that the state had failed to adequately research "appropriate other criteria by which high school students regarded as *highly proficient* but unable to pass CAHSEE could demonstrate their competency and receive a high school diploma" as required by statute. In December of 2005, the Superintendent of Public Instruction (SPI) had announced that a study of the issue identified "no practical alternatives."

Background

In March 1999, the California legislature passed a graduation test statute effective for the Class of 2004. Annual notice of the CAHSEE requirement to students and parents began in Fall 2000 and the statute also required school districts to provide remediation to nonpassing students. The prior spring, the *Williams* lawsuit had been filed charging that disadvantaged schools lacked equal educational opportunity due to inadequate state funding which resulted in outdated instructional materials, less qualified teachers, and facilities in disrepair in their schools. In the summer of 2003, the State Board of Education (SBE) voted to postpone the graduation test requirement to the Class of 2006. The *Williams* litigation was settled in August 2004 when the state agreed to provide additional funding to improve instructional materials, teacher quality, and facilities in the poorest districts but timing prevented this settlement from being fully implemented for students in the Class of 2006. In the fall of 2005, the legislature appropriated $20 million in supplemental funds to be distributed to districts with schools having the highest proportions of seniors who had not yet passed the CAHSEE. An additional supplemental funding allocation of approximately $85 million was provided for the Class of 2007.

Procedural History

The *Valenzuela* lawsuit was filed February 2006. On May 12, 2006 the trial court granted the requested preliminary injunction barring the state from imposing the testing requirement on *any student* in the Class of 2006. On May 24, 2006, the California Supreme Court stayed the injunction pending review and decision by the Appeals Court. Plaintiffs' request to the Appeals Court for an immediate hearing was denied and the graduation test requirement remained in place for the Class of 2006.

Appeals Court Decision

Oral argument was held on July 25, 2006 and the Appeals Court issued its decision on August 11, 2006 vacating the preliminary injunction issued by the lower court. In reviewing the trial court's action, the Appeals Court considered the Plaintiffs' likelihood of success on the merits, the relative interim harms to the parties, and the appropriateness of the remedy (ordering the state to suspend passage of the CAHSEE as a graduation requirement for *all students* in the Class of 2006).

With respect to success on the merits, the Appeals Court did not review the factual findings of the trial court but noted that neither the graduation test requirement nor the validity of the CAHSEE was being challenged. Further, assuming Plaintiffs' equal protection claim was legally proper and Plaintiffs' asserted facts were correct, the Appeals Court held that the remedy fashioned by the trial court was improper. In addition, disagreeing with the trial court's findings, the Appeals Court stated that the legislative allocation formula for supplementary funding, which gave priority to schools with the greatest need, was reasonable and properly within legislative discretion.

Relative Harms

On the issue of relative harm to the parties, the Appeals Court held that the trial court had overemphasized the harm to Plaintiffs while failing to adequately consider public interest. The Appeals Court acknowledged that for some plaintiffs, failing CAHSEE delayed but did not permanently deny them a diploma, and noted that students had nine options for remediation including summer school with July retesting, tutoring, online education, remaining in school for an extra year, adult education, passing the GED test, and community college (high school diploma not required for attendance). The Court also noted that the trial court's acceptance of Plaintiffs' assertions of harm was not supported by any specific factual findings and that the trial court failed to balance the alleged harm to Plaintiffs against the substantial harm to others and to the public interest from the state's inability under the injunction to enforce a statutory scheme designed to raise academic standards in public schools.

Appropriate Remedy

Because the Appeals Court found that the potential injury to Plaintiffs was the loss of educational opportunity to learn the tested skills and not the denial of the diploma, it held that the appropriate remedy was provision of the missed instruction, not removal of the test requirement or the award of diplomas by court order. The Court stated:

> Within the borders of California, until our schools can achieve the academic parity envisioned by the *Williams* litigation and settlement, the CAHSEE also provides students who attend economically disadvantaged schools, but who pass the exit exam, with the ability to proclaim empirically that they possess the same academic proficiency as students from higher performing and economically more advantaged schools. Granting diplomas to students who have not proven this proficiency debases the value of the diplomas earned by the overwhelming majority of disadvantaged students who have passed the exit exam. . . . We believe the trial court's [order] erred by focusing its remedy on equal access to *diplomas* rather than on equal access to *education* (and the funding necessary to provide it). . . . The purpose of education is not to endow students with diplomas, but to equip them with the substantive knowledge and skills they need to succeed in life.

The Appeals Court also found that the scope of the remedy (removal of the test requirement for *all students*) was overbroad because it provided a potential windfall to plaintiffs who could not trace their test failure to inadequate school resources or lack of access to the supplemental funding. The Court stated:

> [T]he ostensibly interim relief of forcing the "social promotion" of plaintiffs, by ordering that they be given diplomas, in fact does not maintain the status quo of the litigation, but ends it. Surely the trial court did not expect that if defendants ultimately prevailed in the litigation, plaintiffs would give back the diplomas they had received under the mandate of the court's preliminary injunction. . . . [D]irecting defendants to give plaintiffs diplomas . . . would inadvertently have perpetuated a bitter hoax: that the [court-ordered diplomas] somehow would have equipped them to compete successfully in life, even though they had not actually acquired the basic academic skills measured by the CAHSEE. . . . Plaintiffs virtually concede the overbreadth of the trial court's injunction in their argument that some [class members] "actually know the material, but do not pass the exit exam due to test anxiety." But plaintiffs have not argued, much less established, that there is any *constitutional* violation involved in depriving a student of a diploma when he or she has in fact received the educational resources required to pass the CAHSEE, but has not been able to do so because of "*test anxiety.*"

Conclusion

In addition to vacating the injunction, the Court's opinion concluded by urging the parties, with the active assistance of the trial court, to work together to provide all seniors in the Class of 2007 and beyond who have not passed the graduation test an equal access to effective remedial assistance.

EVENTS OF REGIONAL INTEREST

Washington Educational Research Association (WERA) Spring Conference

The annual Washington Educational Research Association (WERA) Spring Conference will be held at the Seattle Airport Hilton Hotel on March 28 – 30, 2007. The conference theme is "Making the Grade—Gathering and Reporting Evidence of Student Learning." Keynote Speakers will be Rick Stiggins and Ken O'Connor. Over 35 breakout sessions will be available Thursday and Friday, with half-day training sessions on Wednesday. Additional details will be posted on the WERA Web site as they become available. Find WERA at www.wera-web.org.

Southwest Educational Research Association Meeting

The Southwest Educational Research Association will meet in San Antonio February 7 – 10, 2007. Training sessions include a one-day session on Mixed Methods, a half-day session on Effect Sizes, and shorter sessions on (a) using Excel to teach/understand IRT, (b) using propensity scores to create pseudo-experimental designs, and (c) how to publish from the perspective of four Editors. More information is available at: http://www.sera-edresearch.org

Scott Bishop
Data Recognition Corporation
13490 Bass Lake Road
Maple Grove, MN 55311

Phone: 763.268.2029
Fax: 763.268.2529
e-mail: SBishop@DataRecognitionCorp.com