



FROM THE PRESIDENT

Anne R. Fitzpatrick, Educational Testing Service

Greetings, Everyone!

I hope that you all had a wonderful Thanksgiving.

I want to update you on the activities being carried out in preparation for the 2008 annual meeting to be held in New York City in late March. Also I have news and notes to share with you about other matters of interest to NCME members.



Activities Related to the 2008 Annual Meeting in New York City

Graduate Student Issues Poster Session

The Graduate Student Issues Committee (GSIC) received 96 proposals for poster sessions this year! The committee thanks all students who took time out of their busy schedules to prepare proposals. I have heard that many of the proposals were very interesting and very well done. All proposals have now been evaluated by outside reviewers, and the committee is now making its acceptance decisions. I know that the committee will accept as many as it possibly can; it is well aware that the poster sessions are exciting and rewarding for presenters. My only caution to the committee is that it must avoid filling a room so full of people and posters that we attract the attention of New York City's fire chief...(!)

Activities of the Annual Meeting Program Committee

Tasha Beretvas (tasha.beretvas@mail.utexas.edu) and Karen Barton (karen_barton@ctb.com), co-chairs of the Program Committee, received many, many fine proposals for papers and coordinated sessions. They have told me that making the selection decisions was very, very difficult because there were so many fine proposals. To make their decisions, Karen and Tasha considered not only reviewers' ratings but also the content of the proposed research, because they wanted to make sure that the annual meeting program included research on a broad array of topics. Therefore some excellent proposals could not be accepted simply because their ratings were slightly lower than those of other proposals addressing the same topic. I encourage those of you with proposals that were not accepted to submit them again next year. Trends in the topics addressed by proposals change from year to year, as do the preferences of the program co-chairs.

A list of the invited sessions planned by Karen and Tasha for the 2008 meeting is provided elsewhere in this newsletter.

As a reminder to presenters, chairs/moderators, and discussants, I want to mention again the possibility that NCME will not be able to reconcile its schedule of your activities with the annual meeting schedule that AERA establishes. Therefore it is very possible that when both meeting schedules are published, some of you will discover that you have to be in two places at once!

The best advice we can give to those of you who have papers accepted by NCME and AERA or other involvements in both annual meetings is to identify a back-up person who can handle one of your obligations if by chance you are double-booked.

We are trying our best to identify those of you who are vulnerable to double-booking! Linda Cook, Vice President of AERA's Division D, and her program co-chairs graciously responded to our request for information by providing a list of all first authors of papers and symposia accepted by their division as well as chairs and discussants they have selected; this list will allow us to identify the individuals who will appear on both annual meeting schedules. Although AERA has not finished its scheduling, we will make sure that these particular individuals are prepared in case a schedule conflict arises.

As a reminder to presenters, I want to note again that liquid crystal devices (LCD) will be used instead of overhead projectors at all sessions. Presenters will need a personal computer to link to the LCD projectors. We will be asking the session chairs to

make sure that at least one computer is available for the session participants to use. If a presenter prefers to use an overhead, a special arrangement will have to be made for this with NCME's Central Office.

Activities of the Training and Development Committee

Alina von Davier (AVonDavier@ets.org), chair of the Training and Development Committee, along with Jodi Casabianca and Jazzme Blackwell, who has been assisting Alina, have scheduled 17 (!) pre-conference training sessions. A list of the sessions is provided in this newsletter. As you will see, the sessions cover a wide range of topics that we hope will be of interest to many NCME members. I want to thank Alina, Jodi, and Jazzme for working so hard to put together such a fine training program. We all thank the session presenters for offering such wonderful learning opportunities to conference attendees.

How to Register for the Training Sessions and the Annual Meeting

NCME's website (www.NCME.org) will contain information about registering for the training sessions and the Annual Meeting; we will send you an email as soon as the website registration system is live. On-line registration for the training sessions is encouraged, since the sessions will fill up fast. A brochure describing this information also will be mailed to you very soon.

How to Make Hotel Reservations

The NCME website also will tell you how to make hotel reservations. The reservation service is managed by AERA's Housing Bureau. The Crowne Plaza Times Square Hotel in New York City will be the site for NCME's headquarters and also the site for all NCME training sessions and most NCME paper sessions and symposia.

News and Notes

Website

A new and improved NCME website has been launched! The website has a new look and improved navigational tools that makes the site very user-friendly. It also has a members-only section that will permit you to renew your memberships online. As I noted in the last newsletter, when you renew, you will be asked to provide some demographic information that will help us to know more about the characteristics of our membership. Also we will ask for suggestions from you about possible nominees for awards and elected offices, and we will request your ideas about how NCME can improve things. Finally, we will continue to seek your help on NCME committees.

We also have a newly created Website Management Committee. George Engelhard will chair this committee, which will manage the look and content of the website. Its responsibilities will include proposing materials for the website, organizing the website, reviewing submitted materials, updating existing materials, and considering new additions to the website. If you have suggestions about ways to improve the website, please contact George Engelhard (gengelh@emory.edu).

Publications

As I noted in the previous newsletter, the three-year term for Scott Bishop, current editor of this fabulous newsletter, will end in December, 2008. We are soliciting nominations for Scott's replacement; we would like to have the new editor appointed by June, 2008. Please send nominations that you have to Steve Sireci (sireci@acad.umass.edu); Steve is the NCME Board member responsible for the Publications Committee.

Meeting of the NCME Board of Directors

The Board of Directors held its fall board meeting on November 1 and 2, 2007, in Washington, D.C. As usual, it was a very productive meeting. Perhaps one of the most interesting discussions focused on the ways in which NCME might become more involved with and influential in legislative policy-making at both the federal and state level. One of the goals of NCME's new strategic plan is for the organization to take an active role in the public policy arena. At the Board's invitation, Jerry Sroufe, AERA's Director of Governmental Relations, kindly agreed to give a presentation describing possible approaches. Jerry was very informative, and we learned a great deal. Details about other matters considered by the Board will be provided in the minutes for the meeting, which will be posted on NCME's website after they are approved by the Board.

Our next Board meetings will occur at 4 PM on Monday, March 24, 2008 and Thursday, March 27, 2008, at the Crowne Plaza Times Square Hotel in New York City. As usual, NCME members are always welcome to attend.

I send my best wishes for a joyful holiday season to all of you.

CURRENT ACTIVITIES OF THE STANDARDS MANAGEMENT COMMITTEE

David Frisbie, NCME Representative to the Committee

A website was established by the Standards Management Committee to gather comments about aspects of the 1999 *Test Standards* that are in need of revision. The purpose of this website was to provide the committee with an easy and efficient mechanism for obtaining information about the scope of the revisions that would be needed. The website went live in early June and stayed open through mid-November.

Comments from the website were sent to the Standards Management Committee for analysis. Synopses of the comments will be integrated and shared in early 2008 with the sponsoring organizations, which are NCME, AERA, and APA. Based on its analysis, the Standards Management Committee, in consultation with the co-chairs of the Joint Committee, Barbara Plake and Laurie Wise, will identify the areas of expertise that need to be represented by members of the Joint Committee, which will make the revisions. The members then will be selected. It is expected that the Joint Committee membership will be finalized in the spring/summer of 2008 with the first full committee meeting occurring in the fall, 2008.

ADVISORY BOARD MEETS TO PLAN NEW GRADUATE PROGRAM AT MORGAN STATE UNIVERSITY

Kurt Geisinger, NCME Representative to the Advisory Board

The Advisory Board for the doctoral program in Psychometrics met for the first time on June 5, 2007, on the campus of Morgan State University in Baltimore, Maryland. The Advisory Board has permanent members and rotating members. The permanent members are Dr. Maurice Taylor, Dean of the Graduate School of Morgan State University; Dr. Pamela Scott-Johnson, Chair of the Department of Psychology at Morgan State University; Dr. Ida Lawrence, Senior Vice President for Research and Development at ETS; and Dr. Michael Nettles, Senior Vice President for the Policy Evaluation & Research Center at ETS. The rotating members are Dr. Paul Gammill, Baltimore City Public Schools; Dr. Kurt Geisinger, Director, Buros Center for Testing & the University of Nebraska-Lincoln, representing the National Council on Measurement in Education; Dr. Lisa Harlow, University of Rhode Island, representing the American Psychological Association, Division 5; and Dr. Hariharan Swaminathan, University of Connecticut, representing the American Educational Research Association, Division D. Dr. Scott-Johnson chairs the Advisory Board.

At the meeting the Board discussed that the primary goal of the Psychometrics program will be to develop scholars who possess sophisticated statistical and analytical capabilities and the quantitative and methodological skills (e.g., measurement theory, statistical analysis, research design, evaluation, and qualitative tools) needed to design, develop, interpret and use valid, reliable and fair measurements and assessments of what and how individuals learn. Graduates from the program will have the analytical skills and cultural competence to effectively yield innovative interventions that address issues within the field itself as well as inform policies that influence minority or special populations (e.g. African Americans) and those within urban environments in the United States. An important outcome of the proposed program will be the preparation of individuals from populations under-represented in the educational system to take leadership roles, which potentially offers a needed and unique view to addressing how the academic achievement gap can be closed.

Also discussed was the purpose of the Advisory Board, the role of ETS in the support of and development of the program, and the fact that there would be both masters and doctoral programs offered to students. The curriculum for each program was discussed, and the status of the current search for a program director and for students was briefly described. The possibility of a certificate program was also considered. Finally, plans for future meetings were made. The second meeting of the Advisory Board will be a conference call meeting on December 12, 2007.

Questions about the new graduate program or the application process can be addressed to Dr. Pamela E. Scott-Johnson, Chair of the Psychology Department (443-885-3508) or Dr. Stephen L. Koffler, (732-329-6518), who is coordinating the new program for Morgan State University.

Three Directions for Improving Assessments

Lynn Olson, Education Week

Much of the newspaper coverage of testing these days is negative: How many schools have failed to make adequately yearly progress? How much have state tests narrowed the curriculum to focus on reading and math at the expense of other subjects? Are teachers teaching too much to the test?

But while testing, these days, is often portrayed as a hammer to punish students and schools, it doesn't have to be that way. I'd like to suggest three issues that, if addressed, could improve the tone of the debate: quality, transparency, and accountability.

Paying Attention to Quality

By quality, I don't mean the technical quality of commercially produced tests as defined by reliability studies, bias studies, and equating studies. Rather, I mean, are today's tests measuring what matters most?

For example, the ACT has released a number of reports about the gap between students' high school preparation and their readiness to do college-level work. Yet many state tests require students to write content-free five paragraph essays, when college freshmen are expected to write carefully reasoned and referenced research papers.

Similarly, while there's a substantial body of research suggesting that high-quality formative assessments can dramatically improve student achievement, by providing students with feedback and helping teachers to modify instruction, most attention today centers on high-stakes, large-scale, annual assessments that have limited utility for improving teaching and learning in classrooms. Moreover, much of what passes for "formative assessments" are actually mini-versions of multiple-choice, end-of-the-year tests. They provide teachers with limited evidence for improving instruction.

And we all know that when states self-report that their tests are "aligned" with their content standards, it doesn't mean nearly as much as we'd like it to mean.

The way out of this is two-fold: We need to focus on much richer, curriculum-based assessments that ask students to do ambitious work, including revisiting the idea of portfolios and performance assessments. And we need to help teachers use more fine-tuned, ongoing assessments to deduce whether students are learning from moment to moment and, if not, what to do next.

Greater Transparency

Second, if tests are going to be used to improve instruction they can't be a black box. They need to be much more transparent. Teachers, students, and parents have to understand what the concepts are that the test is meant to measure. They should have access to lots of test questions and answers—including examples of adequate and exemplary work—to understand what the target is. And they should understand when the difference between scores comes down to the difference between answering only one or two questions correctly.

Unfortunately, all of the pressures today—for more tests, delivered cheaper, and scored more quickly—push in the opposite direction of such reforms. And the bad news for the testing industry is that repeated newspaper coverage of the almost inevitable errors that follow on the heels of such pressures means that assessment itself becomes suspect.

I'd suggest we need better accountability on two fronts. First, we need more balanced accountability systems that rely upon tests as an important measure, but not the only measure, of whether schools and students are performing adequately. I'm not necessarily in favor of New York City's new A-to-F report cards, but I think the city's efforts to introduce multiple measures of performance—including school quality reviews; teacher, parent, and student surveys; and the use of growth as well as status measures—is a step in the right direction. The same is true of the "balanced score cards" that some districts are now using.

Similarly, I think we could learn a lot from European countries that ask students to complete curriculum-based tests in a choice of subjects to demonstrate their readiness for college or technical training. Such tests serve more as a gateway to a goal that students care about than as a gatekeeper, which is the case with many of today's low-level, high school exit exams.

Eva Baker at CRESST, for example, has broached the idea of students earning the equivalent of "merit badges" to demonstrate their knowledge and skills in areas they care about, rather than passing a single, uniform test.

Responsible Test Use

Finally, I think the testing industry needs to step up to the plate in terms of the responsible development and use of tests. While the joint standards are quite explicit on such topics, there's essentially no enforcement mechanism. And over the years I've talked to a number of test developers who've essentially said a state or a district is not using our test the way it was intended to be used, but it's out of our hands. In fact, I can't think of an instance where a testing company has withdrawn from a contract because it felt that its assessment was being used inappropriately, in a way that violated the joint standards.

I don't think we're going to get away from performance-based accountability. That's the name of the game. But I do think we need a better balance in the measures we use and how responsibly we use them.

Can—or should—the federal government drive such efforts? Earlier this year I attended a meeting in Chicago where the general consensus was that real improvements in standards and assessments in the United States are going to come from the states, working in collaboration, and not from the federal government. In fact, the prospect of an increased federal role in this area seemed to send shudders around the room.

The federal government could provide incentives, however, in the form of competitive grants and regulatory flexibility—to encourage states and testing companies to pursue such work. Similarly, private philanthropies could sponsor substantial, national competitions to develop a new generation of standards and assessments that would be far better than what we have now and that could take advantage of burgeoning technology platforms.

That, at least, would give reporters something new and exciting to write about—instead of this year's school rankings.

The views expressed in this article are solely those of the author and not of *Education Week*.

GROWTH MODELS UNDER NCLB: BACK TO BASICS

Andrew Ho, *University of Iowa*

The term “growth model” has hit the big time. Its newsworthiness spiked in November 2005, when Secretary of Education Margaret Spellings invited states to submit “Growth Model Pilot Proposals” that would incorporate individual growth measurement into school accountability calculations. Twenty states applied, and nine have had their growth models approved. Approved growth models are having an impact on current accountability decisions in these states. Though reauthorization proposals for the No Child Left Behind (NCLB) Act have been pushed back to 2008, growth models are already a prominent feature of the House committee's draft legislation. The term is out of the psychometric cloister and into the mainstream.



Given the dramatic rise in stakes on interpretations and applications of “growth models,” this short paper attempts to distinguish between specialized uses of the term and the increasing number of variants that are now being incorporated under the “growth model” umbrella. There is an important distinction between a *statistical model* for growth, one that predicts and tries to explain growth, and a *policy model* for growth, one that builds standards, incentives, and sanctions around growth results. I suggest that discussions about growth models are largely proceeding within and not between these two spheres. Statisticians and psychometricians continue to develop sophisticated models for predicting growth, but the models of the Growth Model Pilot Program are simple and, largely by design, use little advanced statistical machinery. Meanwhile, an important debate is occurring around policy models for growth, but very little attention is being paid to fundamental psychometric concerns underlying these simple statistical models. In conclusion, I recommend that a few basic psychometric principles should be revisited in the midst of our headlong rush towards widespread application of growth models to high-stakes educational policy.

This paper draws upon a large body of previous work, much of it conveniently listed under a Council of Chief State School Officers' website, here: http://www.ccsso.org/projects/Accountability_Systems/Resources/#growth. I am also grateful to Jennifer Dunn and Scott Marion, who wrote about the then current status of growth models in last December's issue of the NCME newsletter (Dunn & Marion, 2006). This piece raises some complementary and contrasting points.

Statistical Models Versus Policy Models

When it comes to growth models, the approaches of statisticians and policy makers are so different that they may be usefully described as motivating two separate models altogether. For statisticians, growth models can predict the future growth of

students based on a number of possible student-, teacher-, and school-level predictors. These models help to support inferences about the correlates of individual student growth and, sometimes inappropriately, inferences about the causes of growth.¹ For psychometricians, defining the vertical scale on which growth is measured is a primary concern, as the final choice represents a judgment about which students grow faster and how variability in student achievement changes over time.

In contrast, policy decisions around growth models begin with the results of growth measurement—change scores, predicted change scores, residual gain scores, or the like – and proceed to a system of standards, sanctions, and rewards around these results. Policy models for growth ask questions like, how much growth is enough growth, and, how should school-level growth scores be incentivized? The statistical models underlying these policy models are simple and lack the large number of covariates generally associated with statistical growth models. This is fully intended, as reflected by the second core principle guiding growth model proposals: The policy model must not set expectations for annual achievement based upon student demographic characteristics or school characteristics (U.S. Department of Education, 2006). This requirement, and the unspoken requirement that policy calculations should at least seem straightforward, leads to very simple statistical approaches.

The distinction between a statistical model and a policy model helps to explain why conversations about NCLB growth models may overlook fundamental psychometric problems: Statisticians and policy makers may both assume extension of simple statistical models to these compelling policy problems without revisiting the validity of statistical assumptions in this new context. Value tables, projection towards proficiency, adequate yearly growth, and grade level growth are all policy terms requiring policy judgments, but the simple statistical models underlying them are also judgmental even though they do not come close to approaching the complexity of state-of-the-art statistical growth models.

To transition, I offer two reminders to measurement specialists. First, NCLB growth models are best thought of, not as new statistical growth models, but as new policy models layered on some very basic statistical growth models. Second, the differences between these basic statistical growth models are consequential, sometimes more consequential than differences between policy models that are currently receiving most of the attention. I do not mean to downplay the seriousness of practical policy concerns. Among the biggest obstacles to responsible growth model implementation is the many students missing part of their growth trajectories. An accountability policy must deal with these students in a thoughtful way to avoid incentivizing missing data or disproportionately affecting schools with mobile populations. However, the remainder of this paper will focus on the most fundamental of psychometric problems that, I believe, must not be overlooked as the stakes on growth models rise.

The Basics: Three Old Psychometric Problems to Revisit for NCLB Growth Models

Gain Scores Versus Residual Gains

The much maligned individual gain score has been enjoying a resurgence for some time, as its unreliability has often been shown to be exaggerated or beside the point. Edward Haertel's (2006) *Reliability* chapter in the 4th edition of *Educational Measurement* provides an overview of the longstanding work of David Rogosa and others on the subject (e.g., Rogosa, 1995). When scores over time can be located on a common scale, models for gains are usually intuitive, statistically defensible, and relevant. Where the use of residual gain scores and projections has been motivated solely by the purported unreliability of gain scores, the selection of residual gain models should be reevaluated. Further contrasts between the implications of these two approaches are warranted, especially given the simplicity of the underlying statistical models and the high stakes on their results.

Scaling and Vertical Scaling

Though a vertical scale is often associated with growth models, and the modeling of growth seems to presume some vertical scale on which average growth is meaningful and nonzero, some growth models navigate around this requirement by using covariance methods (e.g., Wright, Sanders, & Rivers, 2006) or categorical procedures like value tables (Hill, Gong, Marion, & De Pascale, 2006). However, scaling decisions are central to growth models regardless of whether a vertical scale is used. Correlations are still dependent on nonlinear scaling decisions, and projections using these methods rely even further on an implicit scale advocacy. Categorical methods that use weights on transitions between categories are akin to scaling decisions, where large weights on transitions are analogous to a stretched scale in a particular latent score region. Extant scales, transition weights, and vertical scales may be usefully described as a selected member of a family of plausible nonlinear scale transformations, and the implications of other members of these families frame a necessary sensitivity study.

Full vertical scales, like value tables, are upfront in the sense that the final selection of scale is explicit, but plausible nonlinear transformations are rarely defined or considered. The consequences of alternative scaling decisions are crucial to growth models, as the potential vagaries within a scale at one time point compound across scaling decisions at other time points. Additionally, a choice of scale presumes that a defensible scale exists through the content domain from which the items are

¹ Yeow Meng Thum (2005) helpfully reminds us that, without experimental designs, there can be no Value-Added Models, just Value-Added Hypotheses.

sampled (Kolen & Brennan, 2004). This vertically structured content domain is required for defensible interpretation and use of growth models regardless of whether a growth model claims to use a vertical scale per se. The selection and defense of scale, these most fundamental of psychometric problems, are hidden yet consequential issues in all NCLB growth models.

Standard Setting

Standard setting is at the intersection of the policy model and the statistical model for growth, as the rendering of a judgment (pass/fail, proficient/basic, adequate/inadequate) can both be informed by psychometric results and, once the decision is rendered, define psychometric features of the classification itself. However, the cut score that defines students as proficient is more thoughtfully selected than the implicit cut score defining adequate or acceptable growth. Given the controversy over cross-state variability of proficiency cut score decisions (Braun & Qian, 2006), standards for growth should be subjected to at least as much scrutiny. For projection-based models, which dominate the current NCLB landscape, the cut score for growth is defined by some function of the following: the cut score for proficiency, one or all of the examinee's previous scores, a linear projection on some scale, and the number of years required to reach proficiency, which is usually less than or equal to four. The concern is that this problem has not been framed as the act of setting a cut score for growth. Instead it relies on a large number of other judgmental decisions: the choice of cut score for proficiency, the choice of scale, the number of years to proficiency, the 2014 deadline for NCLB, and the like. Thus, the definition of sufficient growth is operationally rather than judgmentally defined, and indirectly at that. An understanding of growth should motivate the standard for growth in the way that an understanding of proficiency should motivate the cut score for proficiency.

Conclusions: Communicating the Dependencies of High-Stakes Results on Judgments

I have argued that the implementation of NCLB growth models represents a new direction for policy but generally not for psychometrics. Indeed, as psychometricians, NCLB growth models force us to return to some of our most basic and longstanding research topics. It would be responsible of us to find new ways to communicate the impact of judgments to policy makers, who tend as a matter of their job requirements to be more interested in results from a single decision rather than in the variability of results under a family of reasonable decisions. As I have summarized, growth models under NCLB demand a large number of judgments, most of them implicit, overlooked, and lurking. There are already enough elephants in the room. They have been there a long while. It is time to introduce them around.

References

- Braun, H.I., & Qian, J. (2007). An enhanced method for mapping state standards onto the NAEP scale. In N.J. Dorans, M. Pommerich, & P.W. Holland (Eds.), *Linking and Aligning Scores and Scales* (pp. 313-338). New York: Springer.
- Dunn, J.L., & Marion, S.F. (2006, December). NCLB growth: What are we learning as reauthorization approaches? *NCME Newsletter*, 14(4), 4-6.
- Haertel, E.H. (2006). Reliability. In R.L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 65-110). Westport, CT: American Council on Education and Praeger Publishers.
- Kolen, M.J., & Brennan, R.L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices*, (2nd ed.). New York: Springer-Verlag.
- Rogosa, D. R. (1995). Myths and methods: "Myths about longitudinal research," plus supplemental questions. In J. M. Gottman, Ed., *The Analysis of Change* (pp. 3-66). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Thum, Y.M. (2005, November). *Bayesian Inference for Some Value-Added Productivity Indicators*. Paper presented at the Conference on Longitudinal Modeling of Student Achievement, University of Maryland, College Park, MD.
- 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>
- Wright, S.P., Sanders, W.L., & Rivers, J.C. (2006). Measurement of academic growth of individual students toward variable and meaningful academic standards. In Robert W. Lissitz (Ed.), *Longitudinal and Value Added Models of Student Performance* (pp. 385-399). Maple Grove, MN: JAM Press.

THE ROLE OF COGNITION IN EDUCATIONAL MEASUREMENT RESEARCH AND PRACTICE

Joanna S. Gorin, Arizona State University

In the current age of accountability in education, test score uses are almost as varied as the constructs and content measured by the tests. By and large, however, inferences about student cognition are of some interest. Achievement tests assess students' knowledge in particular content areas and students' level of skill in applying that knowledge. Aptitude tests measure students' ability to process and manipulate information. Even behavioral rating scales commonly used in classroom assessments are used to make inferences regarding students' ability to regulate their attentional and physical systems.



Regardless of the testing purpose, any test is a more effective tool to the extent that it is an accurate and complete representation of the student cognition of interest. Yet historically, surprisingly few educational testing programs examine score meaning in cognitive terms. Even fewer have employed cognitive methods in developing tests and establishing score properties.

A review of recent measurement literature suggests that a change is afoot. An increasing number of educational measurement researchers and practitioners advocate that cognition is a critical component of assessment design and analysis. The presumed benefits include increased evidence of construct validity, improved score interpretations, and increased efficiency in test design with automated procedures. Their general rationale is that when one begins the test development process with well defined, targeted inferences about student cognition, then item writing and scoring, as well as ability estimation and reporting are better connected – with inferences about student cognition as the common thread linking all aspects. Researchers have paid specific attention to the potential role of cognition at two stages of test development: construct definition and score reporting.

The Role of Cognition in Construct Definitions

The use of cognitive psychology in assessment and psychometrics has added a new dimension to construct definitions. Detailed construct definitions including lists of individual cognitive processes and relationships among processes can provide a strong foundation for test development and score interpretation. *Cognitive models*, a representation commonly used in experimental, developmental, social, and cognitive psychology, can arguably provide this level of detail. Cognitive models for assessment specify learners' representations of a domain in terms of requisite *knowledge*, *skills*, and *abilities* (KSAs). The models are developed similarly to model development and testing in social sciences. In the beginning, a review of the literature on associated skills and processes in a domain can provide an initial hypothesis. Next, experimental and observational data can be collected to evaluate and suggest revisions to the model. Once developed, a cognitive model delineates those processes that are construct-relevant and their relationships with one another (see NRC, 2001, Gorin, 2006 or Leighton & Gierl, 2007, for further discussion of the properties of cognitive models). The initial specification of construct-relevant versus irrelevant sources of variance provided by cognitive models becomes the basis for later examinations of test score meaning. Specifically, the validity of score interpretations can be judged relative to the structure and components of the cognitive model provided at the outset of the test development process.

One might wonder why I give such weight to cognitive models as advantageous construct definitions. Clearly, the items developed for a test are of greater interest than the cognitive model in and of itself, and I would not argue that it should be otherwise. But attention to the theory underlying assessment items as they are generated improves the quality of the items in the longer term. To produce the most informative and useful test scores, tasks should engage individuals in processes relevant to and representative of the construct. At first, the cognitive model provides a point of departure. Next, test developers must design test items that provide observable behaviors linked to the construct-relevant cognitive processes. The cognitive model is a testable hypothesis of student processing and construct meaning, and the test item is the tool by which data is gathered to test the hypothesis. Principled test design frameworks such as Evidence Centered Design (Mislevy, 1995) and the Cognitive Design System (Embretson, 1994) describe item development procedures that centralize the role of cognition. Applications of these procedures in computer networking, curriculum based science education, and spatial and abstract reasoning tests have illustrated many of the benefits of cognitively-based assessment design for improved efficiency in test development, stronger validity arguments, and availability of diagnostic score information (see Gorin, 2007, for a review).

The Role of Cognition in Psychometric Modeling and Score Reporting

Although trait estimates generated from traditional item scoring systems and psychometric estimation (e.g., unidimensional theta estimates) are well suited for normative and standards-based score interpretations, they are too broad to be of much use for cognitively-based score interpretations such as diagnosis; they “miss” the available cognitive information that might be available in a test. Psychometric models that produce diagnostic information regarding student skill strengths and weaknesses can augment traditional test scores. These models, sometimes called cognitive-diagnostic models (CDMs), are essentially data analysis techniques designed to link cognitive theory with items' psychometric properties (see Rupp, 2007 for a taxonomy of CDMs). The skills needed to correctly solve problems are specified, and their impact on responses is estimated. Information about skill mastery can be estimated and reported by examining the skills required for each item and the student's response. Incorrect responses suggest that an individual has not mastered the skills required by an item; correct answers suggest that the skills have been acquired. Based on student response patterns and information about the skills required by each item, a skill profile can be estimated for each student or groups of students. CDMs show potential for large scale diagnostic score reporting that is aligned with the mandate of NCLB legislation. If these models can be applied to appropriately designed tests, educational assessment could become a more integral part of instruction and the educational process.

Conclusion

In summarizing the roles of cognition in educational assessment, I have discussed only two points in test development—construct definition and score reporting. In selecting both early and later steps in the process, however, I hope to emphasize the importance of cognition *throughout* the measurement process. Increasing the role of cognition in test development through cognitive theory and methods requires additional time, expertise, and resources. The payoff can be a wealth of student-level information that can be “mined” for score reporting and interpretation *if* the appropriate analytic tool is applied to the test data for ability estimation. A cognitively-based test is useless without appropriate statistical methods. Similarly, CDMs can only generate cognitive information that is available in student responses to test items. Hence, research must be generated to support cognitively-based assessment at all stages in test design. The existing cognitively based assessment research consists predominantly of mathematical, theoretical, and simulation studies of CDMs—and important piece of the puzzle. However, far fewer studies have addressed the development of cognitive-models of construct definitions or cognitively-based item design. Consequently, practitioners interested in incorporating cognition into the assessment process have limited examples of specific procedures to guide them. I would encourage researchers interested either cognitively-based test development or psychometric modeling to work more collaboratively on efforts with a dual focus on issues that inform all steps in the assessment process. It is likely that the most informative research for either cognitive-model based item development or CDM parameterization will consider both issues simultaneously.

References

- Embretson, S. E. (1994). Application of cognitive design systems to test development. In C. R. Reynolds (Ed.), *Cognitive assessment: A multidisciplinary perspective* (pp. 107–135). New York: Plenum Press.
- Gorin, J. S. (2007). Test Construction and Diagnostic Testing. In J. P. Leighton & M. J. Gierl, Eds. *Cognitive diagnostic assessment in education: Theory and practice* (pp. 173-201). Cambridge University Press.
- Gorin, J. S. (2006). Item Design with cognition in mind. *Educational Measurement: Issues and Practice*, 25(4), 21-35.
- Leighton, J.P. & Gierl, M.J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, 26(2), 3-16.
- Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59, 439–483.
- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Rupp, A. (2007). Unique Characteristics of Cognitive Diagnosis Models. Paper presented at the 2007 Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

IN MEMORIAM

Donald Ross Green (1924-2007)

GETTING READY FOR THE NCME ANNUAL MEETING: SOME TIPS FOR NEW PRESENTERS

Susan L. Davis, & Chad W. Buckendahl, *Alpine Testing Solutions*

Congratulations to everyone whose proposals were accepted for the upcoming annual meeting in New York City! As you begin preparations to present your research in this professional forum, we wanted to share some strategies for making the most of this opportunity. Communicating the depth of your study in the typical time allotment (i.e. 12-15 minutes) is not feasible, but there are effective strategies to pique audience members' interest in reading your full paper.



Renfrow and Impara (1989) provided some practical advice for how to develop your presentation, how the presentation should differ from your paper, and how to tailor your presentation for the audience. Their recommendations continue to be useful for new and experienced presenters alike. However, as technologies for supporting presentations have changed, there has been a shift in the last decade in the dominant presentation mode for communicating research.



Validity Theory and PowerPoint

Use of presentation software such as PowerPoint has become commonplace at many professional meetings. However, its uses have gone beyond the original intent of its developers (Gomes, 2007). The intended use of PowerPoint was to support the presentation (not be the presentation) by providing a quick summary of a longer, more in-depth document. There is an analogy here to validity theory. Although we regularly remind test developers and users that validation begins with clearly defined intended uses and interpretations of test scores, misuses still occur. As applied to presentation software, one increasingly common misuse is that presentation slides may serve as a substitute for a full paper. We might interpret this as an unintended consequence.

The following are examples of common mistakes that technology facilitates, but may hurt your presentation:

- Too many slides: Trying to tell the audience everything. . .using 40 slides. . .in 15 minutes.
- Text density: Using presentation slides as an extended abstract and then reading directly from them.
- Overuse of features: Animation, sound, video, etc. overwhelm the presentation.

Despite these potential pitfalls, presentation software can greatly enhance the effectiveness of your presentation if used appropriately. Here are some suggestions for using such software effectively:

- Develop one (1) slide for every 2-3 minutes of allocated presentation time.
- Keep the amount of text on a slide to a minimum. These should be talking points, not complete sentences.
- Slide designs, fonts, and features should consider the size and lighting of the room.
- Graphs, tables, formulas, and figures should be used to highlight key points.

Tips for Maximizing your Presentation Time

Given the limited time you will have to present your paper, it is important to prioritize information you want to share with the audience. Your full paper will include more complete descriptions of the context of the study, an analysis of previous literature on the topic, descriptions of the methodology, complete results, and an in-depth discussion of the meaning of the findings for theory and/or practice. Think about being able to answer these questions through your presentation:

- Why is my study important to the field?
- What findings of my study are novel and go beyond what is already known about the topic?
- Why do the findings matter to theory and/or practice?
- What are the notable limitations and future research that could extend this work?
- What are the 1-2 ideas that the audience should take from my presentation?

Though far from an exhaustive list, we hope that these tips serve as a useful starting point for your presentation development.

References:

Gomes, L. (2007, June). *PowerPoint turns 20, as its creators ponder a dark side to success*. The Wall Street Journal Online. Article can be accessed at: <http://online.wsj.com/article/SB1182281116940840904.html>
Renfrow, D. & Impara, J. C. (1989). Making Academic Presentations – Effectively! *Educational Researcher*, 18(2), 20-21. Article can be accessed at: <http://edr.sagepub.com/cgi/content/abstract/18/2/20>

2008 Annual Meeting • March 25-28 • New York City

Meeting and program information will be available at the NCME website www.ncme.org by December 3, 2007. The NCME Annual Meeting & Training Sessions Brochure will be mailed to NCME members the week of December 10, 2007. For more information, please call the NCME office at +1-608-443-2487, ext. 145.

PROGRAM HIGHLIGHTS

Presidential Address:

The Impact of Anchor Test Configuration on Students' Proficiency Classifications, *Anne Fitzpatrick*

Career Award Recipient Address:

Schrödinger's Cat, Rasch's P and the Most Dangerous Equation, *Howard Wainer*; Moderator: *Stephen Sireci*; Discussant: *Andrew Gelman*

Invited Symposia:

Open Hearing on Revising the 1999 Standards for Educational and Psychological Testing; Organizers/Moderators: *Wayne Camara, David Frisbie*; Organizer/Presenter: *Suzanne Lane*; Presenters: *Bob Brennan, Eva Baker, Michael Kane, Michael Kolen, Robert Linn, Randy Bennett, Martha Thurlow, Joan Herman, Dan Eignor*; Co-sponsored by NCME and AERA Division D

Policy, Technical and Operational Issues and School Accountability for Growth; Organizer/Presenter: *Huynh Huynh*; Presenters: *Charity Smith, Robert Kennedy, Anita Rawls, Do-Hong Kim, Eugene Kennedy*; Discussants: *Peter Behuniak, William Schafer*

Accessibility, Comparability and Validity: Evaluating Test Score Equivalence Across Test Forms and Subpopulations; Organizer/Discussant: *Stephen G. Sireci*; Presenters: *Neil Dorans, Walter (Denny) Way, Chow-Hong Lin, Katie Larson McClarty, Jadie Kong, Linda Cook, Cara Cahalan Laitusis, Ronald Hambleton, Avi Allalouf, Jamal Abedi*; Moderator: *Kadiye Ercikan*

The Intersection of Accountability and Measurement: Policy and Psychometrics; Organizer/Moderator: *Phoebe Winter*; Presenters: *Kerri Briggs, David Abrams, Joseph Martineau, Rebecca Kopriva, Karen Barton*; Discussant: *Robert Linn*

Professional Development Programs in Formative Classroom Assessment: Do Changes in Teacher Practice Improve Student Achievement?; Organizer/Presenter: *Christina Schneider*; Presenters: *Caroline Wylie, Courtney Bell, Dylan Wiliam, Shelley Ragland, Ching Ching Yap, Pamela Kaliski, Susan Brookhart, Connie Moss, Beverly Long*; Discussant: *Bruce Randel*

Dualities of Validity in an International Context; Organizer: *William Lorie*; Presenters: *Michelle Boyer, Juan Enrique Froemel, Vyjayanthi Sankar, Sridhar Rajagopalan, Eugenio Gonzalez*

Future Directions for the Field of Educational Measurement (Graduate Student Issues Committee); *Gregory Cizek, Howard Wainer, Catherine Welch, Daniel Koretz*

English Language Learners and NCLB 101 (Diversity Issues in Testing Committee); Organizer: *Cara Cahalan Laitusis*; Presenters: *Jamal Abedi, Micheline Chalhoub-Deville, Charlene Rivera, John Young*

Cutting Edge Indicators of Educational Quality for Predicting Global Competitiveness (National Association of Testing Directors); Organizer: *Bonnie Strykowski*

NCME Fitness Run/Walk: Thursday, March 27, 5:45 a.m. - 7:30 a.m.

- Run 5K or walk 2.5K course • Commemorative t-shirts for all participants (even if you don't wake up in time to make it!)

New York City has many of the world's most recognizable buildings and some of the world's most celebrated architecture.



2008 NCME Annual Meeting Pre-Conference Training Sessions

Sunday, March 23, 2008

Developing Noncognitive Assessments (Full Day)

Patrick Kyllonen, Educational Testing Service; Richard Roberts, Educational Testing Service

Student Involvement and Formative Feedback in Classroom Assessment: Measurement Concepts and Issues (Full Day)

Jeffrey Beaudry, University of Southern Maine; Leslie Lukin, Lincoln Public Schools; Lori Nebelsick-Gullet, Lincoln Public Schools

Item Response Theory: Parameter Estimation Techniques (Full Day)

Seock-Ho Kim, University of Georgia

Linking and Aligning Scores and Scales (Half Day)

Neil Dorans, Educational Testing Service; Jinghua Liu, Educational Testing Service; Mary Pommerich, Defense Manpower Data Center; Michael Walker, Educational Testing Service

Test Security: Practices, Policies, and Punishment (Half Day)

James Impara, Caveon Test Security; Ardeshtir Geranpayeh, University of Cambridge Esol Examinations; Jamie R. Mulkey, Caveon Test Security

Nonlinear Mixed Models Approach to Item Response Theory (Half Day)

Paul De Boeck, K.U. Leuven; Frank Rijmen, Educational Testing Service; Francis Tuerlinckx, K.U. Leuven; Mark Wilson, University of California-Berkeley

An Introduction to the Application of BMIRT: Bayesian Multivariate Item Response Theory Software (Half Day)

Lihua Yao, CTB/McGraw-Hill; Daniel M. Lewis, CTB/McGraw-Hill

Monday, March 24, 2008

Test Equating Methods and Practices (Full Day)

Michael J. Kolen, University of Iowa; Robert L. Brennan, University of Iowa

Applying Hierarchical Models to Causal Inference (Full Day)

Guanglei Hong, OISE/University of Toronto; Stephen Raudenbush, University of Chicago

Considerations in Setting Performance Standards (Full Day)

Mary Pitoniak, Educational Testing Service; Michael Zieky, Educational Testing Service

Bayesian Networks in Educational Assessment (Full Day)

Russell G. Almond, Educational Testing Service; Robert J. Mislevy, University of Maryland; David M. Williamson, Educational Testing Service; Duanli Yan, Educational Testing Service

Writing Diagnostic Items (Half Day)

Dylan Wiliam, Institute of Education, University of London; Caroline Wylie, Educational Testing Service

Skills Diagnosis with Latent Variable Models (Half Day)

Jeffrey Douglas, University of Illinois, Urbana-Champaign; Hua-Hua Chang, University of Illinois, Urbana-Champaign; Jimmy de la Torre, Rutgers University; Robert Henson, University of North Carolina-Greensboro; Jonathan Templin, University of Georgia

ICL and ETIRM: Open Source IRT Estimation Software for Researchers (Half Day)

Alan D. Mead, Illinois Institute of Technology; Werner Wothke, American Institutes for Research; Yanwei Zhang, American Institute of Certified Public Accountants

Building and Documenting a Valid Assessment System for Students with Disabilities: Psychometric and Practical Considerations for Alternate and Modified Assessments (Half Day)

Karen Barton, CTB/McGraw-Hill; Lara Osleson, CTB/McGraw-Hill; Dianne Lefly, Colorado Department of Education

Exploring the Validity of State Accountability Systems (Half Day)

Brian Gong, Center for Assessment; Marianne Perie, Center for Assessment

Tips for Graduate Students: Advice for Finishing School, Obtaining a Job and Starting a Career (Half Day)

Deborah J. Harris, ACT; Julio Sanclemente, CTB/McGraw-Hill; Andrew Ho, University of Iowa



2008 Annual Meeting March 25-28 New York City Division D Program Highlights

Invited Sessions

The Big Challenges and Research Opportunities in Testing and Measurement

Session Organizer and Chair: Ronald K. Hambleton

Session Participants: Joanna Gorin, Suzanne Lane, W. James Popham, Howard Wainer, Bruno D. Zumbo, Rebecca Zwick

Does Small Size Matter? A Symposium on Findings from New York City's Small High Schools

Session Organizer and Chair: Jane Delgado

Session Participants: Cheri L. Fancsali, Eileen M. Foley, Susanne Harnett, Helen Santiago

Discussant: John Q. Easton

Quality in Formative Assessment Practice

Session Organizer and Chair: Joan L. Herman

Session Participants: Joseph Ciofalo, Margaret Heritage, E. Carolyn Wylie

Discussants: Dylan R. Wilam, W. James Popham

Reflections on the Civic Obligations of Qualitative Inquirers

Session Organizer: Gretchen Rossman

Session Chair: Sharon Rallis

Participant: Thomas A. Schwandt

Discussant: Elizabeth A. St. Pierre

Generalizing from Educational Research: Beyond the Quantitative—Qualitative Opposition

Session Organizer and Chair: Kadriye A. Ercikan

Session Participants: Lyle F. Bachman, Deborah L. Butler, Margaret A. Eisenhart, Robert J. Mislevy, Pamela A. Moss, Wolff-Michael Roth, Guillermo Solanes-Flores, Kenneth G. Tobin, Noreen M. Webb

Comparative Perspectives on Classical Psychometrics and Item Response Theory

Session Organizer and Chair, Kurt F. Geisinger

Session Participants: Robert L. Brennan, Ronald K. Hambleton

Discussant: Chad W. Buckendahl

Especially for Graduate Students

Division D Graduate Student Seminar (Breakfast Meeting)

The Road Less Traveled: Transitioning from Graduate Student to Professional

Session Organizer and Chair: Linda Chard

Participants: H. Gary Cook, Arturo Olivarez, and Hao Song

Division D Mentoring Session Luncheon

Session Organizer: Allan Cohen

Participants: Linda L. Cook, Gregory J. Kelley, Robert J. Mislevy, Stephen W. Raudenbush, Mark, R. Wilson, Bruno D. Zumbo, Rebecca Zwick

Division D Fireside Chat: The "S" Factor: How Researchers, Methodologists and Measurement Scholars Build Capacity with a Foundation of Service

Session Organizers and Chairs: Jade Caines and Jill Adelson

Session Participants: Michael Nettles, Jennifer Bell-Ellwanger, Ellen Harrison, Leanna Stiefel, Lisa Colarossi

Discussant: Linda L. Cook

Important Event for Everyone

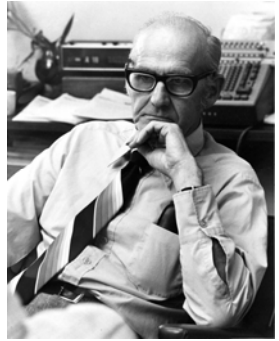
Division D Measurement and Research Methodology Business Meeting and Reception

The agenda for this important meeting will include the 2008 Award for the Outstanding Quantitative Dissertation and the 2008 Award for Significant Contribution to Educational Measurement and Research Methodology. The agenda will also include a discussion of the draft Division D Bylaws and reports of all Division D Standing Committees. Don't miss the reception with great food and drinks that will follow the Business meeting.

AL HIERONYMUS, TESTING PIONEER (1917-2007)

The following remembrance of Albert Nathan Hieronymus (MA '46/PhD '48) is by Professor Emeritus H. D. Hoover, who was first a student and then a colleague of Professor Hieronymus. Hieronymus, 89, died peacefully at his home September 3, 2007.

In the fall of 1964, I became a graduate research assistant to Al Hieronymus, professor of Educational Psychology and director of the Iowa Basic Skills Testing Program. Due to a shortage of office space for graduate students, I was assigned a spot in Al's office. Our desks were shoved right up against each other, and I spent the next year literally face-to-face with the most amazing person I have ever known. I learned more sitting at that desk, watching and listening to Al, than I have learned at any other time in my life.



What I observed is eloquently expressed by Leonard Feldt, one of Al's students, who went on to become the senior author of the Iowa Tests of Educational Development, the first author of the ACT, and director of the Iowa Testing Programs.

"It is not uncommon for a Ph.D. candidate to hold his advisor as a role model when he begins his academic career," Feldt said. "Though Al was my advisor, I could not adopt him as my model. He was deeply involved in so many projects simultaneously—development or revision of the ITBS (Iowa Tests of Basic Skills) and several other major tests, item tryout for future editions of the ITBS, direction of the annual testing program in Iowa, serving as a confidant to E.F. Lindquist on problems relating to the test scoring equipment, advisor to hundreds of undergraduates and many graduate students, teaching both undergraduate classes and graduate seminars, directing doctoral dissertations. How could a new Ph.D. dream of emulating a man of such prodigious talents and productivity?"

Al never closed the door to his office, and I never saw him turn away a student wanting to see him. He had endless patience with undergraduates but expected and demanded a great deal from graduate students and other faculty. When his children stopped at his office on their way home from school, he would give them his full attention—as well as money for treats.

Al's phone number was included on all communications to the hundreds of public and private schools participating in Iowa's testing program, and he received many calls every day. He always answered his own phone, patiently responding to questions from school administrators, teachers, parents, and the press. One of my most vivid memories is of the day Al was to discuss two journal articles at a faculty seminar. Thirty minutes before the meeting, he had yet to read either because of numerous student drop-ins and calls from schools. He gave me one of the articles and asked me to read it aloud. He simultaneously read the other, taking notes on both and only occasionally asking me to repeat a sentence.

Al worked closely with E. F. Lindquist, the founder of the Iowa Tests, to develop the optical scanner that revolutionized test scoring and helped to make Iowa City a world-renowned center for educational testing. In the late 1950s, he worked with Lindquist, James Van Allen, and others on the design and assembly of the campus's first computer. Al introduced numerous innovations in test design and score reporting, and under his leadership, the ITBS grew from a test battery created for use primarily in Iowa to one marketed by a major publisher and taken by millions of students worldwide. The manuals he created to accompany the ITBS reflected his day-to-day involvement with the schools. Al's manuals and their successors present such clear models for the proper use of tests and are so well respected that they are often used in university courses in educational measurement.

As Leonard Feldt recently remarked, "If Lindquist is to be remembered as the CEO of the testing empire that grew out of the Iowa Tests, then Al should be considered the Chief Operating Officer."

In 1991 the National Council on Measurement in Education honored Al with their Career Contributions to Educational Measurement award.

Al distinguished himself outside the field of education as well. During World War II, he served as a radar officer in a battalion that landed on Omaha Beach seven days after D-Day. He received a Bronze Star for his efforts to keep guns firing during the campaign that liberated the port of Cherbourg in late June 1944. He traveled by jeep to the five positions of the 115th Artillery Gun Battalion and made sure the radar and aircraft identification systems were working properly. Later in the war, he was promoted to captain and made commanding officer of a battalion. He traveled across Europe after D-Day with the Fourth Armored Division under General George Patton's command. When Al received the prestigious Croix de Guerre from Charles de Gaulle, General Patton was there to congratulate him and shake his hand.

The war had interrupted an earlier career in education. Before the war, Al had taught high school music, shop, science, math, and physical education in Opdike, Illinois, after graduating from Illinois State University. His future wife Wilfreda (Freda) was the commerce and English teacher in Opdike. After the war, Al and Freda moved to Iowa City where Al received his master's

degree and Ph.D. He joined The University of Iowa faculty in 1948 and immediately assumed responsibility for the ITBS and the state testing program.

Al was more than a pioneering figure in educational measurement and a war hero. He had a number of other passions. Al was a talented, self-taught musician who worked his way through college as a pianist. He continued to play, both for his own pleasure and as part of a jazz group composed of fellow U of I professors. The group often performed at University functions and at parties at Al and Freda's home.

Sports were another of Al's interests, and he especially enjoyed following all Hawkeye teams and his beloved Chicago Cubs. I joined Al's large family and numerous friends every year at his home to watch the baseball playoffs and World Series. It was also a tradition for more than forty years to gather at Al and Freda's home on New Year's Day to watch football from morning until midnight and to eat one of Freda's fabulous dinners.

Al was a gardener of unsurpassed enthusiasm. The several-acre area around the Hieronymus home was a veritable arboretum, containing trees and plants found nowhere else in Iowa. After his retirement, Al spent much time working in his yard and could often be found down on all fours tending his plants, usually with an open can or bottle of beer in the leg pocket of his overalls. How he kept it from spilling no one knows. He and Freda were founding members of Iowa City's Project GREEN, whose volunteers have planted thousands of trees and shrubs in the city's public spaces.

Of Al's many intense interests, his deepest devotion was reserved for his family. Al and Freda were married for 63 years. Freda, who died in 2005, was a well-known businesswoman. They are survived by their five children, seventeen grandchildren, and eight great-grandchildren.

Memorial contributions can be made to the Blommers, Hieronymus, and Feldt Scholarship for graduate students in educational measurement or the Holden Cancer Center, both through The University of Iowa Foundation, or to Project GREEN in Iowa City.

NEWSLETTER ADVISORY BOARD

MARY LYN BOURQUE, Mid-Atlantic Psychometric Services
SUSAN BROOKHART, Brookhart Enterprises LLC
SUSAN L. DAVIS, Alpine Testing Solutions
ELLEN FORTE, edCount LLC
SARA S. HENNINGS, Pearson Educational Measurement
JOAN HERMAN, CRESST/UCLA
THEL KOCHER, Edina Public Schools, Minnesota

WENDY MCCOLSKEY, SERVE
GERALD MELICAN, College Board
CAROL S. PARKE, Duquesne University
S.E. PHILLIPS, Consultant
DOUGLAS RINDONE, Consultant
GARY SCHAEFFER, CTB/McGraw-Hill
DONNA SUNDRE, James Madison University

SCOTT BISHOP, Editor, Data Recognition Corporation

Send articles or information for this newsletter to:

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

The *NCME Newsletter* is published quarterly. The *Newsletter* is not copyrighted; readers are invited to copy any articles that have not been previously copyrighted. Credit should be given in accordance with accepted publishing standards.