

PEARSON



TEST, MEASUREMENT, & RESEARCH SERVICES

Quarterly Newsletter

VOL 3 | NO 1 | 2010

INSIDE

- Editor's Note
- Announcements
- Seminars & Training
- Publications
- Conference Participation
- TrueScores

EDITOR'S NOTE

by Jason Wachsmuth

WELCOME TO THE EIGHTH ISSUE OF THE PEARSON TEST, MEASUREMENT, AND RESEARCH SERVICES (TMRS) NEWSLETTER WHICH IS AIMED AT PUBLICIZING THE ONGOING RESEARCH EFFORTS OF OUR GROUP TO THE MEASUREMENT COMMUNITY, BOTH WITHIN AND OUTSIDE PEARSON.

This issue includes an outstanding number of contributions from Content Support Services (CSS), Psychometric and Research Services (PRS), and Knowledge Technologies. Please read the Announcements section if you are interested in the enhanced features on the Assessment and Information Research & Resources website, staff awards, new staff hires, or career opportunities. If you are curious about the different types of internal training and seminars that have been presented or about literature that has been accepted for publication, please visit the Seminars & Training or Publications sections. Also included is an exhaustive catalogue of conference participation consisting of 59 presentations across 18 conferences. The newsletter concludes with a TrueScores blog entry provided by Ed Wolfe.

I would like to express my gratitude for being awarded the TMRS Newsletter Editor position. Being an editor is different than what I expected. I learned that it takes as much time to collect, organize, and format the content as to actually edit the content. I also have a new appreciation for the different functional areas that offer their support in the production

of this newsletter. Creative Services and CSS Copy Edit take the draft and transform it from a plain word document into the eye-pleasing publication that you see before you. I would like to acknowledge the previous editors, Dr. David Shin and Dr. Jason Meyers, for documenting and refining the process of publishing the newsletter. And I certainly must thank my colleagues for taking time to send me their contributions, and the Advisory Board for their suggestions. I look forward to learning as much as I can from this professional experience, and I hope you find the content to be interesting, informative, and useful.

We aim for widespread dissemination of this newsletter. If you or someone you know would like to be added to our distribution list, or if you require a printed version of the newsletter, please contact me directly. I also welcome questions, comments, and suggestions, as support for a continual effort to improve the newsletter. Back issues can be downloaded from the [Research Publications section of the Assessment & Information website](#).

Jason Wachsmuth, Editor
Assessment & Information, Pearson
jason.wachsmuth@pearson.com
319.339.6400 Ext 216005

**NEWSLETTER
ADVISORY BOARD**

Paul Nichols
Vice President of Psychometric &
Research Services

Mary Kino
Principal Research Scientist

Bob Dolan
Senior Research Scientist

David Shin
Senior Research Scientist

Edward W. Wolfe
Senior Research Scientist

Kelly Burling
Research Scientist

Yuehmei (May) Chien
Research Scientist

Jason L. Meyers
Research Scientist

Jason Wachsmuth
Editor, Research Assistant

*The Pearson Test, Measurement,
& Research Services 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.*

ANNOUNCEMENTS

Research & Resources Website Enhanced

TMRS launched an enhanced version of its **Research & Resources section of the Assessment & Information website**. TMRS worked with Marketing Services to ensure that visitors to the site will now be able to do the following:

- » Search for educational assessment research publications by keywords, topic, or type of publication.
- » Sort publications by title, abstract, author, and date.
- » Sign up to be notified as new research publications are posted to the site.
- » Submit a research question to the new AskAResearchScientist@pearson.com email box. The “Ask a Research Scientist” form on the site allows anyone from the public to submit a question automatically. TMRS staff will be monitoring and replying to the questions as appropriate. The team will also keep metrics including the number of inquiries received and common themes for questions.
- » Link to the refreshed **TrueScores blog**.
- » **Read bios of the TMRS leadership team**.
- » Learn about TMRS capabilities including item and test development, scoring and reporting, and research and analysis.
- » Learn about Pearson’s Global Psychometric Centers to showcase our collaborative relationships with other researchers and agencies to enhance our work in educational assessment.
- » Apply for the 2010 Pearson Research Fellowship.

TrueScores Blog to Re-launch

The **TrueScores blog** will resume with new entries for 2010. The blog will include biweekly updates from TMRS staff voicing insights and perspectives on timely industry trends and topics. Previous entries will be archived and accessible from the TrueScores home page. At some point later this year, we intend to turn on the comments feature of the blog to facilitate dialogue with psychometric peers and customers. All comments will be moderated before posting to the site.

New TMRS Bulletins Available

Catch up on the latest TMRS Bulletins posted to the website: *What is a Balanced Assessment System? What is a Learning Progression?* The bulletins are written for the general public, using easy-to-comprehend descriptions about educational assessment issues affecting our nation’s students.

Awards

IAN LITTLE IS SELECTED PRS EMPLOYEE OF THE YEAR

Ian Little was recognized as the 2009 Pearson PRS Employee of the Year during the TMRS Town Hall Meeting held in March. Little’s leadership and contributions to the Maryland customer this past year were impressive and a testament to his talent and hard work. Going into 2009, Ian accepted the challenge of leading the psychometric work for four major Maryland testing programs, MSA 3-8, MSA Science, Alt-MSA and mod-MSA. Ian successfully integrated the work across these four programs, skillfully facilitating and coordinating the efforts of staff working in San Antonio, Iowa City, and remotely. Ian oversaw a cross-training of the senior staff and strengthened the team’s ability to collaborate together on diverse projects with disparate psychometric processes. Ian developed an extremely positive relationship with the program team and the client, earning their trust and the full respect of the Maryland Technical Advisory Committee. Ian’s work ethic and commitment to quality is unparalleled which contributes to his skills as a leader and garners him the respect of his colleagues. Congratulations, Ian!



Ian Little (left) and Denny Way

ANNOUNCEMENTS

Continued from page 2

ALL-STAR AWARD RECIPIENTS

Julie Grisham

For outstanding contributions on AIMS

Will Steele

For outstanding contributions on FCAT Reading

Amy Hathorn

For outstanding contributions on the Delta Interim Process Initiatives and Training

Marc Johnson

PRS Training and Test Construction Initiatives, November 2009 to present

Daeryong Seo

PRS Equating Training Initiative, October 2009 to February 2010

TEAM PLAYER AWARDS

Joe Libby

Outstanding contributions on MC MCT2

Kellie Crain

Outstanding contributions on South Dakota STEP

Karen Leach

Outstanding contributions on New York WAR

Matt Malay

Outstanding contributions on New York WAR

Charlie Sims

Outstanding contributions on Item Writer's Guide (Mississippi Curriculum Test)

Sarah Schuster

Outstanding contributions on EOC/ELA items for external review

Andrea Swehosky

Outstanding contributions on EOC/ELA items for external review

Deanna Thomann

Outstanding contributions on Maryland Science iCal

Ian Little

Outstanding contributions on Oklahoma Standards Review Meeting

Anita Harvin

Outstanding contributions on South Carolina Math

Angie Tackett

Outstanding contributions on the Oklahoma RFP

Alice Tebbens Rutledge

Outstanding contributions on New York WAR

Hua Wei

Outstanding contributions on PRS Equating Training Initiative, January 2010 to February 2010

Beverly Ramirez

Outstanding contributions on CSS Admin Support for Additional Locations

TOEFL DISSERTATION AWARD

The TOEFL Program at ETS announced that this year's recipient of the Jacqueline Ross TOEFL Dissertation Award is Alistair Van Moere of Pearson's Knowledge Technologies. His dissertation, *Group oral tests: How does task affect candidate performance and test scores?* was supervised by Charles Alderson at the University of Lancaster. As the Ross Award recipient, Dr. Van Moere received a check for US \$2,500 and will be ETS' guest at the upcoming Language Testing Research Colloquium conference in Cambridge, England. An award plaque will be presented to him at the conference banquet.

New Staff Hires

APRIL SU AND AYUSH PATHAK JOIN PSYCHOMETRIC SERVICES IN IOWA CITY

The Iowa City office announces the hiring of April Su, Research Assistant and Ayush Pathak, Research Associate. They joined the PRS team in January and February, respectively, to help support Test Map Team activities. The Test Map Team functionality has recently expanded to include RAs from both Austin and San Antonio in order to fully integrate the processing of test maps across sites and systems.

YING-JU HSU JOINS RESEARCH SERVICES

Ying-Ju Hsu has joined Research Services in Iowa City as an intern. She received her Master's Degree in 2009 from the University of Iowa and is now a full-time student in the Educational Measurement and Statistics program. By working at Pearson she will advance her knowledge of large-scale testing while pursuing her interest in computer adaptive testing.

SERGIO LOZANO JOINS PSYCHOMETRIC SERVICES IN SAN ANTONIO

The San Antonio office welcomes Sergio Lozano to PRS. Sergio joined our team in December, 2009 as a Research Associate. He primarily supports the Florida FCAT and the City University of New York (CUNY) programs. Prior to joining Pearson, Sergio served as a Programmer Analyst at the University of Texas at San Antonio and as a Biostatistician at the University of Texas Health Science Center.

ANNOUNCEMENTS

Continued from page 3

Career Opportunities

Pearson is an Equal Opportunity Employer M/F/V/D, a member of E-Verify, and a drug-free workplace. [Please visit our Careers page for more information about our available positions.](#)

Title: Senior Research Scientist

Location: Austin, TX

Job Type: Regular Full Time

Req Number: 30637

We are looking for a Senior Research Scientist to work onsite in Austin, Texas to support the Texas assessment program. We would like to find a Senior Research Scientist who has experience in designing and implementing the technical solutions for large scale assessment programs. Experience in working with custom K-12 assessment implementation (including peer review response) is strongly desired. Strong verbal and written communication skills and ability to serve as a consultant to clients on technical matters are a requirement.

Title: Research Scientist

Location: Iowa City, IA

Job Type: Regular Full Time

Req Number: 30243

We are looking for a Research Scientist to contribute to our team developing computer adaptive testing and other psychometric algorithms. We would like to find a Research Scientist who understands JAVA and C++ or is willing to learn these programming languages.

Title: Research Scientist

Location: Iowa City, IA

Job Type: Regular Full Time

Req Number: 30241

We are looking for a Research Scientist with a background in models of student learning and teacher practice. Ideally, this Research Scientist would be familiar with qualitative approaches including focus groups and protocol analysis.

Title: Research Scientist

Location: Tulsa, OK

Job Type: Regular Full Time

Req Number: 30312

We are looking for a Research Scientist to work onsite in Tulsa, Oklahoma to support the psychometric development for the Oklahoma and Mississippi assessment programs. We would like to find a Research Scientist who has experience in supporting the design and implementation of the technical solutions for large-scale assessment programs. Experience in working with custom K-12 assessment implementation (including peer review response) is strongly desired. Strong verbal and written communication skills and ability to support clients on technical matters are a requirement.

Title: Associate Research Scientist

Location: San Antonio, TX

Job Type: Regular Full Time

Req Number: 30604

We are looking for an Associate Research Scientist who performs psychometric and research duties under guidance of more senior research scientists. This Associate Research Scientist demonstrates good communication skills (oral, written, and business) and works effectively with team members.

Title: Associate Research Scientist

Location: Iowa City, IA

Job Type: Regular Full Time

Req Number: 30387

We are looking for an Associate Research Scientist to support a statewide assessment program with responsibilities including literature reviews, research reports, item-level data analyses, calibration, scaling, equating, standard setting and technical report activities. We would prefer an Associate Research Scientist with strong SAS programming skills and experience with calibration and scaling software such as Multilog and STUIRT.

Career Opportunities

Continued from page 4

Title: Research Associate
Location: Tulsa, OK
Job Type: Regular Full Time
Req Number: 30222; 30754

We are looking for two Research Associates to work onsite in Tulsa, Oklahoma to support the psychometric development for the Oklahoma and Mississippi assessment programs. We would like to find Research Associates who have experience in SAS programming and who have demonstrated skills in cleaning, organizing, managing, and merging data sets. Strong verbal and written communication skills are a requirement.

Title: Research Associate
Location: Austin, TX
Job Type: Regular Full Time
Req Number: 30226

We are looking for a Research Associate to work onsite in Austin, Texas to support the psychometric development for the Modified (2%) and Alternate (1%) versions of the Texas Assessment of Knowledge and Skills (TAKS) for students with disabilities. We would like to find a Research Associate who has experience in SAS programming and who has demonstrated skills in cleaning, organizing, managing, and merging data sets. Strong verbal and written communication skills are a requirement. Interest in assessment issues for students with disabilities is a plus.

Title: Research Associate
Location: San Antonio, TX
Job Type: Regular Full Time
Req Number: 30692

We are looking for a Research Associate with good technical and analytical skills, knowledge of descriptive statistics, and programming experience in a language such as SAS, JAVA, VB, VBA, R, etc. Our ideal candidate possesses Basic and Advanced SAS certifications and is detail oriented, well organized, team-oriented, and has good problem solving and communication skills.

Title: Research Assistant
Location: San Antonio, TX
Job Type: Regular Full Time
Req Number: 30229

We are looking for a Research Assistant with good technical and analytical skills, knowledge of descriptive statistics, and some exposure to programming languages such as SAS, JAVA, VB, VBA, R, etc. Our ideal candidate is detail-oriented, well organized, team-oriented, and has good problem solving and communication skills.

SEMINARS & TRAINING

In January, Bob Parker delivered a one-day training in Austin for PRS staff from Austin and San Antonio. More than 40 people attended. Presented were some advanced SAS topics including: advanced macros, reading and writing Excel with DDE, executing Winsteps from within SAS, and various approaches to scoring items using SAS.

In February, Bob Parker delivered a two-day training in Tulsa for Tulsa PRS staff. The training included group and individual sessions, and topics covered advanced macros and a variety of SAS programming tasks, processes, and methodologies.

In late February, Brian Wrobel delivered a Test Map Team training for the San Antonio RAs. A test map is an electronic representation of the set of facts needed to build and score a test. The Test Map Team serves as the central authority for the management, dissemination, and archiving of test maps. During this training, he covered two different projects and detailed the steps necessary to process test maps from initial receipt to providing files for IT. He also reviewed possible problems that could be encountered along the way and touched upon some new initiatives for the Test Map Team.

In early March, Jason Wachsmuth delivered a seminar to Iowa City PRS management and the Virginia Program Team on utilizing an Audience Response System. He discussed the advantages and practicality of the technology for data collection at standard settings and data review meetings. The Virginia Team is accepting his recommendation and is pursuing this innovative opportunity to increase efficiency and decrease costs.

In March, Michelle Boazeman delivered a training demonstrating how to create tables in Microsoft Excel for reports using SAS DDE code. She covered three objectives: how to set up tables in Excel for reports that DDE can populate, how to write DDE code to populate tables, and running and verifying results.

PUBLICATIONS

Downey, R., Suzuki, M., & Van Moere, A. (2010). High-stakes English-language assessments for aviation professionals: Supporting the use of a fully automated test of spoken-language proficiency. *IEEE Transactions on Professional Communication*, 53, 18-31.

Elvevåg, B., Foltz, P. W., Rosenstein, M., & DeLisi, L. E. (2010). An automated method to analyze language use in patients with schizophrenia and their first degree relatives. *Journal of Neurolinguistics*, 23, 270-284.

Fuhrken, C., & Roser, N. (2010). Nurturing readers in an age of testing. *English in Texas*. Austin, TX: Texas Council of Teachers of English.

He, W., & Wolfe, E. W. (2010). Item equivalence in English and Chinese translations of a cognitive development test for preschoolers. *International Journal of Testing*, 10, 80-94.

Nichols, P. D., Twing, J., Mueller, C. D., & O'Malley, K. (2010). Standard setting methods as measurement processes. *Educational Measurement: Issues and Practices*, 29, 14-24.

Van Moere, A., Suzuki, M., Downey, R., & Cheng, J. (2009). Implementing ICAO language proficiency requirements in the Versant Aviation English Test. *Australian Review of Applied Linguistics*, 32, 27.1-27.17.

Wolfe, E. W., Hickey, D. T., & Kindfield, A. H. C. (2010). An application of the multidimensional random coefficients multinomial logit model to evaluating cognitive models of reasoning in genetics. In M. Garner, G. Engelhard, Jr., M. Wilson, and W. Fisher, Jr. (Eds.), *Advances in Rasch Measurement* (pp. 266-281). Maple Grove, MN: JAM Press.

Wolfe, E. W., & VanDerLinden, K. E. (2010). Development of scales relating to professional development in community college administrators. In M. Garner, G. Engelhard, Jr., M. Wilson, and W. Fisher, Jr. (Eds.), *Advances in Rasch Measurement* (pp. 242-265). Maple Grove, MN: JAM Press.

CONFERENCE PARTICIPATION

TEXAS COUNCIL OF TEACHERS OF ENGLISH

Fuhrken, C. (2010, January). *Think characterization is a big deal?: Reading tests do*. Paper presented at the Texas Council of Teachers of English Language Arts, Austin, TX.

CONSORTIUM OF SCHOOL NETWORKING

Foltz, P. W., & Rothenberg, L. (2010, March). *New technologies for automated assessment of 21st century skills*. Paper presented at Consortium of School Networking, Washington, D.C.

PITTSBURGH SCIENCE OF LEARNING CENTER

Bob Dolan represented the Assessment & Information group of Pearson at the Pittsburgh Science of Learning Center (PSLC) Industrial Affiliates Meeting on March 16 and 17 at Carnegie Mellon University.

NATIONAL SCIENCE TEACHERS ASSOCIATION

Mary Muehl, Chris Rozunick, and Barbara Tracy attended the National Science Teachers Association annual conference in Philadelphia from March 18-21. The conference included seminars, workshops, and presentations on such topics as the current thinking around core standards for science as well as best instructional/assessment practices for science educators.

CONFERENCE PARTICIPATION

Continued from page 6

Upcoming Conference Participation

SOCIETY FOR INDUSTRIAL & ORGANIZATIONAL PSYCHOLOGY

Murphy, S., Little, I., Kirkpatrick, R., Lin, C., Fan, M., Ragland, S., & Way, D. (2010, April). *Reducing error in equating: Anchor item drift analysis*. Paper to be presented at Society for Industrial & Organizational Psychology, Atlanta, GA. April 8, 12:00pm – 1:50pm.

Little, I., Murphy, S. Ragland, S., & Way, D. (2010, April). *The impact of different anchor stability methods on equating results and student performance*. Paper to be presented at Society for Industrial & Organizational Psychology, Atlanta, GA. April 8, 12:00pm – 1:50pm.

Murphy, S., Little, I., Lin, C., & McClarty, K. (2010, April). *Online versus paper: A comparability study*. Paper to be presented at Society for Industrial & Organizational Psychology, Atlanta, GA. April 9, 2:00pm – 2:50pm.

LANGUAGE TESTING RESEARCH COLLOQUIUM

Thomas, R., & Van Moere, A. (2010, April). *Validation of a benchmarking study by means of repeated measures classification consistency*. Paper to be presented at the Language Testing Research Colloquium, University of Cambridge, UK. April 15, 10:15am – 11:15am.

Van Moere, A. (2010, April). *Group oral tests: What kinds of tasks are optimal for eliciting and measuring functional interactions and interactional competence?* Paper to be presented at the Language Testing Research Colloquium, University of Cambridge, UK. April 16, 10:30am – 12:30pm.

Bernstein, J. (2010, April). *The continued high cost of industrial testing*. Paper to be presented at the Language Testing Research Colloquium, University of Cambridge, UK. April 16, 4:10pm – 6:10pm.

NATIONAL COUNCIL OF SUPERVISORS OF MATHEMATICS

Moyer, E. L., & GreedBurg, R. (2010, April). *Beyond average: Analyzing assessment data*. Paper to be presented at the National Council of Supervisors of Mathematics, San Diego, CA. April 20, 8:45am – 10:15am.

AMERICAN ASSOCIATION OF APPLIED LINGUISTICS

Downey, R., & Van Moere, A. (2010, April). *The comprehensibility of Indian call center agents for expert and naïve listeners*. Paper to be presented at the American Association of Applied Linguistics Texas Foreign Language Education, Austin, TX. April 23 – 24.

Suzuki, M., Ikari, Y., & Yokokawa, H. (2010, April). *Gains in speaking skills after three weeks abroad*. Paper to be presented at the American Association of Applied Linguistics Texas Foreign Language Education, Austin, TX. April 23 – 24.

TEACHERS OF ENGLISH TO SPEAKERS OF OTHER LANGUAGES

Luis Perea (2010, April). *Teacher input on a K-12 English language proficiency assessment*. Paper to be presented at the Teachers of English to Speakers of Other Languages, Boston, MA. April 27, 2:00pm – 2:45pm.

INTERNATIONAL READING ASSOCIATION

Fuhrken, C. (2010, April). *Read! Write! Film!: Using filmmaking to bring together all the language arts in elementary classrooms*. Paper to be presented at the International Reading Association, Chicago, IL. April 27, 10:15am – 11:15am.

Fuhrken, C. (2010, April). *Preparing students for the demands of reading tests*. Paper to be presented at the International Reading Association, Chicago, IL. April 28, 8:35am – 10:15am.

INTERNATIONAL OBJECTIVE MEASUREMENT WORKSHOP

Wolfe, E. W., & McVay, A. (2010, April). *Rater effects as a function of rater training context*. Paper to be presented at the International Objective Measurement Workshop, Boulder, CO. April 28, 11:00am – 12:15pm.

SOCIETY OF EXPERIMENTAL PSYCHOLOGISTS

Landauer, T. K. (2010, April). *A conjecture on space, spacing, randomness and brain evolution*. Paper to be presented at Society of Experimental Psychologists, Philadelphia, PA. April 29 – May1.

AERA DISCUSSANTS

C. David Shin

Using Item Response Theory to Investigate Various Challenges in Current Measurement Practice
Monday, May 3, 8:15am – 9:45am
Sheraton Denver / Governor's Square 17

AERA CHAIRS

Denny Way

Measuring College Readiness: Validity, Cut Scores, and Looking to the Future
Saturday, May 1, 8:15am – 10:15am
Sheraton Denver / Director's Row H

Jon Twing

Rasch Issues of Dimensionality, Scaling, and Fit
Saturday, May 1, 4:05pm – 5:35pm
Sheraton Denver / Plaza Court 6

Alvaro Arce-Ferrer

Validating Educational Assessments
Sunday, May 2, 8:15am – 9:45am
Sheraton Denver / Governor's Square 16

Upcoming Conference Participation

Continued from page 7

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

Susan Beretvas & Daniel Murphy

An Evaluation of the Performance of Information Criteria for Correct Cross-Classified Random-Effects Model Selection
Friday, April 30, 12:00pm – 1:30pm
Sheraton Denver / Governor's Square 17

Jie Lin, Kwang-Lee Chu, & Ying Meng

Distractor Rationale Taxonomy: Diagnostic Assessment of Reading With Ordered Multiple-Choice Items
Friday, April 30, 2:15pm – 3:45pm
Sheraton Denver / Grand Ballroom Section 2

Ha Phan, Christina Sentovich, Jeffery Kromrey, John Ferron, & Robert Dedrick

Correlates of Mathematics Achievement in Developed and Developing Countries: An Analysis of TIMSS 2003
Saturday, May 1, 8:15am – 9:45am
Colorado Convention Center / Room 710

Jon Twing

Commentary Paper
Saturday, May 1, 8:15am – 10:15am
Sheraton Denver / Director's Row H

Raymond Brown, Thanh Nguyen, & Agnes Stephenson

Examining the Effect of Title I and Other Federal Programs on the English-Language Proficiency Scores of Hispanic and Non-Hispanic Students Using Hierarchical Linear Modeling
Saturday, May 1, 10:35am – 12:05pm
Colorado Convention Center / Korbel Ballroom 2

Kimberly O'Malley

Comparison of Texas Schools Meeting Status Targets With Those Meeting Growth Targets
Saturday, May 1, 10:35am – 12:05pm
Sheraton Denver / Plaza Ballroom E

Edward Wolfe

A Comparison of the Distributional Properties of Four Mean-Square Fit Indexes Utilized in Commercial Rasch Measurements Software
Saturday, May 1, 10:35am – 12:05pm
Sheraton Denver / Grand Ballroom Section 2

Edward Wolfe

Analysis of Unidimensionality Testing Procedures in Item Response Data via Principal Component Analysis of Residuals from the Rasch Model
Saturday, May 1, 4:05pm – 5:35pm
Sheraton Denver / Plaza Court 6

Edward Wolfe

Performance of the Chi-Square Test for Detecting Unidimensionality in Applications of the Multidimensional Rasch Model
Saturday, May 1, 4:05pm – 5:35pm
Sheraton Denver / Plaza Court 6

Edward Wolfe

The Hazards of Newness: Connecting the Successful Retention of Students and Teachers in High School English Classrooms
Sunday, May 2, 8:15am – 9:45am
Colorado Convention Center / Korbel Ballroom 2

Husein Taherbhai

Assessing the Impact of Anchor (Linking) Items on Linear Equating in the Maintenance of the Vertical Scale
Sunday, May 2, 10:35am – 12:05pm
Sheraton Denver / Grand Ballroom Section 2

Tracey Magda

An Investigation of NCLB Trend Patterns and School Characteristics
Sunday, May 2, 12:25pm – 1:55pm
Sheraton Denver / Governor's Square 16

Tsung-Hsun Tsai, C. David Shin, Jason Meyers, & Hong Wang

Score Comparability of Paper-and-Pencil and Computer Versions of a Licensing Examination
Sunday, May 2, 12:25pm – 1:55pm
Colorado Convention Center / Room 608

Upcoming Conference Participation

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

Continued from page 8

Alvaro Arce-Ferrer

Investigating Approaches to Estimate an Individual's Strand-Objective Score Profile Reliability: A Monte Carlo Study

Sunday, May 2, 2:15pm – 3:45pm

Sheraton Denver / Governor's Square 16

Laurie Davis

GSC Chair Fireside Chat: Best Practices in Finding, Obtaining, and Maximizing Internship Opportunities in Measurement and Educational Research

Monday, May 3, 8:15am – 9:45am

Colorado Convention Center / Room 207

Ian Little & Daeryong Seo

Multigroup Comparability Testing for Latent Variables of the Profile Analysis via Multidimensional Scaling Model Using Confirmatory Factor Analysis

Monday, May 3, 8:15am – 9:45am

Sheraton Denver / Grand Ballroom Section 2

Edward Wolfe

A Comparison of Global Fit Indexes in a Multidimensional Rasch Analysis of Polytomous Data

Monday, May 3, 8:15am – 9:45am

Sheraton Denver / Governor's Square 17

Dennis Fulkerson, Paul Nichols, & David Mittelholtz

What Item Writers Think: Toward a Theory of Item-Writing Expertise

Monday, May 3, 10:35am – 12:05pm

Colorado Convention Center / Rooms 109, 111, 113

Daniel Murphy

The Effects of Serial Correlation on the Curve-of-Factors Growth Model

Monday, May 3, 12:25pm – 1:55pm

Sheraton Denver / Plaza Ballroom E

C. David Shin & Yuehmei Chien

Comparison of the Classification Accuracy of Subscale Score Estimation Methods

Monday, May 3, 2:15pm – 3:45pm

Colorado Convention Center / Korbel Ballroom 3

Laine Bradshaw, Jie Lin, Michael Young, & Kwang-Lee Chu

An Examination of Linguistic Modifications on a Large-Scale Assessment

Monday, May 3, 2:15pm – 3:45pm

Sheraton Denver / Plaza Court 7

Yuehmei Chien, C. David Shin, & Denny Way

An Alternative to Stocking-Lewis Conditional Item Exposure Control in CAT

Tuesday, May 4, 12:25pm – 1:55pm

Sheraton Denver / Governor's Square 17

NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Stephen Murphy, Ian Little, Robert Kirkpatrick, Meichu Fan, & Chow-Hong Lin

The Impact of Different Anchor Stability Methods on Equating Results and Student Performance

Saturday, May 1, 10:35am – 12:05pm

Majestic Ballroom, B6

Denny Way

Update on the Revisions to the Standards for Educational and Psychological Testing

Saturday, May 1, 4:05pm – 6:05pm

Denver Room, E1

Soo Ingrisone & James Ingrisone

A Joint Distribution Model Using Item Response and Response Time

Sunday, May 2, 10:35am – 12:05pm

Century Room, F2

Ye Tong & Michael Kolen

IRT Proficiency Estimators and Their Impact

Sunday, May 2, 12:25pm – 1:55pm

Silver Room, G4

NCME MODERATORS

Stephen Murphy

Item Response Models and Estimation

Saturday, May 1, 12:25pm – 1:55pm

Savoy Room, C5

Julie Miles

Licensure and Certification

Saturday, May 1, 2:15pm – 3:45pm

Denver Room, D1

Alvaro Arce-Ferrer

Reliability in Measurement

Sunday, May 2, 10:35am – 12:05pm

Gold Room, F3

Kelly Burling

Test Use with Special Populations

Monday, May 3, 12:25pm – 1:55pm

Savoy Room, L5

Qing Yi

Smoothing Techniques in Equating

Monday, May 3, 2:15pm – 3:45pm

Denver Room, M1

Kelly Burling

Cognitive Interviews in the Item Design Process for Assessments with Modified Achievement Standards

Monday, May 3, 4:05pm – 6:05pm

Denver Room, N1

NCME DISCUSSANTS

C. David Shin

Item Response Models and Estimation

Saturday, May 1, 12:25pm – 1:55pm
Savoy Room, C5

Ye Tong

Issues in Equating

Saturday, May 1, 12:25pm – 1:55pm
Century Room, C2

Robert Dolan

Innovative Considerations in Computer Adaptive Testing

Sunday, May 2, 10:35am – 12:05pm
Savoy Room, F5

Paul Nichols

Reliability in Measurement

Sunday, May 2, 10:35am – 12:05pm
Gold Room, F3

Denny Way

Making the Most of Innovative Items

Sunday, May 2, 4:05pm – 6:05pm
Savoy Room, I5

Denny Way

Quality Assurance and Control in Automated Scoring

Monday, May 3, 8:15am – 10:15am
Majestic Ballroom, J2

Upcoming Conference Participation

NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Continued from page 9

Paul Nichols

Are You Being Served? Operational Difficulties in Serving Real and Perceived Needs of State Assessment Clients

Sunday, May 2, 12:25pm – 1:55pm
Majestic Ballroom, G6

Anli Lin, Qing Yi, & Michael Young

Improving the Post-Smoothing of Test Norms with Kernel Smoothing

Sunday, May 2, 2:15pm – 3:45pm
Majestic Ballroom, H6

Julie Miles, Denny Way, & Jennifer Beimers

The Modified Briefing Book Standard Setting Method: Using Validity Data as a Basis for Setting Cut Scores

Monday, May 3, 8:15am – 10:15am
Silver Room, J4

Alvaro Arce-Ferrer

Derivation of a Reliability Index for an Individual Profile: A Multi-Factor Congeneric Approach with Guttman Error Structures

Monday, May 3, 10:35am – 12:05pm
Denver Room, K1

C. David Shin & Yuehmei Chien

An Exploration of Methods for Evaluation of Individual and School Progress at the Subscale Level

Monday, May 3, 10:35am – 12:05pm
Century Room, K2

Daeryong Seo, Ian Little, & Insu Paek

Multilevel Growth Mixture Analysis of Developmental Trajectories of Achievement Motivational Beliefs across Subject Areas

Monday, May 3, 10:35am – 12:05pm
Century Room, K2

Jason Meyers, Ahmet Turhan, & Steven Fitzpatrick

Performance of Ability Estimation Methods for Writing Assessments under Conditions of Multidimensionality

Monday, May 3, 2:15pm – 3:45pm
Gold Room, M3

Kelly Burling & Robert Dolan

Using Cognitive Interviews to Design Instructionally Relevant Simplifications and Supports for an Interactive Computer-Based AA-MAS

Monday, May 3, 4:05pm – 6:05pm
Denver Room, N1

Thakur Karki, Stephen Murphy, & Kevin Fatica

Comparisons of Test Characteristic Curves Alignment Criteria of the Anchor Sets and the Total Test for Maintaining Test Scale and Impact on Students' Performance

Monday, May 3, 4:05pm – 6:05pm
Century Room, N2

Hua Wei

Impact of Non-Representative Anchor Items on Scale Stability

Monday, May 3, 4:05pm – 6:05pm
Majestic Ballroom, N6

Kelly Burling

Cognitive Interviews in the Item Design Process for Assessments with Modified Achievement Standards

Monday, May 3, 4:05pm – 6:05pm
Denver Room, N1

EUROPEAN ASSOCIATION FOR LANGUAGE TESTING AND ASSESSMENT

Van Moore, A. (2010, May). *Usability testing with stakeholders in the development of a formative assessment of oral reading fluency*. Paper to be presented at European Association for Language Testing and Assessment, The Hague, Netherlands. May 28, 11:15am – 12:15pm.

Upcoming Conference Participation

Continued from page 10

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

Wolfe, E. W., Matthews, S., & Vickers, D. (2010, June). *Training/scoring context effects in large-scale writing assessment*. Paper to be presented at the Council of Chief State School Officers National Conference on Student Assessment, Detroit, MI. June 20 – 23.

Zhang, L., Xu, M., Lau, A., Wang, S., & Du, Y. (2010, June). *Identifying extraneous threats to test validity for improving tests and using of tests*. Paper to be presented at the Council of Chief State School Officers National Conference on Student Assessment, Detroit, MI. June 20 – 23.

Carling, D. L., Wangberg, L., Burling, K., Susbury, S., Domaleski, C. S., & King, K. (2010, June). *Comparability: What, why, when and the changing landscape of computer-based testing*. Paper to be presented at the Council of Chief State School Officers National Conference on Student Assessment, Detroit, MI. June 20 – 23.

EUROPEAN COMMISSION'S JOINT RESEARCH CENTRE

Alistair Van Moere has been invited to present at a workshop entitled *Computer-Based Testing of Speaking Skills* in Brussels, June 2010, by the European Commission's Joint Research Centre (JRC). The workshop is related to the European Commission's work on The European Survey on Language Competencies: Challenges for Research and Implementation.

INTELLIGENT TUTORING SYSTEMS

Bob Dolan is a coauthor on a paper entitled *The Effect of Motivational Learning Companions on Low-Achieving Students and Students with Learning Disabilities* that has been accepted at the Intelligent Tutoring Systems 2010 conference in Pittsburgh in June. This work results from an ongoing collaboration with Beverly Woolf in the Computer Sciences Department at the University of Massachusetts at Amherst.

TRUESCORES

Each issue of the Pearson TMRS Newsletter includes a recent entry from the TrueScores blog. For more information on TrueScores, please visit truescores.com.

Some Thoughts About Ratings...

By **Edward W. Wolfe**, Senior Research Scientist, Research Services

I spend a lot of time thinking about ratings. One reason I spend so much time thinking about ratings is that I've either assigned or been subjected to ratings many times during my life. For example, I review numerous research proposals and journal manuscripts each year, and I assign ratings that help determine whether the proposed project is funded or manuscript is published. I have entered ratings for over 1,000 movies into my Netflix database, and in return, I receive recommendations for other movies that I might enjoy. My wife is a photographer, and one of my sons is an artist, and they enter competitions and receive ratings through that process with hopes of winning a prize. My family uses rating scales to help us decide what activities we'll do together—so much so that my sons always ask me to define a one and a ten when I ask them to rate their preferences on a scale of one to ten.

In large-scale assessment contexts, the potential consequences associated with ratings are much more serious than these examples, so I'm surprised at the relatively limited amount of research that has been dedicated to studying the process and quality of those ratings over the last 20 years. While writing this, I leafed through a recent year of two measurement journals, and I found only three articles (out of over 60 published articles) relating to the analysis of ratings. I've tried to conduct literature reviews on some topics relating to large-scale assessment ratings for which I have found few, if any, journal articles. This dearth of research relating to ratings troubles me when I think about the gravity of some of the decisions that are made based on ratings in large-scale assessment contexts and the difficulty of obtaining highly reliable measures from ratings (not to mention the fact that scoring performance-based items is an expensive undertaking).

TRUESCORES

Continued from page 11

Even more troubling is the abandonment, by some, of the entire notion of using assessment formats that require ratings because of these difficulties. This is an unfortunate trend in large-scale assessment, because there are many areas of human performance that simply cannot be adequately measured with objectively scored items. The idea of evaluating written compositions skills, speaking skills, artistic abilities, and athletic performance with a multiple-choice test seems downright silly. Yet, that's what we would be doing if the objective of the measurement process was to obtain the most reliable measures. Clearly, in contexts like this, the authenticity of the measurement process is an important consideration—arguably as important as the reliability of the measures.

So, what kinds of research need to be done relating to the analysis of ratings in large-scale assessment contexts? There are numerous studies of psychometric models and statistical indices that can be utilized to scale ratings data and to identify rater effects. In fact, all three of the articles that I mentioned above focused on such applications. However, studies such as those do little to contribute to the basic problems associated with ratings. For example, very few studies exist that examine the decision making process that raters utilize when making rating decisions. There are also very few studies of the effectiveness of various processes for training raters in large-scale assessment projects—see these three Pearson research reports for examples of what I mean: *Effects of Different Training and Scoring Approaches on Human Constructed Response Scoring*, *A Comparison of Training & Scoring in Distributed & Regional Contexts - Writing*, *A Comparison of Training & Scoring in Distributed & Regional Contexts - Reading*. Finally, there are almost no studies of the characteristics of raters that make them good candidates for large-scale assessment scoring projects. Yet, the basis of most of the decisions that are made by those who run scoring projects focus on these three issues: Who should score, how should they be trained, and how should they score? It sure would be nice to make better progress toward answering these three questions over the next 20 years than we have during the past 20.

Send information to:

Jason Wachsmuth

2510 North Dodge Street

Iowa City, IA 52245

jason.wachsmuth@pearson.com

319.339.6400 Ext 216005