

PEARSON



TEST, MEASUREMENT, & RESEARCH SERVICES

Quarterly Newsletter

VOL 2 | NO 1 | 2009

INSIDE

Editor's Note
Announcements
Advisory Board
Seminars & Training
Recent Publications
Upcoming Conference
Participation
TrueScores

EDITOR'S NOTE

by Jason L. Meyers

WELCOME TO THE FOURTH INSTALLMENT OF THE PEARSON TEST, MEASUREMENT, AND RESEARCH SERVICES (TMRS) NEWSLETTER, A QUARTERLY PUBLICATION DEVOTED TO PUBLICIZING THE ONGOING RESEARCH EFFORTS OF OUR GROUP TO THOSE BOTH WITHIN AND OUTSIDE THE PEARSON COMMUNITY.

This issue marks the first publication in my year long appointment as the Editor of the newsletter. I would like to start by thanking outgoing Editor David Shin for his excellent work during the newsletter's inaugural year and for showing me the ropes, so to speak. I can now see more clearly the amount of time and effort required to get this newsletter published.

I'd also like to thank those of you who have sent me your contributions to the newsletter and commend you on your outstanding research. Finally, I'd like to express my gratitude to the Advisory Board who provided priceless feedback and guidance on this effort. I recognize that everyone involved in this process has to devote time above and beyond his or her primary work obligations and that contributing to the newsletter is difficult, especially this time of year.

In this issue, we are pleased to announce the 2008 Psychometric and Research Services Employee of the year, the most recent Research Grant Award recipient and we welcome a handful of new employees to our group. We also provide an update on our 2009 Pearson Fellowship Program and a special thought-provoking blog by Dr. Jon Twing. Finally, this newsletter contains the times and locations of all NCME and AERA presentations involving Pearson TMRS staff. We are pleased to bring you this detailed program for the first time and hope that it will serve as a handy pocket guide for those of you traveling to San Diego. This program truly illustrates the wealth of research activities conducted by our group. By my count we have 23 NCME presentations, and 25 AERA presentations. In addition to presentations, our staff was heavily involved in these conferences acting as reviewers, moderators, and discussants and leading training sessions. In this issue we also detail the journal articles being

published and the seminars and training sessions conducted by our staff during the first quarter of 2009.

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 in a continual effort to improve the newsletter. Back issues can be downloaded from www.pearsonedmeasurement.com/research/newsletter.htm

ANNOUNCEMENTS

Mary Kino joins Research Services

In early 2009, Mary Kino transferred from Psychometric Services to the Research Services group. She currently provides psychometric leadership as Principal Research Scientist for the National Board of Professional Teaching Standards (NBPTS), a new venture that crosses several Pearson business units. Mary's transition was made, in part, because the NBPTS project contains an extensive research component. Before transferring to Research Services, Mary was most recently Lead Psychometrician for the Washington Assessment of Student Learning program, an inaugural full-service development contract for TMRS. In the seven years prior to joining Pearson in 1998, Mary worked with Advanced Systems, Harcourt, and the Michigan Department of Education. Her research interests include the analysis of quantitative data, and psychometric applications for test development and large-scale assessment.

Continued on page 2

Dr. Jason L. Meyers, Editor
Assessment & Information, Pearson
512.989.5300 ext 224004
jason.meyers@pearson.com

**NEWSLETTER
ADVISORY BOARD**

Paul Nichols

Vice President of Psychometric & Research Services

Mary Kino

Principal Research Scientist

Bob Dolan

Senior Research Scientist

Edward Wolfe

Senior Research Scientist

Kelly Burling

Research Scientist

Yuehwei (May) Chien

Associate Research Scientist

Jason L. Meyers

Editor, Research Scientist

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.

Send information for this newsletter to:

Dr. Jason L. Meyers

Phone: 512.989.5300

Ext 224004

Pearson

jason.meyers@pearson.com

400 Center Ridge Drive

Austin, Texas 78753

ANNOUNCEMENTS (CONT.)

Continued from page 1

Rob Kirkpatrick joins Psychometric Services in San Antonio

At the end of 2008, Rob Kirkpatrick relocated from Iowa City, Iowa to San Antonio, Texas.

Daniel Murphy joins Psychometric Services in Austin

Daniel Murphy is a doctoral candidate in the Educational Psychology Department at University of Texas at Austin. His research interests include multilevel modeling, longitudinal data analysis, and the evaluation of growth in academic achievement. Dan is currently working from Austin and providing assistance on the Texas Growth Project. Prior to joining Pearson Dan was a teacher and guidance counselor in Austin area public schools.

Phyllis Garrett joins Psychometric Services in Austin

Phyllis Garrett was hired as an Associate Research Scientist in Austin and works on the alternate assessments for the Texas program (TAKS-M and TAKS-Alt). She successfully defended her dissertation in December and will be graduating in May 2009. Her research interests include psychometrics related to testing students with disabilities, differential item functioning, effect size measures for polytomous items, and reliability. She will be presenting a paper at AERA entitled "Factors influencing conditional reliability in the classical test theory and item response theory frameworks." Prior to joining Pearson, Phyllis worked as a special education teacher for seven years and worked with Georgia State University to apply reading techniques in the public schools to help elementary and middle school students with disabilities increase their reading skills.

James and Soo Ingrisone join Psychometric Services in Iowa City

James joined the Iowa City Psychometric Services team in January 2009 as an Associate Research Scientist working on the Virginia assessment program. He graduated from Florida State University in December 2008. His research interests include classroom assessment, item response theory, second language acquisition research, and modeling the joint distribution of response accuracy and response time. He will be presenting his dissertation entitled "Modeling the Joint Distribution of Response Accuracy and Response Time" at the annual AERA conference in San Diego.

Soo also joined the Iowa City Psychometric Services team as an Associate Research Scientist working on the Minnesota program. She graduated from Florida State University in December 2008. Her research interests include modeling the joint distribution of response accuracy and response time as well as investigating procedures to estimate ability and item parameters. She will be presenting her dissertation entitled "An Extended Item Response Theory Model Incorporating Item Response Time" at the annual AERA conference in San Diego.

Holly Zhang wins PRS Employee of the Year

Holly Zhang was recognized as the 2008 Pearson PRS Employee of the Year during the Learning at Lunch presentation on March 13, 2009. Although Holly had a very heavy workload in 2008, supporting three Research Scientists in delivering three full-scale state projects, she always managed to find bandwidth to take on additional responsibility or tasks. Holly's talent in data analysis, programming, and implementation of psychometric specifications is unparalleled. She is dedicated to quality and develops repeatable processes for fulfilling analysis work. Holly's attention to detail and dedication was instrumental in rooting out potential defects in our inputs, and ensuring that psychometric work was done correctly the first time. In addition to being a selfless contributor in 2008, Holly completed the remainder of her Ph.D. coursework in Statistics at UTSA. Congratulations, Holly.

Pearson Fellowship Program

Recipients of the 2009 Pearson Fellowship will be awarded on March 31, 2009. This year we received over 50 applications for the 4 available positions. A summary of the Pearson Fellowship program is provided below.

Pearson will offer an 8-week fellowship in summer of 2009 for five doctoral students with outstanding psychometric skills who wish to gain experience in educational testing. The fellowship will include hands-on experience in the development and analysis of data for K-12 assessments. Specific activities may include item analysis/key check procedures, test construction, scaling and equating, technical writing, comparability studies, attendance/facilitation of item review committees, and other research projects. Fellows will also have an opportunity to learn about item content development, scoring and processing, and other aspects of educational testing.

DATES

The 8-week fellowship will take place from June 8, 2009 through July 31, 2009.

LOCATIONS

Two fellowships are available in San Antonio, Texas and one fellowship is available in both Austin, Texas and Iowa City, Iowa.

FELLOWSHIP GOALS

Fellows will have the opportunity to work closely with a team of Pearson research scientists in order to

- 1) Gain experience in many of the tasks involved in a K-12 testing contract.
- 2) Collaborate on a research paper, abstract, or presentation for national dissemination.
- 3) Better understand ways to collaborate and communicate about educational measurement.

AWARD

The award includes a \$5000 stipend. Transportation costs to and from Austin/Iowa City/San Antonio will be provided. Corporate housing will be provided for fellows commuting more than 50 miles.

QUALIFICATIONS

Fellows should be currently enrolled in a doctoral program in educational measurement, educational statistics, educational psychology or similar program. Two to three years of doctoral-level coursework is required. Fellows should have strong written and verbal communication skills. Fellows may not participate in other summer internship or externship programs during the Pearson fellowship program.

Pearson Research Grant Program

During this quarter, a research grant was awarded to Anli Lin, Qing Yi, and Michael Young for their application entitled "Improving the Post-Smoothing of Test Norms with Kernel Smoothing." Congratulations on an excellent proposal.

Pearson Research Grants are reviewed quarterly. The application due date for the upcoming quarter will be announced soon. A summary of the grant program is listed below and more information is available from the second installment of the Pearson TMRS newsletter:

www.pearsonedmeasurement.com/downloads/research/2008_v1n2_researchServ_news_092608.pdf

The Pearson Research Grant Program is an effort to improve the implementation of the researcher-practitioner model in Pearson Psychometric and Research Services.

The application process consists of the following steps:

- 1) A Psychometric and Research Services staff person, or team, will draft an application following the guidelines described in the Pearson Research Grant Application Guidelines.
- 2) The application will be submitted electronically to the committee chair before the quarterly due date.
- 3) The Pearson Research Grant Committee will meet quarterly to consider applications.
- 4) Applications will be announced within one week following the committee meeting.

Research Grant recipients will be provided a partial reprieve from their operational responsibilities to conduct the research described in the application.

SEMINARS & TRAINING

On March 12, 2009 **Bob Dolan** presented a seminar entitled "Psychometric Challenges and Opportunities in Implementing Formative Assessment" to the Computer Science Department at Worcester Polytechnic Institute in Worcester, Massachusetts.

As a part of the Pearson training initiative Bob Parker began giving monthly one hour presentations via WEBEX focusing on SAS system topics called "SAS Learning at Lunch". The objective of these presentations is to focus on some area of the SAS system and give the psychometric staff tools for working with SAS more effectively and efficiently. The first presentation was given in January and covered SAS ODS (Capturing Proc output as RTF, PDF and HTML). February's presentation was "Using Proc Format for more than just formatting". The topics of future presentations have not yet been decided, but Bob is open to suggestions for topics that can be effectively covered in the WEBEX hour long format. Please contact Bob at bob.parker@pearson.com for further information.



RECENT PUBLICATIONS

Arce-Ferrer, A., & Martinez, E. (in press). An investigation of test mode effects of paper-and-pencil and online administrations of the Raven Standard Progressive Matrices test. *Educational and Psychological Measurement*.

Murphy, D. L. & Pituch, K. A. (in press) The Performance of Multilevel Growth Curve Models under an Autoregressive Moving Average. *Journal of Experimental Education*.

Nichols, P. D., & Williams, N. (2009). Consequences of test score use as validity evidence: Roles and responsibilities. *Educational Measurement: Issues and Practice*, 28, 3-9.

Yi, Q. (2009). Book Review. N. J. Dorans, M. Pommerich, & P. W. Holland (2007, Eds) Linking and Aligning Scores and Scales. *Psychometrika*, 74(1), 175-177.

UPCOMING CONFERENCE PARTICIPATION

NCME

TUESDAY, APRIL 14

Um, K., Way, W. D., Fitzpatrick, S. J., & Kreiman, C.

- » The Effects of Response Probability Criteria on the Scale Location Estimation and Impact Data in Standard Setting.
- » Time: Tuesday, April 14, 8:15 a.m. – 10:15 a.m.
- » Location: Hard Rock Hotel, Legends 4, A3

Wan, L., & Henly, G.

- » Measurement Properties of Innovative Item Formats in a Computer-Based Science Test
- » Time: Tuesday, April 14, 10:35 a.m. – 12:05 p.m.
- » Location: Hard Rock Hotel, Legends 6, B5

Arce-Ferrer, A.

- » An investigation of traditional and alternative approaches to vertically scale Modified Angoff cut scores
- » Time: Tuesday, April 14, 12:25 p.m. – 1:55 p.m.
- » Location: Hard Rock Hotel, Legends 2, C4

Seo, D., Shin, S., Taherbhai, H., & Sun, Y.

- » Exploring and Explaining Gender Format Differences in English as a Second Language Writing Assessment Using Logistic Mixed Models
- » Time: Tuesday, April 14, 12:25 p.m. – 1:55 p.m.
- » Location: Hard Rock Hotel, Legends 4, G3

Shin, C. D., Ho, T., Chien, Y., & Deng, H.

- » A Comparison of Person-Fit Statistics in Computerized Adaptive Test Using Empirical Data
- » Time: Tuesday, April 14, 12:25 p.m. – 1:55 p.m.
- » Location: Hard Rock Hotel, Legends 4, C2

Turhan, A., Courville, T., & Keng, L.

- » The Effects of Anchor Item Position on a Vertical Scale Design
- » Time: Tuesday, April 14, 12:25 p.m. – 1:55 p.m.
- » Location: Hard Rock Hotel, Legends 5, C3

Tong, Y. & Kolen, M.

- » A further look into maintenance of vertical scales
- » Time: Tuesday, April 14, 2:15 p.m. – 3:45 p.m.
- » Location: Hard Rock Hotel, Legends 6, D5

Turhan, A., Lin, C., O'Malley, K. & Kolen, M.

- » Vertical Scaling for Paper and Online Assessments
- » Time: Tuesday, April 14, 2:15 p.m. – 3:45 p.m.
- » Location: Hard Rock Hotel, Legends 6, D5

WEDNESDAY, APRIL 15

Kahraman, N., & Thompson, T.

- » Relating Unidimensional IRT Parameters to a Multidimensional Response Space: A Comparison of Two Alternative Dimensionality Reduction Approaches
- » Time: Wednesday, April 15, 10:35 a.m. – 12:05 p.m.
- » Location: Hard Rock Hotel, Legends 5, F4

Meyers, J. L., Turhan, A., & Fitzpatrick, S. J.

- » Interaction of Calibration Procedure and Ability Estimation Method for Writing Assessments under Conditions of Multidimensionality
- » Time: Wednesday, April 15, 10:35 a.m. – 12:05 p.m.
- » Location: Hard Rock Hotel, Legends 5, F4

Shin, C. D., Chien, Y., & Way, W. D.

- » The Weighted Penalty Model and Conditional Randomesque Method for Item Selection in Computerized Adaptive Tests
- » Time: Wednesday, April 15, 12:55 p.m. – 1:55 p.m.
- » Location: Hard Rock Hotel, Legends 2, G1

Yi, Q.

- » The Impact of Ability Distribution Differences between Beneficiaries and Non-Beneficiaries on Test Security Control in CAT
- » Time: Wednesday, April 15, 2:15 p.m. – 3:45 p.m.
- » Location: Hard Rock Hotel, Legends 6, H5

THURSDAY, APRIL 16

Mao, X. & Fitzpatrick, S. J.

- » An investigation of the Linking of Mathematics Tests with and without Linguistic Simplification
- » Time: Thursday, April 16, 8:15 a.m. – 10:15 a.m.
- » Location: Hard Rock Hotel, Legends 6, J5

Ming, X., Wang, J., & Wu, S.

- » A Predictive Validity Study of an English Language Proficiency Test
- » Time: Thursday, April 16, 8:15 a.m. – 10:15 a.m.
- » Location: Hard Rock Hotel, Legends 6, J5

Arce-Ferrer, A.

- » Anchor-Test Design Issues in Equating
- » Time: Thursday, April 16, 10:35 a.m. – 12:05 p.m.
- » Location: Hard Rock Hotel: Legends 5, K4

Chien, Y., Shin, C. D., & Way, W. D.

- » Weighted Penalty Model for Content Balancing in CAT
- » Time: Thursday, April 16, 10:35 a.m. – 12:05 p.m.
- » Location: Hard Rock Hotel, Legends 6, K5

Keng, L. & Dodd, B. G.

- » A Comparison of the Performance of Testlet-Based Computer Adaptive Tests and Multistage Tests
- » Time: Thursday, April 16, 10:35 a.m. – 12:05 p.m.
- » Location: Hard Rock Hotel, Legends 6, K5

Song, T. & Arce-Ferrer, A.

- » Comparing IPD detection approaches in common-item non-equivalent group equating design.
- » Time: Thursday, April 16, 10:35 a.m. – 12:05 p.m.
- » Location: Hard Rock Hotel: Legends 5, K4

Wang, C., Wei, H., & Gao, L.

- » Investigating the effects of speededness on test dimensionality
- » Time: Thursday, April 16, 10:35 a.m. – 12:05 p.m.
- » Location: Hard Rock Hotel, Legends 4, K3

Wei, H.

- » The effect of test speededness on item and ability parameter estimates in multidimensional IRT models
- » Time: Thursday, April 16, 10:35 a.m. – 12:05 p.m.
- » Location: Hard Rock Hotel, Legends 4.

McClarty, K., Lin, C., & Kong, J.

- » How Many Students Do You Really Need? The Effect of Sample Size on the Matched Samples Comparability Analysis
- » Time: Thursday, April 16, 4:05 p.m. – 6:05 p.m.
- » Location: Hard Rock Hotel, Legends 3, N1

Thompson, T.

- » Scale Construction and Conditional Standard Errors of Measurement
- » Time: Thursday, April 16, 4:05 p.m. – 6:05 p.m.
- » Location: Hard Rock Hotel, Legends 6, N5

Ye, F., You, W., & Xu, T.

- » Multilevel Item Response Model for Longitudinal Data
- » Time: Thursday, April 16, 4:05 p.m. – 6:05 p.m.
- » Location: Hard Rock Hotel, Legends 5, N2

NCME TRAINING SESSIONS

Kolen, M. & Tong, Y.

- » Vertical scaling methodologies, applications and research
- » Time: Monday, April 13, 8:00 a.m. – 12:00 noon
- » Location: Hard Rock Hotel, Legends 2, II

NCME DISCUSSANTS

Shin, C. D.

- » Standard Errors of Equating
- » Time: Tuesday, April 14, 10:35 a.m. – 12:05 p.m.
- » Location: Legends 5, B4

Way, W. D.

- » Current Practices in Licensure and Certification Testing
- » Time: Tuesday, April 14, 12:25 p.m. – 1:55 p.m.
- » Location: Hard Rock Hotel, Legends 3, C1

Tong, Y.

- » Test Equating with Constructed-Response Items and Mixed-Format tests
- » Time: Thursday, April 16, 12:25 p.m. – 1:55 p.m.
- » Location: Hard Rock Hotel, Legends 6, L5

Continued on page 6

CONFERENCE PARTICIPATION (CONT.)

Continued from page 5

NCME MODERATORS

Dolan, R. (also a Discussant)

- » Comparability of Paper-and-Pencil and Computer-Based Exams
- » Time: Thursday, April 16, 4:05 p.m. – 6:05 p.m.
- » Location: Hard Rock Hotel, Legends 3, N1

Nichols, P. D.

- » Modifications of Traditional Methods of Setting Standards
- » Time: Tuesday, April 14, 12:25 p.m. – 1:55 p.m.
- » Location: Hard Rock Hotel Legends 2, C4

AERA

MONDAY, APRIL 13

Fulkerson, D., Nichols, P. D., Mislevy, R., Liu, M., Zalles, D., Fried, R., Villalba, S., Debarger, A., Cheng, B., Mitman, A., Haertel, G., & Cho, Y.

- » Research Findings: Leveraging ECD in Scenario-Based Science Assessments
- » Time: Monday, April 13, 12:00 noon – 1:30 p.m.
- » Location: San Diego Marriott Hotel & Marina, Salon D

Dolan, R., Way, W. D., & Nichols, P. D.

- » Technical Quality of Formative Assessments Within Online Instructional Tool
- » Time: Monday, April 13, 2:15 p.m. – 3:45 p.m.
- » Location: Omni San Diego, Salon C

Fulkerson, D., Mittelholtz, D., & Nichols, P. D.

- » The Psychology of Writing Items: Improving Figural Response Item Writing
- » Time: Monday, April 13, 2:15 p.m. – 3:45 p.m.
- » Location: San Diego Marriott Hotel & Marina, Mission Hills

Lau, C., Zhang, L. & Jiang, X.

- » Using Pass/Fail Pattern to Predict Students' Success for Standards: A Longitudinal Study with Large-Scale Assessment Data
- » Time: Monday, April 13, 2:15 p.m. – 2:55 p.m.

Stephenson, A. & Song, T.

- » Using HLM to Investigate Longitudinal Growth of Students' English Language Proficiency
- » Time: Monday, April 13, 2:25 p.m. – 2:55 p.m.
- » Location: San Diego Marriott Hotel & Marina, Marriott Hall Salon 4

Beretvas, S. N., Chung, H., & Meyers, J. L.

- » Modeling Rater Severity Using Multiple Membership, Cross-Classified, Random-Effects Models
- » Time: Monday, April 13, 4:05 p.m. – 6:05 p.m.
- » Location: San Diego Convention Center, Room 3

TUESDAY, APRIL 14

Harrell, L. & Wolfe, E. W.

- » A Comparison of Global Fit Indices as Indicators of Multidimensionality in Multidimensional Rasch Analysis
- » Time: Tuesday, April 14, 8:15 a.m. – 10:15 a.m.
- » Location: Omni San Diego, Gallery 3A

Yue, J., Creamer, E., & Wolfe, E. W.

- » Measurement of Self-Authorship: A Validity Study Using Multidimensional Random Coefficients Multinomial Logit Model
- » Tuesday, April 14, 8:15 a.m. – 10:15 a.m.
- » Location: Omni San Diego, Gallery 3A

Murphy, D. L.

- » Multilevel Growth-Curve Modeling: A Power Analysis of the Unstructured Covariance Matrix
- » Time: Tuesday, April 14, 1:15 p.m. – 1:55 p.m.
- » Location: San Diego Convention Center, Ballroom 6A

Arce-Ferrer, A., Song, T., & Sullivan R.

- » Linking Strategies and Item Screening Approaches: A study with Augmented Nationally Standardized Tests Informing NCLB
- » Time: Tuesday, April 14, 2:15 p.m. – 3:45 p.m.
- » Location: Omni San Diego, Salon E.

Shin, C. D. & Chien, Y.

- » Conditional Randomesque Method for Item Exposure Control in CAT
- » Time: Tuesday, April 14, 4:05 p.m. – 5:35 p.m.
- » Location: Omni San Diego, Salon E

Thompson, T. & Way, W. D.

- » Using CAT to Achieve Comparability with a Paper Test
- » Time: Tuesday, April 14, 4:05 p.m. – 5:35 p.m.
- » Location: Omni San Diego, Salon E

Yang, Z. & Wang, S.

- » Calibration Methods Comparison with the Rasch Model
- » Time: Tuesday, April 14, 4:05 p.m. – 6:05 p.m.
- » Location: Omni San Diego, Balboa 1

WEDNESDAY, APRIL 15

Shin, C.D.

- » Using Bayesian Sequential Analyses in Evaluating the Prior Effect for Two Subscale Score Estimation Methods
- » Time: Wednesday, April 15, 10:35 a.m. – 12:05 p.m.
- » Location: Omni San Diego, Salon D

Beimers, J.

- » Consistency of District Annual Yearly Progress (AYP) Determinations Across Three Types of NCLB Growth Models
- » Time: Wednesday, April 15, 12:25 p.m. – 1:55 p.m.
- » Location: Omni San Diego, Salon E

Wang, Z., Taherbhai, H., Xu, M. & Wu, S.

- » Modeling Growth in English Language Proficiency with Longitudinal Data Using the Latent Growth Curve Model
- » Time: Wednesday, April 15, 12:25 p.m. – 1:55 p.m.
- » Location: Omni San Diego, Salon E

Way, W. D.

- » Increased Use of Technology in Testing
- » Time: Wednesday, April 15, 2:15 p.m. – 3:45 p.m.
- » Location: Omni San Diego, Salon D

THURSDAY, APRIL 16

McGill, M. & Wolfe, E. W.

- » Assessing Unidimensionality in Item Response Data via Principal Component Analysis of Residuals from the Rasch Model
- » Time: Thursday, April 16, 11:15 a.m. – 12:05 p.m.
- » Location: San Diego Convention Center, Ballroom 6A

Ingrisone, J.

- » Modeling the Joint Distribution of Response Accuracy and Response Time
- » Time: Thursday, April 16 - 2:15 p.m. – 3:45 p.m.
- » Location: Omni San Diego, Salon E

Ingrisone, S.

- » An Extended Item Response Model Incorporating Item Response Time
- » Time: Thursday, April 16 - 2:15 p.m. – 3:45 p.m.
- » Location: Omni San Diego, Salon E

Sung, H.

- » Developing a Short Form of the Enright Forgiveness Inventory Using Item Response Theory
- » Time: Thursday, April 16, 2:15 p.m. – 3:45 p.m.
- » Location: Omni San Diego, Salon E

Arce-Ferrer, A., Wang, Z., & Xue, Q. (2009)

- » Applying Rasch Model and Generalizability Theory to study Modified Angoff cut scores for reporting with vertical scales
- » Time: Thursday, April 16, 3:05 p.m. – 3:45 p.m.
- » Location: San Diego Marriott Hotel & Marina/ Marriott Hall Salon 4

McGill, M., Wolfe, E. W., & Jarvinen, D.

- » Validation of Measures of the Quality of Mentoring Experiences of New Teachers
- » Time: Thursday, April 16, 3:05 p.m. – 3:45 p.m.
- » Location: San Diego Marriott Hotel & Marina, Marriott Hall Salon 4

Jiao, H., Wang, S., Wan, L. & Lu, R.

- » An Investigation of Local Item Dependence in Scenario-Based Science Assessments
- » Time: Thursday, April 16- 4:05 p.m. – 5:35 p.m.
- » Location: Manchester Grand Hyatt, Manchester Ballroom Section D

Tsai, T. & Shin, C. D.

- » Generalizability Analyses of a Case-Dependent Section in a Large-Scale Licensing Examination
- » Time: Thursday, April 16, 4:50 p.m. – 5:35 p.m.
- » Location: San Diego Marriott Hotel & Marina, Columbia 2

AERA DISCUSSANTS

Tong, Y.

- » Factors influencing equating accuracy
- » Time: Thursday April 16, 10:35 a.m. – 12:05 p.m.
- » Location: Omni, Balboa 3

Nichols, P. D.

- » Standards, Proficiency Judgments, and Norms
- » Time: Thursday, April 16, 12:25 p.m. – 1:55 p.m.
- » Location: Omni San Diego, Salon E

AERA CHAIRS

Mueller, C.

- » Studies Examining Achievement Gaps
- » Time: Tuesday, April 14, 4:05 p.m. – 5:35 p.m.
- » Location: Omni San Diego, Salon D



TRUESCORES

Each issue of the *Pearson Test, Measurement, and Research Services Quarterly Newsletter* includes a recent entry from the *TrueScores* blog written by Jon Twing. For more information on *TrueScores*, please visit www.truescores.com.

NAEP: Love It or Leave It

by Jon S. Twing

As the national debate about what to do with education reform rages, I hear repeatedly the need to have “national standards”, “common core standards” or at the very least, a more standardized system of national guidelines in order for us to measure, and presumably then improve, education in America. Often this discussion leads to debate about the merits of a national test—which is often assumed to be NAEP—which is sometimes called the “nation’s report card.”

NAEP is well researched, well documented and seems to be well loved if not revered by most psychometricians—other than me and a few others who dare to challenge the status quo. I have questioned the usefulness of NAEP as a “check test” for NCLB at various times in my career, all based on the following premise:

- » Students who take NAEP are essentially unmotivated; while most students are highly motivated to pass mandated state assessments.
- » NAEP essentially measures a “consensus” national curriculum; while state assessments measure very specific content standards, which presumably align or mirror instruction.
- » NAEP is individually administered via a specific student sample where students only take portions of the assessment; whereas all students take the complete statewide assessment.
- » NAEP was targeted to measure at a higher level of proficiency (for example, 39% of all students were at or above proficient on NAEP Mathematics in 2007); whereas for most statewide assessments the percentages were much larger.

When I speak with my colleagues most of them say things like: Lighten up! NAEP is great! Math is math. Etc. Oddly enough, for a group of well-trained colleagues who claim to be scientists, such indefensible positions leave me wanting. Good news—there is some sensible research in the measurement literature. I am referring to the article by Dr. Andrew Ho, from the University of Iowa, published in *Educational Measurement: Issues and Practices*. In this article, entitled *Discrepancies Between Score Trends from NAEP and State Tests: A Scale-Invariant Perspective*, Dr. Ho provides a balanced and well-reasoned argument regarding the usefulness of NAEP for such comparisons. I provide quotes (albeit taken out of context) such that your curiosity will be peaked and you will seek out and read his entire article.

“Given a perspective that NAEP and State tests are designed to assess proficiency along different content dimensions, State-NAEP discrepancies are not cause for controversy but a baseline expectation.”

“Trends for a high-stakes, or ‘focal’ test, may differ from trends on a low-stakes test like NAEP (an ‘audit’ test) for a number of reasons, including different ‘elements of performance’ sampled by both tests, different examinee sampling frames, or differing changes in student motivation.”

“As NAEP adjusts to its confirmatory role, there must be an additive effort to temper expectations that NAEP and State results should be identical.”

My colleagues, who have violated this last conclusion by Dr. Ho, are doing the policy makers and implementers of the NCLB “law of the land” a disservice by suggesting that statewide assessments are somehow inferior simply because their results are not replicated on NAEP. Let’s get back to speaking about the science of assessment and experimental comparison and leave the passion and politics to someone else.

