

PEARSON



# RESEARCH SERVICES

*Quarterly Newsletter*

VOL 1 | NO 1 | 2008

## INSIDE

Editor's Note  
Announcements  
TrueScores  
TMRS Updates  
Awards  
Training & Workshops  
Recent Publications  
Conference Participation  
Seminars  
Psychometric Fellowship

## EDITOR'S NOTE

*by David Shin*

### WELCOME TO THE FIRST EDITION OF THE NEW QUARTERLY NEWSLETTER OF PEARSON'S RESEARCH SERVICES!

The newsletter of the Pearson Test and Measurement Research Service (TMRS) is issued quarterly, in March, June, September, and December. The purpose of this newsletter is to advertise the research activities and efforts of Pearson TMRS staff to the psychometric community including the NCME and AERA division D members, and students on various campuses. As we target the audience in the psychometric community, the content of this newsletter includes announcement (e.g. summer interns), speaker's column, selected articles from TrueScores (Dr. Twing's blog), awards, publications, list of conference papers, seminar information, and some fun facts.

We rely heavily on electronic distribution of our newsletter to the community. If you do not have the capability to access the Internet, please contact the editor ([David.Shin@pearson.com](mailto:David.Shin@pearson.com)) for a text version. We would greatly appreciate your efforts in further distributing the newsletter to your colleagues as our E-mail distribution list is incomplete. If you haven't received the E-version of the newsletter directly from the editor and would like to subscribe to the quarterly newsletter, please forward your E-mail address to the editor. We encourage you to advise us of any changes to the E-mail addresses. If you believe you are outside the psychometric community or received the newsletter in error, accept our apology for the inconvenience and inform the editor to unsubscribe. Please send your comments to the editor. We would like to hear from you.

**Dr. Chingwei (David) Shin**, Editor  
Assessment & Information, Pearson  
319.339.6480  
[david.shin@pearson.com](mailto:david.shin@pearson.com)

## ANNOUNCEMENTS

Please join me in congratulating **Bob Dolan** on his appointment as editor of the Pearson Research Report Series. Bob's appointment will run through July 1, 2008. As editor, Bob will work with researchers from Psychometric Services, Content Support Services and other departments to identify and prepare for publication high quality research papers. Following successful peer review, these papers will be published on the Pearson website [www.pearsoned-measurement.com](http://www.pearsoned-measurement.com) as part of the Pearson Research Report series.

Bob Dolan started in November, 2007, as a Senior Research Scientist in the Research Services group. His work at Pearson focuses on computer-based summative and formative assessment with emphases on cognition, accessibility, and usability. Bob's previous work in education research focused on designing, implementing, and evaluating advanced technology-based learning environments that support diverse learners, including the development of novel methods and technologies for formative and summative assessment. Bob has served as principal investigator on research projects funded by the National Institutes of Health, the National Science Foundation, the U.S. Department of Education, and private foundations. Over the past 25 years, Bob has pursued parallel tracks in research and development in systems and cognitive neuroscience, and the development and engineering of technology-based tools for educational, scientific, and medical research. He received a B.S. in Biology at Cornell in 1986 and a Ph.D. in Brain and Cognitive Sciences at MIT in 1992.

**NEWSLETTER  
ADVISORY BOARD**

**Paul Nichols**  
Director of Research Services

**Bob Dolan**  
Senior Research Scientist

**Kelly Burling**  
Research Scientist

**Jason Meyers**  
Research Scientist

**Yuehmei (May) Chien**  
Research Associate

**David Shin**  
Editor, Research Scientist

*The Pearson Test, Measurement, and Research Service 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:

**David Shin**  
Phone: 319.339.6480  
Pearson  
david.shin@pearson.com  
2510 North Dodge Street  
Iowa City, IA, 52246

## TRUESCORES

*Each issue of the Pearson Research Services Quarterly Newsletter will include a recent entry from the TrueScores blog written by Jon Twing. For more information on TrueScores, please visit [www.truescores.com](http://www.truescores.com).*

### *Griddable Items Get No Respect... No Respect at All!*

*by Jon S. Twing*

While most people argue that you have to earn the respect you are given, this is not always the case. Take for example the hard working, informative, creative and open-ended test item type commonly known as the “griddable item.” This item type gets no respect. In fact, my guess is you don’t even know what I mean when I refer to a griddable item. Let me elaborate.

When criterion-referenced and mastery testing was all the rage back in the late 60’s and early 70’s, most bashers of multiple-choice or supply-only assessment items came crawling out of the woodwork. Now remember, this was prior to high-stakes assessment so most tests were loved by all! In response to this, assessment developers looked to “enhance” objective measures by making them more “authentic.” One way to do this and still keep the advantages of machine scoring was to ask an open-ended item (say a multiple-step mathematics problem) and to place a grid on the response document similar to how you might grid your name or date of birth. Once the student solved the math problem and presumably reached one correct answer in one format, he or she could grid the answer on the document. What a great idea! Boy, did people hate it and, as far as I can tell, people still hate it today.

Pearson has conducted research in all manners of investigations regarding the griddable item (see Pearson Research Bulletin #3, [http://www.pearsonedmeasurement.com/bulletin/Bulletin\\_3\\_FINAL.pdf](http://www.pearsonedmeasurement.com/bulletin/Bulletin_3_FINAL.pdf)), and very little of which has generated much interest. For example, when Pearson was advising the Florida Department of Education in this regard, the griddable item was perceived by the program’s critics as an “ineffective” attempt to “legitimize” a large-scale objective assessment as measuring “authentic” and meaningful content (i.e., including performance tasks) when it did not. This really seemed to be a policy and/or political battle which positioned the proponents

of performance tasks, who wanted rich embedded assessments, against the policy makers, who wanted economical and psychometrically defensible measures. It is too bad griddable research did not carry the day.

Another issue with griddables seems to be their content classification. Multiple step mathematics problems, for example, are likely to match more than one cell of a content classification. Furthermore, depending on how they are classified, substeps are not likely to reach a Depth of Knowledge (DOK) of 3 even if the total item does. Finally, some concerns have been raised from psychometricians using IRT to calibrate griddable items. Under IRT the argument goes, unless you are using the Rasch Model, a 3PL model will be required for traditional multiple-choice type items, but there will be no guessing associated with a griddable item. Hence, a 2PL model will be required to calibrate these items with no pseudo-guessing parameter. (We will save the argument of forcing the c-parameter to zero and not going to a mixed model for another blog.) Add to this the inevitable sprinkling of two and three category open-response items and the mixed model becomes a burden that might not be justified given the relatively few gridded items. Other attributes of the griddable item are delineated in the Pearson Research Bulletin #3.

The point of this blog (clearly a failure given that I feel the necessity to remind you of the point I was making) is to get assessment specialists, psychometricians, policy makers and teachers to objectively evaluate the merits of this item type. Another goal is to have my readers consider how the use of griddable items might help assessment become more of a driving force for good instruction. These are the goals of the blog despite the fact that griddable items get no respect.

## TMRS UPDATES

### *From the integration of former Harcourt Assessments staff into Pearson to the future Pearson research management plan*

*by Walter (Denny) Way*

The integration of former Harcourt Assessments staff into Pearson and has significantly increased our psychometric and research expertise. Our staff will now be even more prominent at conferences like NCME, AERA, and CCSO and we will have more published articles, book chapters, and professional presentations than ever. The integration of staff provides us with both the opportunity and challenge of considering how we will support research efforts going forward. During the expansion of Test, Measurement and Research Services over the past five years, research and professional visibility has been encouraged by Pearson's upper management. This will continue under the integration. In fact, Dr. Jon Twing, Executive Vice President of TMRS, has the following objective as part of his 2008 goals: "Lead efforts to build Pearson's reputation and credibility in the educational measurement and research community, continuing the transformation of Pearson's market position as the leading testing organization and a full-service measurement solutions provider."

For Jon – and all of us in TMRS – to achieve this goal, we must continue to encourage and support research efforts. However, because of our growth and the constant pressures of delivering on our customer commitments, we must do more than encourage and support research – we must actively manage our research efforts. Towards this end, Dr. Paul Nichols, Director of Research Services will be working with the TMRS leadership to develop and implement a research management plan. This plan will address mechanisms by which we can continue build our reputation as a leading educational measurement and research organization. It will address funding and allocating time for research activities, mechanisms by which research projects will be proposed, evaluated and monitored, and implications of supporting research on staffing levels.

Paul's plan will be developed by the end of June and will be communicated with staff in TMRS in July and implemented as we enter the budgeting cycle for 2009.

## AWARDS

### *Psychometric & Research Services Employee of the Year: Kelly Burling, Ph.D.*

Kelly joined the staff of Research Services at the end of 2006 as a consultant across projects and departments on alternate assessment issues. Kelly's work has made a significant positive impact across Pearson in a number of areas. In the area of research, Kelly worked with our customer in Virginia to win funds from the Technical Assistance on State Data Collection General Supervision Enhancement Grant. Kelly is working with Virginia to support an expansion of the commonwealth's assessment options for students with disabilities through the development and implementation of an alternate assessment with modified achievement standards in Grade 8 Reading and Math. In addition, Kelly has published three Pearson white papers to help customers understand alternate assessments: NCLB Regulations for Alternate Achievement Standards (1%), NCLB Regulations for Modified Achievement Standards (2%), and The Legislative History of Alternate Assessments. Kelly also coauthored the paper, Universal Design for Computer Based Testing, posted on Pearson's research page. In addition to research, Kelly has consulted on operational projects for a number of states including Maryland, New Jersey and Texas.





## TRAINING & WORKSHOPS

### *Seminars & Research Work*

#### **Alvaro Arce-Ferrer**

The training on the Latent Class Rasch Modeling will be delivered on August 6, 2008 to the Pearson psychometric fellows in San Antonio. The goals are: (1) to introduce the essentials of the model, (2) demonstrate its application, and (3) provide hands-on practice.

### *Standard Setting Workshop at ATP in Dallas*

The Pearson Psychometric and Research Services team presented a workshop at the annual conference for the Association of Test Publishers (ATP) held in Dallas, Monday, March 3, 2008. The title for the workshop was “Setting Performance Standards on High Stakes Tests.”

Pearson has arguably more experience setting performance standards under NCLB than anyone. Most of this research is not published in peer-reviewed journals but becomes part of statewide technical reports. The workshop was a great opportunity for customers, researchers and other practitioners to see what standard setting is all about for large-scale, high-stakes assessments required under NCLB.

The conference this year was held at the Gaylord Texas resort near Grapevine. The workshop took place on Monday, March 3, 2:00 - 4:30 p.m.

The presenters from Pearson included:

- » **Dr. Scott Davies**
- » **Dr. Erika Hall**
- » **Dr. Paul Nichols**
- » **Dr. Kimberly O’Malley**

The team described basic activities used under common standard-setting methodologies, including:

- » item mapping
- » modified Angoff
- » body of work
- » ID matching
- » contrasting/borderline groups
- » judgmental procedures

Facilitators described the roles played by psychometricians, meeting coordinators, and data analysts. Then, attendees participated in a sample item mapping standard setting in which they set a cut point and described reasons for their judgments. Throughout the workshop, seasoned facilitators shared lessons learned and distinguished what should happen in theory from what does happen in practice. Attendees left the workshop with a set of practical materials that will help them plan a future standard-setting meeting.

## RECENT PUBLICATIONS

**Arce-Ferrer, A.** (2008). Comparing Screening Approaches to Investigate Stability of Common Items in Rasch Test Equating. *Journal of Applied Measurement*, 9 (1), 57-67.

**Arce-Ferrer, A. & Borges, I.** (2007). Investigating postgraduate college admission interviews: Generalizability Theory Reliability and incremental predictive validity. *Journal of Hispanic Higher Education*, 6(2), 118-134.

**Chen, Y., Chang, H., & Yi, Q.** (2007). Flexible content balancing in computerized adaptive testing. *Applied Psychological Measurement*, 31(6), 467-482.

**Davis, L.L. & Dodd, B.G.** (2008). Strategies for controlling item exposure in computerized adaptive testing with the partial credit model. *Journal of Applied Measurement*, 9(1).

**Keng L., McClarty, K.L. & Davis, L.L.** (in press). Item-level comparative analysis of online and paper administrations of the Texas Assessment of Knowledge and Skills. *Applied Measurement in Education*.

**Meyers, J.L., Miller, G.E., & Way, W.D.** (in press). Item position and item difficulty changes in an IRT-based common item equating design, *Applied Measurement in Education*.

**Nichols, P. & Joldersma K.** (in press). Review of Cognitive Diagnostic Assessment for Education: Theory and Applications. *Journal of Educational Measurement*.

**Prince-Embury, S. & Courville, T.** (2008). Comparison of one-, two-, and three-factor models of personal resiliency using the Resiliency Scales for Children and Adolescents. *Canadian Journal of School Psychology*, 23(1).

**Prince-Embury, S. & Courville, T.** (2008). Measurement invariance of the Resiliency Scales for Children and Adolescents with respect to sex and age cohorts. *Canadian Journal of School Psychology*, 23(1).

**Tong, Y. & Brennan, R.L.** (2007). Bootstrap Techniques for Estimating Variability in Generalizability Theory. *Educational and Psychological Measurement*, Vol. 67, No. 5, 804-817.

**Tong, Y. & Kolen, M.J.** (2007). Comparisons of Methodologies and Results in Vertical Scaling for Educational Achievement Tests. *Applied Measurement in Education*, Vol. 20, No. 2, 227-253.

**Wang, S., Jiao, H., Young, M.J., Brooks, T., & Olson, J.** (2008). Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, 68(1), 5-24.

**Wang, S., Jiao, H., Young, M.J., Brooks, T., & Olson, J.** (2007). The Effects of Computer-based Test versus Paper-and-Pencil Test on K-12 Student Mathematics Tests: A Meta-Analysis. *Educational and Psychological Measurement*, v37, 219-238.

**Wang, S., Wang, N., & Hoadley, D.** (2007). Construct equivalence of a National Certification Examination that uses dual languages and audio assistant. *International Journal of Testing*, v7, No3, 355-368.

**Yi, Q., Harris, D., & Xiaohong, G.** (2008). Invariance of equating function across different subgroups of examinees taking a science achievement test. In A. A. von Davier and M. Liu (Eds.), Population invariance [Special issue]. *Applied Psychological Measurement*, 32(1), 62-80.

**Yi, Q., Zhang, J., & Chang, H.** (2008). Severity of organized item theft in computerized adaptive testing: A simulation study. *Applied Psychological Measurement*.



## SAS Global Forum

by Robert D. Parker

The 2008 SAS Global Forum was held mid-March in San Antonio, Texas. Attendees from all across the globe attended this year's conference, with a record 48 countries represented. More than 3,700 attendees came to San Antonio, and of those, more than 22 percent came from outside the United States. The Pearson Test and Measurement Research Services and the Clinical Assessments groups were well represented by 16 attendees. More than 400 presentations were given during the two and a half day conference. This year's conference included 12 pre-conference seminars and statistical tutorials, 43 Coders' Corner sessions, 13 Hands-on Workshops, 37 Tutorials, 44 presentations in six industry segments, 67 papers and presentations from international presenters, 25 poster presentations, and much more. In the SAS Demo Room, there were 111 stations and 34 super demos.

## CONFERENCE PARTICIPATION

### AERA

#### **Arce-Ferrer, A., Shin, S., & Lau, A.**

Evaluating Equivalence of Test Forms in Test Equating with the Random Group Design: Practical Effects of Unnecessary Equating.

#### **Chu, Kwang-lee & Lin, Serena Jie**

Distracter Rationale Taxonomy: A Formative Evaluation Utilizing Multiple-Choice Distracters

#### **Harms, Michael**

An Introduction to User-Centered Design in Large-Scale Assessment

#### **Jirka, Stephen**

Test Accommodations and Item-Level Analyses: Mixture DIF Models to Establish Valid Test Score Inferences

#### **Keng, Leslie; Leite, Walter L.; & Beretvas, Natasha**

Comparing Growth Mixture Models when Measuring Latent Constructs with Multiple Indicators

#### **Keng, Leslie; Miller, Edward; O'Malley, Kimberly; & Turhan, Ahmet**

Composite Score Reliability Given Correlated Measurement Errors between Subtests and Unknown Reliability for Some Subtests

#### **Kirkpatrick, Rob & Way, Denny**

Field Testing and Equating Designs for State Educational Assessments

#### **Lau, Allen**

Evaluating Equivalence of Test Forms in Test Equating With the Random Group Design

#### **Lin, Serena Jie**

Examining the Impact of Omitted Responses on Equating

#### **Meyers, Jason; Kong, Xiaojin; & McClarty, Katie**

An Investigation of the Changes in Item Parameter Estimates for Items Re-field Tested

#### **Nichols, Paul**

The Role of User-Centered Design in Building Better Assessments

#### **Seo, Daeryong**

Exploring the Structure of Achievement Goal Orientations Using Multidimensional Rasch Models

#### **Stephenson, Agnes**

Examining Individual Students' Growth on Two States' English

Language Learners Proficiency Assessments

Using HLM to Examine Growth of English Abilities for ELL Students and Group Differences

#### **Strain-Seymour, Ellen**

A User-Centered Design Approach for the Refinement of a Computer-Based Testing Interface

#### **Tong, Ye; Wu, Sz-Shyan; & Xu, Ming**

A Comparison of Pre-Equating and Post-Equating Using Large-Scale Assessment Data

#### **Wan, Lei & Ching-Chow Wu, Brad**

Person-fit of English Language Learner (ELL) Students in High-Stakes Assessments

#### **Wang, Jane**

Modeling Growth: A Longitudinal Study Based on a Vertical Scaled English-Language Proficiency Test

#### **Wang, Shudong**

Vertical Scaling: Design and Interpretation

The Sensitivity of Yen's Q3 Statistics in Detecting Local Item Dependence

#### **Wilson, Jeff**

A User-Centered Design Approach to Developing an Assessment Management System

### NCME

#### **Arce-Ferrer, Alvaro & Diaz, Ileana**

An Experimental Investigation of Rating Scale Construction Guidelines: Do They Work with Spanish-Speaking Populations

#### **Arce-Ferrer, Alvaro & Shin, Seon-Hi**

Three Approaches to Measuring Individual Growth

#### **Hall, Erika & Ansley, Timothy**

Exploring the Use of Item Bank Information to Improve IRT Item Parameter Estimation

#### **Keng, Leslie; Ho, Tusng-Han; Chen, Tzu-An; & Dodd, Barbara**

A Comparison of Item and Testlet Selection Procedures In Computerized Adaptive Testing

#### **Mueller, Canda**

Response Probability Criterion and Subgroup Performance

#### **Nichols, Paul & Williams, Natasha**

Evidence of Test Score Use In Validity: Roles And Responsibility

**Thompson, Tony**

Using CAT To Increase Precision In Growth Scores

**Tong, Ye & Kolen, Michael**

Maintenance of Vertical Scales

**Twing, Jon S.**

Off-the-Shelf Tests and NCLB: Score Reporting Issues

**Wang, Shudong & Jiao, Hong**

Empirical Evidences of construct Equivalence of Vertical Scale Across Grades in K-12 Large-Scale Standardized Reading Assessments

**Wang, Shudong; Jiao, Hong; & Hi, Wei**

Parameter Estimation of One-Parameter Testlet Model

**Wang, Shudong; Zhang, Liru; Kersteter, Patsy; Bolig, Darlene; Yi, Qing**

An Investigation of Linking a State Assessment to the 2003 National Achievement of Educational Progress (NAEP) for 4th and 8th Grade Reading

**Way, Denny; Lin, Chow-Hong; McClarty, Katie; & Kong, Jadie**

Maintaining Score Equivalence as Tests Transition Online: Issues, Approaches and Trends

**Way, Denny; Nichols, Paul; & Vickers, Daisy**

Influences of Training and Scorer Characteristics on Human Constructed Response Scoring

**Ye, F. & You, W.**

Applying Multidimensional Partial Credit Model in a longitudinal Design of Diagnostic Assessment. Paper was presented at the annual meeting of American Educational Research Association, New York, NY.

**Yi, Qing**

Item Pool Characteristics and Test Security Control in CAT

**CCSSO****Lau, C. A. & Arce-Ferrer, A. (2008).**

Quality Control for Large-scale Assessment Equating Practice. Paper to be presented at the 2008 CCSSO Large-Scale Assessment Conference, Orlando, FL.

**Nichols, Paul; Nielsen, Todd; Mislevy, Robert; Haertel, Geneva; Alcaya, Cheryl**

Using Evidence-Centered-Design in Building States' Large-Scale Assessments. Session to be presented at the 2008 CCSSO Large-Scale Assessment Conference, Orlando, FL.

**Wang, S. & Jiao, H. (2008).**

Construct Validity of Assessing Students with Accommodated State Testing program. Paper to be presented at the National Conference of the Large-scale Assessment, Orlando, FL.

**Wang, S. & Jiao, H. (2008).**

Growth Model in large scale assessment. Paper to be presented at the National Conference of the Large-scale Assessment, Orlando, FL.

**ATP****Davis, L.L., Strain-Seymour, E., Lin, C., & Kong, X. (March, 2008).**

Evaluating the comparability between online and paper assessments of essay writing in the Texas Assessment of Knowledge and Skills. Presentation given at the annual meeting of the Association of Test Publishers, Dallas, TX.

**Others****Arce-Ferrer, A. (Summer 2008).**

Assessing mode effects on computer based testing: Effects of four data collection designs on comparability results. 6th Conference of the International Test Commission. Liverpool, U.K.

**Arce-Ferrer, A. (Summer 2008).**

Measuring reliability of individually reported score profiles. 6th Conference of the International Test Commission. Liverpool, U.K.

**Feifei, Ye & Wenyi, You (2008).**

Comparing Multilevel IRT model and Multilevel SEM in Estimating the Effect of Multilevel Covariates on a Latent Trait Measured by Dichotomous Items. Presented at the Psychometric Society conference.

**Ornelas, Dyanne (Aida) & Muehl, Mary (2008).**

The ABCs of State Assessment Data. Workshop presented at the National Science Teachers Association conference.

**SEMINARS****Len Swanson: Pearson Visiting Scholar<sup>1</sup>**

by Jon Twing

Dr. Len Swanson from ETS was the most recent keynote speaker during the Pearson Visiting Scholars program in Iowa City. Dr. Swanson talked about Computer Adaptive Testing (CAT) and the history of

how we got to where we are. Dr. Swanson was a particularly good choice for this presentation as he has worked in CAT since its inception and was “on the floor” when most of the ground work was laid for what we take for granted today. Pearson was very lucky to have Dr. Swanson’s expertise, which he also shared with students and faculty from the University of Iowa.

Len pointed out that the desire to tailor testing toward individuals was really enabled by the proliferation of IRT methodology, as well as the continued improvements in technology. Early research was provided by think tanks like ETS with funding coming from the Office of Naval Research.

Len provided the following timeline as a backbone to anchor CAT development to:

- » 1980-1984: Computerized college placement tests
- » 1987-1988: Computerized mastery testing (NCARB)
- » 1990-1993: Praxis exam operational
- » 1990-1993: Graduate Records Exam (GRE) CAT Version
- » 1993-1994: NCLEX (Nurses’ Certification and Licensing Exam) CAT Operational
- » 1994-2008: Statewide CATs implemented both district and state level

*Continued on page 8*

## SEMINARS (CONT.)

Continued from page 7

When asked about the challenges encountered on the road to operational CAT exams, Dr. Swanson responded that quality item pools, infrastructure, and exam security were the big issues of the day. Funny, isn't it? Almost fifty years later the same issues are still roadblocks to fully realizing the potential of both computer-based, as well as computer adaptive testing.

<sup>1</sup> This article can be found from the following website: <http://www.truescores.com/2008/04/len-swanson-pearson-visiting-scholar.html>

### Harvey Goldstein: Pearson Visiting Scholar

On Friday, May 30, Dr. Harvey Goldstein, Professor of Social Statistics at the University of Bristol and Project Director for the Center for Multilevel Modeling Team, presented a workshop entitled "Theory and Methods of Multilevel Modeling" on multilevel modeling as part of the Pearson Visiting Scholars program in Austin, TX. This six hour workshop, organized by Dr. Jason Meyers and led by world renowned British statistician

Dr. Harvey Goldstein covered the following general topics:

- » The need for multilevel modeling (general theory, effects of not using these models, examples of data structures necessitating multilevel modeling)
- » How to estimate basic multilevel models (using MLwiN)
- » How Pearson staff might use multilevel models in their work
- » Multi-level item response models

## PSYCHOMETRIC FELLOWSHIP

*Pearson offers an 8-week fellowship each summer for doctoral students with outstanding psychometric skills who wish to gain experience in educational testing. The fellowship includes hands on experience in the development and analysis of data for K-12 assessments. Fellowships are available in Austin, TX, Iowa City, IA and San Antonio, TX. The following students have been awarded Pearson Psychometric Fellowships for 2008.*

### San Antonio

#### Jiseon Kim

University of Texas at Austin  
Advisor: Dr. Barbara Dodd

Jiseon's primary areas of interest are IRT and computer adaptive testing. Jiseon recently presented a paper at AERA, "A comparison of the test design variations in panel structures of the computerized adaptive sequential testing system under the partial credit model."

#### Sangwook Park

Florida State University  
Advisor: Dr. Akihito Kamata

Sangwook has broad research interests. Sangwook has worked on research projects including studies regarding item parameter drift, ideal test characteristic curves and ideal test information curves for assembling tests, and the effect of content representation for anchor items in IRT equating.

#### Tian Song

Michigan State University  
Advisor: Dr. Mark Reckase

Tian recently presented a paper at AERA, "Teacher qualification and student reading, growth in early childhood." Tian has worked in several large research projects including the large-scale IEA study in teacher education (TEDS-M), an NSF funded project in science education (Deep think), and the English certification testing programs at the University of Michigan.

#### Kirsten Hochstedt

Penn State University  
Advisor: Dr. Pui-Wa Lei

Kirsten's primary area of interest is the application of factor analysis to assess the structures of test responses, particularly, how such procedures can be used to provide information about the validity of test scores. In addition, she is studying IRT test equating and linking methods.

### Iowa City

#### Tsung-Han Ho

The University of Texas at Austin  
Advisor: Dr. Barbara G. Dodd

Han is a doctoral student at the Quantitative Methods program. He has worked as a graduate research assistant for the National Center for Educational Accountability (NCEA) and two NIH funded projects. He participated in the MCAT graduate student research program in 2007 sponsored by the Association of American Medical Colleges. He is now working on his dissertation addressing CAT supervised by Dr. Barbara G. Dodd. His current interests focus on item selection, ability estimation and exposure control procedures in CAT, computer-based testing as well as the applications of polytomous IRT models.

### Austin

#### Phyllis Garrett

Georgia State University  
Advisor: Carolyn Furlow

Phyllis has a background with and interest in assessment for students with disabilities. She will be working with the TX alternate assessments team on the TAKS-Alt and TAKS-M assessments.