# Pearson's Research Newsletter

Published Quarterly
Vol. 5, No. 3, 2012

INSIDE



# Editor Note

by Aline Orr

Welcome to the fall edition of Pearson's Research Newsletter.

With the creation of the Research & Innovation Network, the newsletter has shifted focus from reporting on Test, Measurement, and Research Services news to reporting more broadly on research across Pearson. We have introduced updates on several groups such as the Pearson Assessment Community, the Knowledge Technologies group, and the Language Testing group, and have strived to represent the broad scope of research and activities taking place at Pearson. As the Network grows, collaborations between groups are evolving and research across Pearson continues to flourish and expand. As a consequence of this evolution, we need a more dynamic method of communicating research progress and therefore are moving to a digital format. Research news will be shared and disseminated on the Research & Innovation Network website, which will be launched by the end of 2012. A link to the site will be sent to all of our readers once the site is operational.

We are excited to be moving to a completely digital platform and believe this will be a more accessible and dynamic way to broadcast information about our research. Due to the extensive overlap in content between the website and the Newsletter, this will be the final issue of Pearson's Research Newsletter. Please note that we will continue to deliver research news, the only change will be in the platform.

I would like to take this opportunity to thank our contributors for providing the material that made the Newsletter successful and interesting. I would also like to thank the Newsletter review board members for their time and advice, our readers for their support and feedback, and the previous editors, who created and helped shape this editorial endeavor. It has been a privilege to serve as the editor since spring 2012. Thank you for sharing your time, knowledge, and wisdom!

In celebrating five years of publications, I have invited the previous editors to participate in this note. Their impressions about their tenure as editors are presented next.

**David Shin, Ph.D.** , Senior Research Scientist
Editor from spring to winter 2008

In one of the Research Service meetings in 2008 the group decided to publish a quarterly newsletter to announce research Pearson staff had being doing. As the first editor of the Newsletter, I was able to influence the content and editorial mechanics and was always inspired by the quantity and quality of the work being done. But what I enjoyed the most was the feedback I received from the readers.

I feel that the newsletter provided a space for compiling and reporting our research publications and conference participation when another easily accessible platform didn't yet exist. In addition, it provided an opportunity for our summer interns to reflect on their experience at Pearson.

I cherish that experience as editor and appreciate the opportunity to contribute to this last issue of the Newsletter!

**Jason Meyers, Ph.D.** , Senior Research Scientist
Editor from spring to winter 2009

As the second editor of the Pearson Test, Measurement, and Research Services Newsletter, I served a one-year term. During this time, as I compiled the research

accomplishments of my fellow colleagues, I became impressed by the prolific contributions Pearson was making to the measurement community. While I had often thumbed through the meeting program at the AERA and NCME meetings looking for presentations by my Pearson co-workers, I had never been fully exposed to the vast body of research being produced by my organization. Serving as the editor helped me gain a deeper appreciation of Pearson's presence in the research arena.

*Continued on next page*



PEARSON

*Editor's Note, continued from page 1*

**Jason Wachsmuth**, Associate Systems Product Support Analyst
Editor from spring 2010 to winter 2011

The most rewarding aspect about being the editor was thinking of ways to make the newsletter better for the audience. Whether it was implementing a small detail like disseminating the publication via a hyperlink instead of a 1 MB attachment, or making the newsletter more eye-catching by adding photos, my goal was to always add value. I am appreciative for the knowledge and skills I learned as editor, and thankful for the people that assisted me along the way.


Aline Orr, Ph.D.
Associate Research Scientist, Editor
Psychometric & Research Services
Assessment & Information
Pearson
aline.orr@pearson.com
512.989.5300 Ext 224116


# Carl Hyman: In Memorium

by Denny Way, Ph.D., Senior Vice President, Psychometric and Research Services

The Pearson community was recently saddened by the loss of Carl Hyman, who passed away on September 5, after an extended battle with cancer. He was 57.

Carl joined Pearson from Sylvan Learning and worked as a program director delivering large state assessments for several years. Carl moved into Organizational Quality in 2008, where he worked both domestically and internationally as International Director of OQ Initiatives. In that capacity he was a founding member of the Pearson Assessment Community (PAC), a global shared services organization or community of interest, that was established in 2010 to address emerging needs for assessment services throughout the world. It was in this role that I got to know Carl, and I was privileged to work closely with him during the last years of his life.

Carl approached his work on PAC as he did all things in his life: he was wholly committed to the charter of the group and unwavering in his dedication to achieving tangible results. Carl's passion was project management (he was both PMP and Prince2 certified), and he gladly offered consulting services to Pearson colleagues throughout the world. Even as he fought cancer, Carl traveled to England, India (twice), and Australia to advance the mission of PAC and provide project management support to Pearson colleagues.

Carl's interests were remarkably broad. He earned his bachelor's degree and a master's degree at the University of Maryland, College Park, where he worked in urban studies, methodology, and demography. He was elected to the Phi Beta Kappa honors society. Prior to working in assessment, Carl spent a number of years working for the Baltimore City Department of Education as its director of the Office of Project and Grant Management. Carl's academic background prepared him well to understand the research and statistical aspects of assessment, and he was on equal footing with his more specialized colleagues when these topics were addressed.

Carl was a lifelong resident of Baltimore. His obituary in the *Baltimore Sun* highlighted his neighborhood activism and his involvement in a number of community and educational groups.

As a colleague, Carl was incredibly supportive and had a "can do" attitude that was infectious. He was principled to a fault, and I confess to rolling my eyes more than once listening to Carl's impassioned critiques when some person or group fell short of his ideals. Yet setbacks never dissuaded Carl, and he remained positive and optimistic through both the drudgeries of work and the more profound challenge of fighting the terrible disease that ultimately took his life. Carl's wife Meg sometimes referred to him as "Carl the Ironman." Indeed, his spirit and character will persist in the hearts and minds of those who knew him.

# Research in Action

## ADAPTIVE LEARNING: FORMATIVE ASSESSMENT FOR THE DIGITAL AGE?

by Bob Dolan, Ph.D., Senior Research Scientist

Adaptive learning is big. This artificial intelligence-based approach to e-learning has received a wave of recent attention from educators, administrators, policy makers, researchers, and the learning industry. So what is it? Broadly speaking, an adaptive learning system is an online instructional or tutoring solution that modifies instruction and/or supports as a function of individual student needs. At its simplest, an adaptive learning solution might provide a student with a customized lesson plan consisting of topics he or she hasn't yet mastered. However, adaptive learning is in its glory at greater levels of sophistication, ones rooted in more constructivist models of learning. This type of adaptive learning derives from the discipline of intelligent tutoring systems (ITS). ITS design tells us that in order to deliver effective content and supports to students, we must model the **domain** to be taught, the **instructional** approach for teaching it, and what the **student** currently knows and can do. The more sophisticated and detailed these models, the better the adaptive learning system can deliver appropriately tailored content and supports for the student.

So why frame adaptive learning in terms of formative assessment? Before answering this, we must establish a proper definition for formative assessment, as the term tends to be used quite loosely. First and foremost, classroom-based formative assessment is a *process* for improving learning opportunities for students based upon their instructional needs. What formative assessment is *not*, therefore, is an instrument alone. Rather, information gathered through instruments—formalized or not—can support students' and teachers' instructional decision-making processes at the core of formative assessment. Thus, a test cannot be a "formative assessment," any more than a boat can be a voyage. Central to classroom formative assessment is the teacher, who must be well-versed in formative assessment practices. In fact, both teacher and student must play an active role in successful formative assessment practice.

I'd like to argue that adaptive learning is an approach to formative assessment in which technology assumes some of the role of the teacher. Specifically, instruction and/or supports are provided to students as a function of what they need in real time. Thus, a middle school student learning to expand algebraic expressions can be presented with instruction and practice appropriate for his or her current level of topical knowledge and skills. In one case, this might include prerequisite instruction on order of operations; in another, it might be a complex problem with available hints. In any case, the goal of adaptive learning is to maximize a student's opportunity to learn based upon ongoing evaluation of his or her knowledge, skills, and other pedagogically relevant attributes. This is a core attribute of formative assessment.

That said, a major challenge—and opportunity—still exists for adaptive learning. Current adaptive learning solutions focus on modeling and responding to students "self-sufficiently," with no or limited active involvement by teachers. In reality, there are things that technology is good at doing, such as continuous estimation of multiple students' fine-grained knowledge and skills, data mining-based analytics to uncover non-obvious patterns and relationships, and providing highly individualized interventions. Humans, on the other hand, are adept at understanding and responding to non-cognitive factors, such as affect and learning preferences, at working with students on setting and maintaining learning goals, and at knowing when to let up on a student who's struggling versus providing additional challenge. As such, an ideal adaptive learning solution is one in which the teacher is kept in the loop regarding a student's progress. There are many ways of accomplishing this. Adaptive learning, for example, can be a component of a balanced assessment system used to provide teachers with actionable information to support data-driven decision making, especially since it can measure not only knowledge and skills but evidence of learning itself. Adaptive learning can be at the heart of instructional improvement systems to provide teachers with instructional recommendations. Also, adaptive learning systems could be "tweakable" by teachers—and by students—allowing sophisticated and highly personalized overrides or preferences that modify their behavior.

I entitled this article provocatively by intention. The short answer is, "not yet, but getting there." It's important to reiterate that adaptive learning is posited as an *augmentation* to traditional classroom formative assessment. Effective teaching remains as much art as it is science. But as long as we

*Continued on next page*

*Research in Action, continued from page 3*

can agree on what formative assessment is, respect the limitations of technology, understand the importance of putting students in charge of their learning, and include teachers directly in the design process, then adaptive learning systems are likely to radically enhance formative assessment. This will translate into improved opportunities to learn by increasingly diverse students.

# Education Technology

## AUTOMATED ITEM GENERATION

by Kirk A. Becker, Ph.D., Senior Research Scientist

I am currently one of several scientists within Pearson working on research related to automated item generation. This exciting area focuses on different ways to improve and/or automate the process of creating test items in all areas of assessment (educational, licensure and certification, clinical, admissions, etc.). This article provides to the Pearson research community an overview of the field and the current capabilities for automated item generation.

We are interested in automated item generation primarily for reasons of security, efficiency, and quality. The possibility of creating every possible test item in a given construct domain diminishes the security needs of individual items (or even of entire tests). Regarding efficiency, the effort involved in creating a family of items or an item template is indeed greater than the effort needed to create a single item. More crucially, this effort will typically be less than that required to create all of the items a program needs in a given area. Finally, automated item generation can improve quality by improving our understanding of individual test items and of the influence of the specific elements of those test items on their difficulty and quality.

Within automated item generation, one major distinction differentiates template-based item generation and item generation from source material. Another distinction distinguishes between on-the-fly item generation, in which items generated in real time are administered on the spot to candidates, and "cranking," in which generated items are reviewed, edited, and pretested in a process similar to that used by humans in generating items. Pearson's MyLabs electronic textbooks generate multiple-choice items on-the-fly from templates for the homework section, while the GRE math (I've been told) uses templates to crank out items that are then reviewed, edited, and pretested. Metametrics generates cloze items on-the-fly from source material, while programs that create multiple-choice items are currently generating items for review (Karamanis, Ha, and Mitkov, 2006; Gütl et. al., 2011).

At VUE we are currently exploring several research areas related to automated item generation.

- We are in the planning stages for a project that will use supervised learning to model the selection of effective and ineffective item distractors.
- We recently presented research on the creation of enhanced item writing materials—targeted information for human item writers to use in writing items—which led to increased item production compared with a control group (Becker & Olsen, 2012).
- We have constructed and tested clones of high-performing items for admissions tests, and we have found similar statistical performance for the clones.
- We have demonstrated the automated creation of millions of stems, keys, and distractors for certain admissions test reasoning items.
- We are reaching out to several research groups in Europe that have created algorithms that actually generate test items.
- We are employing methods that automatically classify items into test plan content areas and that flag potential content overlap and cuing (enemy items).

In a fully automated environment, algorithms would process source materials relevant to the construct (textbooks, audio podcasts, websites, course materials, etc.) to define the test content outline. Based on the construct definition, items automatically generated from a combination of item models, templates, and text generation algorithms would completely cover the construct domain. Generated items would be automatically referenced to source materials and classified on the content outline. All pairs of items would be automatically compared to evaluate content

*Continued on next page*

*Education Technology, continued from page 4*

overlap, cuing, and other content features that lead to local item dependence. Characteristics of the newly generated items would accurately estimate their statistical quality. All of these ideas based on current research in the fields of psychometrics and computer science have very exciting practical outcomes.

Automated item generation truly has the potential to revolutionize the field of assessment. While we have much work to do in refining the technologies, and we may never see a fully automated system, we are quite close to being able to produce an environment for item writers that provides numerous assistive features. Imagine an item development system that offers real-time feedback to item writers if their items do not match certain specifications or if the topic of their item is relatively trivial. Current prototypes can produce lists of plausible distractors for creating multiple-choice items as part of an item banking (content-management) system.

**Becker, K. A., & Olsen, J. B.** (2012, April). Generating Enhanced Item Writing Materials With Natural Language Processing. Paper presented at the annual meeting of the National Council on Measurement in Education, Vancouver, BC. Canada.

**Gütl, C., Lankmayr, K., Weinhofer, J., & Höfler, M.** (2011) Enhanced automatic question creator – EAQC: Concept, development, and evaluation of an automatic test item creation tool to foster modern e-Education. *Electronic Journal of e-Learning*, 9(1), 23-38.

**Karamanis, N., Ha, L. A. and Mitkov, R.** (2006). Generating Multiple-Choice Test Items from Medical Text: A Pilot Study. *In Proceedings of INLG 2006*. Sydney, Australia

# Language Testing Update

## THE PEARSON SCALE OF ENGLISH

by Glyn Jones, Senior Researcher

Work continues on the development of the Pearson Scale of English (PSE). A study was undertaken under the direction of Ian Wood, Editorial Director for test preparation materials based in Harlow, UK. More than 100 level descriptors (CAN DO statements such as "can understand short, simple letters" or "can express a relationship of cause and effect") were formulated, drawing on syllabi and courses throughout Pearson and elsewhere. These were presented to 89 experienced Pearson staff—editors and courseware developers—who, after appropriate training in the PSE, were asked to rate each descriptor on the scale (i.e., to allocate a number between 10 and 90 to each descriptor as an estimate of its degree of difficulty). At the same time an open poll was conducted via an online survey in which language teachers were invited to rate the same descriptors according to the Common European Framework (CEFR), i.e., to allocate each descriptor to one of the six levels of the CEFR. Several hundred teachers responded, of whom more than 300 professed to be very familiar with the CEFR. When the results of the two studies were analyzed a very high degree of agreement was found. Not only did each group—the Pearson experts and the CEFR-experienced teachers—show a remarkable degree of consistency in the way they placed the descriptors on the respective scales, when correlated with each other they indicated a concordance between the PSE and the CEFR that was very close to the original mapping that Language Testing carried out when developing Pearson Test of English Academic.

The future development of the scale will be subject to regular scrutiny by a specially convened Technical Advisory Group (TAG). This is a committee of internationally prominent language learning experts from outside Pearson, who will meet periodically to advise us on the further development of the PSE. The TAG will meet for the first time in November 2012.

*Continued on next page*

*Language Testing Update, continued from page 5*

Other News: Chinese test taker tracking project Ying Zheng, Director of Research, and Glyn Jones, Senior Researcher, are conducting an internal research project aimed at tracking the progress of Chinese test takers. We have been contacting international students of Chinese origin who took PTE Academic and who are now studying at institutions in the UK and seeking their feedback regarding the challenges they face, from a linguistic perspective, as students in an English medium academic environment. By analyzing these data, together with reports from faculty members and objective measures such as essay grades, we aim to discover how far PTE Academic trait scores reflect the typical language skills profiles of Chinese students as well as their academic performance.

# Knowledge Technologies Update

## THE SPOKEN CHINESE TEST

by Alistair Van Moere, Ph.D., VP of Product & Test Development, & Masanori Suzuki, Principal Test Development Manager

Pearson's Knowledge Technologies group, the publisher of the Versant suite of language tests, is pleased to announce the completion of the automated Spoken Chinese Test. The project was a two-year development effort with Peking University, one of China's leading universities. The Spoken Chinese Test is delivered and scored automatically via a telephone or a computer delivery system. The test takes approximately 20 minutes and reports spoken Chinese proficiency on a granular scale of 20–80 for an overall score and for each of five analytic subskills: Grammar, Vocabulary, Fluency, Pronunciation, and Tone.

As a precursor to launching the test, Peking University and Pearson organized a meeting in Beijing on August 29, 2012, with a panel of eight Chinese experts from relevant academic disciplines (Chinese linguistics, Teaching Chinese as a Second Language, and Educational Measurement). They received presentations about the development and validation of the test, which included the following information:

- Reliability for the Overall test score is 0.97 (split-half method) and 0.95 (test-retest method).
- The automatically generated scores and human judgments correlate at 0.98 for the Overall score.
- The test correlates with two other well-established oral proficiency tests in Chinese at 0.85 (China's HSK Oral Intermediate level) and 0.86 (the US government's Oral Proficiency Interview test, or ILR OPI).

At the end of the meeting, the expert panel unanimously endorsed the Spoken Chinese Test and acknowledged its usefulness for the learning and assessment of Chinese as a second or foreign language.

With the introduction of Chinese, the Knowledge Technologies group can now deliver immediate, automatically scored language tests in six languages. The other five languages are: Arabic, Dutch, English, French, and Spanish. The Spoken Chinese Test will be available soon for sale outside of China; availability inside China is pending distributor arrangements.

# Pearson Assessment Community Updates

## TEST, MEASUREMENT, AND RESEARCH SERVICES CREATES NEW INTERNATIONAL ASSESSMENT UNIT

by Michael Young, Ph.D., Vice President Psychometrics and Research Services

Last July, Denny Way, Ph.D. and Senior Vice President of Measurement Services, announced a series of changes that would take place to make our organization more responsive to the evolving needs of our business. One such change was the creation of the new International Assessment Unit.

The new group has its origins in the pioneering work of the Pearson Assessment Community, or PAC. That group was formerly charged with developing a global community of interest in assessment, content, and project management expertise to assist our Pearson colleagues worldwide in creating assessments in countries where none had previously existed.

*Continued on next page*

*Pearson Assessment Community Updates, continued from page 6*

Many of you are still members of this vital community, as evidenced by your continued participation in the PAC Neo sites, your interest in global testing opportunities, and your nominations of qualified individuals to the Pearson Associates program.

What necessitated the creation of the International Assessment Unit within Test, Measurement and Research Services (TMRS) was a global realignment of the manner in which Pearson has decided to manage the growth of assessment activity around the world. In response to Pearson CEO Marjorie Scardino's emphasis on rapidly expanding opportunities in key regions such as sub-Saharan Africa, Latin America, and India, it was decided in January of this year to create an International Business Unit (the IBU), located in the UK and headed by **Peter Miller**. The purpose of the IBU is to coordinate the various growth opportunities in developing regions of the world, provide strategic guidance and support, and locate the most appropriate Pearson shared services worldwide to help fulfill programs.

TMRS and other Assessment & Information (A&I) shared services such as information technology play key roles in this new effort. As the largest organization in Pearson with assessment, technology, and content services, the A&I group will be providing under contract to Pearson UK and other regional units much of the support that they will need to both deliver assessments and develop their own capabilities in this realm.

To this end, **Dr. Michael J. Young** has taken on a new role as Vice President of International Assessment, reporting to Denny Way. Michael will work closely with Peter Miller's International Business Unit and **Sir Michael Barber's** Research and Innovation Centers. Michael will serve as a point of contact and bridge for TMRS efforts in support of international programs and opportunities, and will be directly involved in and lead the solution of complex psychometric needs. Michael has spent several months this past spring and summer in India and Australia, assisting Pearson colleagues with national assessments in those countries. As the new International Group evolves, it will be involving TMRS colleagues from all over North America in projects that support the IBU and Pearson.

# Food for Thought

## THE COLLEGE BOARD, ETS, AND READINESS FOR COLLEGE

by Tom Brooks, Ph.D., Senior Research Scientist

The material I am using for this article is based on a 1987 book (*The College Board and the School Curriculum*) written by John Valentine and published by The College Board (full citation appears at the end of this article). This is not intended as a review of the book, other than to say that the book is a very readable and informative discussion of the history of college entrance examinations in the 20th century (from 1900 to 1980). Valentine's book is a history of The College Board (aka College Entrance Examination Board). Full discussion of The College Board's history would of necessity include the history of the Scholastic Aptitude Test (SAT) and the Educational Testing Service (ETS) because of their close association. This essay, however, deals only with the founding of the College Board and its initial testing program. Since Valentine's book has as its central theme, stated in his Preface, the "conflict between willing acceptance and utter rejection of a standard-setting role in secondary school education," it also defines a historical context for Pearson's own efforts regarding the College Readiness movement, in which so much of the educational community is now involved.

The College Board was founded in November 1900 and gave its first round of examinations in June of the following year. Before that, the use of standardized college entrance exams or secondary school completion exams was non-existent. There was no framework within which colleges could cooperate with each other to define a common level of preparation for college, nor was there agreement on what subjects students should have studied or on the level of performance that they should demonstrate in those subjects. Secondary education involved public and private high schools and endowed academies. The private schools and academies were essentially all college prep. The public high schools served many purposes, including college preparation, but also included curriculum addressing more general education and vocational needs. In any case, what entrance exams that existed were specific to the institutions that administered them.

*Continued on next page*

*Food for Thought, continued from page 7*

In the absence of systematic entrance exams, there were several alternative approaches for certifying schools and/or students for having proper preparation for college. One of these alternatives, practiced at the University of Michigan, included the "accreditation" of high schools based on the judgment of university faculty members. For students from such accredited high schools, all that was necessary for admission to the university was a recommendation from their school; they did not have to take entrance exams that other students had to take. Many private colleges in the eastern United States accepted students on the basis of certificates issued by the secondary schools they had attended. Students who were certified in this manner were admitted without need of further documentation.

The impetus for moving toward more standardized and rigorous ways of selecting college applicants was provided by Charles William Eliot, then President of Harvard University, and Nicholas Murray Butler of Columbia University. Eliot was pushing for common entrance examinations as early as 1890, in a speech he gave to the National Education Association (NEA). Butler was largely responsible for an 1893 report from the Committee of Ten of NEA that laid out the principles that largely shaped the American high school curriculum. That document defined the nine traditional liberal arts subjects, to be included: chemistry, English, French, German, Greek, history, Latin, mathematics, and physics.

The work of creating the examinations started after the November 1900 founding meeting. The colleges and universities making up the membership of the board included Barnard, Bryn Mawr, Columbia, Cornell, Johns Hopkins, New York University, Pennsylvania, Rutgers, Swarthmore, Union College, Vassar, and Woman's College of Baltimore, plus five secondary school representatives. The board immediately set to work developing its first exam, set for administration in June 1901. Each subject had a "Chief Examiner," all of whom were college teachers. Then two associate examiners—one from a college and the other from a secondary school—were designated for each subject. An oversight committee (The Committee of Revision), consisting of the nine chief examiners and the five secondary school representatives of the board, was appointed to provide close review of all the examinations.

The tests were given at 60 sites during the third week of June, 1901. Supervisors were appointed for each site. Students registered to take the tests and paid their fees, readers were appointed to grade the answer forms, and provisions were made to mail the results to the candidates. Direct reporting of results to colleges did not come about until many years later.

A total of 978 students took the first examinations. Of these, 738 (75 percent) were Columbia applicants. In all, there were 39 readers (Valentine did not specify whether the papers were read by only one or by multiple readers). All the scoring was done at facilities provided by Columbia. Papers were scored on a 0 to 100 point scale, with scores of 60 to 69 regarded as "passing." The standards applied to the scoring were demanding, but varied from one test to another. Forty percent of all exams received failing (below 60) scores. More than 50 percent taking Greek failed; fewer than 30 percent failed math.

Despite the 100-year-plus difference between the first College Boards and now, much of the process is familiar. The tests were essentially tests of achievement of secondary school curriculum. The items were constructed response. Whether there were formal scoring rubrics or not, the readers went through a process of reading sample papers and discussing them to establish a consistent scoring standard. The scoring appeared to be very much the standards-based approach commonly used now. All that is missing are multiple-choice items, machine-scored answer sheets, and score reports sent directly to those who use them to make decisions.

Valentine, J. A. *The College Board and the School Curriculum*. New York: College Entrance Examination Board, 1987.

# Summer Inter Reflections

## JASON KOPP (INTERN IN SAN ANTONIO, TX)

My summer internship with Pearson has been an educational and rewarding experience. I had the opportunity to work with a number of knowledgeable and helpful research scientists at the San Antonio office. I was able to learn new research techniques and operational processes, and get a taste for a career in psychometric research. I received superb mentoring from my intern supervisors, Dr. Allen Lau and Dr. Michael Young. Their guidance made my time at Pearson very enjoyable and instructive.

During my time in San Antonio, I focused primarily on applying diagnostic models to mathematics testing data. These models involve giving information on specific cognitive skills to students and teachers in order to shape learning and instruction. The project involved data management and compiling as well as writing code and specifying models. Much of the project involved thinking critically about methods to modify the modeling process to accommodate operational testing data. Michael and Allen, as well as the other research scientists in Test, Measurement and Research Services, were instrumental in this process. They contributed their ample skills, knowledge, and experience in solving various challenges. Overall, the project allowed me to apply my training in an operational environment with supportive colleagues.

I found the weekly summer intern seminars extremely enlightening. Every week, the summer interns attended seminars by various research scientists presenting the innovative research being developed at Pearson. This research distinguished itself by being focused primarily on improving testing for teachers and students. I was impressed by the immense knowledge of the speakers in the seminars. The series culminated with all of the interns presenting their work from the summer. This allowed us to share our experiences and see the excellent work that was being done by the other interns. I want to give special thanks to Dr. Hua Wei for organizing this outstanding seminar series.

Overall, the internship was a deeply worthwhile experience. I want to give my thanks again to Michael and Allen, and to all of the research scientists and staff who helped me with my project and helped make this internship possible.

## DANIEL JURICH (INTERN IN TULSA, OK)

By offering a glimpse into the operational world of K-12 educational testing, the Pearson internship proved to be an enlightening experience. In my brief two months at the PRS office in Tulsa, Oklahoma, I was able to partake in a wide variety of projects and tasks that strengthened my psychometric skill set and, perhaps more importantly, allowed me to obtain a greater understanding of the field. Before detailing my specific experiences at Pearson, I would like to acknowledge the support I received from the Tulsa research scientists and associates. In particular, I must thank Dr. Stephen Murphy and Dr. Mike Clark for serving as my mentors, providing invaluable guidance, and making me feel like an integral part of the Tulsa office.

For the first major project of my internship, I served as a data analyst for the Oklahoma Alternate Assessment Program (OAAP) standard setting in Oklahoma City. Being involved in a standard-setting process for a large-scale high-stakes test was a unique and eye-opening experience. In addition to serving as a data analyst, I had the opportunity to observe Pearson employees facilitate the standard-setting process with Oklahoma school teachers. Being a part of this process was particularly educational as the facilitators used a combination of two cutting-edge standard-setting methods: reasoned judgment and body of work. Thus, I was able to not only see how standard settings are applied in a large-scale environment, but also expand my repertoire of standard-setting techniques.

When I returned to the Tulsa office, my time was primarily spent assisting a test security initiative, researching state policy for test security, and conducting a study on cheating detection methods. As my master's thesis dealt with effects of cheating on equating procedures, this work felt like a natural continuation of my research agenda and was relevant to my personal interests. Because the focus of my previous research had primarily been statistical in nature, scrutinizing state test security policies was extremely informative, giving me perspective on what methods states have in place to prevent and deter cheating on educational tests. The technical research component dealt with examining the robustness of answer similarity detection methods to false positives. This project exposed me to a variety of answer similarity detection techniques and has required me to think critically about both statistical and political issues associated with identifying cheaters. This project is currently submitted to

*Continued on next page*

*Daniel Jurich (Intern in Tulsa, OK), continued from page 9*

NCME as part of a coordinated symposium with projects from various other educational testing companies.

Overall, this summer internship experience was an invaluable look into the K-12 educational testing field. Through the internship I gained both technical and practical skills that will serve me well in finishing my degree and entering the field. In closing, I am grateful for this wonderful opportunity and I look forward to continuing relationships with my colleagues at the Tulsa office.

## AMANDA SOTO (INTERN IN AUSTIN, TX)

When applying for internships this summer, my primary goal was to broaden and enhance my academic experiences by learning about operations and day-to-day management of a testing program. My time with Pearson in the Austin office provided invaluable insight into the way that research and operations blend together to refine and improve testing. I was fortunate to work with talented (and patient!) individuals on a range of projects during the psychometric fellowship. Overall, my internship provided an enriching, challenging, and inspiring experience. During the eight-week internship, I was involved in several projects. Through these I learned about the modified and alternative assessments offered to students in Texas, the political demands on a state testing program, the challenge of being accountable to different stakeholders, and the operational demands of second-language testing as well as developing and testing linguistic accommodations for non-English speakers.

My work struck a balance between research and practice in writing a review of the literature on the development of learning progressions and possible validation methods for learning progressions/trajectories. This project examined the different learning progressions designed to complement the Common Core State Standards. Learning progressions describe hypothesized trajectories through related concepts within a subject area (like number sense within mathematics). By establishing a pathway that places fundamental concepts before more complicated variations, a student's level of mastery can theoretically be assessed with greater precision. Though learning progressions are becoming more popular in practice, there have been few research studies on their validity or reliability for shaping instruction or assessment. This is an increasingly important area of study as states align their curricula to the Common Core Standards and develop related assessments.

I also participated in a study of the effect of linguistic accommodations for English speakers and learners. This involved large-scale data management and programming in SAS—tremendous learning opportunity for me as well as a chance to collaborate with colleagues at Pearson. Examining the effect of accommodations at the item level (on both World History and Geometry tests) provided interesting insights into which items were more challenging for less proficient students as well as which accommodations were more helpful for students learning English. This study served to answer the overarching question of which accommodations helped students with limited English proficiency to access the content without impacting the performance of proficient speakers. This study was unique in that it compared the performance of English speakers to learners; in this area it is difficult to gather large samples from both groups to test the effect of accommodations in a controlled way. Participating in this work was a tremendous learning experience for me and hopefully made an important contribution to the ongoing development of linguistic accommodations and dual-language testing in Texas.

This summer I was fortunate to learn from some tremendously gifted psychometricians and researchers at Pearson. My mentors Sonya Powers and Laurie Davis both went out of their way to welcome me to the office and to provide ongoing support during my internship. I also enjoyed collaborating on a paper proposal and research project with Melinda Taylor, whose intelligence and good humor were invaluable. Katie McClarty, Ha Phan, Phyllis Garrett, Matt Gaertner, and Leslie Keng all helped and guided me this summer. It was a privilege to have access to their ideas and support as I learned about the Texas testing program.

## YUN TANG (INTERN IN IOWA CITY, IA)

During In my application narrative essay, I wrote that I expected that the internship program would give me a platform to let me strengthen my psychometric knowledge and skills, experience day-to-day testing operations, and learn more about the current developments and practical issues related to educational assessment. Ultimately, I aspire to pursue a career as a psychometrician in an educational testing organization. As I planned to complete my dissertation in the summer of 2013, I wanted to use this invaluable opportunity to be better prepared for my career pursuits. Now, looking back at the past eight weeks, I can confidently say that this experience rewarded me beyond my expectations.

*Continued on next page*

*Yun Tang (Intern in Iowa City, IA), continued from page 11*

First, I was very fortunate to have the opportunity to participate in the operational projects of Pearson. During this internship, I have primarily worked on the Virginia Modified Achievement Standards Test (VMAST) project under the guidance of my mentors, Tony Thompson, James Ingrisone, and Xia Mao. This project examined the technical adequacy of VMAST and compared the performance of students on VMAST against that for the corresponding statewide general assessment. With the help of my mentors, I drafted a technical report and developed a NCME submission. I gained invaluable operational skills, learned how to work with large-scale assessment datasets, and improved my SAS programming skills. I also gained further insight into different aspects of operational activities needed to construct quality K-12 assessments. Most importantly, I came to appreciate the unparalleled importance of "QC" (quality control) in the operational practices at Pearson.

One of the highlights of my internship experience was the chance to interact with and learn from a group of excellent research scientists and statistical analysts. Through valuable informal exchanges, I learned about their roles within the company, received advice regarding my career development, and expanded my knowledge of psychometric services. Another benefit of this internship was the weekly seminars given by the Psychometrics and Research Services professionals. Seminar topics included computerized adaptive testing, technology-enhanced items, building SAS macros, student growth percentile models, and much more. These seminars gave interns further insight into the research of a large testing company like Pearson, and they updated my own knowledge of the important current issues in educational assessment.

I left the program with cherished memories and with more confidence and desire to pursue my career goals. My internship could not have been more successful, and I am grateful for the amazing support I received from everyone. Thank you. I would like to express a very special thanks to my mentors for giving me an excellent opportunity to work closely with them and learn. They always treated me as a trusted colleague, letting me resolve difficulties when I encountered them and explore solutions by myself. When I sought help, they always worked around their busy schedules to make time for me. In addition, they shared their own experiences working in the field and provided professional guidance as I conducted research and planned my own career path.

Finally, I want to express my gratitude to my manager, Julie Miles, and the entire PRS staff at the Iowa City office. They made me feel welcome from the very first day and always offered their gracious support and friendship. I am especially grateful to Yuehmei Chien, Xia Mao, David Shin, Tian Song, and April Su for helping me adjust to life in Iowa City, and to the internship program coordinator, Hua Wei, who wonderfully cared for my internship needs. I feel privileged to be part of this rewarding fellowship program, and I would recommend it to any student who aspires to work in the testing industry.

# Publications

**Dietrich, C.B., Wolfe, E.W., & Vanhoy, G.M.** (2012). Cognitive radio testing using psychometric approaches: applicability and proof of concept study, *Analog Integrated Circuits and Signal Processing, 72,* 1-10.

**He, W., & Wolfe, E.W.** (2012). Treatment of not-administered items on individually administered intelligence tests. *Educational and Psychological Measurement, 72,* 808-826.

**O'Malley, K.J., Lai, E., McClarty, K., & Way, W.** (in press). *Marrying formative, periodic, and summative assessments: I do.* In R. Lissitz (Ed.) Informing the practice of teaching using formative and interim assessment: A systems approach. Charlotte, North Carolina: Information Age Publishing.

**Pae, Hye K.** (2012). Convergence and discriminant: assessing multiple traits using multiple methods. *Educational Research and Evaluation, 18(6),* 571-596.

**Tsai, T. H., & Shin, C. D.** (2012). A Score Comparability Study for the NBDHE: Paper-Pencil Versus Computer Versions. *Evaluation & the Health Professions* published online 22 May 2012, DOI: 10.1177/0163278712445203

**Van Moere, A.** (2012). A psycholinguistic approach to oral language assessment. *Language Testing, 29(3)*, 325-344.

**Wolfe, E.W., & McVay, A.** (2012). Applications of latent trait models to identifying substantively interesting raters. *Educational Measurement: Issues and Practices, 31,* 31-37.

# Conference Participation

## Annual Meeting of the International Association of Teachers of English as a Foreign Language,

Prague, Czech Republic, October 2012

**De Jong, J.H.A.L.,** Extending the CEF and its usability.

## Association for Language Testing and Assessment of Australia and New Zealand,

Sydney, Australia, November 2012

**Jones, G.**, Are we on track? A study of Chinese test takers.

## California Educational Research Association Conference,

Monterey, CA, November 2012

**Frantz, R.,** Operationalizing academic language in English language proficiency assessments.

**McClarty, K., & Gaertner, M.,** Evidence based standard setting: Using empirical evidence to set college ready cut scores.

**O'Malley, K., & Starr, L.,** Top ten ways for better reporting for English learners.

**Seo, D., & Taherbhai, H.,** Student growth percentiles as a formative tool in assessing English learners' progress in language acquisition.

## International Meeting of the Psychometric Society,

Lincoln, NE, July 2012

**Chien, Y. & Shin, C. D.,** A Recursive Algorithm for IRT Observed Score Equating.

## INTERSPEECH,

Portland, OR, September 2012

**Cheng, J.,** Automatic Tone Assessment of Non-Native Mandarin Speakers.

**Ivanov, A., & Chen, X.,** Modulation Spectrum Analysis for Speaker Personality Trait Recognition.

# Announcements

**Julie Miles, Ph.D.**, Vice President, has become the President-Elect 2013 for the Iowa Educational Research and Evaluation Association (IEREA).

**Mat Gaertner, Ph.D.**, Research Scientist, is joining the advisory board for the journal *Educational Measurement: Issues and Practice* (EM:IP) starting in spring 2013.

The Research and Development team in Pearson's Knowledge Technologies group has won a speech recognition competition at the 2012 INTERSPEECH conference.