Welcome to the summer edition of Pearson’s Research Newsletter, a quarterly publication for discussing prominent topics in education and for sharing research accomplishments across Pearson. I would like to thank our writers for their contributions. In addition, I would like to thank the members of the newsletter advisory board for helping identify and secure content for this issue. I recognize and appreciate that they must take time out of their busy schedules to make this newsletter a reality.

In this issue, we start with an exciting discussion of standardized testing merits, which is followed by reports on research being performed on computerized adaptive testing, automated test assembly, and grading systems. In addition, we have a very enlightening piece on career readiness, and news from the Pearson Center for Applied Psychometric Research at the University of Texas at Austin, the Language Testing Group, and the Pearson Assessment Community. Finally, this issue also introduces the 2012 summer interns and provides the schedule for the Summer Intern Seminars Series.

I would like once again to thank you for your contributions and to encourage you to send us your comments, suggestions, and thoughts. If you would like to provide a discussion topic or an essay for one of the future issues, please contact me about it.

We aim for widespread dissemination of this newsletter. If you or someone you know outside Pearson would like to be included in our distribution list, please follow the link to sign up or send me an e-mail. (http://www.pearsonassessments.com/hai/forms/CRS.aspx). Back issues can be downloaded from http://education.pearsonassessments.com/pai/ai/research/resources/TMRS_Newsletters?ResourceId=newletter2_4.

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

“STANDARDIZED TESTING” ARE NOT BAD WORDS, NO MATTER HOW YOU SAY THEM

by Kimberly O’Malley, Ph.D.

My two sons, Jace (9 yrs.) and Luke (6 yrs.), have always loved words. They question me about what words mean, how they are spelled, and ways they are used. Their newest obsession is about “bad” words. They hear them in songs and see them on TV, and they have taken to asking me about them. It is not uncommon for them to come to me, shoulder to shoulder, and nudge each other to encourage the other to ask me questions about whether a word is “bad” and whether or not they are allowed to say it. Recently, for example, Luke—who must have drawn the short straw—came to me and asked, “If we say the word ‘hello’ and do not say the ‘o’ sound at the end, or if we say it too quietly for people to hear, is it a bad word? Will it get us in trouble?”

Trying not to smile or laugh, my response was that they were not to say the word “hello” with a silent “o” and that saying a bad word is inappropriate. I went on to say that the word “hello” is a great word. It denotes a greeting that opens the door to learning about other people. It welcomes people. It is a universal concept with similar words in many other languages. Disappointed that they did not get more of a rise from me, they went off to debate words like “damage,” “scrap,” and “glasses.” It goes without saying that these conversations are not those that rank in the “pride mom moments.” But curious kids will be curious kids.

It seems my sons are not the only ones questioning whether good words might be turned into bad words. Take the words “standardized testing.” Many around the nation have been debating whether these have become bad words. A quick Internet search brings up articles titled “Standardized Testing as Child Abuse” (San Diego City BEAT, 2012), “Standardized Testing and Its Victims” (Education Week, 2000), and “Stop the Standardized Test Tyranny” (BusinessWeek.com Debate Room, 2009).

Though most aspects of standardized testing deserve debate—such as how standardized tests are designed, the extent to which they cover the curriculum, the costs, and the uses of the scores—it is shortsighted to define the activity itself as a bad thing and put standardized testing in the bad word category. As a parent, former teacher, and psychometrician who currently develops assessments and conducts research on test score uses, I welcome an open discussion of the issues around standardized assessment.

The concerns around standardized testing are well publicized. The advantages, however, are not being discussed much, so the debate has become one-sided (Cizek, 2001). To balance the debate, some have proposed positive aspects of standardized testing (Cizek, 2001; Phelps, 1997, 1998, 2005). In his 2001 article in Educational Measurement: Issues and Practice titled “More Unintended Consequences of High-Stakes Testing,” Cizek describes and supports ten unintended positive outcomes of increased standardized testing. As a parent, former teacher, and measurement specialist, I have also witnessed positive aspects of standardized testing. I share three reasons that reflect my parental perspective for why “standardized tests” should not be considered bad words.

1) A lens into the classroom—Standardized testing provides a view, a lens if you will, into what happened over the approximately 180 school days that our children learned and what our teachers taught. The knowledge and skills that students are supposed to learn, the content standards, are published before the school year begins. Standardized testing provides information about how our children performed on that knowledge and those skills after a year of instruction. Parents do not have much visibility into what happens each day our children are in school. We hear updates from our children, visit the classroom on party or presentation days, and receive teacher reports, but the view we have into the day-to-day learning can be murky. Standardized testing enriches our understanding about how our children learned during those approximately 180 school days.

2) External verification—Without standardized test results, parents and administrators must rely solely on grades and reports from the teacher to gauge a student’s progress. Though grades and teacher reports are valuable sources of information, they vary from teacher to teacher.

Continued on next page
The quality of work that earns an A from one teacher may not be the same as that from another teacher. Standardized testing by its nature means that the tests are given in a consistent manner from student to student and school to school. The standard way that the tests are given and designed allows score information to be interpreted the same way for all students who take the test. Parents do not have to interpret a score based on the particular teacher a student had or what questions appeared on the test. Instead scores can be interpreted similarly for all students. In addition, the information teachers send home tends to focus on specific skills or sets of skills. That information does not directly represent student performance against a representative sample of the content standards. Standardized test scores provide that external verification of how students are performing on a state’s set of predefined knowledge and skills. The information from the standardized tests offers parents, teachers, and administrators a view into how much students learned that year that is not dependent on the teacher’s way of grading.

3) An additional source of information—Standardized testing is not supposed to provide the only indicator about how students performed during the year. It is designed to be one of many sources of information that, taken together, describe student performance against the expected grade-level knowledge and skills. The optimal way for standardized test scores to be used is as an additional piece of information that can be combined with all other information provided about a student during the school year. By combining standardized test scores with grades, scores on daily practice activities, reports from the student, and teacher observations, students, parents, teachers, and administrators can more comprehensively understand where a student started, the progress the student made, and how well the student learned the information expected that year.

The future our children and our students face and the world in which they will compete for jobs will not be the world we live in today. The world is changing rapidly, and competition for jobs and resources in the future will be global and dramatically more competitive than today. To best prepare our children and our students for a globally competitive world, we need to set high expectations, define the knowledge and skills needed to meet those high expectations, understand how well students are learning today, compare our student performance nationally and internationally, and provide them with learning environments that stretch and challenge them.

Standardized testing does not do all of those things. But, because standardized tests are standard and not dependent on the specific coursework, teacher, or classroom environment in which the student learned, they offer interpretations and comparative information that is not possible with coursework or grades tied to a particular teacher or classroom. Standardized tests scores are key to understanding how well our students are learning today, the progress they are making toward college and career, and how well the performance of our students compares with that of students across the nation and world.

Used appropriately, standardized testing offers a lens into the classroom: external verification for those of us not in the classroom each day that our students are progressing toward the goal of college and career, and an additional source of information that can be combined with other sources to enrich our understanding of student progress. Unlike the word “hello” that my boys tried to make a bad word, the words “standardized testing” are not bad words, no matter how you say them.


Research in Action

LESSONS LEARNED FROM OUT-OF-LEVEL TESTING IN COMPUTERIZED ADAPTIVE TESTING RESEARCH
by Jie Lin, Ph.D., and Hua Wei, Ph.D.

The purpose of the study was to investigate how the use of out-of-level testing affects measurement accuracy and test efficiency in computerized adaptive testing. Out-of-level testing refers to administering a test above or below a student’s grade level. Although on-level items are adequate to correctly classify the highest- and lowest-achieving students, they often fail to accurately assess what those students know and are able to do or to appropriately gauge their growth from year to year. Out-of-level testing continues to be a controversial topic, especially in the context of standard-based and accountability assessment programs. Our study simulated student responses to computerized adaptive tests with either fixed or variable lengths, and compared the performance of on-level and out-of-level testing for accuracy of ability estimation and number of items required to arrive at a specified level of measurement accuracy. The test items were selected from a vertically scaled empirical mathematics item bank for grades 3–8.

Results from our study showed that allowing out-of-level items for administration generally produced lower standard errors of measurement and bias and higher correlations between true and estimated abilities. Including lower- and upper-level items significantly improved measurement accuracy and resulted in much shorter tests for the highest 10% and lowest 10% of students at each studied grade level. The study provided supportive evidence to increase the accuracy of ability estimation for higher- and lower-performing students by including out-of-level items in computerized adaptive testing. Promoting the alignment between teaching, learning, and assessment can inform the selection of out-of-level items to maintain assessment validity and to fortify the advantages of computerized adaptive testing applications.

NEW METHOD FOR AUTOMATED TEST CONSTRUCTION: THE TWO-PHASE METHOD
by Gerald Griph, Ph.D.

My primary research interest lies in the area of computer-assisted test assembly and revision, which subsumes that of automated test assembly (ATA). The title of this essay (and upcoming paper) emphasizes both the name of the algorithm/heuristic that I’ve developed and the fact that this is different from other commonly used ATA methods.

Currently, the most common methods being used for ATA are those based on mixed integer programming (MIP). MIP is able to find the optimal form that can be constructed from a given pool of items meeting a given set of constraints. The form will be “optimal” in the sense that it has the minimum or maximum possible value for some mathematical or statistical property of the form (for example, maximize the value of the form’s IRT test information function at the test cut points; this is termed the “objective function”). While this is a valuable approach to ATA, specialized MIP “solver” software is required to carry out this analysis. As is common with most things technical, the good software costs thousands of dollars to license while the free software is very slow. Additionally, by definition there is only a single “optimal” form, and so constructing a given form from a given set of items with a given set of constraints multiple times will always result in the same optimal form.

As a solution, the two-phase (2P) method does not require specialized software. It is currently implemented as a standalone Visual Basic .NET program and will be ported into Pearson’s systems at some point in the future. In addition, it is fast—a moderately complex passage-based test can be constructed in around 10–20 seconds, while a simple test composed of discrete independent items typically takes only a few seconds to build. Also, the 2P method builds locally optimal test forms, meaning that it selects each item from the pool that maximizes/minimizes the form’s objective function compared to its current value. Since a pool of items will have many possible locally optimal forms, constructing a given form from a given set of items, with a given set of constraints multiple times will almost certainly NOT result in the same optimal form. If the test author’s idea of a “good” form doesn’t correspond to that constructed by the software, another run of the software will produce a different locally optimal form that may be more acceptable. Finally, the 2P method easily extends from ATA to computer-assisted test form revision. The
data structures used in implementing the 2P method can help identify all possible replacements for an item or item set flagged for removal from the current form, ranking them in order of their statistical similarity to the item or item set being replaced. Thus, the author is able to see all available options and select the item that best meets the content requirements associated with the replacement while best preserving the current match to the form’s psychometric targets.

This study compared forms produced by the 2P method and the MIP-based approach. Two open-source MIP software packages were used—the GNU Linear Programming Kit (GLPK) and LPSolve. An item pool for a math computerized adaptive test (CAT) that Pearson is developing using Rasch IRT and a pool composed of synthetic items with randomly generated 3-PL IRT item parameters were used. The results indicated that both the MIP- and 2P-based methods were able to construct forms that similarly met the content blueprints. The single optimal form produced by MIP had a closer match to the psychometric targets than any individual form produced by the 2P method. However the 2P method produced multiple forms that were comparable in their match to their psychometric targets and were close to the match produced via MIP.

The software implementation of the 2P method has been used to build forms for the last few years for several programs. Experience with those programs has been used to improve the software in preparation for integrating it into Pearson’s systems. It currently supports building single forms to specific content, item type, and psychometric targets for tests composed of discrete items or passage-based item sets. After the draft form has been built, the software allows the test author to add, remove, or replace individual items or passage-based item sets. When replacing items/item sets, the software will identify all possible items/item sets that could be swapped onto the form without upsetting the current match to the blueprint. Items are sorted according to their psychometric similarity to the item being replaced, thus allowing the author to know which items will perturb the current match to the psychometric targets least.

Computer-assisted test forms revision has proven to be almost more important than the ATA capabilities of the software in terms of improving the test construction process. An MIP-based solution would almost certainly have been much less capable of being extended to computer-assisted forms revision. Future directions for research and extension of the 2P method include building multiple forms simultaneously, item pool analysis, and application to a fixed length content-constrained CAT.

GRADING SYSTEMS
by Jim Tognolini, Ph.D.

Most education systems have moved from a norm-referenced (grading on a curve) to a standards-referenced system of reporting their educational outcomes. Instead of a fixed proportion of a cohort being assigned the top grade, judgments need to be made about where to place the cut-score on a distribution of marks to indicate achievement of the required grade standard for each grade level awarded. A grade is only assigned to those students who have demonstrated the standards criteria for the grade.

Systems differ in the scope of their standards, ranging from narrow definitions of behavior (as in the United States) to more holistic descriptors of performance (as in the United Kingdom). The potential downside of standards-referenced reporting is that there is often subtle and or more explicit pressure on the judgment process to inflate student achievement. Grade inflation is often claimed to be occurring in many systems. Jim Tognolini (Senior V-P Pearson and Senior Research Fellow at Oxford University) and Gordon Stanley (Pearson Consultant and Senior Research Fellow at Oxford University) have been studying quality assurance processes and measures to validate whether changes in the distributions of results with standards-based reporting of grades are real or inflated.
University Collaborations

THE PEARSON CENTER FOR APPLIED PSYCHOMETRIC RESEARCH AT THE UNIVERSITY OF TEXAS AT AUSTIN

by Barbara Dodd, Ph.D.

In this piece, I would like to briefly highlight the accomplishments of the UT Austin Pearson Center during the third year of the five-year endowment period. Since its creation, the Center has hosted several highly renowned researchers. For example, in September 2011, we hosted Joshua McGrane as a visiting scholar from the Pearson Center Laboratory at the University of Western Australia. Dr. McGrane gave a presentation entitled “Empirically reconciling cumulative and unfolding item response models” and discussed research with many faculty and students during his two-week visit. In addition to visiting scholars, the center hosts the Distinguished Lecture Series, which provides attendees with state-of-the-art information on various topics given by renowned researchers in the field. In the spring of 2012, the series featured Wayne Camara, who presented “Defining and measuring college and career readiness: Establishing validation evidence to support the use of new assessments.” The talk was well attended by college faculty and students as well as by Pearson employees from the Austin and Iowa offices. Pictures from the event can be viewed on the UT Pearson Center Web site: http://www.edb.utexas.edu/UTPearson/lecture_series.html. The Web site for the UT Pearson Center was redesigned this year. Please see http://www.edb.utexas.edu/UTPearson/ for the latest version.

The Pearson Research Center at UT has greatly impacted the training of graduate students for a career in research as well as for applied work in industry. Under supervision of the Center, our graduate students have analyzed a new placement test for college calculus for the College of Natural Sciences, and as a result a technical report was given to that college to assist it in the development of its test. In addition, staff from the College of Education’s Institute for Public School Initiatives and staff from the School of Undergraduate Studies also have received assistance with various applied psychometric research issues. It is deserving of notice that the services provided by the UT Pearson Center include work external to university. For example, Wanchen Chang, a graduate research assistant, has worked with Pearson’s staff on various validity studies for the Texas educational assessment contract. In addition, the Center’s services are projected to continue expanding as the end of the endowment phase approaches.

Finally, the Center was well represented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education this past April in Vancouver, British Columbia. A list of those presentations appears below. (Note that four of the six papers are first-authored student papers.)


Language Testing Updates

THE PEARSON SCALE OF ENGLISH

by Kirsten Ackermann

The Pearson Scale of English (PSE) was originally developed as the reporting scale for marking the Pearson Test of English Academic (PTE Academic). It is a numeric scale that ranges from 10 to 90. The PSE has been aligned to the Common European Framework of References for Languages (CEFR), as this is an internationally recognized standard. Each CEFR level is thus equivalent to a range of values on the PSE, and each PSE value will be meaningful to the widest possible audience.

The PSE has been empirically validated in the course of the PTE Academic Field Tests and is being underpinned by a set of descriptors that are more highly specified than the CEFR. Ultimately, the PSE has the potential to replace the CEFR in some domains and applications.

The PSE has since not only been used for other Pearson Tests of English, such as PTE General, but Pearson has fully embraced the idea of establishing the PSE as a global benchmark for English achievement. As part of Pearson’s efficacy objective outlined by Dame Marjorie Scardino in her New Year’s message, the PSE is being related to a wide range of tests, examinations, and course books, as well as to the levels of English attainment needed for various professional and academic purposes. These initiatives will lead to a new Pearson Syllabus aligned to the PSE providing a unified structure for instruction and assessment that will fill a huge unmet market need, and which none of our competitors can match.

In Bill Anderson’s words, this will enable us to offer “an end-to-end service/solution of diagnose, deploy, test and certify which will be difficult for any current competitor to match, and which will change the nature of the competitive discourse in the institutional ELT market.”

For more information on the PSE, contact Prof. Dr. John H.A.L. de Jong, Senior Vice President at john.dejong@pearson.com

PEARSON EXTERNAL RESEARCH PROGRAM 2012

Language Testing invited researchers in the field of language testing and assessment, second/foreign language education, linguistics, applied linguistics, and other related areas of research to participate in our 2012 external research program.

On behalf of the 2012 Pearson External Research Program Committee, we are delighted to announce that three proposals have been selected and will be funded as part of this year’s program.

(a) Faculty research - Treffers-Daller, J., Prof., Williams, S., Prof., Parslow, P., Slaght, J., & Brewer, B., PhD, University of Reading, UK.

Automated assessment of lexical diversity and n-grams in essays of test takers of PTE Academic — This is an interdisciplinary project in which an applied linguist, engineers, and specialists of EAP from the University of Reading, UK work together to obtain a comprehensive picture of the vocabulary found in essays of PTE Academic test takers, using a wide range of new technological tools for automated data analysis. The overall aim of the project is to find out how automated analyses of vocabulary usage in essays of English learners can further our understanding of students’ achievement at different levels of PTE Academic. In addition, the project will investigate the extent to which vocabulary used in academic writing differs from that found in the Pearson International Corpus of Academic English (PICAE) and will make use of the Academic Collocation List (ACL) in the analysis of academic writing.

(b) Doctoral student research - Ma, J., PhD candidate, Queen’s University, Canada

The effects of test preparation on Chinese test takers’ PTE Academic performance, English language proficiency, and motivation of English language learning — This is part of a bigger project sponsored by the Canadian federal government. The study aims to examine the effects of test preparation courses on PTE Academic Chinese test takers’ performance, their self-perceived English language proficiency, and their motivation of English language learning. It will also provide evidence on washback effects of PTE Academic on English language learning.

Continued on next page
Language Testing Updates, continued from page 7

(c) Masters student research - Palani, M., MA candidate, Warwick University, UK

Investigating the predictive validity of PTE Academic listening for understanding university lectures—This project aims to investigate how effective the PTE Academic Listening test can predict the ability of students to comprehend daily university lectures. The results of the project can add evidence for predictive validity of PTE Academic.

PAC News

THE ROLE OF TECHNOLOGY AND ASSESSMENT IN SYSTEM-WIDE IMPROVEMENT – FREMANTLE, AUSTRALIA

The 2012 Pearson Global conference was held in Fremantle, Western Australia at one of the four funded university research centers around the world. The conference program integrated global assessment themes and Australian educational policy and research initiatives, and the general theme was the emerging role of technology at system-wide levels and its integration within assessment, teaching, and learning pedagogies.

Four keynote addresses explored the importance of worldwide educational system change and how assessment and technology are key drivers. Sir Michael Barber, Pearson’s Chief Education Advisor, spoke about the challenge of transformation and reform in education. Reinforcing transformation of educational systems around the world is a prerequisite to meeting major global challenges and dealing with the increasingly rapid pace of social and economic change. His talk outlined the key features of successful educational systems—high levels of standards and accountability, significant investment in developing human capital, and effective and efficient educational structures and organizations.

The second keynote speech was given by Professor Gordon Stanley, who focused on the integration of technology into global examination systems. He described how there have been a number of false dawns over time regarding the integration of technology into examination systems, in part due to risk aversion and in part due to lack of infrastructure. However, assessment is rapidly reaching a tipping point in terms of the rising scale of assessments and accountability systems, the need for system efficiencies, the requirements to assess broader ranges of knowledge and skills, the efficiencies of automated marking technologies, and the diagnostic power of data-capture systems.

The third keynote speech was given by Dr. Peter Hill, CEO of the Australian Curriculum and Assessment Reporting Authority (ACARA), and addressed evolutionary and revolutionary reform and thinking at a national level. More specifically, Dr. Hill spoke about how Australia is working toward creating a national curriculum and online testing program across multiple curriculum areas and multiple year groups. In this new program, teachers will have access to rich curriculum materials, students will be tracked online throughout their schooling, and intervention, remediation, and extension materials will be linked into diagnostic analysis of assessment data.

Greg Whitby, executive director for the Catholic Diocese of Parramatta, gave the final keynote speech, which focused on the importance of teachers in educational transformation. His speech highlighted the fact that rich technological learning and assessment tools must be combined with teachers’ expert use of informed cognitive learning approaches to transform student outcomes. In such a way, schools must become a place for student learning and teachers’ professional development.

Combining these different but cohesive keynote addresses, and the various talks that took place during the meeting, the Pearson Global Research conference offered a broad range of insights into how innovations in technology and assessment can enhance learning.
Food for Thought

THE LEFT SIDE OF THE
(COLLEGE READINESS) EQUATION
by Tom Brooks, Ph.D.

The Unless you have spent the past few years on a deserted Pacific island talking to a volleyball, you have heard people talking about college readiness as a big deal for America in general and for elementary and secondary education in particular. Although there is a broad consensus that American high school students currently are not being adequately prepared to succeed in higher education, there is also a lot of disagreement about just what college education is and what it should be. What is defined as a college education includes not only four-year colleges, with or without accompanying graduate programs, but also two-year community colleges, which may or may not be stepping stones to four-year undergraduate degrees. Is it necessary that two-year college studies need to fit hand-in glove with four-year college studies? We might even ask, “Why college?” Is the real goal to prepare all students for college or to prepare them for the world of work? What, specifically, do high school graduates need to know and do to thrive in the modern world and to meet whatever new needs may develop over time?

The issue of college readiness has been part of a general education reform movement since the late 1950s. After the Soviet Union launched Sputnik, there was a lot of concern that the United States had fallen behind in education and technology, prompting President Dwight D. Eisenhower to speak of a need for national goals and standards in education. In modern parlance, the idea didn’t gain any traction. Presidents Richard Nixon and George H. W. Bush also attempted to establish national education standards, also to no avail.

President Bill Clinton’s administration saw the release of voluntary national standards, as well as the beginning of efforts to develop national tests at selected grade levels. This effort, too, came to naught, collapsing after former head of the National Endowment for the Humanities Lynne V. Cheney (yes, that Cheney) severely criticized a draft of the history standards. Underlying all these setbacks was the American tradition of local control of schools and concern that national standards could lead to federal control of schools.

Finally, a number of factors came together to create an environment in which national standards became possible. First was the movement in the 1990s toward standards-based education and assessment. This culminated in the No Child Left Behind Act of 2002 and the legal requirement that each state develop its own standards and measure the “proficiency” of its students by the 2013–14 school year.

Much of the impetus for bringing college readiness standards to the forefront came from the Bill and Melinda Gates Foundation, which supported the work of David Conley. This body of work culminated in the publication of a comprehensive conceptual definition of college readiness, embodied in the Redefining College Readiness series of reports. There are also publications associated with the ACT and SAT that provide conceptual definitions of college readiness consonant with the respective publishers’ visions about college readiness. Finally, and potentially the most influential of all these efforts, is the joint efforts of the National Governors Association and the Council of Chief State School Officers. As part of the Common Core Standards that accompany the No Child Left Behind Act, this group produced comprehensive standards for college readiness appropriate for different levels of elementary and secondary education across content areas.

So, we have comprehensive standards and two national college entrance exams (ACT and SAT) supposedly defining what it means to be ready for college. And whether these ultimately become the operational definition of readiness for college or whether some completely new measures are created, a whole lot of thought and work have gone into defining college readiness as thoroughly defined independent variables; in other words, essentially the right side of an immense, all-encompassing, regression equation. But what about the left side of the equation, the dependent variables?

Frankly, if one were a faculty member at a postsecondary institution, one might not know that any of this dialogue was happening at all. Just about everything there is to read about college readiness is written from the standpoint of what is missing from, or needs to be added to, elementary and secondary education to ensure that graduating students are ready.
Food for Thought, continued from page 9

to perform at acceptable levels in college. Although this certainly seems like a good place to start, it is also true that the need to improve college readiness comes from, and is ultimately defined by, the expectations of postsecondary institutions themselves.

The response of the colleges and universities is typically to provide placement tests and remedial coursework for entering students who don’t pass them. Entering students take a test or tests to determine the appropriate course level. Those who don’t get a score that makes them eligible for entry-level courses (i.e., they are “underprepared”) are required to enroll in non-credit remedial course(s), which they must pass before they can take credit courses. According to a 2003 NCES report, nearly one-third of first-year students are required to take remedial courses in reading, writing, or mathematics. Although research indicates that remediation consistently has statistically significant, if rather modest, positive effects on the likelihood that students will ultimately complete a bachelor’s degree, the annual cost is over $1 billion at public colleges alone. The great majority of this cost is borne by students and their families.

At one time, secondary schools and colleges collaborated in this effort. College Entrance Examination Board (CEEB) and participating secondary schools together defined college preparatory curriculum and developed, administered, and scored tests to measure students’ mastery of it. The coalition finally unraveled to lead to the present situation of a nearly complete lack of coordination between high schools and colleges on exactly what constitutes readiness for college. So now we have colleges pretty much developing or otherwise buying their own readiness tests. Some states have statewide services that provide tests to all public colleges and universities; most do not. If there are no statewide tests, then each college has its own unique test(s), which may or may not (most likely not) be measuring the same aspects of readiness. In either case, it is the test itself that defines readiness or preparedness.

2012 Pearson Summer Research Internship

Pearson’s Psychometric and Research Services offers an eight-week summer internship for doctoral students in education or psychology. Interns have an opportunity to learn about item content development, scoring and processing, and other aspects of educational testing.

We are pleased to announce Pearson’s 2012 summer interns:

Yun Tang (Iowa City, IA)—Yun is a doctoral student in the Research Methodology program at the University of Pittsburgh. She is currently working on her dissertation, which focuses on comparing performances of different propensity scoring methods used to estimate causal effects for large-scale educational datasets. During her summer internship at Pearson, she hopes to expand her psychometric knowledge and skills by participating in a research study on the Virginia 2% reading assessment under the supervision of Dr. Tony Thompson.

Daniel Jurich (Tulsa, OK)—Daniel is in his second year of the Assessment and Measurement doctoral program at James Madison University (JMU). He received his Master’s degree in Psychological Sciences from JMU, where his research focused on the effects of cheating on test equating methods. Under the supervision of Dr. Stephen Murphy, Daniel’s internship with Pearson will involve exploring practical applications of vertical scaling methods for Oklahoma state testing and investigating the potential of domain scoring using item response theory procedures. In addition, he will serve as a data analyst for the Oklahoma Alternate Assessment Program (OAAP) standard setting.

Jason Kopp (San Antonio, TX)—Jason received a Master’s degree in Psychological Sciences with a concentration in quantitative psychology from James Madison University. He is currently enrolled as a doctoral student in the Assessment and Measurement program at JMU. During his fellowship, Jason will be working under the supervision of Dr. Allen Lau and Dr. Michael Young. Jason will be applying various diagnostic models to statewide mathematics testing data in an effort to design new methods for modeling skill growth in a diagnostic framework.

Continued on next page
SUMMER INTERN SEMINARS

During the summer internship, research scientists in Psychometric and Research Services present seminars covering aspects of educational testing. The seminar topic, time, date, and presenter of each seminar are shown in the table below. Seminars are presented at the facility where the speaker is located. In addition, all seminars are broadcast via WebEx so that all our Pearson colleagues may attend. If you would like to have more information about the seminars or the Webex set up, please contact Hua Wei, hua.wei@pearson.com.

<table>
<thead>
<tr>
<th>Session</th>
<th>Presenter</th>
<th>Topic</th>
<th>Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SONYA POWERS</td>
<td>LINGUISTIC ACCOMMODATIONS TO ELL STUDENTS</td>
<td>WEDNESDAY, JUNE 6</td>
<td>1–2:30 PM</td>
</tr>
<tr>
<td>2</td>
<td>STEPHEN MURPHY</td>
<td>DATA MINING: INNOVATIVE ITEMS</td>
<td>WEDNESDAY, JUNE 13</td>
<td>1–2:30 PM</td>
</tr>
<tr>
<td>3</td>
<td>DAVID SHIN</td>
<td>CAT</td>
<td>WEDNESDAY, JUNE 20</td>
<td>1–2:30 PM</td>
</tr>
<tr>
<td>4</td>
<td>CINDI KREIMAN</td>
<td>SAS MACRO PROGRAMMING</td>
<td>WEDNESDAY, JUNE 27</td>
<td>1–2:30 PM</td>
</tr>
<tr>
<td>5</td>
<td>ALLEN LAU</td>
<td>SPRT (SEQUENTIAL PROBABILITY RATIO TESTING)</td>
<td>THURSDAY, JULY 5</td>
<td>1–2:30 PM</td>
</tr>
<tr>
<td>6</td>
<td>BOB DOLAN</td>
<td>ADAPTIVE LEARNING</td>
<td>WEDNESDAY, JULY 11</td>
<td>1–2:30 PM</td>
</tr>
<tr>
<td>7</td>
<td>DAERYONG SEO</td>
<td>ELL RESEARCH</td>
<td>WEDNESDAY, JULY 18</td>
<td>1–2:30 PM</td>
</tr>
<tr>
<td>8</td>
<td>INTERNS</td>
<td>INTERNSHIP EXPERIENCES</td>
<td>WEDNESDAY, JULY 25</td>
<td>1–2:30 PM</td>
</tr>
</tbody>
</table>

Publications


Career Opportunities

Pearson is an Equal Opportunity Employer M/F/V/D, a member of E-Verify, and a drug-free workplace. Please visit our Careers page for more information about our available positions:
http://www.pearsonassessments.com/pai ai/about/careers.htm

Title: Senior Research Scientist
Location: Iowa City, IA
Req Number: 2012-8512
The successful candidate will have experience in designing and implementing innovative technical solutions for large-scale assessment programs that could include computer adaptive testing, diagnostic assessments, growth models, and alternate assessments. Experience in working with custom K–12 assessment implementation (including peer review response) is required.

Title: Research Scientist
Location: Iowa City, IA
Job Type: Regular Full Time
Req Number: 2012-10483
The successful candidate will have experience in designing and implementing innovative technical solutions for large-scale assessment programs. Experience in working with custom K–12 assessment implementation is required.
Conference Participation

Language Testing Research Colloquium, Princeton, NJ

Suzuki, M. Are predictor variables for item difficulty of L2 sentence repetition consistent between English and Chinese?

Van Moere, A. Elicited imitation: What can it tell us about oral language proficiency?

International Conference and Workshops on Technology and Chinese Language Teaching in the 21st Century, Honolulu, HI

European Association for Language Testing and Assessment (EALTA) conference, Innsbruck, Austria

Bosker, H.R., Quené, H. & Sanders, T. The effect of . . . silent pauses . . . on native and non-native fluency perception.

De Jong, N. Singling out L2-specific disfluencies for valid fluency assessment.

Lam, J.F. & Downey, R. The validity of the reading comprehension test for immigrants: the Dutch case.

McCray, G., Alderson, C. & Brunfaut, T. Validity in reading comprehension items: triangulation of eye-tracking and stimulated recall data.

Van Moere, A. & Bernstein, J. Technology, test design, and validity: Which one is in the driving seat?


Upcoming Conference Participation

Annual International Meeting of the Psychometric Society, Lincoln, NE

Awards

Kimberly O’Malley was appointed to the American Association of Community Colleges (AACC) Implementation Task Force on K12 Relations, May 2012; Washington, D.C.

Mitch Benson and Kimberly O’Malley were appointed to TechAmerica Foundation’s Recommendations for Education and the Advancement of Learning (REAL) Commission. The commission made recommendations to the national Digital Promise research center: January 2012; Washington, D.C.

Career Opportunities, continued from page 11

Title: Test Development Manager
Location: Palo Alto, California or Boulder, Colorado
Req Number: 2012-9168

Title: Senior Test Development Manager
Location: Palo Alto, California or Boulder, Colorado
Req Number: 2012-9167

The Knowledge Technologies group of Pearson has two positions open. The role involves test development and validation research. Expertise in linguistics, statistics, and research methodology are required. Applicants should have an M.A. or PhD in applied linguistics, educational measurement, or a related field. Fluency in a second language and experience in language teaching, assessment, or research are desirable.

The Knowledge Technologies group of Pearson has two positions open. The role involves test development and validation research. Expertise in linguistics, statistics, and research methodology are required. Applicants should have an M.A. or PhD in applied linguistics, educational measurement, or a related field. Fluency in a second language and experience in language teaching, assessment, or research are desirable.