Welcome!

Thank you for participating in the Educational Technology and Computational Psychometrics Symposium!

ACTNext, the Research, Development and Business Innovation Center at ACT, is pleased to continue to support this successful symposium. Until now, this gathering has primarily been for an internal ACT audience. However, this year we are excited to open this opportunity to the public, with the promise of providing a forum that will promote the free exchange of ideas with some of the best minds in our industry on a grander scale than ever before.

Iowa City has come of age as an innovation hub, a truly unique place where academia and industry are working together to not only solve our current challenges, but to anticipate and shape the way we will meet them in the future.

The goal of this symposium is to build on current momentum and heighten awareness concerning the dynamic changes sweeping the educational landscape and to provide an environment for fostering discussion about the implications of these changes.

In particular, the symposium focuses on the role of technology in educational practices and the role of computational psychometrics in educational measurement, and the ways in which they are intertwined.

Over the next two days, you will hear from speakers passionate about education, research, computational psychometrics, and technology, and I encourage you all to engage with them throughout the event.

I anticipate the interactions amongst our participants and speakers and the ensuing dialogues will provide a powerful spark for fresh ideas and new opportunities for innovation, and I am so pleased to be able to host this most exciting enterprise.

Again, thank you for attending our symposium. I am certain you will find it to be an informative and valuable experience.

Sincerely,
Alina A. von Davier, PhD
Vice President, ACTNext
8:00  Registration & Continental Breakfast
8:45  Welcome  Alina von Davier
8:55  Where are We Going with Learning and Assessment?  Dror Ben-Naim
9:55  Break
10:05  Entertainment Education and Media Impact Assessment  Johanna Blakley
11:05  Towards Strengthening Links Between Learning Analytics and Assessment  Dragan Gašević
12:05  Lunch Break
1:20  Psychometric Models of Small Group Collaborations  Peter Halpin
2:20  Break
2:30  Lessons Learned and Future Directions for Game-Based Assessment  Kristen DiCerbo
3:30  Some Assembly Required: Collaboration in the 21st Century  Noshir Contractor
4:45  Poster Session and Event Reception  Hilton Garden Inn Downtown Iowa City

WEDNESDAY

November 15
8:00  Registration & Continental Breakfast

9:00  Welcome
     Marten Roorda

9:10  Augmented Intelligence and the Future of Item Development
     Donna Matovinovic and Mark Gierl

10:15  Break

10:25  Inq-ITS And Inq-Blotter: Real-Time Assessment and Alerting On Students’ Inquiry Practices
     Janice Gobert

11:30  How to Use Practical Virtual Reality for Computational Psychometrics
     Pietro Cipresso

12:35  Closing Remarks
     David Kuntz
MEET THE SPEAKERS
Dr. Dror Ben-Naim is the founder and CEO of Smart Sparrow, an edtech startup based in San Francisco and Sydney pioneering personalized and adaptive learning technology.

Prior to founding Smart Sparrow, Dror led The Adaptive eLearning Research Group, working in field of Intelligent Tutoring Systems and Educational Data Mining at The University of New South Wales in Sydney. In 2011, Dror and his team founded Smart Sparrow to take Adaptive Learning to the next level and secured backing from venture capital investors. Today, Smart Sparrow is adopted by hundreds of leading academic institutions worldwide.

Dr. Ben-Naim is a Professor of Practice at Arizona State University’s teachers college and Associate Director of the university’s new Center for Education Through eXploration (ETX). He is also an Adjunct Academic at the UNSW School of Computer Science and Engineering, from where he received a B.Sc. in Physics and Computer Science and a PhD in A.I. and Education.

Johanna Blakley is the managing director at the Norman Lear Center, a research and public policy institute that explores the convergence of entertainment, commerce and society. Based at the University of Southern California’s Annenberg School for Communication and Journalism, Johanna performs research on a wide variety of topics, including global entertainment, cultural diplomacy, entertainment education, and digital media. She has two talks on TED.com and she speaks frequently in the U.S. and abroad about her research. Johanna is co-Principal Investigator on the Media Impact Project, which seeks to better understand the role that media plays in changing knowledge, attitudes and behavior among individuals and communities. Before arriving at USC, Johanna worked on dozens of entertainment-education game titles at Davidson & Associates and Knowledge Adventure, which evolved into Vivendi-Universal Games. Johanna received a PhD in English from the University of California, Santa Barbara, and she teaches classes on transmedia storytelling at USC.
Noshir Contractor is the Jane S. & William J. White Professor of Behavioral Sciences in the McCormick School of Engineering & Applied Science, the School of Communication and the Kellogg School of Management at Northwestern University, USA. He is the Director of the Science of Networks in Communities (SONIC) Research Group at Northwestern University. He is investigating factors that lead to the formation, maintenance, and dissolution of dynamically linked social and knowledge networks in a wide variety of contexts, including communities of practice in business, translational science and engineering communities, public health networks and virtual worlds. His research program has been funded continuously for over 15 years by major grants from the U.S. National Science Foundation with additional current funding from the U.S. National Institutes of Health (NIH), NASA, Air Force Research Lab, Army Research Institute, Army Research Laboratory, the Gates Foundation and the MacArthur Foundation.

Professor Contractor has published or presented over 250 research papers dealing with communicating and organizing. His book titled *Theories of Communication Networks* (co-authored with Professor Peter Monge and published by Oxford University Press, and translated into simplified
Dr. Kristen DiCerbo is the Vice-President of Education Research at Pearson. She leads a team of researchers conducting and translating research about learners and learning to influence the development of curricula and digital tools.

Dr. DiCerbo’s personal research program centers on interactive technologies, particularly the use of games and simulations to understand what learners know and can do. She was a partner in GlassLab Games’ development of SimCityEDU and Mars Generation One: Argubot Academy, both examples of game-based assessment. Prior to joining Pearson, Kristen was a researcher with the Networking Academies at Cisco and was a school psychologist in a local school district in Arizona. Kristen received her Bachelor’s Degree from Hamilton College and Master’s Degree and Ph.D. in Educational Psychology at Arizona State University.
Dr. Mark J. Gierl is Professor of Educational Psychology and the Director of the Centre for Research in Applied Measurement and Evaluation (CRAME) at the University of Alberta. His specialization is educational and psychological testing. He received his PhD from the College of Education at the University of Illinois, Urbana-Champaign. Professor Gierl’s current research program focuses on automatic item generation. He co-edited the most recent book on this topic, *Automatic Item Generation: Theory and Practice* (Routledge, 2013), with Professor Thomas Haladyna from Arizona State University.

Professor Gierl works closely with government agencies, testing companies, and academic publishers to implement item generation principles produced from his research into their test development practices. For example, he received a grant for three consecutive years (2012-2014) to conduct item generation research for the testing company CTB/McGraw-Hill in Monterey, California through their Innovation Research and Development Program. He has also worked with ACT Inc. (USA), the American Dental Association (USA), the Australian Council for Educational Research (Australia),...
Cengage Learning (USA), the College Board (USA), the Educational Records Bureau (USA), the Medical Council of Canada (Canada), the National Board of Medical Examiners (USA), and Reed-Elsevier (The Netherlands) to implement his research on automatic item generation into their operational test development process.

Professor Gierl has written 4 books, 22 book chapters, 73 refereed journal articles, 186 conference papers, and 58 technical reports—many of which were co-authored with his graduate students. He has also delivered 49 invited lectures throughout the world. Professor Gierl is the recipient of the 2009 American Educational Research Association (AERA) Division D Significant Contribution to Educational Measurement and Research Methodology Award for his co-edited book, "Cognitive Diagnostic Assessment for Education: Theory and Applications", and the 2016 National Council on Measurement in Education (NCME) Annual Award for his research program on automatic item generation. He received the Distinguished Alumni Award in 2017 from his Alma Mater, the University of Illinois, Urbana-Champaign. Professor Gierl has been awarded over $9.4 million in external research funding from agencies in Canada and the United States—including 6 grants as the principal investigator and 9 grants as a co-investigator from the Social Sciences and Humanities Research Council of Canada (SSHRC). His most recent SSHRC grant was awarded in April 2016. It is titled "Using Automated Processes to Generate Educational Test Items and their Associated Solutions and Rationales to Support Formative Feedback".

Professor Gierl holds the Tier I Canada Research Chair in Educational Measurement.
Peter Halpin is an Assistant Professor of Applied Statistics at New York University. His research focuses on psychometrics (e.g., confirmatory factor analysis, item response theory, latent class analysis), as well as statistical methods for complex and technology-enhanced educational assessments. His work has been published in leading methodological journals including Psychometrika, Structural Equation Modeling, and Journal of Educational Measurement, as well as general interest journals including Educational Researcher. Notable funding sources include NAEEd / Spencer Foundation (Postdoctoral Fellowship); the Statistics and Research Methodology Program of the Institute of Education Sciences (Early Career Grant); and the Natural Science and Engineering Council of Canada (Postdoctoral Fellowship).

Janice Gobert is a Professor of Learning Sciences and Educational Psychology at Rutgers Graduate School of Education. Formerly, she co-founded and was the co-director of the Learning Sciences and Technologies Program at Worcester Polytechnic Institute. Her Ph.D. from the University of Toronto (1994) is in Cognitive Science; her Masters is from McGill University, also in Cognitive Science. Her specialty is in technology-based learning in scientific domains. Her research areas are: intelligent tutoring systems for science, skill acquisition, performance assessment via log files and educational data mining, and eye tracking. Gobert holds four patents for her work in the areas of assessment and eye tracking.
Ms. Matovinovic has more than 25 years of experience in education and educational assessment. As the Senior Vice President of Test Development at ACT, Ms. Matovinovic provides expertise, leadership, and strategic direction for the design and development of products and services across the K through career continuum designed to support claims related to college and career readiness. She has spent more than a decade as an educator, curriculum developer, and formative and summative assessment leader in Alberta, Canada and more than a decade leading assessment design, content development and publishing systems development for CTB, including leading the team responsible for developing test content for both the Smarter Balanced and PARCC consortia. Previous positions held by Ms. Matovinovic at CTB include: Senior Publishing Director, Director of the Publishing Project Management Office, Director of Business and Customer Management for the McGraw-Hill publishing system, and Science Content Development Manager. Prior to CTB, she was Biology Examinations Manager at the Alberta Department of Education, and taught middle school to university level science and mathematics for 10 years. Ms. Matovinovic has a M.Ed. in Science Curriculum and Instructional Technology; a B.Ed. in Science and Mathematics Education and a B.Sc. in Biological Sciences, all from the University of Alberta, Canada.
Where Are We Going With Learning And Assessment

8:55

Dror Ben-Naim

Virtually every aspect of education is transforming at an accelerating pace, but how to make sense of it all? Where are we going? How will we get there? What is the impact on learning and assessment?

In this talk I will share insights and learnings from my experience researching and designing rich, interactive and adaptive learning experiences that authentically blend learning and assessment. These innovative new environments provide tremendous opportunities for 21st century education systems, as well as posing fundamental challenges to education and assessment providers.

Entertainment Education and Media Impact Assessment

10:05

Johanna Blakley

Media has long been utilized to communicate information to mass audiences, but measuring the effectiveness of those media interventions has been challenging, to say the least. In partnership with the Centers for Disease Control and Prevention, the Norman Lear Center has generated media impact studies for 16 years. With the founding of its Media Impact Project five years ago, the Center expanded its focus from scripted TV to news, documentaries, narrative films, online video projects, and virtual reality. This presentation will explain the entertainment education model that we’ve used in our Hollywood, Health & Society program, which bridges the gap between health experts and TV storytellers, as well as the work of our Media Impact Project, which develops and gathers best practices in media impact assessment. Methods and findings from a range of media impact studies will be summarized, including results from studies on changes in awareness, knowledge, attitudes and behavior among viewers of scripted entertainment (soap operas, primetime dramas, feature films), as well as consumers of journalistic content on broadcast TV, film, online, and in virtual reality environments.
In this talk I’ll argue that the social combination approach to group problem solving provides a suitable framework for extending existing psychometric models to collaborative settings. The literature on social combination theory is selectively reviewed to motivate an overall modeling framework. I then propose a model of pairwise group work, which leads to a number of results about the design of group assessments. In particular, I consider how to select items and group members such that the expected performance of a dyad can be empirically distinguished from that of either individual. I then address estimation of the model, and the
results are illustrated with data simulation and an empirical example in which pairs of respondents work together to complete a twelfth-grade level mathematics assessment. In concluding, attention is given to specific avenues of future research that seem most fruitful for advancing current initiatives concerning the assessment of collaboration, teamwork, and related domains.

Lessons Learned and Future Directions for Game-Based Assessment
Kristen DiCerbo
This session will describe findings and lessons learned from a research program investigating the use of games as both learning and assessment tools. Key issues to be addressed include: specification of learning progressions, task design to align with learning progression stages, identification and summarization of evidence from log files, and reporting to inform instructional decision-making. In addition, evidence of validity and reliability will be discussed in the context of how we think about these concepts with game-based assessments. Finally, challenges of scaling the psychometric design process from a research project to production will be described.

Some Assembly Required: Collaboration in the 21st Century
Noshir Contractor
Recent technological advances provide comprehensive digital traces of social actions, interactions, and transactions. These data provide an unprecedented exploratorium to model the socio-technical motivations for creating, maintaining, dissolving, and reconstituting into teams. Using examples from research on collaboration, Contractor will argue that Computational Social Science serves as the foundation for the development of social network theories and methods to help advance our ability to understand the emergence of effective collaborations. More importantly, he will argue that these insights will enable the assembly of effective “dream teams” by building a new generation of recommender systems that leverage our research insights on the socio-technical motivations for creating ties.
Augmented Intelligence and the Future of Item Development
Donna Matovinovic and Mark Gierl
Automatic item generation (AIG) is a rapidly evolving research area where cognitive theories, computer technologies, and psychometric practices are used to create models that produce test items with the aid of computer technology. The purpose of our presentation is to describe two AIG methods; logical and feature, provide an example of each, and share the business benefits of combining technical expertise in item automation and content generation with subject-matter expertise (augmented intelligence) for the purpose of producing large numbers of high-quality, content-specific, test items. The methods and processes will be used to help transform item and passage development at ACT Inc. from what is currently a manual, labor intensive, non-scalable process to a specification driven, automated, highly-scalable process.

Inq-ITS And Inq-Blotter: Real-Time Assessment and Alerting On Students’ Inquiry Practices
Janice Gobert
Inq-ITS is a platform for science inquiry. Instead of taking multiple choice tests on rote science knowledge, students conduct and learn real science inquiry with interactive, virtual labs in the Inq-ITS system. While kids “show what they know”, educators get real-time, actionable reports and alerts on their mobile devices. These reports and alerts enable them to pinpoint which students need help most, and on what skills. Students can also get real time feedback from Rex, a cartoon dinosaur, in real time as they work on inquiry tasks. All reports, alerts, and scaffolds are driven by data-mined algorithms, on which we have two patents.
Computational sciences have become quite extensively used in the last few decades in several disciplines. For example, computational biology and computational neuroscience have strong traditions, and many models of human science rely on crucial questions. In this view, it is not surprising that psychometrics is also adopting computational techniques and models to better understand quantitative psychology and measurement. In particular, computational psychometrics can be used to develop new solutions to understand and measure behavior. This presentation reviews the use of Virtual Reality to elicit emotions and specific behavior to be included and measured in computational models. In particular, specific solutions will be proposed for integrating VR platforms with artificial simulations and making decisions based on possible scenarios. The approach will be presented in terms of its analytical structure (also with nonlinear system dynamics) as well as computational simulation models.
ABOUT
THE
POSTERS
In conjunction with the symposium, ACTNext accepted submissions for a limited number of poster presentations relating to outstanding research associated with educational technology and computational psychometrics. Finalists were selected by a panel of experts for today’s session. ACTNext is extremely excited to provide this showcase for researchers to share their best work.
The Expanded Evidence-Centered-Design (e-ECD) Framework for Learning & Assessment Systems

This poster presents a new framework for the design and development of learning and assessment systems. The new framework, titled e-ECD is an expansion on an existing framework - the Evidence-Centered Design (ECD; Mislevy, Almond & Lukas, 2003). The contribution of the e-ECD is in incorporating aspects of learning into each of the three core models of ECD, as well as integrating methodology from computational psychometrics.

Many Educators Are Trying to Innovate in Digital Learning, but Results Often Fail to Meet Expectations

BioBeyond is a Smart Course developed by the Inspark Science Teaching Network through a grant by the Bill & Melinda Gates Foundation. It was built using the Smart Sparrow Platform, and stands as a prime example of an online course that is successfully improving student understanding and performance. There are six dimensions of innovation brought to life with BioBeyond: 1. Learner-Centered Design guiding the creative process; 2. “Education Through eXploration” as a new pedagogical theory; 3. The degree and type of interactive student engagement; 4. The approach to personalization and adaptively to support diverse student populations; 5. The use of authentic assessment and analytics to understand student behavior and mastery of knowledge; 6. Scaling efforts through a collaborative teaching network. Over the last few years, BioBeyond has been evaluated by independent researchers and instructors, and has continuously shown to improve student outcomes compared to traditional teaching.

Topic Modeling and Creativity of Writing Responses

The purpose of this project was to explore the use of LDA topic modeling to contribute to automated scoring of writing responses in the creativity domain. A writing prompt was designed to elicit creative responses and administered to participants through Amazon Mechanical Turk. All responses
Validity of Computer-Based Accommodations for English Language Learners in Assessment

A fair test should be sensitive to the characteristics of subgroups in the test-takers throughout all its phases of design, administration and interpretation while yielding valid interpretations, according to American Educational Research Association (2014). The validity of test score interpretation could be interfered with the characteristics of individual test-takers from diverse groups as defined by ethnicity, gender, language or so. Accommodations are needed to respond to the individual characteristics.

English language learners (ELL) limited English proficiency could restrict their access to the construct(s) to be measured, which, in turn, compromises the validity and fairness of test score interpretations. The total ELL population in the United States has increased both in quantity and diversity, and almost every state in the United States has witnessed the increase of ELLs (ETS, 2016). Meanwhile, there is a large performance gap between ELL and non-ELL students, as found by analyses of standardized test scores across multiple states (Abedi, 2014). To resolve the obstacles in the testing environment, one of the most commonly used accommodations adopted for ELLs is making linguistic modifications of the test items (Abedi, 2012). Computer-based testing that adopts modern technology can facilitate the provision of such accommodations.
A promising edTech trend is the move from a generalized, discrete, fixed time/place delivery approach towards a personalized, continuous, mobile anytime/anywhere one. Universities, school classrooms and edTech companies are offering mobile solutions that allow students to consume lecture materials, hold learning sessions, and take tests – all from the palm of their hand. Mobile devices hold the promise for richer, intelligent interactions that can be adaptively tailored to each learner. ACTNext – a disruptively innovative research arm of ACT, Inc. is working on a new app called the Educational Companion App (ECA) to change that status quo. ECA has the promise to deliver an integrated, intelligent, personalized guide for mobile learning.

To handle unfinished CAT tests, a penalty function was devised to meet two criteria: 1) the greater the omit rate, the greater the penalty, and 2) examinees with the same ability and the same omit rate should receive the same penalty. This study investigated the impact on theta estimates resulting from the approach to handling unanswered items in a fixed-length CAT.

In response to the emerging demand of valid accommodation for ELLs, this poster presents a meta-analysis study in process that aims to examine the validity of the computer-based accommodations for ELLs, with a focus on the accommodations that make linguistic modifications in test items. Based on a review, analysis and critique of the existing literature, the study compares the advantages and disadvantages of the computer technologies that help provide accommodations, and examines the validity of four fully-developed computer-based testing models that drive to address the language barriers for ELLs. The poster concludes with implications of the research findings to the states and districts for applying computer-based assessment with accommodations for ELLs.

Introducing the ACTNext Educational Companion: An Intelligent, Personalized Guide for Mobile Learning

An Imputation Approach to Handling Incomplete Computerized Tests
comprehensive mobile learning experience. Students can review a fused perspective of their abilities drawn from a range of inputs including formative and high-stakes test results (The ACT, PreACT), social-emotional assessments (e.g. ACT Tessera), skill practice (Benchprep) and targeted quizzes. These results are seamlessly linked to a suite of Skill Up activities and open educational resources (OER) from ACT’s OpenEd that allows students to practice the skills they are to master yet. ECA leverages the ACT Holistic Framework – that covers all aspects of development: cognitive, emotional, cross-cutting and navigational. This work presents our ECA research and design approach, along with results from a pilot study which we are conducting with a cohort of high school students in South Carolina, USA.

Exploring Device Form Factors in the Educational Assessment Setting

This poster discusses the importance of understanding device form factors for supporting score comparability in computer-based testing. While the development of new technologies creates new measurement possibilities, we should begin to explore technology related factors that may create construct-irrelevant variance as well as influence test score comparability when different devices are used in assessment. Device form factors may affect student performance and the comparability of test scores across devices. The purpose of the study was to investigate student-device interaction in an educational assessment. A think-aloud study was conducted with students across four different device conditions—desktop, laptop, Chromebook, and tablet. Students in this study read and responded to test questions in each of the four subject areas—English, Reading, Math, and Science. Multimodal data were captured and synchronized to support analyses. The findings suggested that overall students were able to interact with the test content reasonably well across all devices; however, the approaches they took to accomplish the tasks varied somewhat from device to device with some approaches arguably requiring more steps and/or greater cognitive demand. While this study does not quantify to what degree these differences would impact student performance and resulting score comparability, it does highlight the challenges likely to be faced when delivering test content in a multiple device setting.
Modeling Response Accuracy and Response Time in Cognitive Diagnosis

At present, examinees classifications using cognitive diagnosis models (CDMs) are typically based solely on response accuracy. However, response time is another source of information that contains potentially valuable information, and may be easily obtained in computer-based testing settings. Although a continuous response CDM that can be used in response time modeling has been recently proposed, examinee classifications are still made on the basis of a single response type – either response accuracy or response time. This paper proposes a CDM in which both response accuracy and response time can be used jointly to classify examinees. A simulation study is conducted to determine the extent to which incorporating response time to response accuracy under the proposed model can improve examinee classification. Results indicate that the proposed model can offer improvement in classification accuracy, particularly when the items for response accuracy are of low quality.

Task Loop Adaptivity for Bayesian Knowledge Tracing Using Computer Adaptive Testing

Curriculum sequencing is the selection of items or questions in an optimal sequence based on learners’ actions, answers, and other observable data in an adaptive environment. Standard methods for curriculum sequencing have traditionally not taken into account learner proficiency and item difficulty, but rather rule-based systems and Bayesian networks have focused on the knowledge relationships between items (i.e. prerequisites). Concepts from computer adaptive testing (CAT), which take into account learner proficiency and item difficulty, have been utilized to recommend items in an adaptive learning framework. However, CAT methods choose items for which a learner has 50% probability of getting right, which is optimal for efficiently estimating proficiency, but not optimal for learning. We present a method to recommend items based on CAT concepts that take into account learner proficiency, item difficulty, but also allow more flexibility with how likely a learner is to answer an item correctly.
A Framework for Hypothesis-Driven Approaches to Support Data-Driven Learning Analytics in Measuring Computational Thinking in Block-Based Programming

There is a growing need to measure students’ learning of Computational Thinking in the context of introductory block-based programming so as to be able to support all K-12 learners through this process of learning computational problem solving. The goal of this research is to explore hypothesis-driven approaches that can be combined with data-driven ones to better interpret student actions and processes in log data captured from block-based programming environments with the goal of measuring and assessing students’ CT skills. Informed by past literature and based on our empirical work examining a dataset from the use of the Fairy Assessment in the Alice programming environment in middle schools, we present a framework that formalizes a process where a hypothesis-driven approach informed by Evidence-Centered Design effectively complements data-driven learning analytics in interpreting students’ programming process and assessing CT in block-based programming environments. We apply the framework to the design of Alice tasks for high school CS to be used for measuring CT during programming.

Moving Schools’ Existing Homework Online Can Double Student Learning

A team of software developers at WPI have created an online tool that has proven to drastically effect student outcomes. Two studies are highlighted, one that lasted one night and another that lasted one year. These studies showed that when the tool, ASSISTments, is used for homework it can double student learning. ASSISTments is not a curriculum. It is a tool that allows students to receive feedback and teachers to get reports. The one night study showed significant gains and also explored the roll of homework review. The one year study was conducted by outside evaluator, SRI, and had three findings. 1. Teachers changed homework review; 2. ASSISTments increased student learning by 75%; and 3. ASSISTments closed achievement gaps.
Bayesian and Shadow-Test Approach to Item Calibration Embedded in Adaptive Testing

Field-test item calibration typically precedes operational adaptive testing, often in the form of a separate study where a group of examinees are assigned to respond to fixed sets of field-test items. In this research, a Bayesian and shadow-test approach is proposed to efficiently embed field-test item calibration in adaptive testing. Specifically, the shadow-test approach is used to assign a few field-test items to each examinee adapting to the current posterior distributions of their ability parameters. In this way, field-test items are embedded in the operational adaptive testing with respect to the test blueprint, exposure rate control, and other statistical/nonstatistical aspects. For the examinee’s ability estimation and field-test item parameter update, a Markov chain Monte Carlo (MCMC) algorithm is implemented with a Gibbs sampler and a Metropolis-Hastings step. The purpose of this presentation is to show how the adaptive testing and field-test item calibration can be integrated in one framework while meeting all the requirements. Simulation studies are carried out based on a realistic item pool with different algorithm settings. Simulation results are presented and discussed to show the effectiveness of the proposed approach for various adaptive testing scenarios.

Evidence About Evidence: Bridging The Gap Between Multimodal Data and Assessment of Complex Skills

There is a growing need for assessment tools that capture a broad range of learner behavior necessary for the evaluation of skills such as problem solving, communication and collaboration. A key feature of such assessments is the use of interfaces that enable rich, immersive interactions and can capture student data in a multitude of sensory modalities. However, the analysis of multimodal data poses a significant challenge: how do we extract meaningful evidence of construct competency from such unstructured data? To address this challenge, I present a hierarchical inference framework to model temporal dynamics and integration of multiple data modalities. This approach exploits concept hierarchies that reflect the nature of the data and goals of the assessment to identify ecologically valid evidence of the
Eye-tracking data were analyzed to provide understanding into the cognitive processes used to solve Graphic Literacy tasks. Thirty-eight individuals participated, of which 16 were high school students, four were college students, and 18 were working adults. The eye-tracking study collected gaze data as participants worked through the tasks. Gaze data included fixations, saccades, sequence, heat maps, and pupillometry. The findings demonstrated that high-performing graphic literacy examinees used qualitatively different strategies for approaching the graphics and associated items from low-performing examinees. Further, the findings indicated that low-performing examinees struggled with specific features of graphics. The findings not only provided evidence supporting ACT’s construct definition of graphic literacy, but they also provide insights for assisting individuals to learn better methods for utilizing graphic information.

Peer assessment encompasses processes whereby students evaluate or are evaluated by their peers. Recent years have seen increasing use of peer assessment in classrooms and other learning settings. In addition to increasing teachers’ efficiency in grading, peer assessment is advocated as a pedagogical strategy for facilitating students’ learning. For example, peer assessment can foster students’ evaluative and critical ability, promote students’ engagement, metacognitive awareness, and their social-affective development. Despite the prevailing view that peer assessment has a positive effect on learning, mixed results have been reported in empirical studies. In this study, we conducted a meta-analysis to examine the general effect of peer assessment on learning outcomes. The results of the study are expected to inform researchers and teachers of how to make effective use of peer assessment as a learning tool.
Semantic Similarity of Examinee Written Responses to Exemplar Training Responses: Impact on Automated Scoring Accuracy

Using word2vec word embeddings, the semantic similarities between examinees’ written short responses and exemplar training responses were quantified, and the similarity values were added to the response feature set in a machine learning model for automated scoring. Accuracy improvements were noted for many content areas.

A Dynamic Perspective on Collaborative Learning Processes

Learning can be considered as an adaptive developmental process that seeks functional alignment between internal and external representations of the environment. Collaborative learning additionally creates a sub-ecological system of which an individual is an integrated part, and partially through which he/she interacts with the environment. Whether this sub-system facilitates or hampers one’s learning process depends on the structure of the sub-system, and the interactions therein. This project presents ways to represent collaborative learning processes as differential equation models, and illustrates their modeling behavior through simulations. Implications on next-generation educational measurements and interventions will be discussed.


ALOSI is an open source adaptive engine developed by Harvard and Microsoft learning sciences teams aimed to power individualized learning and assessment pathways. The ALOSI architecture integrates seamlessly with multiple LMS’s and independently with content repositories. ALOSI uses Bayesian Knowledge Tracing—a machine learning algorithm—to power learning pathways and assessments. The ALOSI engine has been recently deployed in a Microsoft Course on edX “Essential Statistics for data Analysis using Excel”, with forthcoming deployments in HarvardX MOOCs, residential on-line learning at Harvard via Canvas LMS and more. To
utilize the engine, the course team significantly enhanced the assessment scope, and included over 35 knowledge components and 400 assessment items tagged to those knowledge components. Learners were randomly assigned to three independent cohorts, two experimental and one non-adaptive group, so we can measure the effects of adaptive pathways on learning gains, persistence, time-to-mastery, and completion rates using different tuning parameters in the adaptive engine against a standard non-adaptive learning experience. The presentation will provide a deeper dive into the ALOSI architecture, study design, considerations for efficacy measures and illustrate preliminary findings from the pioneer adaptive course.

**Half-Life Regression: A Personalized Statistical Model of Forgetting**

We present half-life regression (HLR), a novel model for spaced repetition practice with applications to second language acquisition. HLR combines psycholinguistic theory with modern machine learning techniques, indirectly estimating the "half-life" of a word or concept in a student's long-term memory. We use data from Duolingo — a popular online language learning application — to fit HLR models, reducing error by 45% compared to several baselines at predicting student recall rates. HLR model weights also shed light on linguistic concepts that are systematically challenging for second language learners. Finally, HLR was able to improve Duolingo daily student engagement by 12% in an operational user study.

**The Duolingo English Test: Assessment Design & Development via Machine Learning**

We describe our process for developing a computer-adaptive, criterion-referenced English language assessment using machine learning and computational linguistics. The underlying construct is comprised of statistical models fit to text passages and dictionaries that were annotated with levels from the Common European Framework of Reference (CEFR). Test items are then semi-automatically generated in large quantities and projected onto the scale using CEFR-aligned model predictions. Combined with statistical item scoring, this approach produces a test with high reliability (.96+) and significant correlation with other major English
Identify Targets for Cognitive Training to Improve Fluid Reasoning in Children and Young Adults

There is significant current interest in whether high-level cognitive capacities, such as fluid reasoning, can be improved via dedicated training. Working memory (WM) has been a target of these investigations, due to the well-established correlations between performance on WM tasks and real-world outcomes such as scholastic achievement and general cognitive functioning. However, games (e.g., Cogmed) and mobile applications (e.g., Lumosity) that attempt to train WM to improve higher cognitive capabilities have yielded inconsistent results. The majority of these approaches have used methods that are not optimal for producing substantial learning and transfer – for example, by training adults rather than children, or by using only one or a few highly-similar training tasks.

The current study sought to identify a wider range of tasks that predict fluid reasoning in children and adults, with the long-term goal of developing a battery of training tasks with sufficient variety to provide significant learning and transfer. We report data from 46 children (age range 7.1-9.2 years) and 97 young adults (age range 18.0-23.1) across 13 lab tasks measuring WM and attentional control, as they predict fluid reasoning. Results showed similar patterns between the age groups, and combined data indicated multiple tasks that are not commonly considered to tap the same underlying constructs overlapped considerably in predicting fluid reasoning. This suggests that the same range of tasks could be targeted for intervention from early childhood to adulthood. We conclude by considering how such training tasks may be integrated into educational technology to provide general improvements in performance.

Game-Based Assessment of Collaborative Problem Solving

A broad range of skills (e.g., creativity, communication, collaboration, problem solving) are required to thrive in a 21st century society rich with both information and opportunity. These same skills empower people to fulfill their potential as effective and creative knowledge seekers and collaborative
problem solvers. Measurement and training of such varied skills present challenges for assessment and education professionals, but also opportunities. This project focuses on collaborative problem solving, which is part of ACT’s Holistic Framework—a comprehensive description of the knowledge and skills individuals need to know and be able to do to succeed at school and at work. The skills required to effectively combine problem solving, communication, and behavioral strategies to successfully solve a problem within a team context are outlined in ACT’s Collaborative Problem Solving construct (CPS). We chose five of these facets for exploration and prototyping. The innovation of the prototype resides in designing an interactive educational game according to a comprehensive framework for assessing and teaching difficult-to-measure skills, the CPS. Our research objective was to explore the feasibility of measuring these facets in an interactive and engaging way that will eventually allow for the ability to provide valuable and authentic insights and feedback to participants.

Using Open Source Machine Learning to Holistically Score Composite Items based on NGSS

Learning Progressions (LP) in science pose several major challenges. First, it is difficult to gather data from enough students to validate the progression. Second, it can be costly and time consuming to score responses that include written explanations that get at key practices like arguing from evidence and cross cutting concepts in patterns, cause and effect, and matter cycles and energy fluxes. To meet these challenges, we have used LightSide Researcher’s Workbench, an open source machine learning engine, to score composite items holistically according to the LP. The engine is trained to score the item to the subscale level which increases accuracy as student responses at sublevel 2.3 are similar to each other but very different from responses at 2.2, which represents a different misconception along the LP. Using a mix of decision tree and logistic regression methods, 30 different items have been scored for the past two school years with over 565,000 student responses scored from the Carbon TIME LP. Additionally, the computer scoring has served as a feedback loop for LP validation and item development. Items that the computer is unable to create a working model for (QWK<.7) often have problems with either human scoring or item mechanics that require one or both to be revised. This work addresses the issue of scalability.
ACT Tessera is designed to measure six social and emotional learning (SEL) skills: Organization/Responsibility, Tenacity/Grit, Teamwork/Cooperation, Composure/Resilience, Curiosity/Ingenuity, and Leadership/Communication Style. Research demonstrates that are SEL skills are important. SEL skills predict a variety of important outcomes, including academic performance, academic retention, and job performance.

Tessera assesses each of the six SEL skills with three methods: self-report Likert-type scales, forced choice, and situational judgment tests. Each is subject to its own biases or weaknesses, and Tessera employs multiple methods to minimize these effects. This is known as a multi-trait multi-method design.

Tessera was administered to 2852 middle school students in the fall of 2016 and Spring of 2017. Outside of a few marginal scales, SJT and SR scales were reliable (α = .67-.84). Each of the 6 skills significantly predicted self-reported grades (r = .20-.46). Importantly, each assessment method predicted grades incrementally over the other.

Tessera was administered to 2142 high school students in the fall of 2016 and Spring of 2017. Results were similar to the middle school version. Supplemental analyses on a subset (N = 800) of high school students revealed that Tessera predicted grades incrementally over achievement test scores. Finally, white/non-white subgroup differences were less than half as large for Tessera as they were for academic achievement.

Developing Authentic Digital Math Assessments

With the shift to next generation digital assessments, increased attention has focused on technology-enhanced assessments and items. However, many technology enhanced items lack construct fidelity relative to the
degree to which they mirror an authentic application of the measurement construct or capture student responses using the tools they use in the classroom. This study evaluates the feasibility of a high fidelity digital assessment item format, which allows students to solve mathematics questions on a tablet using a digital pen. This digital ink approach allows students to hand write their responses while showing their work and intermediate steps in a digital format. Responses obtained using the digital pen will be compared to both paper- and type-written response formats in a repeated measures design. This presentation will showcase the digital ink application for mathematics, share examples of student work across the different response formats, and describe the recent data collection experience with 100 high school students at a school in Ohio which has been using touch screen tablets and digital ink applications for a number of years. Researchers will also reflect on lessons learned from working with students in a pilot of the study and discuss next steps relative to scoring student responses and data analysis plans.
Special Thanks

It takes a fair amount of heavy lifting to produce the Educational Technology and Computational Psychometrics Symposium. ACTNext would like to extend a thank you to all the team members involved in producing the ETCPS 2017 including:

- The Englert Theatre
- Hilton Garden Inn
- Sheraton Hotel
- Bread Garden Catering
- ACT Legal
- ACT Events Team
- ACT Facilities & Security
- ACT Marketing Automation
- Short’s Travel
- Aramark Catering

We would like to extend a special thank you to ACT CEO Marten Roorda, our distinguished speakers, poster authors, and attendees for participating in the ACTNext Educational Technology and Computational Psychometrics Symposium!
1 - Java House
2 - Heirloom Salad
3 - Iowa Chop House
4 - India Cafe
5 - Fieldhouse Iowa City
6 - Uncle Sun
7 - Baroncini Ristorante
8 - Formosa
9 - Bread Garden Market
10 - Share Wine Lounge & Small Plate Bistro
11 - Graze
12 - El Patrón - Taco & Tequila Bar
13 - Joseph's Steakhouse
14 - The Mill
15 - Zombie Burger
16 - Soseki Café
17 - Harold's Chicken #25
18 - Szechuan House
19 - Cactus Mexican Grill
20 - Thai Flavors
21 - Mami's Authentic Mexican Food Service
22 - Molly's Cupcakes
23 - Short's Burger & Shine
24 - The Airliner
25 - Pancheros Mexican Grill
26 - Mesa Pizza
27 - Food Republic
28 - Bo-James
29 - Osaka Japanese
30 - One Twenty Six
31 - Jimmy John's
32 - Coldstone Creamery
33 - Which Wich
34 - Z'Mariks Noodle Cafe
35 - Pullman Bar & Diner
36 - Dessert Shop
37 - Subway
38 - Forbidden Planet
39 - Aspen Leaf Frozen Yogurt
40 - Brothers Bar & Grill
41 - Saloon
42 - Donnelly's Pub
43 - Yotopia Frozen Yogurt
44 - PepperJax Grill