

**NCME Special Conference on Classroom Assessment  
and Large-Scale Psychometrics:  
The Twain Shall Meet**

**Conference Program  
September 12-14, 2017  
The University of Kansas  
Lawrence, KS**

**Table of Contents**

Welcome- 1

General Conference Information- 3

- Shuttle Bus Schedule
- WiFi
- Conference URL
- Media Permission
- Twitter
- Name Badges
- CEUs
- Print Services
- Lactation Room
- Gender Neutral Restrooms
- Reflection Room
- Lost and Found
- Local Time
- Wednesday Night: Downtown Lawrence
- Learn more about Lawrence

Schedule At-A-Glance- 10

Session Descriptions- 11

# Welcome

As President of the National Council on Measurement in Education, it's my honor and pleasure to welcome you to NCME's conference on classroom assessment. The conference came from Past President Mark Wilson's insight that most assessment occurs within the classroom context, that the NCME membership had much to learn from what was occurring in that context, and that classroom assessment practice might benefit as well from the principles and, perhaps, some of the methods the NCME membership had to offer. We owe a great debt to NCME's Classroom Assessment Task Force for their vision and dedication, and to Neal Kingston and the staff of the University of Kansas Achievement and Assessment Institute for being our host. To each of you, thanks for joining us in what we hope will be the start of an exciting direction for NCME!

Randy Bennett

President, NCME

\*\*\*

Welcome all to the inaugural National Council on Measurement in Education (NCME) conference on classroom assessment. The theme of this conference addresses the similarities (and differences) and synergies among classroom and large-scale assessment approaches. In addition, welcome to Lawrence, Kansas and the University of Kansas (Rockchalk, Jayhawk, KU!). I hope you enjoy your time here and find the conference informative and thought provoking.

It takes a village to raise a child, but also a village to host a conference. While I am honored to have been asked to organize and host this conference it would not have occurred without the efforts of many. The idea of NCME adding focus to classroom assessment was one of Mark Wilson's presidential initiatives. NCME Board members Dale Whittington and Kristen Huff were asked to lead an NCME Classroom Assessment Task Force, whose members include Heidi Andrade, Alison Bailey, Susan Brookhart, Caroline Wiley, Mark Wilson, and me.

The conference was guided by a planning committee whose members included Alexander Naumann, Alicia Alonzo, Angela Broaddus, Amy Clark, Melissa Collins, Jose Guadarrama, Leanne Ketterlin-Geller, Brad McMillen, Bryan Drost, Bruce Frey, Chad Gotch, Fengyi Hung, Susan Kahn Ehsan Kattoula, Dan Lewis, Christine Lyon, Emily Lai, Catherine McClellan, Jim McMillan, Michael Remillard, Melissa Strouth, Jon Twing, Monica Washington, and Anita Welch. The committee also served as proposal reviewers.

In addition, to her role on the conference planning committee, Amy Clark coordinated the technical aspects of the proposal review and session scheduling. As the conference project manager, Michelle Wilson, with additional conference support from Amy Jaimez and Ronda Consolver, provided the logistical support, without which this conference would not have occurred.

Special thanks are due to our platinum sponsors, ACT, edCount, Educational Testing Service, Pearson Educational Measurement, and Renaissance Learning; our gold sponsor, Curriculum Associates; and our silver sponsor: Houghton Mifflin Harcourt. Financial support for classroom teachers presenting at the conference was provided by our platinum sponsors. Finally, thanks to the Achievement and Assessment Institute at the University of Kansas, which covered the time of the conference support staff and me.

Your host,

Neal Kingston

# General Conference Information

## Shuttle Bus Schedule

Tuesday, September 12 reception service		
Comfort Inn	DoubleTree	Springhill Suites
5:30	5:35	5:30
5:50	5:45	5:50
6:10	5:55	6:10
6:30	6:05	6:30
6:50	6:15	6:50
7:10	6:25	7:10
7:30	6:35	7:30
7:50	6:45	7:50
8:10*	6:55	8:10*
	7:05	
	7:15	
	7:25	
	7:35	
	7:45	
	7:55*	
	8:05	
	8:15*	
	8:30	

\*Bus will take attendees to needed hotel

The conference reception is being held at the DoubleTree Hotel, beginning at 6:00 p.m.

### Bus identification

All buses will say "Private Shuttle" and are the large blue buses.

### Where to find the bus at each location

**Double Tree:** Pick-up/drop-off under the awning, by the front doors.

**Springhill Suites:** Pick-up/drop-off will be on the street at the 6th/New Hampshire cutout. Buses will not enter the parking lot.

**Comfort Inn:** Pick-up/drop-off will be on the north side of the hotel, on the side street off of N. Iowa. Buses will not enter the parking lot.

<b>Wednesday, September 13</b>				
<b>morning service</b>				
<b>DoubleTree</b>	<b>Comfort Inn</b>	<b>KS Union</b>	<b>Springhill Suites</b>	<b>KS Union</b>
			8:00	8:15
			8:20	8:35
8:10	8:15	8:25	8:40	8:55
8:30	8:35	8:45	9:00	9:15
8:50	8:55	9:05	9:20	9:35
9:10	9:15	9:25	9:40	9:55
9:30	9:35	9:45	10:00	10:15*
				11:00*
				12:00*
				13:00*
				14:00*
				15:00*

\*Bus will take attendees to needed hotel

Wednesday, September 13 post-event service			
Springhill Suites	DoubleTree	Comfort Inn	KS Union
			16:00
			16:20
16:10	16:20	16:25	16:40
16:30	16:40	16:45	17:00
16:50	17:00	17:05	17:20
17:10	17:20	17:25	17:40
17:30	17:40	17:45	18:00*

\*Bus will take attendees to needed hotel

The conference is at the Kansas Memorial Union (KS Union). You will enter the Union on the 4<sup>th</sup> floor. Registration is on the fifth floor and conference events are held on the fifth and sixth floors.

**Bus identification**

All buses will say “Private Shuttle” and are the large blue buses.

**Where to find the bus at each location**

**Double Tree:** Pick-up/drop-off under the awning, by the front doors.

**Springhill Suites:** Pick-up/drop-off will be on the street at the 6th/New Hampshire cutout. Buses will not enter the parking lot.

**Comfort Inn:** Pick-up/drop-off will be on the north side of the hotel, on the side street off of N. Iowa. Buses will not enter the parking lot.

**KS Union:** Pick-up/drop-off will be right outside the front of the Union at stop 173 on Jayhawk Blvd. Look for the buses that say “Private Shuttle” since there will be many buses going by the Union.

Wednesday, September 13 evening service			
DoubleTree	Comfort Inn	10 <sup>th</sup> /New Hampshire	7 <sup>th</sup> /Vermont
18:00	18:05	18:15	18:20
18:30	18:35	18:45	18:50
19:00	19:05	19:15	19:20
19:30	19:35	19:45	19:50
20:00	20:05	20:15	20:20
20:30	20:35	20:45	20:50
21:00	21:05	21:15	21:20
21:30	21:35	21:55*	22:00*

\*Bus will take attendees to needed hotel

Buses will run the evening of Wednesday, September 13, to shuttle guests from the DoubleTree and Comfort Inn to downtown Lawrence. There will not be a shuttle stop at the Springhill Suites since it is located downtown.

There will be two drop-off/pick-up locations downtown. Massachusetts Street is the main street in downtown Lawrence and is located between New Hampshire and Vermont streets.

### Bus identification

All buses will say “Private Shuttle” and are the large blue buses.

### Where to find the bus at each location

**Double Tree:** Pick-up/drop-off under the awning, by the front doors.

**Comfort Inn:** Pick-up/drop-off will be on the north side of the hotel, on the side street off of N. Iowa. Buses will not enter the parking lot.

**10<sup>th</sup>/New Hampshire:** Pick-up/drop-off in front of Sunfire Ceramics on New Hampshire Street.

**7<sup>th</sup>/Vermont:** Pick-up/drop-off will be at the bus hub across from the Public Library on Vermont Street.

### WiFi

- 1) View available wireless networks on your device
- 2) Select KU Passport
- 3) Enter the following login when prompted
  - Guest User Name: NCME
  - Password: H5nqlXob (middle character is a lower case L, not the number 1)

## **Conference URL**

<https://aai.ku.edu/ncme17-special-conference/registration>

## **Media Permission**

By registering for this conference, you grant permission to the National Council on Measurements in Education (NCME) and Achievement and Assessment Institute (AAI) at The University of Kansas to **photograph, videotape, audiotape, and otherwise record images** as part of the program identified below. You release NCME, AAI and the University from any and all liability resulting from the publication of said photos, videos, or remarks. Publications may include but are not limited to **website, publications, social media, and presentations**.

## **Twitter**

#NCME17

Join the Conversation

## **Name Badges**

Name badges are not only a distinctive fashion statement, they are an important way for conference staff to identify conference attendees. Please wear your name badge in all sessions.

## **CEUs**

If you paid to receive CEUs at the conference, please stop by the Conference Registration Check-in desk, located on the 5<sup>th</sup> floor, to sign the CEU Information sheet.

## **Print Services**

Jayhawk Ink is located on the 2<sup>nd</sup> floor (the trade bookstore on the west side) can make copies and print from a flash drive (see Kansas Union Map- 9).

## **Lactation Room**

The Lactation Room is located on the 5<sup>th</sup> floor in the Regionalist Room (see Kansas Union Map- 45).

## **Gender Neutral Restrooms**

Gender Neutral Restrooms are located in the Student Involvement and Leadership area on the 4<sup>th</sup> floor. You may ask for directions at the Welcome Desk on the 4<sup>th</sup> floor (see Kansas Union Map- 30).

## **Reflection Room**

There is a Reflection Room located on the 3<sup>rd</sup> floor (see Kansas Union Map- 22 a).

## **Lost and Found**

If you find an item or are missing an item, please come to the Conference Registration Check-in desk, located on the 5<sup>th</sup> floor, turn in an item or inquire about a missing item.

## **Local Time**

Kansas time zone is: Central Time Zone

## **Wednesday Night: Downtown Lawrence**

Sometimes the best kind of breakout session is the one you didn't have to plan. After getting down to conference business Wednesday, you'll have the opportunity to hop on a shuttle bus and experience an evening in our historic downtown, a lively, easily walkable district packed with food, fun, shopping, visual art, live music, and more.

Downtown Lawrence Passport (found in bag at conference registration check-in)

- Deals and discounts at over 25 Downtown Lawrence Inc. member businesses
- Present passport at time of purchase. Offers good through 9/30/17 & are one-time only use.

Pub Crawl

- Jefferson's Restaurant, 743 Massachusetts St.:  
For the drink side of specials, Jefferson's would like to offer NCME Conference Attendees their happy hour specials (usually from 3pm to 6pm) from 6pm to 10pm. \$4 23oz Domestic Drafts, \$3 Glasses of wine, \$3 craft bottles, \$2 Domestic Bottles, and \$2 Single Wells.
- Red Lyon Tavern, 944 Massachusetts St.:  
For specials on Wednesdays, they offer a pitcher special from 4:00pm-7:00pm that all pitchers (55oz.) are \$10.00. From 7:00pm until closing time, they have a rotating pint special on a few taps that varies week to week. They will make sure to have a few craft beer taps on special this Wednesday night.
- The Sandbar, 17 E 8th St.:  
\$2 wells and they will be having the wheel of happiness that spins every 30 minutes for shot specials.

### **Wednesday Night: Downtown Lawrence Shuttle Bus Service**

Wednesday evening shuttle bus service from the DoubleTree and Comfort Inn & Suites to Downtown Lawrence will begin at 6:00 p.m. and continue through 10:00 p.m. Downtown drop-off/pickup will be at two locations: at 10<sup>th</sup> & New Hampshire and 7<sup>th</sup> & Vermont. The drive downtown is about 15 minutes one way; the loop from the DoubleTree and Comfort Inn & Suites to the two Downtown Lawrence drop-off/pickup points and back takes approximately 35 minutes to complete.

**Learn more about Lawrence**

- Local food, arts, and entertainment: [LAWRENCE.COM](http://LAWRENCE.COM)
- Visitors guide: [EXPLORELAWRENCE.COM](http://EXPLORELAWRENCE.COM)
- Local news & weather: [LJWORLD.COM](http://LJWORLD.COM)

## Schedule At-A-Glance

### Tuesday, September 12

5:00 pm – 6:30 pm Conference Registration Check-In at DoubleTree  
6:00 pm – 8:00 pm Opening Reception at DoubleTree

### Wednesday, September 13

8:00 am – 5:00 pm Conference Registration Check-In  
9:00 am – 10:00 am Opening keynote  
10:00 am – 10:30 am Break  
10:30 am - noon Session 1  
Noon – 1:00 pm Lunch  
1:00 pm - 2:15 pm Session 2  
2:20 pm – 3:35 pm Session 3  
3:35 – 4:00 pm Break  
4:00 – 5:15 pm Session 4

### Thursday, September 14

8:00 am – 10:00 am Conference Registration Check-In  
8:30 am – 9:30 am Session 5  
9:35 am – 10:35 am Session 6  
10:35 am – 11:00 am Break  
11:00 am – 12:15 pm Session 7  
12:15 pm – 1:15 pm Lunch  
1:15 pm - 2:15 pm Session 8  
2:20 pm – 3:30 pm Closing plenary session

# SESSION DESCRIPTIONS

Wednesday, September 13

**9:00 am – 10:00 am**                      **Opening Keynote: Classroom Assessment and Large-Scale Psychometrics: Shall the Twain Meet?** Margaret Heritage & Neal Kingston  
Location: Woodruff Auditorium

Classroom assessment and large-scale assessment have, for the most part, existed in mutual isolation. Some experts have felt this is for the best and others have been concerned that the schism limits the potential positive impact on practitioners of both forms of assessment. Margaret Heritage has long been a champion of best practices in classroom assessment. Neal Kingston has been involved with the application of psychometrics to large-scale assessments for 39 years. Together they will ask what commonalities and differences exist between these two assessment contexts, whether the twain should meet, what impediments or concerns exist, and whether they expect the status quo will change at all in the near future.

**10:30 am - noon**                      **Session 1**  
**1A** Preparing Teachers to Use Evidence Centered Design in the Classroom- Min Li, Susan Elizabeth Cooper, & Catherine Taylor  
Location: English Room

In this session, three teacher educators will discuss the following question:  
How do we prepare teachers to consider (the relevant aspects of) psychometrics when selecting, developing, and using assessments?

We address these questions from three points of view: a classroom assessment specialist, a teacher researcher, and a psychometrician.

Background:

Lately psychometrics has been equated with the mathematical aspects (scaling, equating, calibration, etc.) rather than the full breadth of a psychometrician's work. From an evidence centered design (ECD) view, every step– from construct definition to reporting – is an aspect of psychometrics.

Figure 1 shows the typical pathway of work in large scale psychometrics.

(Insert Figure 1 here)

Figure 1: Test Development Stages for Large Scale Tests

The most critical task in ECD is construct or domain definition. The second most critical task in ECD is defining how we want assessment results to be interpreted and used.

When engaged in ECD, every action in the cycle is purposefully driven by construct definition and intended interpretation and uses of results. We gather evidence during every stage to support (or refute) our claims about score meaning and whether scores can

be used for intended purposes.

What stages of this work apply to classroom assessment?

Figure 2, shows the tasks from Figure 1 redefined to be relevant to classroom assessment. We phrase them as questions teachers can ask themselves as they plan, develop/select, and use assessments and assessment results.

(insert Figure 2 here)

Figure 2: Classroom Assessment Cycle

The tasks in Figure 2 guided the design of an assessment course at the University of Washington and are elements of the edTPA (a teacher performance assessment now being used in 20 states). We will briefly 1) show the connection between the classroom cycle, course assignments, edTPA tasks. We present data from two instances of tasks in this cycle. The first involves teachers' evaluations of assessments from 5 elementary reading curricula and 7 elementary mathematics curricula. Teachers found published assessments were mostly poorly aligned with standards and instruction and of poor technical quality. We will also share the results of a recent study on the dilemmas AP teachers in low income schools face during grading – in particular, how to achieve conflicting grading purposes between communicating college expectations and encouraging low income students to remain in the courses.

**1B Using Learning and Motivation Theories to Coherently Link Formative Assessment, Grading Practices, and Large-Scale Assessment-** Lorrie Shepard, Jim Pellegrino, Bill Penuel, Lizette Burks, Suzanne E. Myers, & Danira Fernandez-Flores  
Location: Kansas Room

In *Knowing What Students Know* (Pellegrino, Chudowsky, & Glaser, 2001), the NRC committee argued that to be effective assessment observations and inferences must be grounded in a theory of learning. Penuel and Shepard (2016) identified four different theories of learning being used (implicitly or explicitly) to further formative assessment practices. To extend this work and to follow KWSK's call for systems that are vertically coherent, authors Shepard, Penuel, and Pellegrino are developing a new paper to examine how the learning theories underpinning formative assessment can be made coherent with classroom grading practices and with large-scale assessment. They focus in particular on the research evidence supporting sociocognitive and sociocultural perspectives on learning and motivation.

For this conference session, we propose a modified version of the “Book Club” format. Lead presenters Shepard, Penuel, and Pellegrino will distribute their paper (same title as this session) in August to participants and panelists. The three leads will make brief 10-minute presentations addressing formative assessment to enhance learning, grading, and large-scale assessment (for both program evaluation and accountability purposes). Then audience small-group discussions will be facilitated addressing questions such as the following:

- Should formative “assessments” be graded, and if so, when?
- What are the pros and cons of discipline-specific versus generic formative assessment

professional development?

- What does research say about grading for effort and/or improvement?
- How do arguments for norm-referenced versus criterion-referenced (so-called “standards-based”) grading practices play out differently for kindergarteners versus 11th graders?
- What should be done when state tests are misaligned with local district curricula?
- How could states support more ambitious teaching practices and correspondingly rich assessments without encroaching on local control?

Following group discussions, invited panelists will give their responses to these questions. Audience table groups will then be called upon to give their responses if their arguments are different from what has already been said. The Shepard, Penuel, and Pellegrino paper was already under development for submission to EM:IP. The manuscript will be revised based on conference critique and input.

**1C Pre-/Post-Tests for Data Driven Decision-Making: Perspectives from Teachers, Administrators, and Psychometricians- Alicia C. Alonzo, Alexander Robinson, Samantha Duwe, Malcolm Bauer, & Derek Briggs**  
Location: Malott Room

As data driven decision-making gains traction in school districts across the country, one common way of obtaining data about student learning is pre-/post-unit tests. From tests and/or assessment questions drawn from commercially-available curricula and assessment materials to instruments developed by individual teachers or groups of teachers (e.g., in professional learning communities), these pre-/post-tests may be derived from a variety of sources. Once students have responded to these pre-/post-tests, the data may be represented in a variety of forms, including automatically-generated score reports included with assessment materials or generated by grading software. This session will bring together two groups of people with critical roles in the development and use of pre-post-tests and associated score reports: 1) psychometricians who focus on the development of assessments and score reports and 2) teachers and administrators who develop and use assessments for school-based data driven decision-making. The session will be divided into two parts to maximize participants’ engagement with each other. First, a moderated panel will model a discussion between psychometricians and teachers/administrators about pre-/post-tests in science. Second, participants will be a) broken into roundtable groups in order to discuss pre-/post-tests in grade-level and subject-area-specific groups and then b) brought back together to synthesize ideas across groups. In both parts of the session, we expect psychometricians to share principles of test development and ideas for different formats for looking at pre-/post-test data and teachers/administrators to share the features of pre-/post-tests and score reports that are most useful to them and their “wish lists” for future development of pre-/post-tests and associated score reports. Thus, both groups of participants should come away from the session with deeper insights into the school-based use of pre-/post- tests for data driven decision-making.

**1D** Guidance in the Standards for Classroom Assessment Practices to Support Instructional Decisions: Useful or Irrelevant?- Steve Ferrara, K. Maxey-Moore, & Susan Brookhart  
Location: Centennial Room

The Standards for Educational and Psychological Testing (2014) “provide criteria for the development and evaluation of tests and testing practices and...guidelines for assessing the validity of interpretations of test scores for their intended purposes” (p. 1). The Standards are developed for testing situations where intended interpretations and uses of test scores require significant psychometric and procedural rigor. Although the Standards explicitly do not extend to classroom assessment practices (p. 2; see also Ferrara, 2014), selected standards may provide guidance to help teachers ensure that the information they gather via classroom assessment practices supports the interpretations and uses they intend.

In this panel discussion, a psychometrician and a school district administrator will discuss translations of selected psychometric standards for application for two classroom assessment situations: (a) Teacher-made and selected instructional unit tests; (b) Teacher inferences about student understanding of instructional concepts and procedures, based on observations and discussions with students during instruction; and (c) Teacher inferences about student progress in relation to student learning objectives (SLOs). The presenters also will consider the selected psychometric standards for commercial formative assessment products (e.g., interim assessments) and services (e.g., formative assessment item banks and testlets). The table excerpt below illustrates an example of translating the Standards to classroom assessment situations. Presenters will consider in detail a small selection of psychometric standards, translated for classroom assessment situations, from a comprehensive collection of psychometric standards. A widely known thought leader in classroom assessment practices will act as discussant for the panel discussion.

Each panelist will have five minutes to make opening comments, followed by discussion among the three panelists of the standards that are selected and translated for classroom assessment. Audience members will be invited to participate in the discussions.

#### References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standard for educational and psychological testing. Washington, DC: American Educational Research Association.

Ferrara, S. (2014). Formative assessment and test security: The revised Standards are mostly fine; our practices are not (invited commentary). *Educational Measurement: Issues and Practice*, 33(4), 25-28.

Table X. Psychometric Requirements for High-stakes Tests and Translations for Classroom Assessments (Table excerpt available from the organizer upon request; I couldn't paste the Word or pdf version into this form)

**1E Navigating Data Use in a Balanced Assessment System- Caroline Wylie & Christine Lyon**

Location: Pine Room

This interactive training session will focus on issues related to the continued growth and desire for balanced assessment systems. We recognize that while there are volumes of data available to all stakeholders in K-12 school systems, stakeholders need mechanisms to help them maximize the value of each data point at the appropriate time in a school year. Different types of data can be used across the year for informed decision making (e.g., using summative assessment formatively for curriculum review and unit planning, using interim assessment data to confirm or adjust long range planning, using formative assessment for short-cycle adjustments). There are a variety of representations of assessment systems: Wiliam's (2006) cycles of assessment; Shavelson et al's (2008) continuum of formal to informal formative assessment; the triangle of balanced assessment (Perrie, Marion & Gong, 2009); Brookhart's (2013) assessment quadrants. We will invite participants to connect these ideas to their own experiences of data use in their schools and districts.

This workshop will engage participants in discussion about appropriate decision-making across components of a balanced assessment system. The learning goals for the session are for participants to understand:

- Appropriate uses of summative, interim and formative assessment information
- Understand different frameworks or representations of how the range of assessment information can be integrated.
- Models of activities that can be used with teachers to engage them in discussion around the appropriate use of assessment data

This session will use a combination of presentation, small group activities, whole group discussion and reflection on video exemplars. The session is intended for a wide range of practitioners including teacher leaders, grade level team leads, department chairs, content supervisors, principals, district administrators with an assessment focus.

**1F Arts Assessment for Learning: Practice and Research- Heidi Andrade, Maria Palma, Joanna Hefferen, & Angela Fremont**

Location: Divine Nine Room

This presentation will introduce a decade-long collaboration on formative assessment in the arts. Supported by a variety of external funders, the Arts Assessment for Learning work provides high-quality professional development to hundreds of music, dance, theater, and visual arts teachers in New York City. The purposes of the professional development include helping arts educators assess their students' arts learning based on local and national standards, promote their students' arts achievement, and share their practices with one another. The structure of the professional development includes workshops about formative assessment, participation in professional learning communities, and action research. The outcomes of this work include a website that

makes best practices available to arts educators everywhere (<http://artsassessmentforlearning.org/>), as well as experimental studies of the effectiveness of formative assessment in promoting student achievement in the arts (Chen, Lui, Andrade, Valle, Mir, 2017; Chen & Andrade, 2016; Mastroilli, Harnett, & Zhu, 2014). In this presentation, we will share the innovative ways in which formative assessment is being implemented in the arts, as well as the empirical evidence of their effects on student achievement and teachers' assessment literacy. Given the collaborative nature of this work, the presentation will be given by a university researcher, a NYC DOE director, a professional development provider, and hopefully an art teacher (depending on the availability of funding).

**1:00 pm - 2:15 pm                      Session 2**

**2A** The complexity of complexity: Disentangling cognitive complexity from item difficulty- Sue Brookhart, Mark Wilson, Pooja Shivraj, Deb Perry & Leanne Ketterlin Geller  
Location: English Room

Few measurement specialists, researchers, and practitioners would disagree with the statement that cognitive complexity is different from item difficulty. However, when faced with the task of disentangling the two, the situation itself becomes considerably more complex. Complexity can be defined as an increase in sophistication of the content along a progression or trajectory, as the depth of knowledge that characterizes the complexity of students' actions, or from a disciplinary perspective where some knowledge, skills, and abilities are pre-requisite to others and/or elicit more or less complex cognitive processes.

Item difficulty, however, can be influenced by a number of different factors that are both relevant to and distal from the assessed construct. In some ways, the distal factors that influence item difficulty (e.g., readability, graphics, visuals) are much easier to spot. However, the impact of content relevant features is more difficult to disentangle. When taken together, the content-relevant and irrelevant item features may make it difficult to determine what is actually contributing to item difficulty estimates. In the context of classroom assessments, the distinction between cognitive complexity and item difficulty grows murkier, with practitioners often assuming that hard items must be testing complex content.

This presentation is divided into three parts. First, two measurement and assessment experts with extensive knowledge in classroom assessments will discuss the issues involved in disentangling cognitive complexity from item difficulty through a moderated discussion, bringing in evidence from cognitive labs and psychometric analyses. Next, two practitioners (one representing the district-level perspective and an experienced practitioner) will focus on instantiated classroom assessment practices and the impact of the confounding nature of complexity and difficulty on classroom level decisions. Third and finally, the panel will engage with the audience to establish a series of guidelines for practitioners to consider when writing and evaluating items for complexity and difficulty.

**2B** Designing and Using Instructionally-Supportive, NGSS-aligned Science Assessments with Diverse Learners- James W. Pellegrino, Brian Gane, Sania Zaidi, Christopher Harris, Kevin McElhany, Nonye Aloze, Phyllis Pennock, Joseph Krajcik, Daniel Damelin, Ann Novak, & Jenny Sarna

Location: Kansas Room

We propose an interactive, structured poster-symposium to illustrate various strands of our work related to developing and validating instructionally supportive assessments aligned with the Next Generation Science Standards. In the process of developing technology-enhanced assessments that can be used in classrooms with diverse learners, we have encountered many of the same issues that the NCME community faces when developing large-scale assessments. These issues include the articulation of an appropriate validity framework, designing rubrics that promote scoring reliability, identifying task features that make assessments fair for diverse learners, and taking advantage of technology affordances for assessment delivery. We have also confronted issues unique to classroom assessment, such as the development of rubrics that can be used by teachers and educative resources that enable the effective integration of assessment and instruction.

In a 90- minute session we will provide a brief introduction to the scope our work to orient the audience to the broad project goals, significant accomplishments, and salient challenges faced over the last 4 years. We would follow with a 1-2 minute summary from each poster presenter on the content of his/her poster and demonstration material. We would then allow 45-60 minutes for attendees to interact with the presenters. Finally, we would reserve 20 minutes at the end of the session for a discussant to moderate interaction among all attendees and presenters.

We plan on approximately 8 posters and demonstrations presented by an interdisciplinary group of junior and senior project personnel, a teacher who has been using the assessment tasks, and a district administrator overseeing creation of a community of practice for teachers who use the assessment tasks. This mix of presenters and topics will advance NCME's efforts to support discussion and collaboration among members of NCME and individuals concerned with the design and use of instructionally supportive classroom assessment.

**2C** Supporting Instruction for Students with Significant Cognitive Disabilities Using Online, Instructionally Embedded Assessments- Meagan Karvonen, Russell Swinburne Romine, Emily Thatcher, Jennifer Denne, & Laurel Friedman

Location: Centennial Room

The Dynamic Learning Maps (DLM) Alternate Assessment Consortium features online, instructionally embedded assessments that are available throughout the year for students with significant cognitive disabilities in participating states. The assessments make use of a flexible blueprint, allowing teachers to choose which content standards to assess their students on from among a list of requirements. DLM Assessments are built using an architecture of learning map models, research synthesis representations that show how students develop knowledge, skills and understandings over time. Assessments are

administered as short three- to eight-item testlets following an engagement activity. Additionally, testlets for each standard are available at five levels of complexity, with a level recommended by the system to ensure all students can access the content. The DLM system was designed to meet states needs for accountability purposes while providing instructionally relevant assessments that provide meaningful information for teachers and IEP teams. The proposed session will include a panel of test developers, state education agency partners, and educators and from two participating states. The panel will discuss the benefits of administering instructionally embedded assessments and how they use results from the assessment to inform instructional decision making. The panel will also discuss ways to use progress reports to develop individual education programs (IEP) and how professional development opportunities, such as professional learning communities, can support teachers' use of assessments to promote student learning.

**2D** We have a report for that: Examining educators' use of interim and summative assessment reports- Brian Gong, Dan Gruman, Darcy Swan, Erin Smith, Kristy Fornal, Krista Carson, & Joanna Roche  
Location: Pine Room

Assessment validity and usefulness stem from the interpretations that can be made from assessment results, typically from assessment reports. Recently much attention has been paid to trying to get more useful information from summative assessments.

This session involves a case study addressing how educators actually use interim and summative assessment reports for a variety of instructional purposes. The session is framed by a presentation by Brian Gong (Center for Assessment) on the characteristics of summative assessments that have kept them from being very useful to educators (e.g., lateness of delivery, dearth of detail, lack of content connections), and the characteristics of interim assessments that make them evidently more useful to educators but which often incur criticism from measurement professionals (e.g., lack of validity evidence, lack of standardization, often easily misinterpreted metrics). Gong will suggest ways that educators can successfully merge information from multiple reports, overcoming the shortcomings of each. He will also suggest caveats.

The case study will be provided by an outstanding classroom teacher/department team leader and a school principal who will demonstrate how they have used their 12+ different reports provided by their commercial vendor, along with their state summative reports, and other school- and classroom-generated information to make decisions that have led to dramatically higher student performance. The case study emphasizes the role that deep knowledge of content, pedagogy, students, and school organization play in using assessment reports well. Fascinating aspects include illustrations of how these experienced educators make judgments involving what measurement experts would recognize as rough equating of tests, analysis of reliability/generalizability, and alignment, while keeping an eye on external and internal validation.

**2E** I've a feeling we're not in Kansas anymore! - Expanding large-scale psychometricians' perspectives on assessment- Karen Barton, Kristen Huff, Paul Nichols, & Charlie DePascale

Location: Malott Room

As with multicultural education, the goal in bringing together professionals engaged in classroom and large-scale assessment is co-existence not confluence; fostering an appreciation of the differences in assumptions, values, behaviors, patterns of thinking, and communicative styles across diverse cultures. Darling-Hammond's (2002) thoughts on multi-cultural education in schools apply equally well to our assessment community, "acknowledgment of diverse experiences creates new associations that helps us ultimately to build the common ground in which a more inclusive and powerful learning community can rest."

This session will engage panelists and audience members in critical discussion about those differences between the purposes and uses of classroom and large-scale assessment that give each unique measurement and psychometric needs. Panelists will offer perspective gained from careers seeking the intersection between the purposes and constraints of large-scale assessment and the need to inform the placement and/or instruction of individual students. We will discuss issues such as

- Avoiding a repeat of the assessment spring of the mid-1990s when great work in performance assessments, portfolios, and projects which were authentic, classroom-embedded, and had the potential to provide good data to inform instruction and improve student learning were tossed to the curb in favor of more reliable, psychometrically sound assessments.
- Promoting the concept that balanced assessment systems are not systems in which large-scale psychometrics are applied to all assessments, but rather systems in which sound assessment and measurement practices and psychometrics are applied in a way that ensures "the alignment of different assessments to different consumer's information needs such that the needs of all consumers are met".
- Understanding the emergence of Big Data and learning analytics and ensuring their peaceful coexistence with classroom assessment and psychometrics
- Improving the assessment literacy of all stakeholders, including psychometricians, to have the best chance of using assessment to promote successful instruction and learning.

**2:20 pm – 3:35 pm                      Session 3**

**3A** Educator Professional Learning in a Balanced Assessment System- Chad M. Gotch, Christine J. Lyon, Cristen McLean, & E. Caroline Wylie

Location: Malott Room

Various models for balanced assessment systems have been proposed in order to help stakeholders identify what to assess, how to assess it, and how to use the assessment evidence to inform decisions (e.g., Herman & Heritage, 2007; Perie, Marion, & Gong, 2009; Brookhart, 2013). While the models provide frameworks that can help stakeholders make decisions about assessment types, they do not provide guidance for how to develop assessment capacity. This session will draw on the existing models, Coburn and Turner's

(2011) framework for teachers' data use, and the literature on teacher expertise to propose a model for sequencing teacher learning and building assessment capacity.

The presenters represent complimentary professional roles (e.g., researchers, university faculty, state agency staff), and have diverse areas of expertise (e.g., pre-service teacher preparation, in-service assessment professional learning, assessment literacy measurement, and program evaluation). They will draw from program evaluations findings, professional development observations, and experiences related to building teacher assessment capacity to support the proposed model.

Participants in the session will be provided access to four examples of publicly available professional development resources including: (1) a website for scoring calibration on state assessment performance tasks; (2) a blended learning course for formative assessment practices; (3) facilitator and participant resources for professional learning communities focused on formative assessment practices; and (4) facilitator and participant resources for a workshop on performance assessments. Participants will first each engage with one set of resources. Small group discussions will focus on how the resources relate to the model, affordances and constraints related to the mode of delivery, and how resource efficacy could be measured. The results of the group discussions will be shared to document overall feedback on the proposed model for building assessment capacity.

**3B** Beyond State-Standardized Assessments: A Large Urban School District's Approach to Measuring and Monitoring Student Learning- Debbie Durrence, Leslie Aiken, & Aminah Perkins

Location: Centennial Room

For Gwinnett County Public Schools (GCPS), serving over 178,000 students, developing, implementing, and sustaining a balanced assessment system is key to our mission of becoming a system of world class schools. The GCPS balanced assessment system includes summative, interim, and formative measures that provide data to various stakeholders to ensure students are mastering GCPS' academic knowledge and skills. An integral component is GCPS benchmark assessments, administered to grade 1-12 students. They are designed to measure student mastery of curriculum during a specified instructional period. GCPS interim assessments provide opportunities for teachers to evaluate the degree to which each student has mastered the curriculum, and if needed, make necessary adjustments in instruction prior to summative assessments.

Every step of the assessment development process involves vital stakeholders including GCPS teachers, curriculum directors, assessment directors, assessment coordinators, and psychometricians. With over 450 district-developed assessments, processes must be comprehensive and detailed in order to maintain a balanced assessment system.

The presentation will illustrate the development process of benchmark assessments including the creation of instructional calendars, designing of blueprints, item development, implementation, and subsequent data analysis and use. As key

stakeholders, teachers involved in the process undergo extensive, specialized training to build capacity in the areas of assessment literacy and item development. The presentation will highlight strengths and weaknesses of the processes, lessons learned, and opportunities for growth from a district perspective.

GCPS is committed to effectively employing a balanced assessment system, and it serves as a strong example of how a large district can successfully implement an iterative and cyclical process that produces data that is used at the district, school, and teacher level to improve student learning. During this panel discussion we will be soliciting input and feedback from the participants to improve the processes within our assessment system.

**3C Performance Assessment in the Classroom, Building, and District-** Susan Brookhart, Rebecca Greer, Shawnee Heights USD 450 Principal & teachers  
Location: Kansas Room

Performance assessment uses observation and judgment to describe (formative) or evaluate (summative) students' knowledge and skills. Performance assessment is well suited to assessing complex learning outcomes, especially those requiring any of the thinking skills collectively known as higher order thinking.

Shawnee Heights Unified School District 450 has focused on performance assessment for 3 years. Currently, classroom teachers are encouraged to use performance assessment as a regular part of their classroom assessment. In addition, teachers are required to use data from at least one performance assessment (and preferably more) as part of the evidence they offer for their annual review. Both the quality of the performance assessment and the student results are important in this review.

Becky Greer, who directs curriculum, assessment, and federal programs for the district, has spearheaded professional development for teachers and administrators in the use of performance assessment. The development began as an outgrowth of common formative assessments and the desire to better indicate what students know and can do regarding complex learning outcomes. The district values robust standards where students build an inquiry based on questions or problems and then go about solving them, while demonstrating strong background knowledge. They report, "We needed to be able to assess what we value in a way not measured by state assessments or standardized tests."

In this team demonstration session, Sue Brookhart will give a brief overview of performance assessment. Becky Greer will describe the performance assessment initiative in Shawnee Heights USD 450. A building administrator will explain how she supports teacher use of regular performance assessment in the classroom and what that has done for student learning and teacher practice in her building. Several classroom teachers will focus on how their use of performance assessment in the classroom has changed and show examples documenting student learning.

**3D Poster Session**

- Use of Technology-based Diagnostic Assessment Tools in the Classroom- Zachary Feldberg & Laine Bradshaw  
Location: Jayhawk Room

The focus of our presentation is the classroom use of technology- based diagnostic assessment systems. Our research questions focus on (a) what diagnostic assessment technology systems are used and how, (b) what are some strengths and limitations of common systems, and (c) what do teachers desire from diagnostic assessment systems. We conducted a survey of districts throughout a southern state to determine what diagnostic assessment systems are widely used. Commonly used systems were analyzed to determine their psychometric properties and to choose a variety of systems for further inquiry. Finally, teachers were interviewed to explore how teachers use diagnostic assessment systems and what opinions teachers have of such systems. Our findings will be summarized and reproduced into a poster for presentation at the NCME special conference.

- Game-based Classroom Assessment of Collaborative Problem Solving- Kristin Stoeffler  
Location: Jayhawk Room

This presentation showcases a subset of the broad range of skills that are required to thrive in a 21st century society rich with both information and opportunity. These same skills empower students to fulfill their potential as effective and creative knowledge seekers and collaborative problem solvers. This presentation focuses on collaborative problem solving and the skills required to effectively combine problem solving and behavioral strategies to successfully solve a problem within a team context. ACT's Collaborative Problem Solving construct (CPS) includes over 20 facets related to CPS. We chose five of these facets: perspective taking, persistence, goal reflection, strategy, and problem feature awareness. A goal of this work was to explore the feasibility of measuring these facets in an interactive and engaging way that will eventually allow for the ability to provide insights and feedback to participants and teachers. In the collaborative game, Circuit Runner, a player navigates a circuit board maze and interacts with a student "bot" to solve a number of challenges presented at locked gates. Varying levels of team effectiveness and task effectiveness skills are required to successfully navigate the challenges and unlock the gates. Various forms of the game have now been played by participants ranging from middle school to the workforce. The game was also accompanied by a set of game usability and research survey questions. We will review game features designed to elicit evidence of CPS skills, insights gleaned from the design and development processes, as well as discuss findings that highlight the importance of these five facets and their interactions. This type of classroom assessment provides rich information for both teachers and students, around skills students use across content domains, and are thus critical to be better understood as students work individually as well as collaboratively.

- Utilizing Common Assessments to Measure and Reduce Knowledge Decay- Jennifer Reimers  
Location: Jayhawk Room

Each year teachers are faced with the task of re-teaching material from preceding grade levels to overcome summer learning loss and knowledge decay. Knowledge decay refers to the fading of neurochemical pathways, “memory traces,” making it more difficult to retrieve information over time (Thorndike, 1914). Several studies have found that assessments may not only be used for directing instruction, setting standards, and evaluating progress, but also for student learning and reducing knowledge decay (Larsen, Butler, & Iii, 2008; Leeming, 2002; McDaniel, Roediger, & McDermott, 2007). The “Testing Effect” or “Test-Enhanced Learning” is one example of how repetitive and spaced assessment may improve knowledge retention (Gates, 1917).

This is a proposal suggesting the use of common assessments (CAs) and collaboration within Professional Learning Communities (PLCs) to measure and reduce knowledge decay among students. This may be accomplished using five steps:

1. Collaborate within department level PLC to create CAs
2. Administer the end-of-year CA at each grade level. Then administer the same or parallel preceding grade level CA at the beginning of each school year
3. Work together with grade level (department specific) cohorts to identify areas at risk of knowledge decay.
4. PLC devises plan to reteach at-risk areas. A report is given to the preceding grade level teachers and the PLC works to brainstorm alternative teaching methods. Additionally, teachers with students who showed high knowledge decay in areas could seek consult from teachers with students who showed low knowledge decay in the same areas.
5. Teachers spiral the at-risk curriculum throughout instruction as well as formative and summative assessments

An overview of how teachers can use assessments for student learning and reducing knowledge decay will be discussed and comments or suggestions to the model proposed will be encouraged.

#### References:

E. L. Thorndike, *The psychology of learning*, N. Y., Teachers College, 1914, p. 4.  
Gates, A. I. (1917). "Recitation as a factor in memorizing". *Archives of Psychology*. 6 (40).

Larsen, D. P., Butler, A. C., & Iii, H. L. (2008). Test-enhanced learning in medical education. *Medical Education*, 42(10), 959-966. doi:10.1111/j.1365-2923.2008.03124.x

McDaniel MA, Roediger HL III, McDermott KB. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev* 2007;14:200–6.

Leeming FC. The exam-a-day procedure improves performance in psychology classes. *Teach Psychol* 2002;29:210–2.

- Qualitative Inquiries of Teachers' Experiences with Instrument Development- Jade Caines Lee  
Location: Jayhawk Room

Formative assessment, or assessment for learning, is a critical component of effective classroom instructional practices (Black & Wiliam, 1998; Daws & Singh, 1996; Gearhart et al., 2006; Pellegrino et al., 2001; Ruiz-Primo, 2011). Within the formative assessment literature, there is typically a focus on formative assessments for students that can be useful in providing teachers with information to assist in adjusting their instruction (Cizek, 2010). There is much less literature, however, on classroom-based assessments for teachers, despite evidence that demonstrates how classroom observations can improve the quality of teachers' instructional practices (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Bill & Melinda Gates Foundation, 2014; Hill, et al., 2012). Observations can clarify expectations for teaching, provide tailored feedback to support teachers in elevating their practice, and provide useful information to inform professional development decisions (Little, Goe, & Bell, 2009). In order to encourage formative assessment practices and maximize the benefits of classroom observations, teachers must have a certain level of assessment literacy skills. There are, however, barriers that exist in developing teachers' assessment knowledge. The focus of this research poster is to highlight these barriers and challenges.

#### The Context

We developed an observational instrument (the Science Practices Observation Protocol, or S-POP) intended to measure middle school science teachers' implementation of science practices aligned to the Next Generation Science Standards. After the initial development phase, we conducted cognitive/think-aloud interviews within a focus group format (n = 8). Research findings suggest that the utility of the S-POP is determined by teachers' assessment literacy skillset.

This research has the potential to enhance dialogue between several education stakeholders. By privileging teacher voices, educators can make significant contributions to research on assessment literacy practices. Also, psychometricians and measurement researchers can gain better insight into the barriers and challenges that educators face in utilizing formative assessment measures.

- Students Performing Above or Below Expectations on High-Stakes Secondary Exams and First-Year University Achievement- Russ Kupperstein  
Location: Jayhawk Room

Motivation amongst students in the second semester of senior year generally is thought to decline as unconditional university offers are made for enrollment the following year. For students enrolled in international curricula with high-stakes culminating exams, such as International Baccalaureate or Advanced Placement, extrinsic motivation may suffer as university signing day precedes the start of the May exam session. This research study serves to explore the relationship between intrinsic motivation, follow-through on external exams sat at the conclusion of secondary school, and student achievement in first-year university classrooms. If students do not need to score well, do they try anyway? Is there any difference between those who are motivated and those who are not as they complete coursework at the next level? A poster presentation of evidence and analysis of data and interviews collected from graduates (2006-2016) of Cairo American College, an independent school delivering IB and AP curricula, will be considered and discussed.

- Integrating Assessment Data from a Variety of Sources- William P. Skorupski & Dianne Henderson  
Location: Jayhawk Room

Although, classroom assessment seems like it is as simple as collecting data, looking for mastery of content and guiding instruction, done well, these things are much more complex than they sound. In addition the teacher must not only collect the data, but also integrate it into actionable information to assist in planning instruction. The purpose of this session will be to describe how the use of technology and advanced psychometric models can be used to integrate data from a variety of sources. This interactive poster session will describe the underlying multi-dimensional item response theory model and the process of adapting this model to actual student data from a variety of assessments. The session will provide an opportunity for participants to better understand the underlying model and the practical issues associated with implementing this approach. The session will also describe the practical application of the results for use in the classroom.

- Measurement in an Online Learning Environment: K-12 Student Engagement- Elizabeth Anderson  
Location: Jayhawk Room

Measurement and assessment in an online learning environment have different accessibility and proctoring requirements yet most of the current measures of student engagement in K-12 education are designed for brick-and-mortar learning environments. The purpose of this research study was to develop a measure of online student engagement for grades 3 through 8 using a partial credit Rasch model and validate the measure using confirmatory factor analysis. This research study used tracked online student behaviors as items. Online student behavior items were converted from continuous to categorical after assessing indicator strength and possible inverted U relationship with academic achievement. The measure development and item categorization processes resulted in an online cognitive engagement measure and an online behavioral engagement measure for grades 3

through 8, with each grade having its own measure. Future research will include measure development specifically for students receiving special education services, comparing measures developed using the original continuous items without categorization, identification of facilitators of online student engagement for grades 3 through 8 and further evaluation of the relationship between online student engagement and academic achievement.

- Effects of Writing Skill on Students' Explanations of Their Mathematical Understanding- Tim Hazen, Heather Helmke, & Tania Moore  
Location: Jayhawk Room

Curricular standards requiring students to demonstrate their mastery of Mathematical concepts through justification or explanation have become common. For example, the Common Core State Standards in Mathematics contain 22 specific standards that require such explanations. This study 1) examines the effect of writing on student Mathematical explanations then 2) explores the implications of these findings for Mathematics classroom assessment and instruction.

To investigate construct-related issues, both Mathematics and Writing scores on constructed-response Mathematics explanation items as well as scores on a standardized Mathematics achievement test for students in grades 3-11 were analyzed to determine their measurement properties (i.e., the extent to which they could be judged to represent measures of the same construct in a test theoretic sense.)

Correlations were used to estimate a one-factor congeneric model for these scores, and the loadings and unique variances for each manifest variable in each grade were explored. A key finding of this study is that the residual variance associated with writing is by far the largest of the dimensions included in the model.

The results of this study have direct implications for Mathematics classroom instruction and formative assessment, and their intersection with large-scale assessment. For example, students who are asked to explain their understanding of mathematical concepts through means other than writing (e.g., verbally) may be disadvantaged when called to communicate that same understanding through writing (as is the case with today's large-scale assessments). Relatedly, this study supports the call for inclusion of writing instruction as a dedicated part of the Mathematics curriculum. Such a program would add to the debate regarding shared responsibility for students' literacy development across all subject areas.

- Using Science Simulations in the Classroom for Learning and Assessment- Kaylie McGrath, Kristen DiCerbo & Amy Reilly  
Location: Jayhawk Room

Innovations in assessment design and uses of technology are allowing for digital observations of student performances that are more similar to human observations; that is, we can actually digitally observe student processes and the sophistication of those processes as they work to solve problems. Further, by collecting evidence of student interactions with digital tools to set-up scientific experiments and introduce variables to solve problems, we can more closely mimic what may be done in a

science laboratory. This has many potential benefits. First, laboratory activities are limited to those experiments that are 1) safe, 2) possible to control, and 3) affordable. With simulations, we can make experiments previously thought of as impossible, come to life for students to interact with. Secondly, observing student interactions in a laboratory setting is time consuming. Teachers often do not have the bandwidth to observe small groups or individual students as they work through experiments. By collecting the data digitally, we can provide teachers access to reports that summarize student performances and their strengths and weaknesses.

The Insight Science System was created to research the use of these types of simulations and technologies for both learning and assessment and understand the potential benefits to teachers and students. This poster session will allow participants to interact with two pilot simulations as well as hear from the simulation designers and a classroom teacher who piloted them with her students.

- Understanding patterns of students' science learning using log data from visualizations in technology-enhanced instruction- Emily Toutkoushian & Kihyun (Kelly) Ryoo  
Location: Jayhawk Room

With the advent of the Next Generation Science Standards (NGSS), it becomes critical to capture not only students' content knowledge, but also how they engage in the content within science practices (Pellegrino, 2013). Technology-enhanced classroom learning environments offer the potential to collect vast quantities of information about students' learning processes (Rupp, Nugent, & Nelson, 2012). However, making sense of the available data in ways that are meaningful for teachers can be daunting and difficult (Ferguson, 2012). Analysis of the fine-grained log data of student clicks and actions within visualizations can provide insight into student practices (i.e., Gobert, Sao Pedro, Razuiddin, & Baker, 2013). This poster will demonstrate how data from summative and embedded assessments, along with log data can be utilized to better capture student learning and engagement in NGSS science practices and content. The project, centering on properties of matter and chemical reactions, featured three dynamic visualizations embedded in an online learning environment which allowed students to manipulate variables, such as controlling amount of thermal energy, while capturing their log data, including the sequence and counts of actions within the visualization. Specifically, this poster will focus on: 1) how cluster analysis of log data can be leveraged to identify and define key factors of student interactions within the project's visualizations and 2) how these factors can then be linked to scored data from formative and summative assessment items to provide information about students. Examples will highlight relevant inferences from this type of analysis, such as helping teachers target subgroups of students who need support on a specific science practice or concept based on their actions. The need for usable inferences from log data is especially relevant in this study because of the different stakeholders involved, including classroom teachers, district partners, simulation designers, and educational researchers.

- Presentation Accommodations for Students with Disabilities During Computer-based Science Assessments- Paul Meng  
Location: Jayhawk Room

The present systematic review of the literature identified 16 papers relevant to the current state of knowledge regarding presentation or format accommodations for students with disabilities (SWD) during computer-based science testing. A saturation analysis was conducted to ensure full coverage of content represented in the literature. Overall, these papers included accommodations for only a subset of SWD. The literature reviewed for the present study addressed the needs of learners with: (a) blindness/low vision, (b) deafness/hard of hearing, (c) attention-deficit hyperactivity disorder, (d) emotional and behavioral disorders, (e) learning disabilities, and (f) cognitive disabilities necessitating alternative assessments for modified achievement standards. Some of the research reviewed did not include data collection with SWD, and most did not include data collection with students having the full range of disabilities for which the studied accommodations may be relevant. Format accommodations included in the studies presently under review range from methods of content presentation, to the context of presentation and methods for selecting test items. Very few presentation accommodations are presently supported in the literature with empirical demonstrations. Of these, none identified during the current review of the literature have support sufficient to deem them evidence-based using presently accepted standards. Much more research is needed to validate the promising supports presently recommended in the literature. This poster presents a review of the literature separated by disability category. The perspective of the presenter for this presentation spans across his current role as a doctoral student researcher and his past roles as a special education teacher. Description of the various reviewed accommodation technologies, their applicability to students with various disabilities, questions raised by the findings, and future directions will be presented.

### **3E Demonstration Session**

- A Multi-Tiered Approach for Personalizing Learning- Amy Sandoz & Katie Goddard  
Location: Jayhawk Room

This session begins with a demonstration of the Summit Learning Platform and the multi-tiered approach that is the basis for how we address personalizing learning. This approach consists of four areas: content knowledge, cognitive skills, habits of success, and student purpose. There are 36 cognitive skills, which fall into eight domains: 1) Textual Analysis, 2) Using Sources, 3) Identifying Patterns and Relationships, 4) Inquiry, 5) Analysis and Synthesis, 6) Writing and Composing, 7) Speaking and Listening, and 8) Products and Presentations. By incorporating these domains into a competency-based progression, students can track their own progress in applied situations. Additionally, students are measured in more traditional formats for their content knowledge through classroom assessments that measure content standards. Students are taught Habits of Success, which are the social and emotional learning skills that are distinct from the cognitive skills and content knowledge above. Add in student purpose, which describes students' short and long-term goals, and

together, these four areas provide a fuller perspective of each student.

These areas have data that collectively comprise a holistic view of each student's learning progress in a dynamic fashion. The formative nature of the personalized platform is guided by the triangulation of multiple data sources online from these four areas, with teachers as guides for students to best use this information to plan next steps. We will show how students and teachers use the various sources of data from the platform to update and revise the student's learning path.

After the demo of this multi-tiered approach, the session will open up for Q&A. One question to open up the forum: How and where would formal measurement interact with different aspects of this platform and have outcomes that are usable for teachers and students?

- Using Keystroke Logs for Teaching, Learning, and Assessing Writing- Mo Zhang, Paul Deane, & Heidi Andrade  
Location: Jayhawk Room

Large-scale writing assessment has generally not been able to provide instructionally useful information for the classroom teacher. However, the introduction of technology now allows large-scale writing assessment to report not only a score but also a description of exactly how students wrote their essays. Teachers can use this information to understand if students had difficulty fluently generating text, if they planned before or during writing, where and what they might have edited, and if they read what they wrote before submission. Summarizing this type of information at the group or individual levels should provide teachers a suitable starting point for instructional decision making. In addition, students might benefit from seeing exactly how it is that they composed their essays helping them to become more reflective with respect to their own writing practice.

To our knowledge, research on writing processes in the context of educational assessment is quite sparse, with few exceptions (Zhang et al., 2017). In this mix-format session, we present research and development results accumulated in the past few years, with a focus on how they can help improve teaching and learning of writing in classrooms. We plan to engage the audience in a conversation on the potential applications of writing processes in educational and assessment contexts. A tentative structure of the session is shown below.

- Presentation on the theoretical grounding and research findings (25 minutes)
- A video animation on evaluating writing processes (5 minutes)
- Keystroke replay demonstration (10 minutes)
- Commentary (20 minutes)
- Q&A and audience discussion (15 minutes)

Reference:

Zhang, M., Zou, D., Wu, A. D., Deane, P., & Li, C. (2017). An investigation of writing processes employed in scenario-based assessment. In B. D. Zumbo and A. M.

Hubley (Eds.), Understanding and investigating response processes in validation research. New York: Location: Springer.

- Personalized Learning: An Opportunity for Technology- Dianne Henderson  
Location: Jayhawk Room

We all understand that technology will provide the potential means to improve learning and move us closer to personalized learning. This presentation will describe how technology can be used to increase instructional time, through providing efficient interim and formative assessments; integrating data from across multiple measures, planning differential instruction, as well as setting goals and monitoring. In addition to describing the opportunities, this presentation will also address one of the limitations – demonstrated effectiveness. Although there are a number of approaches and a variety of technological approaches, there is no one single definition of personalized learning. The purpose, tools, and instructional techniques vary considerably and are dependent on the local context. As a result school and district leaders who want to pursue personalized learning need to become more sophisticated consumers. The presentation will address these questions and provide a framework for use in evaluating these technological solutions.

- Envisioning How to Better Connect and Use Assessment Data Within and Across Classrooms and Schools- Pamela L. Paek and Britte Cheng  
Location: Jayhawk Room

This session opens with a demonstration of SimDistrict, a free, online, interactive software tool simulating K-12 education practices, including key players, factors, interactions, and relationships. SimDistrict allows users to consider key players and their interactions across a K-12 system, thus simulating the many complex dynamics that drive implementation of interventions or policies to succeed or fail. To develop SimDistrict, we integrated multiple theoretical and empirical findings from literature as well as expert feedback to determine which components and subsystems to include. The resulting tool includes variables (components) to represent individual and groups of students, teachers, classrooms, schools, and professional learning communities within a district. Including these multiple levels in SimDistrict enables users to visualize how system feedback loops involving evidence of student learning can be used across the many levels of this K-12 system.

A goal in this session is to illustrate the potential value of a systems perspective on K-12 education policies and practices around formative assessment, including implications for test accountability and data use. The framework used to model a system of factors and stakeholders involved in formative assessment including students, teachers, classrooms, school, and professional learning communities. Incorporating multi-level dynamics, we focus more broadly than what has conventionally been labeled “formative” (the moment-to-moment interaction of a teacher and student). We are more interested in the system feedback loops involving evoking and interpreting evidence of student learning needs to make adjustments in

components within and across a K-12 system.

After the demo, we encourage lively discussions on the implications of using this tool to help districts reconceptualize how multiple types of assessment data triangulate versus when they are at odds with one another, and what ways we can use systems thinking to better integrate the multiple levels of assessment data practitioners use.

- Using an Interactive Coherence Map Tool to Develop Tiered Math Assessments- Char Shryock  
Location: Jayhawk Room

Coherence in math allows teachers and students to see connections within the math strands of geometry, algebraic thinking, numbers and operations in base 10, measurement and data and statistics. This coherence extends across grades and within grades. Tiered assessments enable teachers to gather evidence of student learning, and allow for multiple entry points into an assessment. Participants will have the opportunity to investigate the interactive Coherence Map Tool, including discussion of the model problems and math progression documents. Participants will collaborate on how teachers could use the Coherence Map's interactive capability to map a specific standard, looking at related standards within the grade and in prior or future grades. Identifying the necessary prior knowledge a student will be asked to build on when demonstrating mastery of a grade level standard is an important step of developing instruction and assessment that will meet the needs of all students. This identification of prior knowledge is the starting point for creating a tiered assessment with multiple entry points. A simple tiered assessment blueprint will be created by each small collaborative group.

- Equity in Measured Mathematics Performance Outcomes Across Subgroups: An Existence Proof Challenges our Assessment Assumptions- Pamela L. Paek & Andrew Coulson  
Location: Jayhawk Room

The use of differential item functioning (DIF) as the main mechanism for weeding out items that are biased against marginalized subgroups is the norm in the psychometric community, as a means to ensure fairness on high-stakes assessments. Yet, certain subgroups continually underperform on overall results, sometimes with statistical and/or practical significance, and the measurement community passively accepts these differences. The assumption that is never explicitly stated is that students from marginalized subgroups know less, or have less “underlying true ability in the content area being addressed” – so it is valid that they chronically don't perform as well on assessments.

What if we challenged this acceptance? What if we learned that significant factors for these observed subgroup performance differences are due to the way we've developed items?

There is a dearth of evidence showing where marginalized subgroups can perform as well as their peers on assessments, so the measurement community isn't moved to question the results of their own methods. However, here we present a radically different way of measuring math concepts and mastery, through a spatial-temporal approach. Through interaction and assessment with this method, students with disabilities as well as those classified as English Language Learners performed similarly to regular education students or those classified as gifted. In this learning & assessment platform that only progresses students after they solve the interactive visual puzzles of math concepts correctly, there is equitable progress and attainment. Meanwhile, these same student subgroups still underperform on high-stakes assessments that purport to measure the same math concepts.

This presentation will begin with a demo to show how spatial temporal math content differs from conventional ways math is both taught and assessed in classrooms across the US. We will deep dive to focus on how a spatial temporal "assessment item" differs from conventional: what is the underlying true knowledge required to solve the item. Findings on equitable performance across subgroups via spatial temporal assessments will be shared. The largest portion of this session will be to open the floor for discussion. We want to foster conversation about what the measurement community may be inadvertently claiming about student learning, and how we may need to reconceptualize how and what we assess.

**4:00 – 5:15 pm**

**Session 4**

**4A** Accessibility of Classroom Assessments- Karen Barton, Anne Davidson, Melanie Magee, Sharri Zachary, & Leanne Ketterlin Geller  
Location: Centennial Room

Accessibility of educational assessments refers to students' ability to engage with the tested content in a way that allows them to accurately demonstrate their knowledge, skills, and abilities in the tested domain, and is a critical component when considering the fairness of the decisions that emerge from the uses and interpretations of results. Accessibility can be influenced by many factors that are related to both the test or testing situation and students' personal characteristics. Understanding the interaction between these dimensions can provide insights into the causes and possible mechanisms to support accessibility. Much of the published literature on accessibility focuses on summative assessments; however, because of the important role of classroom assessment in guiding instructional decision making, the accessibility of classroom assessments is of paramount importance.

This presentation will be divided into three components. First, accessibility experts will discuss the concept of accessibility within classroom assessments, describe the factors that influence accessibility, and discuss the role of universal design, accommodations, and differentiation supporting accessibility. Experts will present evidence from quantitative and qualitative studies examining the effectiveness of accessibility features, and alignment between classroom practices and summative assessments. Next, two practitioners (a district-level assessment coordinator and an elementary school assistant

principal) will comment on successes and challenges in implementing accessible classroom assessments. These practitioners will focus on the professional development and other school-based structures teachers need to understand, design, and implement accessible classroom assessments. Third and finally, the combined panel will engage with the audience to generate an initial list of resources for schools and knowledge/skills of teachers and leaders to support accessibility of classroom assessments. Examples of accessible and inaccessible assessments will be shared and strategies for improvements will be discussed. The role of educational leaders and higher education faculty as the mechanism for supporting teachers' practices will be discussed.

**4B** The Montessori Approach to Classroom Assessment- Angela K. Murray, Jade Caines Lee, Dayle Dryer, Jennifer Baker-Powers, Laura Hosek, Emily Proffitt Holtzclaw, Kathy Klocke, & Joe Sears  
Location: Pine Room

Montessori teachers guide and monitor students' progress through a graduated curriculum on an individualized basis where the teacher constantly observes the children to identify where they are in their development at any given moment (Lillard, 1996). Teachers gauge understanding by the way materials are handled, accuracy of written work, ability to transfer concepts to new situations, and demonstrating mastery through one child teaching a concept to another (Charlap, 1999). The focus of this session is to demonstrate how Montessori teachers incorporate instructionally embedded assessment in their day to day work with children.

The session will begin with a brief overview of Montessori philosophy related to classroom assessment followed by small break-out groups of participants interacting with multiple Montessori teachers (at both the early childhood and elementary levels) who will demonstrate how they use the specially designed materials to embed assessment in their ongoing interactions with students. Specific materials that will likely be the focus of the demonstrations at the early childhood level include number rods, pink tower, knobbed cylinders, metal insets, puzzle maps, and/or chains (<http://shop.heutink-usa.com>). At the elementary level, materials to be demonstrated will include the checker board, large bead frame, grammar box, box of sticks, landforms booklet, parts of a flower, and/or sentence analysis (<http://shop.heutink-usa.com>).

For over 100 years, Montessori teachers have been implementing unique practices in their classrooms that enable individualized learning and facilitate substantial diversity in student activity on any given day. Sharing these practices can provide a valuable perspective on innovative ways that classroom assessment can be conceptualized in the twenty-first century classroom. The format of hands-on demonstrations delivered in small groups will foster interaction of participants with experienced Montessori teachers to share the valuable perspectives that each can bring to the field of classroom assessment in order to enhance teacher practices and student learning.

Lillard, P. P. (1996). *Montessori today: A comprehensive approach to education from birth to adulthood*. New York: Schocken Books.

Charlap, J. (1999). Montessori for the elementary years. *Tomorrow's Child*, 7(2), pp. 5-7.

**4C** Doing's More Important Than Just Knowing: Assessment of Technical Readiness- Brian Gong, Dan Robbins, & Laura Arnold  
Location: Malott Room

This session presents a case study from Kentucky of how a system of assessments and policy supports is being developed to promote “technical readiness.” Technical readiness incorporates aspects of content knowledge and skills and essential work ethic skills (including personal accountability, punctuality, preparedness, organization, respectful interactions, effective communication, cooperation, teamwork, and community service). Assessment of technical readiness involves demonstration in applied settings, and passing industry certification exams approved by the Kentucky Workforce Innovation Board.

The assessments occur at the classroom level to district standards, and through a set of standardized assessments overseen jointly by the Kentucky Department of Education and the Kentucky Workforce Cabinet. The assessment policies include incorporation into student grades, student graduation options, and school accountability. Making the policies and supports coherent across classroom, district, and state levels has been essential to developing a widely supported and expanding system of assessments. A number of assessment quality control and evaluation processes have been developed.

The case study will be framed by Brian Gong (Center for Assessment), who will draw out the main assessment themes, including the construct definitions combining both academic and non-cognitive knowledge and skills; the validity argument for performance-based and classroom-embedded assessments; the coherence across levels of the assessment and accountability systems; and how measurement challenges have been met and what is still in development. Dan Robbins, principal at Hardin County School's Early College & Career Center will present the assessment plans, along with examples of student work and evidence, as well as describe the development, implementation, and monitoring processes. Laura Arnold, Kentucky Department of Education state CTE director, will explain the state-level supports, especially of state-approved assessments and accepting CTE assessments for student- and school-accountability.

**4D** MyDesign® tool for supporting learning, doing and assessment of the design process and design thinking- Rosemary Reshetar, Leigh Abts, Pamela Kaliski, Jennifer Lee Kouo, Matthey Miller, & Lei Wan  
Location: English Room

The Engineering Design Process Portfolio Scoring Rubric (EDPPSR) was developed with extensive efforts of educators and engineering subject matter experts. The EDPPSR is a suite of rubrics for each of the 12 elements of the engineering design process, each with a score scale (0 to 5) and a set of descriptive criteria (Goldberg, 2014). The EDPPSR was adapted as the model rubric for Project Lead the Way's e-portfolio (Innovation Portal,

2010).

The standardized EDPPSR, was subsequently used as the framework to collect design-based artifacts in MyDesign® learning management system. MyDesign® has been developed as a web-based software platform that links learning applications to seamlessly support the design process, anytime, anywhere by anyone. This platform can personalize, guide, and record the learning and practice of a design or series of designs by customizing the process to a user's knowledge, practice level and setting (formal or informal).

This platform of applications enables users to follow a simple, icon and template guided process to: 1) learn and practice design; 2) create the products of design; and 3) maintain a design 'diary'.

While design is practiced daily, by almost everyone, it is a notoriously difficult topic to teach; and 'feared' as a complicated 'engineering' process requiring specialized instructional skills. Our intent is to 'demystify' design, and to empower teachers to gain the confidence to understand and practice design as: 1) an individual and / or team-centric 'active learning' process; 2) exemplified through everyday activities; and 3) a process to engage, excite, and track students in their exploration of science.

Our diverse panel with expertise in education, engineering, research, and psychometrics will demonstrate MyDesign® and use that as the basis for a lively panel and audience discussion of the teaching of the design process and the collection and evaluation of related research and assessment data.

Goldberg, G. L. (2014). Revising an Engineering Design Rubric: A Case Study Illustrating Principles and Practices to Ensure Technical Quality of Rubrics. *Practical Assessment, Research & Evaluation*, 9(8), 2

Innovation Portal (2010). <http://www.innovationportal.org>.

- 4E** Perspectives on a formative approach to social-emotional learning assessment and development- Angie McAllister, Jason Way, Wade Leuwerke, Norma McCormick & Jeff Allmon  
Location: Kansas Room

Social-emotional learning (SEL) and formative assessment systems are two topics that have recently been gaining traction in the education world. SEL skills are motivational, social, and self-regulatory skills that are critical for college and career readiness and predict a wide range of academic outcomes. Formative assessment approaches involve repeated assessment and intervention over time with professional development for teachers, rather than the stand-alone, one-shot assessment typical of summative assessment approaches. The focus of this panel discussion is to bring these two trends together and discuss why a formative approach to assessing students' SEL is more appropriate and beneficial than a summative approach, as well as best practices for

implementation of an SEL formative assessment system based on participants' direct experience.

The panel members bring a wide variety of perspectives to the discussion. The moderator has experience as a classroom teacher and working in the educational technology field. Other panel members are experts in the various aspects of the formative assessment approach. Two panel members are experts in developing and implementing K-12 curricula for improving student SEL skills. Another panel member is an expert in coordinating SEL assessments across schools and implementing professional development for teachers on how to help underserved student populations improve their SEL skills. The last panel member is an expert on SEL research and assessment.

Panel members will be asked to share their expertise and experience around SEL and formative assessment. To encourage discussion amongst panel members, the panel will be asked to identify strengths and suggest improvements to the implementation efforts that three of the panelists have been involved with. This will culminate in the identification of best practices for developing and implementing SEL formative assessment systems.

#### **4F Paper Session**

- The Use of Evidence: Necessary but not Sufficient- Christine Lyon, Caroline Wylie, & Leslie Nabors Olah  
Location: Divine Nine Room

Current definitions of formative assessment emphasize that multiple actors (i.e., teacher, peers, and self) within the classroom must be considered and the manner in which evidence is collected and its quality matters. Therefore, this presentation will take the perspective that research on formative assessment implementation needs to be situated within the classroom context and the larger domain of instructional practice.

The overall presentation will focus on a study that was conducted in the context of the Cognitively Based Assessment of, for, and as Learning (CBAL) research initiative and examined the implementation of formative assessment tasks that are linked to models of student learning and include domain-specific professional support materials. Seven high-school mathematics and ELA teachers participated in a pilot of the materials. Trained observers scored the teachers' classroom assessment practices using a structured protocol that splits the construct of formative assessment into three hypothesized domains (i.e., structures that facilitate teaching and learning, iterative content development, and activating students) and eight dimensions. The results from the observations were used to create four profiles of formative assessment implementation -- Low-Impact, Low-Evidence Use, Didactic, and High-Impact Practice. These profiles identify important differences among teachers in their ability to use evidence of student learning to inform next instructional steps. The profiles also help identify practices that may be complimentary to (e.g., student involvement) or prerequisites for (e.g., preparation and planning) the collection and use of evidence

for instructional decision making.

Participants will be introduced to the theoretical framework for the observation protocol, provided with a brief overview of each profile, and will engage in a discussion of how this type of tool could be used to evaluate professional development programs focused on formative assessment and inform differential starting points for teachers engaged in these types of professional development programs.

- A homework administration system for personalized learning by using test results-  
Minsung Kim  
Location: Divine Nine Room

The specific purposes of this proposal are to introduce a system to connect large scale assessments and classroom learning by using homework as the medium. The effectiveness of homework in learning has been discussed by many researchers. Both cases for (Copper, Robinson, & Patall, 2006) and against (Kralovec & Buell, 2000; Bennett & Kalish, 2006) have been perennially discussed among educational researchers. Positive relationship between time spent on homework and students' performances have played an important role as the evidence. However, the quality of homework assignments has not been emphasized in the previous studies.

Teachers typically make each homework assignment the same for all students, who then complete the assignment on their own after school. Discordance between individual ability and homework difficulty can occur, and some students may experience homework that is too easy or too difficult due to individual differences in cognitive competency.

There are gaps between state assessments and classroom learnings. State assessments are usually aligned with state standards or consortium standards (i.e. Smart Balance, Common Core). State assessment results are able to provide information about the width and depth of knowledge according to its standards, which can be individualized. This information can provide the direction of future individual learning and class instruction. (Murane, Sharkey, & Boududett, 2005)

A personalized homework system can be built up with the help of educational technologies. Computer Adaptive Homework (just like Computer Adaptive Tests) monitors students' development, and the system assigns appropriate content and difficulty level of homework to individuals. State assessment results can be utilized while designing the individualized homework scaffolding plans based on the learning theory of Vygotsky (1978)

A well-established homework administration system will narrow the gaps between large scale assessments and classroom learning. Also it will enhance the effectiveness of personalized learning in a long term perspective.

- Diagnostic Assessment – A Tool For Quality Control in Education- Oluwatoyin Obadare-Akpata, & Targema Numbe Terwase  
Location: Divine Nine Room

The paper examines the need to integrate diagnostic assessment (DA) in teaching and learning process to serve as quality control measures in the education system. It defines quality control as a system for setting standards in a process and taking appropriate actions to deal with deviations outside permitted tolerance. The need for the use of diagnostic assessment for quality control was discussed, in terms of its appropriateness in assessing students' learning difficulties for remediation measures. The steps in designing and conduction diagnostic assessment for quality control in education identified are; specify learning goals and objectives, plan classroom and assessment, deliver instruction to identify learners' gaps among others. While the models recommended were Item Response Theory for quiz/test, Empowerment evaluation model for conference/interview and self-assessment types of diagnostic assessment, and CIPP for holistic diagnostic assessment of the education system. Challenges of implementing diagnostic assessment identified include large class size, non-inclusion of DA in the curriculum; lack of commitment on the part of teachers as well as lack of motivation of teachers by their employers could hinder the implementation of DA in schools. Recommendations were made. Finally, the paper concludes that integrating diagnostic assessment in the teaching and learning process would improve students' comprehension level, reduces their errors and provide timely remediation measures before they are allowed to participate in standardized examinations.

- Digital Formative Assessment Techniques: Improving Student Motivation- Bryan R. Drost  
Location: Divine Nine Room

This interactive presentation will provide an overview on what formative assessment is and how it differs from traditional summative assessment practices, as conceptualized in an article by the same name (Drost, 2016). The presenter will discuss, model, and allow participants to experience several formative assessment techniques using technology that can be used to track student progress amongst content standards and improve motivation. Each tool describe will capitalize on the student-teacher relationship inherent in classroom assessment.

### Thursday, September 14

#### **8:30 am – 9:30 am                      Session 5**

- 5A** What can we Learn from Game-based Performance Data: A Panel Discussion with a 7th Grade Student, a Classroom Teacher, and an Educational Researcher- Elizabeth Moore, Kaylie McGrath, Jack Reilly, & Kristen Dicerbo  
Location: Malott Room

Using digital education games as assessments of student learning have gained popularity as more students have access to devices. Many claim to provide feedback on student knowledge and skills that can be used to base instructional decisions and to monitor student progress. But to what extent do we rely on that information to make inferences about what students know and can do? In this session, Elizabeth Moore, Director of School Improvement at Pearson, will moderate a panel to explore that question through the lens of an educational researcher in the area of game-based learning and assessment, an elementary school teacher, and a middle school student.

Dr. Kristen DiCerbo, VP of Educational Research at Pearson, will share her expertise based on studies on the uses of this data and the extent to which we can use it to make valid claims on student learning. Kaylie McGrath, grade 3 teacher from Dripping Springs, Texas, will join the panel to discuss her experience with using games in the classroom and the types of features from games and reports that she finds the most reliable. Finally, Jack Reilly, 7th grade student, will share his perspective on how he evaluates his own performance while playing games and how the feedback from games compares to other types of feedback he receives during instruction.

**5B** Strengthening Claims-based Interpretations and Uses of Local and Large-scale Science Assessment Scores: Project Insights- Liz Summers, Valorie Foy, Sara Cooper, Susan Davis-Becker, Ellen Forte, & Howard Everson  
Location: Pine Room

Currently, SEAs and LEAs have abundant options for testing with assessments that are intended to align with academic content standards. These assessments purport to diagnose instructional needs for individual students, gauge progress toward goals within and across school years, offer teachers actionable next steps for instruction, inform parents of whether their children are on the path toward college- and career-readiness, and inform students, parents, K-12 and university educators, that students are prepared to succeed in college and career.

The Strengthening Claims-based Interpretations and Uses of Local and Large-scale Science Assessment Scores (SCILLSS) project, a recently funded Enhanced Assessment Grant (EAG), brings together three states (Nebraska, Montana, and Wyoming) with three independent organizations to develop a comprehensive assessment approach that clarifies and strengthens the connection between statewide assessments, local assessments, and classroom instruction. Through the SCILLSS project, we are developing tools and resources that focus efforts on restructuring and re-engineering both large-scale and local science assessments to ensure standards alignment, support system coherence, and to yield scores with more meaningful information.

In this panel discussion, SEA and LEA representatives from Nebraska will come together with a representative from each of the three partner organizations to share insights on how the project activities will engage state and local educators, along with large-scale assessment developers, in creating a shared understanding of science as it is taught and tested. This discussion will include descriptions of how principled-design elements for

assessment will derive from the same conceptual representation of science used to support instructional design and local assessments. Participants will benefit from hearing about the initial stages of the federally-funded SCILLSS project and will be encouraged to consider how they can leverage similar approaches in connecting instruction and assessment to support far richer standards-based alignment than is possible with most current large-scale assessment designs.

**5C** Introduction to the Classroom Assessment Standards- Don Klinger & Barbara Howard  
Location: Kansas Room

The revised Classroom Assessment Standards (CAS) (McDivitt, Klinger, Howard, Rogers, Wylie, & Munoz, 2015) are the only ANSI approved standards outlining the foundations for effective classroom assessment. The Task Force who developed the Standards were selected by the Joint Committee on Standards for Educational Evaluation (JCSEE). The JCSEE is made up of 14 national and international professional and academic societies, including the AERA, NCME, APA, AEA, CES, CREATE, and CSSE. Of critical importance, the intended audiences for the CAS are classroom teachers, educational leaders, teacher educators, and policy makers. Hence the CAS have been written in a manner that enables these audiences to more readily understand and implement these principles and guidelines for fair classroom assessment practices.

The CAS are based on the premise that the assessment of students in the classroom can be central to student learning. To effectively serve student learning, teachers, instructional leaders, and principals need to understand and apply sound classroom assessment practices. The 16 standards within the CAS fall into 3 broad domains, Foundations, Use and Quality, covering principals for classroom assessment that range from purpose to design to administration to analyses, while also addressing issues of communication, diversity, bias, reliability and accuracy. Each standard is supported by an explanation of the standard and a set of guidelines for effective implementation of the standard.

This session will be hosted by the co-chair of the Task force that wrote the CAS and by the chair of the JCSEE. The purpose of the session will be to provide conference participants an opportunity to see the CAS and discuss how these ANSI approved, NCME supported Standards can be used to guide educators' assessment practices and improve in-service and pre-service instruction about classroom assessment.

**5D** Assessment of Core Engineering Design Competencies Through Notebooks- Kerrie A. Douglas, Hillary E. Merzdorf, Amanda C. Johnston, & Tamara J. Moore  
Location: English Room

While much integrated STEM curricula resources have been published for classroom use, little has been provided to be of practical help for teachers tasked with assessing deeper levels of learning, particularly around engineering. From this practical need to assess engineering competencies across differing middle-school science contexts, we identified core engineering design competencies with specific learning objectives that are reflective

of NGSS, but more specifically based on the Informed-Design framework (Crismond & Adams, 2012). The competencies represent practices of engineering design as problem-scoping, using evidence when developing a solution, and communicating the design solution through evidence-based reasoning. From these competencies, we created Engineering Notebooks for Middle School as a tool for teacher use across integrated STEM lessons.

In this presentation, we will overview our approach to embedded formative assessment of engineering competencies through Engineering Notebooks in middle grades. Our research team comprises former mathematics and science teachers, as well as researchers trained in educational measurement. This work builds from a large Math and Science Partnership where teachers were trained to write integrated STEM curricula units which mapped to their state science standards. The Engineering Notebooks were designed as a means for common assessment across the curricula units of varying contexts. One research question we have considered is the extent to which students' understanding of science concepts can be assessed when students are responding to questions concerning design decisions, and whether the common assessment are sensitive enough to capture context-specific learning. We will present examples from students' notebooks as used in three different lessons. Participant discussion will be centered the following issues: 1) Validation of notebooks and rubrics for classroom use, 2) Training teachers to effectively use notebooks for formative assessment, and 3) Writing effective prompts for capturing students use of science and mathematics concepts to solve design challenges.

### **5E Paper session**

- Comprehensive Assessment System: The View from the Trenches- Fengyi Hung, Zeek Edmond, Taj Jensen, & Christine Kelly  
Location: Centennial Room

Please join the Tacoma Leadership Team as they walk you through the process of strengthening their Comprehensive Assessment System with a focus on aligned, coordinated and systematic support for TPS educators. Topics covered will be: the formation of a 57-member District Assessment Committee to revamp the system, the adoption of the Universal Screener, the organization of professional development for 1,400 educators, and unveiling this year's focus on integrating Smarter Balanced Interim Assessment Resources into high quality teaching and learning. Attendees will

- 
- \* Understanding how to evaluate and revamp a district assessment system
- \* Processes to strengthen teacher buy-in and system/infrastructure readiness
- \* The value of integrating a Universal Screener and Standard-Based Formative Assessment Resources to personalize instruction

- Tests, formative practice, and coherence between classroom and accountability assessments: A Dialogue- William P. Fisher, Jr.  
Location: Centennial Room

Recent decades have witnessed striking developments in the theory and practice of assessment. Many of these developments involve highly technical psychometrics focused on meeting the needs of classroom teachers, but may not be widely known.

This session will be a dialogue in which the presenter will pose questions directly to the audience about assumptions often made about assessment, and will moderate the ensuing discussion. Slides illustrating relevant topics (information hidden within scores in response patterns, learning progressions, kidmaps, Wright maps, construct maps, construct vs content interpretations) likely to emerge will be prepared in advance. Questions will focus on formative assessment and developmental, horizontal, and vertical coherence in assessment theory and practice:

1. Does testing inevitably involve ignoring the uniqueness of individual students?
2. Can formative assessments reveal important qualitative individual differences useful in tailoring instruction?
3. Are test scores (counts or percentages of correct answers) the primary assessment product available for use in formative applications?

These three questions should open up different opinions contingent on familiarity only with test scores or with more advanced evaluations of individual item responses. If teachers comprise a significant portion of the attendees, many may be familiar only with test scores. Those familiar only with sum scores may rightly note that teachers experienced with students in their classes may be aware of differences in the abilities of students obtaining the same scores.

Another domain of responses to these questions will come from those familiar with construct mapping (Wilson, 2005) and associated displays of each student's responses to individual items, ordered by difficulty.

Additional questions will take up topics concerning the time and technical skill needed to integrate assessment and instruction, for meaningfully linking classroom and accountability assessments, as well as item vs information coherence, and teacher-oriented assessment management systems.

- Actionable Assessment Results to Inform Instruction- Stephen T. Murphy  
Location: Centennial Room

The focus of the presentation will be to describe a collaborative effort between a large district and a testing vendor to help guide educators to make better use of assessment results. The session will open with a description of the challenge posed and then a dive into the collaboration that ensued to help drive educators to use assessment results, to make the results of an administration more actionable, and how assessment outcomes can impact curriculum and instruction for each individual student. During the presentation, the approach that was selected will be provided and give detailed information regarding the use of reports and how the reports can be connected to the curriculum standards taught in the district. Perspectives from psychometric/research,

content design and development, curriculum and instruction director from the district, as well as the assistant district superintendent will be shared.

The session will be an open dialogue about how the partnership and collaboration manages both successes and challenges and questions will be asked of the audience throughout the presentation to engage discussions with each other. Example questions will include 1) what challenges do you face in guiding educators to use assessment results?, 2) what improvements in reports can be made to drive action from assessments?, 3) what success/challenges have used faced in using assessment results?, etc. The session will end highlighting the importance of such collaborations and how as testing vendors we need to be sensitive to how we are addressing customer assessment uses and interpretations, i.e., what questions are they trying to answer?

**9:35 am – 10:35 am                      Session 6**

**6A Classroom Assessment: New Rules-** Brett Jenkins, Don Klinger, Caroline Wylie, & Eric Stickney

Location: Centennial Room

Although, classroom assessment seems like it is as simple as collecting data, looking for mastery of content and guiding instruction, done well, these things are much more complex than they sound. Unlike standardized assessments, such as the state summative assessment, SAT or ACT, there are a variety of classroom assessment practices with no single agreed upon set of approaches or methodology. The concept for this session is to provide a forum for the "host" to interact with 3 people with different perspectives to comment and debate current controversial topics associated with classroom assessment. The participants, all well versed in the topics will also serve as double duty as they bring a variety of perspectives including, product management, classroom teacher, policy maker, researchers, and parents.

Entitled "New Rules" the goal is to provide a provocative look at issues in classroom assessment and engage in healthy debate on topics covering policy, practice, and theory of assessment. Three sample topics are:

- 1.) Technology in the classroom: while technology holds the promise to improve engagement, provide immediate feedback and automate tasks, there are also challenges, including, social isolation, expense, the lack of research supporting some products and uneven access for all students. Panelists will be asked to comment on how to manage these challenges and to debate the best use of technology in the classroom.
- 2.) Data Integration: All assessment data from summative, interim and formative assessment activities provide additional information. It is the responsibility of the teacher to interpret the information and provide feedback to students and determine next steps. What are the best methods for helping teachers integrate data from all these varied sources.
- 3.) Quality: With all the information available from classroom activities – what are the associated quality standards that should be applied to classroom assessments and practices.

**6B** Elevating educators as developers and consumers of assessment information- Linda Vanderford, Carol Middleton, & Chad Gotch  
Location: Malott Room

Although substantial media and legislative attention has been focused on assessment at the state level, the majority of assessment decisions happen in the classroom. It is through coordinating the assessment practices and roles from the state to the classroom that a “balanced assessment system” becomes more than jargon and improves the quality of education for individual students. Effective coordination in Oregon has depended upon a strong partnership between the state education agency (ODE) and educational service districts, who can link statewide assessment to local assessment needs and practices.

Participants in this session will learn about four strategies that ODE and regional education service districts have used to elevate the value of educators as developers and consumers of assessment information. These include: 1) incorporating educators in statewide summative assessment development and scoring; 2) incorporating educators in assessment decision making; 3) providing local assessment options when assessments are used in high stakes contexts; and 4) providing assessment professional learning resources.

Findings from multi-method program evaluations and examinations of educators’ use of localized assessment options will frame a discussion of successes, challenges, and lessons learned. Participants will receive access to professional learning resources for their own use/adaptation, and will participate in a discussion about how these four strategies could be deepened for greater impact. The presenters are a university researcher and two regional education service district program staff and the presentation was co-planned with ODE program staff.

**6C** Learning and assessment: progress and diagnostic testing in the classroom- Ardeshir Geranpayeh & Nick Saville  
Location: Kansas Room

This hands-on session will give delegates the opportunity to see how technology can be used to combine learning and assessment in the teaching of English as a second language, within the context of Learning Oriented Assessment (LOA).

The LOA approach is systemic and ecological, and aims to combine informal classroom assessment with formal large-scale assessment. It discounts the false dichotomy of formative and summative assessment. LOA is primarily focused on how assessment can provide evidence of and for learning.

In the context of the classroom the emphasis is on how assessments can be used to by teachers and learners to measure progress, set goals, and provide feedback on the performance of learning tasks, and inform teacher decision making. This approach personalizes learning and puts the learners’ needs at the heart of assessment.

In this session we will demonstrate two different ways of using technology to close the gaps which can occur in the assessment cycle of performance-observation-interpretation-feedback. Technology connects the parts of the cycle in a way not possible due to the challenges, among others, of record keeping and time constraints. We will argue that technology creates a virtuous circle which enables the evidence for and of learning to flow more freely than ever before.

The first method is end of course or teaching unit progress testing. Empower is an English language course which combines learning activities with online in-course assessments. As the test results are recorded digitally, the teacher can easily use the results diagnostically to identify areas of strength and weakness, and tailor the learner's next activities accordingly (and the course includes such supplementary activities).

The second method is the use of artificial intelligence (produced by annotated data driven supervised machine learning) to evaluate learners' writing and provide immediate, in under 15 seconds, feedback and suggestions for improvement.

## **6E Paper session**

- School Professional Learning to Support Assessment Item Design and Student Response: A Case Study- Russ Kupperstein  
Location: Pine Room

Assessment design is a critical component of the learning process, and in High Schools, measurement of course aims and objectives is a high priority particularly in schools adhering to an international curriculum with high-stakes external examinations, such as International Baccalaureate and Advanced Placement. The focus of this session is to share the case study of Cairo American College (CAC), an independent school teaching IB and AP, and the initiative to reorient staff focus on the use of Command Terms in assessment item prompts. Over many professional learning workshops and several years, CAC has been fully committed to exploring the manipulation of assessment questions and the proficient use of Command Terms to probe deeper levels of understanding with precision and purpose. The session will outline the sequence of professional learning opportunities that staff experienced, including workshops oriented towards assessment design, feedback for students on question interpretation, reflection on assessment gaps, analysis of Command Term distributions, and student work analysis from external, high-stakes examinations. Included in the session will be an analysis of data from indicators of staff and student progress related to the goals of the initiative. Finally, session participants, including district and school administrators, professional development consultants, and teachers, will discuss the body of research upon which the initiative was based, consider new directions for the initiative, and examine the contextual limitations of the initiative and the flexibility of its introduction into U.S., public school settings.

- Proposing a New Way to Distinguish Formative and Summative Uses of Assessment Information- Robert Good  
Location: Pine Room

The positive impact that the effective use of formative assessment strategies has on student achievement has been well documented. However many who have advocated the use of such strategies often contrast formative and summative assessment as occurring exclusively during or after an instructional unit, respectively. A new lens is needed that distinguishes formative and summative assessment information in a way that is less limiting and takes a more subtle approach when using both types of classroom assessment information.

The purpose of this session is to propose and discuss the notion of using expected mastery of both broad and narrow learning targets as the primary consideration in determining formative or summative use. Since instructional units span multiple lessons and contain different points in time when students are expected to master narrow skills or broad concepts, casting all information obtained during the instructional unit as formative and all gathered afterward as summative is not an accurate portrayal, nor is it helpful to teachers as they use both types of information. By using expected mastery, teachers can simply ask, “Can I expect the student(s) to have mastered the skills and concepts assessed at the time the assessment is given?” A response of “No” indicates a likely formative use and a “Yes” allows for summative uses depending on the results seen. The perspective this question provides shifts the focus from an instructional point in time (teacher-centered) to the acquisition of skills and concepts (student-centered); this allows the teacher to be more nimble in navigating the moves between formative and summative information that occur during and after an instructional unit.

This session will include 1) an introduction to the proposal with background research and definitions; 2) the use of several guiding questions to elicit audience feedback; and 3) some closing thoughts/reflections (time permitting).

- Instructional Shifts in Math and ELA Assessments- Bryan R. Drost & Tricia Ebner  
Location: Pine Room

This interactive presentation will provide an overview of the three instructional shifts inherent in math and language arts standards and how those shifts impact the types of assessments that can be given to reliably make claims about students' mastery of content standards. Participants will experience aligned assessments, quasi-aligned assessments, and be given the opportunity to produce classroom-based formative assessments that address the instructional shifts.

**11:00 am – 12:15 pm                      Session 7**

**7A** Repositioning classroom assessment within a postcolonial nation state: Benchmarking, model construction, and initial exploration- Jerome De Lisle, Rhoda Mohammed, Susan Herbert, Carla Kronberg, Reina Braithwaite, Neela Premchand, Delia Lalchan, Nicola Mark, & Patricia Claxton

Location: Centennial Room

Assessment systems in the Anglophone Caribbean still place inordinate emphasis upon public examinations administered at 11+, 16+ and 18+. This system imbalance is a colonial legacy in which gatekeeper tests like the College Exhibition and Senior School Certificate managed and restricted access to further schooling. Assessment reform has been slow, with classroom assessment most severely neglected. In this structured poster presentation, we report on (1) an international benchmarking exercise to inform assessment reform; (2) development of a classroom assessment model; and (3) initial exploration of that model in selected classrooms. In the benchmarking exercise, comparator countries were selected by system performance and assessment system architecture. The countries selected were Singapore, Hong Kong, Ontario, England, Chile and Brazil. Based on the process and competitive benchmarking review, theory, and emerging, promising and best practice, a model for classroom assessment in Trinidad and Tobago was developed. This model, called the FPSI, explicitly promotes four assessment/pedagogical elements: formative assessment, performance based, 21st century skills, and integrated learning. These elements capture the current promise of using classroom assessment to promote and support learning. The exploration of the model in selected classrooms by trained practitioners is further detailed in the poster presentations. We envision a flexible model with multiple strategies and outcomes. The five case studies cover a wide range of contexts and disciplines in the differentiated secondary school system of Trinidad and Tobago. The value and utility of the model across these heterogeneous school environments support further development of policy in this area.

FPSI Posters

- 1- Assessing and Learning about the cultural self in English Literature
- 2- Performance Assessment in Industrial Technology
- 3- Learning about alcohol in Form 5 Chemistry
- 4- Promoting global awareness in Form 5 Geography
- 5- Urban overpopulation in Form 5 Geography

**7B** To make a move: Examining how high leverage classroom-based formative assessment practices make a difference- Brent Duckor, Carrie Holmberg, Steven Shirley, Richard Lehrer, Erin Pfaff, Min-joung Kim, Mark Wilson, & Lorrie Shepard  
Location: Kansas Room

Research on teachers and teaching has shown that formative assessment can improve student learning more than most instructional practices (Hattie 2012). Empirical evidence spanning two decades indicates that thoughtfully implemented formative assessment practices improve students' learning, yet little work has been done on the development of teachers-as-formative-assessors as they take up high-leverage practices ("moves") in the domain of classroom-based assessment for learning.

Our session explores implications of teachers becoming formative assessors from pre- and in-service classroom perspectives, centered on how high-leverage instructional and assessment moves can be modeled for study. Wilson, Lehrer, Kim and Pfaff (2009, 2014) argue that changes in classroom assessment practices appear in coordination with teachers' use of content-based constructs (e.g., statistics) as tools for discerning forms of student thinking and of moving that thinking toward higher levels during whole-class

conversations. Specific discourse moves that elicit student thinking to more explicit forms of discipline- and construct-specific forms of thinking can improve teacher efficacy and student achievement.

Duckor and Holmberg (2014, 2017) relax the assumption of content-specific moves and argue for developmentally appropriate “footholds” and “fixed lines” to help teachers with formative assessment practices that are broader and more actionable. They see instructional moves such as posing, pausing, probing, bouncing, tagging and binning as traceable and tractable across different subject matter, including, but not limited to, STEM curricula and classrooms.

New state-led accountability policies that emphasize local control have the potential to distinguish between classroom-based assessment practices and accountability measures. Shepard (2000, 2015) argues that we need to “develop and pursue an agenda of public education to help policymakers and the general citizenry understand the differences between large-scale, system monitoring tests and what we hope for from teachers... daily,” implying classroom assessment-for-learning practices will continue to sharply differ from accountability assessments. How can a moves-based, classroom assessment-for-learning orientation influence these deliberations? How might teachers challenge and affect them?

**7C The Role of Learning Theory in the Design and Utility of Classroom Assessment-**  
Leanne Ketterlin Geller, Michele Carney, Garron Gianopulos, & Jere Confrey  
Location: Pine Room

The focus of this session will be on the role of learning theory in the development of classroom assessments that are designed to give teachers useful, actionable, and accurate feedback for instructional decision making. Three research groups will discuss their perspective on the role of learning theory in classroom assessment of mathematics. Each panelist will provide a brief overview of the learning theory they employ in classroom assessment. The first panelist, Garron Gianopulos, psychometrician in the STEM department at NC State, will describe the role of Learning Trajectories (LT) in the development of the Math-Mapper 6-8, a new digital learning system developed around sixty-two LTs with associated diagnostic assessments. The second panelist, Leanne Ketterlin from Southern Methodist University, will describe how she integrates theories of learning into diagnostic assessments of algebra readiness to support teachers’ design of supplemental instructional opportunities. Michele Carney from Boise State University will draw from her work on diagnostic assessments of student proportional reasoning for grades 6-8 and will speak to the intentional design of items to help teachers identify and scaffold student thinking. The moderator, Jere Confrey, the Joseph D. Moore distinguished University professor at NC State, will explore the differences and similarities of the approaches by posing such questions as, “Why and how did you select your chosen learning theories? What did you expect of the learning theories and in what ways did they meet or disappoint those expectations? How did the learning theories improve the assessment, the feedback, and/or the instructional decision making? What

practical difference does the learning theories make in your implementation of classroom assessment? What types of validity evidence are you collecting to support your claims?"

**7D** Two Examples of Large-Scale Classroom-embedded Assessments Wrestling with New Approaches to Validity, Reliability, and Utility- Brian Gong, April McCrae, & Ellen Hume-Howard

Location: English Room

This session involves multiple case studies addressing how classroom-embedded assessments are being developed for large-scale assessment and state accountability uses. These classroom-based assessments may be less standardized in terms of timing, administration conditions, scoring, and even content. In addition, many involve complex performance assessments, sometimes integrated with curricula that present challenges of knowing the influence of teachers and peers, and generalizability beyond specific contexts. All present challenges of scoring, measurement models, practical development, and potentially fairness. These challenges were raised for performance assessments in the 1990's, but innovators are developing new answers to demonstrate appropriate validity, reliability, comparability, and fairness, as well as utility.

Brian Gong (Center for Assessment) will present a framework (posted prior to the meeting) that summarizes the challenges and how the case studies address the challenges. For example, some users of classroom and performance assessments are changing their claims about quality and generalizability—"best work" may be deep and individual but with less evidence of generalization and less directly comparable to other students' performances. Another interesting approach is for evaluators to combine information from multiple assessments through content-based and judgment-based approaches, to yield more reliable portrayals of student growth and achievement than any of the single assessments would provide.

The first case study will be provided by April McCrae, Delaware Department of Education science assessment specialist. Delaware is developing classroom embedded interim assessments in conjunction with curricular units for the state accountability assessment of the Next Generation Science Standards. The second case study will be provided by Ellen Hume-Howard, Curriculum Director of the Sanborn (NH) District, who has been a key leader of the district's implementation of competency-based performance assessment. Sanborn District is a leading district in New Hampshire's PACE Project, which is in its fourth year of development under a U.S. Department of Education waiver.

**7E** What a Classic Resource Can Tell Us about the Future of Assessment Literacy- Anthony Nitko, Susan Brookhart, & Sarah Bonner

Location: Malott Room

NCME's current initiative on classroom assessment builds on a long history. A classic resource for assessment literacy is the textbook *Educational Assessment of Students*, the first edition of which was published in 1983 by Anthony Nitko, who a decade later

(1993-94) would become president of NCME. Other NCME presidents with special interest in classroom assessment have included Robert Ebel (1957-58) and Brenda Loyd (1995-96), and of course Mark Wilson (2016-17), who constituted the current NCME Classroom Assessment Task Force. We are sure there must be others.

Educational Assessment of Students is one resource that was conceived originally to be useful for classroom teachers. Over the 35 years of its life, it has evolved to be more useful for what would now be called assessment literacy. It has become the only resource of its type that is still actively revised and used as a resource for practitioners both in schools and in other assessment contexts (Davis, personal communication).

This panel discussion session will use the case of Educational Assessment of Students to discuss the past, present, and future of assessment literacy. Anthony Nitko will describe how he conceptualized the book originally to be useful to classroom educators. He will describe how working with teachers in the USA and abroad, as well as changing conceptions of student learning and assessment, shaped subsequent editions. Susan Brookhart will describe her encounter with the first edition as a recent teacher attending graduate school, her current work as one of the authors, and her experiences developing assessment literacy among teachers and administrators. Sarah Bonner, who provided assistance with this resource early in her career, will present the results of her recent research on teacher assessment literacy and teacher beliefs. In dialogue with the audience, all three will provide insights for improving assessment literacy in schools.

**1:15 pm - 2:15 pm                      Session 8**

**8A** Enhanced Learning Maps Project: Roadmaps for Formative Assessment- Holly Dubiel, Shana Poettker, & Nicki Lindner  
Location: Malott Room

The Enhanced Learning Map project is a collaboration between University of Kansas researchers and teachers in four states: Kansas, Alaska, Missouri, and Wisconsin. KU researchers will present the resources developed by the project, which support math and English teachers in their efforts to implement formative assessment and improve student outcomes. The resources include access to the Enhanced Learning Map, as well as evidence-based units of instruction for grades 2-8. This framework is designed to help classroom teachers personalize instruction by making the learning process more transparent and by addressing gaps in student understanding. The Enhanced Learning Map is an organized learning model that depicts different ways students may progress through stages of learning. Researchers will discuss how pathways in the Enhanced Learning Map guide the design of the ELM units, allowing for appropriate differentiation and opportunities for deliberate feedback about students' progress in relation to the learning target. A project teacher will share experiences with the ELM project, discussing how the framework resulted in more effective, reflective, and student-centered teaching practices. During the presentation, participants will collaborate as they explore the Enhanced Learning Map software and embedded instructional resources. Participants will be asked to share the potential benefits of the ELM project components, as well as suggestions for making the framework even more impactful.

**8B ASSISTments: A Teacher Centered Online Assessment System. Students Get Immediate Feedback, Teachers Get Actionable Data-** Neil Heffernan & Cristina Heffernan  
Location: Kansas Room

Last year SRI finished a 3.5 million dollar IES evaluation on ASSISTments, a platform I host here at WPI as a free public service. SRI reported that the use of ASSISTments for homework double the amount of expected student learning. In addition, the intervention helped close achievement gaps, not widen them, as Steenberger-Hu & Cooper (2013) found for most K-12 mathematics intelligent tutoring systems. Students with incoming sixth-grade scores below the median gained even more than those above. Cristina Heffernan and I were not surprised as we had built a tool that is very different than all the other program that emphasizes self-paced study, as ours is mostly teacher paced. We think that teachers learn something about their own lessons by seeing how their students do on their homework, thus improving instruction.

ASSISTments, a free web-based tool for improving homework, has a simple premise: to let computers do what computers do well and free up the teachers to do what teachers do well. In ASSISTments, teachers write and/or select questions (fill in the blank, multiple choice, or open response) from our open libraries of content to assign as homework to their students. The tool is free, as teachers wouldn't be interested in joining an online community were it not free, and is a repository of teacher-shared content and feedback. Teachers benefit from using ASSISTments by receiving actionable information in the form of reports on students' actions. Students benefit from using ASSISTments by receiving a variety of support within the system, including teacher feedback.

We are happy to talk about ASSISTments, show how it work, and explain what is next. Specifically, the US Dept of Ed just funded us with two grants worth over 6 million to test ASSISTments in new context.

**8C Assessing English Learners: Using Appropriate Teaching and Assessment Practices to Support Equity and Access-** Matthew R. Lavery, Florin Mihai, Joyce W. Nutta, Mary Margaret Shimada, Leslie Davis, & additional classroom teacher  
Location: Centennial Room

English Learners (ELs) consistently lag behind English-speaking peers on standardized achievement tests. Accurately assessing the content knowledge of ELs presents unique psychometric challenges, however, as demonstration of content proficiency often relies on both receptive and productive proficiency with English. This session assembles a panel of professionals with expertise from a variety of perspectives to discuss the challenges and successes of assessing and teaching ELs and their experiences with teacher preparation programs that infuse EL teaching skills into educational foundations and content area teacher education courses. Panel members will discuss three broad topics, each from their own unique perspective: 1) Preparing teachers and teacher candidates to teach and assess with supports that provide access to content standards for ELs and non-ELs alike, 2) the inherent challenges of (and successes with) assessing ELs'

content knowledge without measures being confounded by limited English proficiency, and 3) how a reflective exercise using pre/post assessments can be used to reflect on and improve individual teacher practice and can be aggregated and analyzed to evaluate programs across classrooms and/or schools. The panel includes researchers with interest in pre-service and in-service teacher education with a focus on ELs; in classroom assessment and psychometrics, and classroom teachers with various degrees of experience with ELs. This presentation is supported by the Micro-credentialing of English Learner Teaching Skills (MELTS) Project funded by a US Department of Education Office of English Language Acquisition National Professional Development Grant.

**8D** Statewide Kindergarten Formative Assessment Systems: Challenges and Innovative Designs- Christina Schneider, Richard G. Lambert, & Do-Hong Kim  
Location: English Room

Teachers need more sophisticated and nuanced support systems to understand and facilitate student learning. Teachers need more information regarding how a state interprets its standards and what mastery of those standards represents. They need supports to monitor sophistication of student reasoning and content acquisition when learning rather than looking at skills in isolation. Teachers need supports in interpreting student work and using that evidence to take instructional actions.

Recently teachers in Georgia and North Carolina developed statewide kindergarten classroom-based formative assessment processes based on learning progressions. Using assessment tasks targeting the progressions Georgia and North Carolina have the intended purpose of providing a structured and comparable methodology that teachers can use to draw inferences about where a child is in their learning, likely next instructional steps, and an indicator of state level mastery. In the development process, experts in early learning, formative assessment, test development, and psychometrics have collaborated to support the teachers.

This session demonstrates the life-cycle story of supporting teachers in using principled-assessment design and formative assessment processes. The learning progressions, by explicating the intended inferences about what growth and mastery means, bridge processes from large scale assessment to those from the formative assessment area. We propose a coordinated electronic poster/presentation approach in which participants can walk through each presentation in the intended story sequence or (a sequence of their choosing) and have personal interactions and conversations with each presenter as an optimal vehicle for this session.

**8E Paper Session**

- Pedagogical Structures to Develop Teacher Candidate Assessment Literacy Skills-  
Aarti P. Bellara  
Location: Pine Room

This interactive presentation aims to explore the pedagogical and curricular structures that support the development of assessment literacy in pre-service teacher candidates. Various lessons, formative and summative assessments used in an Assessment of Learning Class within a teacher education program will be presented. Lessons and assessments focus on formative assessment, assessment of learning, grading, and communication of learning with students and parents. Presenters are hoping attendees will engage in a discussion to contribute their thoughts about the assessment practices used in an effort to improve efforts to expose teacher candidates to practical assessment strategies.

- Moving Teachers along The Formative Assessment Continuum: Lessons from Rural Africa that Resonate.- Don Klinger  
Location: Pine Room

Imagine a Grade 1 classroom with over 150 students, often with 5-6 students sharing a desk. Pens, pencils, notebooks, and paper are all in high demand, a demand that is rarely if ever met. The teacher works from a single chalkboard at the front of a classroom without electricity. This is a typical primary classroom in rural Africa. The “Strengthening Education Systems East Africa (SESEA) initiative represents a joint government initiative to meet this goal. Our SESEA sponsored research (Queen’s University and Aga Khan University, Tanzania) is a collaborative effort to provide embedded professional learning focused on formative assessment to teachers in Eastern Rural Africa.

Within the context described above, it is easy for teachers to consider such formative assessment practices and use traditional, didactic teaching approaches. Nevertheless, our work with such teachers suggests teachers can learn to use sound formative assessment practices to better meet their students’ needs. An Instructional Rounds (IR) approach to learning enabled these teachers to take “instructional risks”, implementing initial formative assessment practices, including effective questioning strategies, and group learning.

Interestingly, the questions, concerns, and learning we are observing in Africa align closely with those teachers we have worked with in North America. Certainly, teachers find such professional learning challenging; however, the IR approach has helped these teachers begin to move along the continuum of developing formative assessment expertise within their teaching. Our 15 minute video highlights the developing assessment practices of these teachers, while also highlighting common challenges in developing teachers' formative assessment practices, a surprisingly universal challenge. Our interactions with conference participants during or after the video will focus on the “universal truths” we are discovering about the development of educators’ formative assessment practices.

- Having your Cake - Eating it Too: Assessing Skills AND Content in Social Studies Classrooms- Rob McEntarffer, Jaci Kellison, & Kevin Bower  
Location: Pine Room

This discussion will focus on a multi-year collaboration between the curriculum and assessment departments in a K12 school system and a small, local liberal arts university. The ongoing goal of the collaboration is to help teachers assess "historical thinking skills." During our collaboration, we grew tired of the fight between "content" and "skills." So we decided to declare a truce! We collaborated with classroom teachers to develop, field test, and revise classroom "historical thinking" assessments. Innovative, engaging examples will be used to discuss useful relationships (rather than conflicts) between content, skills, and assessment. Participants in the discussion will 1) explore the perceived conflict between emphasizing skills vs. content in social studies courses, 2) participate in practical examples of assessments measuring skills and content (using content to demonstrate social studies thinking skills), and 3) discuss potential implications and how to realistically use this assessment model, including practical and realistic applications in their classrooms/districts.

**2:20 pm – 3:30 pm**

**Closing Plenary Session**

Location: Woodruff Auditorium

**The Confluence of Classroom and Large-Scale Assessment: What Have We Learned?**

In the closing session host Neal Kingston will ask a panel of three distinguished experts to reflect on the conference topic and on the conference itself.

Panelists: Derek Briggs, Margaret Heritage, and Pamela Paek,

# THANK YOU

**for attending the  
2017 NCME Special Conference on  
Classroom Assessment  
and Large-Scale Psychometrics:  
The Twain Shall Meet**