TC Columbia University Roundtable in Second Language Studies, 2014

List of Presentations by Category

Section 1. Theoretical Papers on LOA in Applied Linguistics

Speaker(s)	Title	Abstract
James. E. Purpura[1] & Carolyn. E. Turner[2] 1. Teachers College, Columbia University 2. McGill University	A Learning-oriented Assessment Approach to Understanding the Complexities of Classroom-based Language Assessment	Influenced by research designed to examine the effects of formative assessment on academic content learning (e.g., math) in mainstream classrooms (Black & Wiliam, 1998; Sadler, 1989; Wiliam, 2011), many second and foreign language assessment (L2) researchers (e.g., Genesee & Upshur, 1996; Leung, 2004; McNamara, 2001; Purpura, 2004; Rea-Dickens, 2008; Shohamy, 1998; Turner, 2012) have highlighted the central role that assessment plays in L2 classrooms and have expressed the need to relate assessment principles and practices to teaching and learning in L2 instructional contexts. This interest has generated considerable research relating to: (1) teacher processes in L2 assessment (Colby-Kelly & Turner, 2007; Davison, 2004; Leung & Teasdale, 1997; Rea-Dickins, 2001); (2) specific assessment methods employed by teachers (Brown & Hudson, 1998; Cheng, Rogers, & Hu, 2004); (3) teacher assessment decision-making processes (Brindley, 1998; Chalhoub-Deville, 1997; Turner & Upshur, 2002); (4) the role of teacher knowledge, experience, and beliefs in using assessments (Rea-Dickins, 2004; Yin, 2010); (5) the role of diagnostic or dynamic assessment in promoting teaching and learning (Alderson, 2005; Alderson et al., 2015; Lantoff & Poehner, 2011); (6) the value of self and peer assessment for self-regulation, autonomy, motivation, and learning (Oscarson, 1997; Patri, 2002; Saito, 2008); (7) the role of technology in learning and assessment (Chapelle & Douglas, 2006; Sawaki, 2012), and (8) the effects of standards- and outcomes-based assessment on teaching, learning and policy (Davidson, 2007; McNamara & Roever, 2006; Leung & Rea-Dickins, 2007). These studies highlight the importance of classroom-based assessment as an emerging paradigm within the field of L2 assessment, and underscore the potential that assessments have for advancing learning processes and learning success. These studies have also spawned an approach to assessment, where learners and learning processes (not necessarily *only* teaching and teaching processes) are viewed as central when considering the interrelationships across instruction, assessment, and learning. This approach has been referred to as *learning-oriented assessment (LOA)* (Purpura, 2004, 2009; Turner & Purpura, 2015; Purpura & Turner, forthcoming), and is not to be confused with nor in competition with other approaches such as diagnostic or dynamic assessment (Alderson et al. 2015; Lantolf & Poehner, 2011), even though these approaches share certain characteristics and concerns. These studies have also provided important insights, with findings suggesting that critical factors in the use of assessments to create and support an environment for learning involve: the contextual characteristics of instructional spaces, learner performance levels with respect to instructional targets, evaluation processes and their implementation, the socio-cognitive characteristics of teachers and learners, interactional patterns associated with assessments embedded in instruction, and other factors. What remains unclear is how these dimensions interact within a theory of LOA, and how they might promote a coherent set of localized best practices. Aiming to understand the factors involved in L2 assessment from a learning orientation, the current paper provides a description of the dimensions of LOA, and a tool to help analyze and talk about the role of classroom assessment from a learning-oriented perspective. These dimensions are illustrated through data from a classroom in which ELLs are learning the passive voice as a resource for discussing topical processes like desalination.

| Constant Leung

King's College, London | Learning from Feedback: Conception, Reception and Consequences | Current formative assessment literature foregrounds the significant role played by teacher feedback in student learning. Formulating feedback, however, is conceptually complex in that it has to be pitched appropriately in terms of domains and levels of knowledge and ways of knowing/learning from the student's point of view. Perrenoud (1998) and others have pointed out the importance of recognising situated regimes of teaching, assessment and learning. This talk will be based on a case study of the design, interpretation and use of formative feedback from the standpoints of tutors and students on an MA TESOL programme. The programme team involved is strongly in favour of using assessment as an opportunity to improve learning, and has been trying to develop a set of learning-oriented assessment strategies as part of a continuing effort to improve pedagogic practice. There is a collective desire on the part of the teaching staff to better understand how students, many of whom are from diverse cultural and linguistic backgrounds, understand and respond to their attempts to promote learning though assessment. Drawing on the work in dialogic teaching (Alexander, 2008), formative assessment (Black and Wiliam, 2009), phenomenography (Marton, 1986), and social participation and classroom learning (Leung 2010, 2014), I will examine feedback as an intersection of academic values, disciplinary dispositions, individual perceptions and actions, and curricular processes. The data will comprise student written assignments, tutor written feedback comments, and interviews with tutors and students. The overall aim is to map individual actions and pedagogic activities onto a coherent conceptual frame for assessment and learning that takes account of participant and contextual variations. |

| Matthew E. Poehner

Pennsylvania State University | **Dynamic Assessment and Mediated Learning: Toward a Coherent Theoretical Framework for Developmental L2 Education** | In the L2 field, as in education more generally, the widespread use of standardized tests to inform high-stakes decisions and concerns with the quality of educational outcomes have, in recent years, led growing numbers of researchers, policy makers, and classroom teachers to search for ways of more closely aligning assessment practices with instructional activities. Terms such as assessment for learning (Black & Wiliam, 1998), learning-oriented assessment (Turner & Purpura, forthcoming), and interactive assessment (Hamp-Lyons & Tavares, 2011) compel new ways of understanding how assessment may be leveraged to support student learning, and have been pursued alongside newly invigorated research programs concerned with notions of formative assessment (Leung 2004), classroom assessment (Rea-Dickins 2008), teacher-based assessment (Davison, 2004) and diagnostic assessment (Alderson 2006). Against this backdrop, L2 Dynamic Assessment (DA) has emerged as a dialectical framework for understanding and promoting learner development (Poehner, 2008). What distinguishes DA from the aforementioned perspectives is its commitment to a particular psychological theory of thinking and learning, namely, Vygotsky's Sociocultural Theory. This paper examines DA and its contributions to our understanding of how assessment may be integrated with teaching in order to intervene in and guide processes of learner L2 development.
     The paper first overviews the origins of DA in Vygotsky's (1987) writings, with particular attention to the concept of the Zone of Proximal Development (ZPD). Following Lantolf and Poehner (2014), the ZPD is conceived as an essential feature of developmental education, that is, education that aims not merely to impart knowledge to learners but to lead to new ways of understanding and acting in the world. The ZPD references activity undertaken in cooperation with learners wherein task demands exceed learners' current capabilities and success requires cooperation with a mediator. DA is an especially powerful form of ZPD activity that foregrounds the quality of mediator-learner dialogic interaction, with mediators intervening when learners experience difficulties, probing the underlying sources of poor performance, and noting learner responsiveness during interaction. DA has long been pursued in cognitive psychology and special education as a principled approach to diagnosing learner development according to the degree of explicitness required from mediators to provoke successful responses from learners (Haywood & Lidz, 2007). In the L2 field, DA has been framed as an integration of assessment and teaching as the extent of mediator involvement indicates learner proximity to independent functioning while the use of prompts, leading questions, models, and feedback to learners represents instruction tailored to their emerging abilities. The instructional potential of such cooperative interaction has been systematically elaborated through Reuven Feuerstein's proposal of Mediated Learning (Feuerstein et al., 2003). Mediated Learning offers in-depth psychological analysis of successful independent, or self-regulated, functioning and points toward the various dimensions of such functioning that may become a focus of mediator intervention. In this paper, instances of L2 mediator-learner interactions are analyzed in relation to DA and Mediated Learning. It is argued that, together, these concepts provide a coherent, theoretically motivated framework that may orient practitioners to assessment and teaching as developmental activities. |

Nick Saville & Angeliki Salamoura Cambridge English Language Assessment	Learning Oriented Assessment - A Systemic View from an Examination Provider	For an examination provider the challenge is to combine familiar functions of language assessment, such as placement, diagnosis, monitoring and international certification, into a systemic relationship with formative types of assessment that result from planned or unplanned classroom interactions. How can assessment and teaching expertise be best combined to promote more effective learning? Researchers acknowledge the central role that assessment plays in second language classrooms and the need to relate assessment principles and practices to teaching and learning activities (Turner 2012). Empirical research is now providing useful insights into classroom-based assessment, but it is still not clear how this might fit within a coherent framework linking both formative and summative uses of assessment. In this talk, we propose a systemic approach to Learning Oriented Assessment (LOA) that describes how formal and informal assessments are conceptualized and implemented from a learning perspective. We show how external examinations can be combined with classroom-based assessment and illustrate how planned assessments, together with those occurring spontaneously through social interaction, contribute to the successful attainment of language learning outcomes. Finally, we note the roles that psychometric scaling and digital technology play in implementing LOA, and how the collection of large amounts of information (big data) can contribute to our understandings of learner autonomy and to the individualization of learning pathways within formal educational contexts.

Liz Hamp-Lyons & Tony Green University of Bedfordshire, UK	Applying a Concept Model of Learning-oriented Language Assessment to a Large-scale Speaking test	Although emerging from classroom contexts, the recent movement in language education, towards *assessment for learning* [AfL] (e.g., Black & Wiliam 1998) and *learning-oriented assessment* [LOA] (e.g., Carless, Joughin et al., 2007) is gaining attention from providers of large scale public examinations (Saville, 2009). We present a concept model for *learning-oriented **language** assessment* which emphasizes learning-focused tasks, learning-focused feedback, and active learner engagement through self and peer assessment and acknowledges the contingency of assessment and the centrality of interaction to speaking events. 　　　To explore whether it could be usefully applied to formal exams and to discover whether and how greater pedagogical validity might be brought into a large-scale 'formal' speaking test, we test-drove the conceptual model with one international public examination, Cambridge English: First (FCE). Viewing the examination as a system that incorporates both the test event and the learning processes that lead up to it, we studied (a) a sample of video recordings of FCE speaking test events, and (b) wraparound teacher support material for those preparing students for the examination. We explored the behaviour of the interlocutors during the speaking test and we analysed the teacher support materials. 　　　In the speaking test itself, the interlocutor's role in eliciting language and managing opportunities for examinees to demonstrate their speaking competences was striking. In our data, we found very few instances of **tasks designed to stimulate effective learning,** strategic use of **questioning, scaffold**ing of learners' performance, or opportunities for **learners to engage in self- and peer-evaluation**. We identified specific language and social behaviours of the interlocutors that inhibited examinees from active **engagement** with the interlocutor or each other. Feedback to the learner was notably absent and so could not be taken up. 　　　In the prep materials, we focused on presentation of the test tasks and their authenticity, guidance on the provision of feedback and opportunities offered for learner engagement through peer and self assessment. We were impressed by the extent of teacher support material available and the scope this offers for embedding LOLA principles. However, we found only very limited evidence for LOLA practices in the current provision. We did not find evidence of explicit connections being made between test tasks, test preparation and longer term learning goals, but a dominant emphasis on the test event. Instances of peer and self-assessment and opportunities for feedback to learners were included in some materials, but these were not systematic and were not followed by suggestions on how to encourage or gauge uptake of the feedback on the part of learners. 　　　Our study suggests that opportunities do exist for a greater learning orientation in speaking assessments such as Cambridge English: First. Preparation for being a speaking rater could usefully include awareness training in identifying and using LOLA opportunities as well as substantial exposure to and supported critique of exemplars. Conscious embedding of LOLA principles and practices into support materials could serve to better integrate examinations into lifelong pathways for language learning and reduce perceptions of tension between assessment *for* and *of* learning.

Section 2. Theoretical Papers on LOA in Mainstream Education

Speaker(s)	Title	Abstract
James W. Pellegrino & Susan R. Goldman University of Illinois, Chicago	**Reading for Understanding: A Principled Approach to the Integration of Assessment and Instruction for Reading in the Disciplines**	In this presentation we will focus on the challenges of designing assessments of reading for understanding where the latter is defined as the capacity to engage in *evidence-based argumentation* drawing on multiple text sources. By evidence-based argumentation we mean making a claim or assertion that is supported by evidence that connects to the claim in a principled way. Such a definition of *reading for understanding* must make three critical components clear to teachers and students. First, it must detail what counts as *reading for understanding* in a specific discipline. Second, it must convey both the performances and work products that demonstrate the desired outcome. Third, it must capture the kinds of instruction that can engage students in critical literacy processes in a specific content area. Project READI (Reading, Evidence and Argumentation in Disciplinary Instruction) has been engaged in a principled process of developing a domain analysis and domain models of reading for understanding in the disciplinary areas of literature, history and science for grades 6-12. In this presentation we will describe the results of that process, including examples of assessments developed to support processes of instruction and learning. We will discuss how an evidence-centered design process was used to create a common framework across the three disciplines as well as specific disciplinary instantiations of that framework for purposes of designing instructional interventions and for assessing their efficacy. The assessment framework can function to support both summative and formative assessment functions at the level of the classroom. We will also consider how the challenges of reading for understanding interact with student background knowledge and general reading skills, especially as these issues relate to English language learners and the implications for instruction and assessment.
Howard T. Everson City University of New York	**Developing a Validity Framework for Classroom Language Assessments**	This paper will address the issue of how best to evaluate (validate, in test jargon) the instructional utility of classroom-based language assessments. Whether used as interim measures of student achievement, or as formative assessments to improve the teaching and learning of languages, using tests and assessments in the classroom—embedding them in the instructional sequence—requires language teachers provide an argument in support of the validity of tests for such purposes. Assessments *for* language learning are intended, by design, to transform classroom tests into instructional interventions, i.e., ways of improving student learning. When used in this way, teacher-designed, classroom-based language assessments ought to be closely aligned with instruction, accurately reflect student achievement, and formatively guide and monitor students' learning over time. Unfortunately, very few teachers have had the opportunity to learn about the principles of sound assessment design, especially when assessments are used to promote learning. Traditional measurement theorists established *validity* as the central technical criterion for educational tests, defining it as the "degree to which evidence and theory support the interpretation of the test scores entailed by proposed uses of the test (AERA, APA, & NCME, 1999). Unlike large-scale language assessments, most classroom-based language tests are used for a variety of purposes and in varying instructional contexts, often with little regard to quality of the validity evidence supporting those uses and contexts. Classroom-based assessments, if they are to promote teaching and learning, require a more contemporary view of validity, one based in both an interpretive and a validity argument (Kane, 2013). Collecting on-going evidence in support of the validity of classroom assessments places a substantial burden on otherwise over-worked teachers. This paper will outline an argument-based validity framework intended to support the design and use(s) of language assessments with the aim of improving the instructional role of testing in the classroom.

| Caroline Wylie

Educational Testing Service | **The Role of Learning Progressions to Support the Development and Use of Mathematics Formative Assessment Tasks** | The process of classroom-based formative assessment can be characterized as three questions that both students and teachers must engage in during instructional sequences (Ramaprasad, 1983; Wiliam, 2004):

1. What are the learning goals for my students?
2. Where are the students right now with respect to those goals?
3. How can the gap be closed between where students need to be and their current location?

These questions relate back to aspects of formative assessment that have been identified in the research literature: the role of clear learning goals; the importance of collecting evidence that directly relates to the learning goals; and the importance of both students and teachers using that evidence to adjust teaching and learning (Black & Wiliam, 1998; CCSSO, 2008; ETS, 2008). The second question calls to mind Bennett's observation that formative assessment be considered as "neither a test nor a process, but some thoughtful integration of process and purposefully designed methodology or instrumentation" (p. 7). To understand students' current location with respect to some aspect of learning requires a "purposefully designed methodology" to collect evidence of student understanding. With respect to the third question, there is a small but developing body of research that teachers struggle with, particularly when it comes to determining the next instructional steps they should take based on the formative assessment evidence (Heritage, Kim, Vendlinski, & Herman, 2008).
     In our current work my colleagues and I (Attali, Wylie, & Bauer, 2012) have suggested that learning progressions can play a direct role in supporting teachers' formative assessment process by contributing resources for the first two of these questions, and to some extent the third. The first question requires the identification of student learning goals. Learning progressions contribute by providing both long-term goals through the full scope of a learning progression, and near-term goals by offering a way to characterize students' current level of understanding and the next appropriate step. To identify students' current understanding with respect to those goals, purposefully designed instruments can be built around learning progressions to help both students and teachers identify current learning. Finally, learning progressions can help close the gap between the students' current and intended learning by providing clear descriptions of the conceptual jumps needed to move between levels of learning. This articulation of key ideas will help teachers identify what to do or focus on to close the gap between intended and current learning (Attali, Wylie, & Bauer, 2012).
     In our current IES project we are working with three learning progressions, central to understanding middle school algebra. In this presentation I describe the learning progressions and their role in both the development and use of formative assessment tasks. I will illustrate these ideas drawing on results from our current grant, with some preliminary results on the utility of such an approach. I will also explore how the learning progressions could be adapted or extended to provide greater support for teachers of English Learners and propose some possible task modifications to support English Learners. |

Section 3. Application Papers of LOA

Speaker(s)	Title	Abstract
Tenaha O'Reilly & John Sabatini Educational Testing Service	**Using Cognitive Theory and Technology to Improve Reading Assessment**	Recent large scale initiatives and policy reports have called for a new generation of assessments that are better aligned with theory, research and their practical use in educational settings (Gordon Commission, 2013). Building upon this theme, we present an overview of the research and development from a federally funded project called the Reading for Understanding (RfU), an initiative that is designed to improve reading comprehension through intervention and assessment (IES, 2010). The presentation will focus on the assessment strand of the initiative that is charged with building innovative reading comprehension assessments for students in second through twelfth grade. The assessments are based on a cognitive framework that synthesized the research in the learning sciences, reading, and strategies literatures. Two types of complementary assessments are used to measure reading ability: component assessments and scenario-based assessments (SBA). The component assessments are designed to measure foundational reading skills including word recognition, decoding, vocabulary, morphology, sentence processing, reading efficiency and basic reading comprehension. The component skills are used to help interpret performance on the SBA. The SBA provides a realistic purpose and context for reading a collection of thematically-related source materials. Tasks and activities are sequenced to both build up students understanding of the content and help identify areas that might be problematic for some students. Reading strategies and simulated peer students are included in the design to help model desired performances. Performance moderators such as background knowledge, motivation and engagement are also included in some assessments to help improve the interpretation of test scores. This presentation will provide an overview of the reading framework, the two types of assessments and how they may be useful for education and instruction.

Mikyung Wolf & Alexis Lopez
 	
Educational Testing Service
	The Use of Scaffolding Strategies in the Assessment of English Learners	In K-12 public schools in the United States, there are a growing number of English learners (ELs) who are in need of developing appropriate English language proficiency to meaningfully participate in school settings. To appropriately serve these students, an accurate measure of their English language proficiency (ELP) is essential to identify areas for improvement and determine suitable instructional support. While the current policy in K-12 education has required schools to implement an annual, summative assessment of ELP to measure ELs' attainment of English language development, these accountability assessments have been criticized for the lack of detailed information to help teachers offer more targeted instruction for ELs.
		
		    In addressing the need for an improved ELP measure, the present study focused on devising innovative ELP assessment tasks particularly using scaffolding strategies. In instructional settings, scaffolding is widely used to guide students to complete given tasks and improve student learning. In assessment settings, many ELs are often unable to complete tasks due to language barriers, resulting in insufficient data to make accurate inferences about students' language proficiency. By adopting scaffolding strategies into assessment, we explored ways to increase students' production of language use abilities and, in turn, to make more informed inferences about students' areas of development at finer-grained levels.
		
		    In this study, we developed various scenario-based assessment tasks on tablet computers. Each scenario included a storyline with ample visual images to provide authentic school-based situations and increase student engagement. For example, one scenario depicted a typical school day, including classroom announcements, a reading activity, lunch time, and an art class. Assessment tasks were embedded throughout the scenario to measure both discrete and integrated language skills. A few speaking tasks were designed to also include scaffolding questions to examine the extent to which students were able to complete the task using various degrees of scaffolding support. In these tasks, students were first asked to retell a story independently, and then answer a series of scaffolding questions. The students were then given a second opportunity to retell the story.
		
		    The study design was mainly qualitative with a focus on analyzing students' verbal responses. A total of 140 students in Grades K-5 participated in the study. Researchers conducted one-on-one interviews following observation and interview protocols. Student responses were scored using both holistic and analytic rating scales. Descriptive statistics were computed for each task. Interview transcripts and students' responses were also closely analyzed.
		
		    The results indicated that students tended to perform equally or better on the second retell task compared to the first retell, particularly in Grades K-2. Students were seemingly able to retell the story more effectively after answering a set of scaffolding questions. Even students with low performance on the retell task tended to complete the scaffolding questions. These results suggest that the incorporation of scaffolding into assessment has the potential to improve the measurement of EL students' language proficiency and also provide useful information for teachers' instruction. Further, assessment design with scaffolding offers a promising means of engaging students in learning during assessment.

| Lee B. Abraham, Paul Stengel & Steve Welsh

Columbia University | **Affordances and Constraints of Technology-enhanced Tools for Learning-oriented Assessment in Second Language Learning** | In learning-oriented assessment peer feedback and self-assessment are essential elements for promoting learning in the classroom (Carless, 2007; Liu & Carless, 2006) and in technology-enhanced environments (Keppell, Au, Ma, & Chan, 2006).

A recent research synthesis of technology-enhanced second language (L2) learning identified the need for studies that examine students' progress afforded through the use of Web 2.0 tools (Wang & Vásquez, 2012). This study responds to this gap in L2 research by analyzing students' use of peer feedback and self-assessments embedded in a final project that integrates a blog and an audio recording tool. The study addresses the following research questions:

1) To what extent do L2 learners of Spanish provide different levels/functions of peer feedback (Hattie & Gan, 2011) on drafts of three assignments for a course project?
2) To what extent do learners provide specific feedback (Narciss, 2008) on drafts of three assignments for a course project?
3) Do learners use the peer feedback in order to complete the assignments?
4) What are learners' perceptions of the importance of peer feedback and self-assessment for L2 learning?

Forty-eight students in three different intact sections of a third semester (intermediate-level) university Spanish-language course were randomly assigned to pairs in the same section to use the blog to post drafts and final versions of project assignments and to comment on a peer's assignments. Students responded to open-ended instructor prompts and commented on the following in-progress assignments that were posted on the project blog: (1) the first interview before they conducted a second interview, (2) the first draft of the project essay, and (3) the draft of the oral presentation. Two researchers independently coded 88 comments for the following features: (1) feedback level/type (Hattie & Timperley, 2007) and (2) feedback specificity (Narciss, 2008). They also independently coded the final versions of the essay and oral presentation as well as the second interview for either the presence or absence of peer feedback. Cohen's Kappa for level of feedback was .87, for feedback specificity Cohen's Kappa was .92, and for peer feedback use Cohen's Kappa was .89. Students completed one questionnaire on the importance of peer feedback and two self-assessment questionnaires (Oscarson, 2013) about their project interviews completed in Spanish.

Overall, peers provided process-level (39%) and self-regulation level (47%) feedback, which have been identified as important feedback features for enhancing learning (Hattie & Gan, 2011). For specificity, peers provided feedback not only about missing information but also about how peers should proceed (86%) with final versions. The results indicated that students incorporated peer feedback for the final essay (90%), second interview (70%), and oral presentation (63%). Student perceptions about the importance of peer feedback varied widely. Overall, students perceived both self-assessment questionnaires as valuable tools for improving on the second interview and for developing learner autonomy. Taken together, the findings revealed that planned technology-enhanced learning-oriented assessment tasks embedded in a project afforded opportunities for students to receive effective feedback that enhanced L2 learning and autonomy. |

Section 4. Empirical Papers of LOA

Speaker(s)	Title	Abstract
Margaret Heritage & Alison Bailey University of California, Los Angeles	**The Role of Language Learning Progressions in Formative Assessment for English Learners**	Recently created English language development (ELD) standards (e.g., CCSSO, 2012) take account of newly adopted college and career ready standards and primarily focus on the language used in specific subject areas, for example, mathematics and science. While ELD standards can serve as general guidelines for teachers of ELL students' language growth in content area learning, they do not provide the level of detail related to linguistic content needed for day-to-day instruction and, in particular, for formative assessment. Learning progressions of language development, which are analogous to progressions gaining momentum in mathematics and science (e.g., Corcoran, Mosher, & Rogat, 2009), can provide greater specificity about language growth by tracing its development from rudimentary stages through increasingly sophisticated forms. This presentation reports on an ongoing project, Dynamic Language Learning Progressions (DLLP) that aims to provide teachers with empirically derived language progressions for the purpose of instruction and formative assessment. We view formative assessment as a dynamic, interactive process involving ongoing assessment of how learning is evolving, and subsequent adjustments to teaching and learning to meet students' immediate learning needs. The DLLP project began by focusing on the development of a progression for explanation, a core language function that cuts across domains. To create the progression, we collected an average of six oral and two written explanations in response to a battery of prompts provided by researchers for a personal routine (teeth-cleaning) and an academic task (mathematics) from 325 students enrolled in grades K, 1, 3, and 5 at two time points, fall and spring of the same year. The students were deliberately selected to provide diversity, including ethnicity, family income, L1 literacy and ELL status. A subsample of 100 students was sampled at two further time points after they had entered the next grade. Approximately 1,500 of 4,300 audio-recorded, oral language samples in the corpus have been analyzed to date, through a process of researcher analysis and through a database developed for the project. Cluster analysis is planned on the features generated by the database and on the researcher-coded features. While we recognize that he DLLP language features are still subject to further validity research, we have identified eight high-leverage language features to provide the core of the progression. As part of the development of the DLLP for oral explanation, and to try it out as an interpretive framework for formative assessment, we have engaged, for the past nine months, in a case study with a group of experienced elementary teachers. We were interested to find out if the course of acquisition of the language features in the progression made sense to them, and if they were able to use the DLLP to interpret their observations of students' explanations in academic contexts. Our findings from the case study suggest that the teachers' use of the DLLP has both increased their knowledge of language development and better positioned them to engage in effective formative assessment and support language learning in the content areas.

| Kathryn Hill

La Trobe University, AU | **A Framework for Promoting Teacher Awareness of their CBA Practices and Learning-oriented Assessment** | **Rationale**
While the central focus in LOA is on learners, learning processes and learning outcomes (Purpura & Turner, forthcoming), teachers have a crucial role to play in creating the conditions for learning. According to Scarino (2013), "[n]ot only do teachers need to understand the conceptual bases of different approaches [to assessment], they also need to relate such knowledge to their professional practice in their particular context" (p. 310). However, teachers don't necessarily have the skills and training to reflect in any systematic way on their own assessment practices, let alone on the theories of learning, which underpin them or the contextual forces which shape them.

Purpose
This presentation will consider how a framework originally designed to help researchers understand classroom-based assessment processes has been reframed as a tool to develop teachers' assessment literacy and promote LOA. The original framework was developed as part of an ethnographic study of foreign language classrooms in two Australian schools (Hill, 2012; Hill & McNamara, 2012). Based on observation of what teachers actually do (e.g., Leung, 2005) the framework adopts a definition of CBA designed to account for the full spectrum of CBA practices, including the more incidental types of assessment which occur as part of everyday classroom interactions (Purpura, 2014).

Methodology
Revisions to the original framework were informed by a review of taxonomies of teacher assessment practices, definitions of teacher assessment literacy and the literature on good practice in classroom-based assessment more generally.

Key findings/ Implications for LOA
While not explicitly organized within a LOA framework, the revised tool is nonetheless consistent with the principles of LOA as set out by Purpura and Turner (forthcoming). For example, the tool is designed to facilitate reflection on the nature of the assessments as well as the beliefs and understandings (about the subject, SLA, language teaching and assessment), which underpin them (Assessment Dimension). It also explores the relationship between assessments and the relevant curriculum standards and frameworks and how prior knowledge, ability level, interest level and learning needs are taken into account (Proficiency Dimension). It emphasizes the learner perspective and agency in assessment and highlights the relationship between feedback, motivation and goal orientation (Affective Dimension). Furthermore, by including a specific focus on how practice is shaped by contextual factors at the local (classroom and institutional) level as well as the social and political level, the revised framework acknowledges the inevitable gap between policy and practice as well as the situated nature of CBA (Contextual Dimension). |

| Christian Colby-Kelly

Government of Canada | **A Theoretical Analysis Approach to AFL Pedagogical Materials Development in an L2 Classroom Setting** | Learning-Oriented Assessment (LOA) applications in second language (L2) classroom settings focus on helping learners notice gaps and advance in their learning, and on helping them be more autonomous. LOA challenges teachers to stimulate their students' language learning in more engaged and effective ways than former methods have usually done (Davison & Hamp-Lyons, 2010; Fox & Harwick, 2011). In fact, the challenge of finding interesting and useful ways of applying LOA in L2 classrooms remains one of the key challenges L2 teachers may face. While it is encouraging that the field of language testing/assessment has recently evolved to include the new paradigm of LOA, to make this approach a truly viable alternative to traditional L2 assessment, more research into innovation in LOA applications and their effectiveness is needed.

    With that in mind, the present research will report on the preliminary results of a larger exploratory mixed methods investigation of the interpretation, application, and effectiveness of an Assessment for Learning (AFL) application (an iteration of LOA), in two L2 classes of pre-university students in Quebec, Canada. In particular, it will focus on the process of pedagogical material development, and evidence of the effectiveness of the materials in supporting learning will also be discussed.

    The study took a Vygotskian socio-cultural theoretical approach, and centered on these students' learning of a challenging L2 grammatical feature. In preparation for the development of pedagogical materials, the 10 Principles of AFL were categorized within 4 key characteristics, engaging (1) learner autonomy, (2) the way students learn, (3) teacher and learner assessment goal sharing, and (4) learner motivation. Based on the four features, pedagogical materials were developed incorporating constructive feedback, instances of self- and peer-assessments, learner reflections, teacher guided questioning and knowledge scaffolding, and taking into account individual learning styles. The pedagogical materials included computer-assisted language learning (CALL), an online individual concept mapping (CM) exercise, and peer-group and teacher-class concept mapping exercises. The data collection instruments included the concept maps produced, classroom observation field notes, transcribed group and class discourse, teacher and student survey questionnaires, and pre- and post-treatment tests to indicate trends. The data were analyzed by mixed methods and the results triangulated.

    The results provided strong evidence in favour of the effectiveness of the application in support of L2 learning. The survey data showed strong teacher and student perceptions that learning had taken place in all of the exercises. Evidence in the form of classroom observations and transcribed audio-recordings supports this. The results also suggest that some learners noticed gaps in their actual and target competency in L2 production tasks.

    The results of this investigation contribute to a body of evidence showing that an AFL approach may be effective in supporting learning. The present research has illustrated the process of pedagogical materials development in one setting, and as such it may inform on other iterations of LOA in other applications, in other settings where teachers may challenge their learners towards more autonomy in their learning, and promote ways to advance towards greater learning goals. |

| Ah-Young Kim[1] & Hyun Jung Kim[2]

1. Cornell University

2. Hankuk University of Foreign Studies | **Using an Integrated Reading and Writing Task for Learning-oriented Assessment in an English for Academic Purpose Setting** | Learning-oriented assessment (LOA) has gained much attention for its potential to promote learning through assessment by creating a link between the two in various learning and assessment contexts (Assessment Reform Group, 2002; Black & Wiliam, 1998; Purpura & Turner, 2013). A number of previous studies have examined the effectiveness of feedback, an element of LOA, in L2 writing (e.g., Bitchener & Ferris, 2012; Ferris, 2003; Hyland & Hyland, 2006). Yet, few have explored how LOA could be implemented using an integrated reading and writing task, especially for English for academic purposes. Therefore, the purpose of this study was to examine the effectiveness of LOA of an L2 integrated reading and writing task for the development of EFL learners' academic writing ability.

     This qualitative case study collected multiple sources of data (e.g., writing samples, teacher feedback, interview data) from learners and their instructor for a six-week period. Participants included ten graduate students majoring in TESOL at a Korean university enrolled in a research method course and their professor. Every week, each student completed an integrated reading and writing task as part of their classroom assessment, which required them to first read an academic research paper, and then write a summary and critique on it. On each completed integrated task, the instructor gave formative feedback involving detailed comments on the four areas of content, language, organization, and mechanics. After receiving feedback, students were required to revise their work and resubmit it with their new writing of a subsequent task. In addition, each student participated in a semi-structured interview to share their thoughts on the effectiveness of the formative feedback they received. The analysis focused on (1) the nature of instructor's formative feedback and (2) improvements/changes in the students' writing in accordance with the feedback provided. Students' interview data were also analyzed to account for how the students perceived and used the instructor's feedback to improve their writing.

     Findings indicated that the students displayed diverse developmental patterns in the *content* and *organization* dimensions. With regards to *content*, the first group of students continued to improve the content of summary and critique steadily throughout the six-week period. On the other hand, the second group did not provide any evidence of improvement over time. The last group showed rather unstable changes over the six weeks by displaying strengths and weaknesses in *content* in a random manner and then made progress in the final assessment; for *organization*, a couple of students showed a steady improvement while most other students' performance fluctuated. Interview findings further explained reasons for such differences (e.g., partly due to the nature of the integrated reading and writing task). The study results provide pedagogical implications for using integrated academic reading and writing tasks and sustained formative feedback for LOA. |

Payman Vafaee, Yuichi Suzuki & Eric Pelzl
 	
University of Maryland
	How Aptitude-treatment-interaction Studies can Benefit Learning-oriented Assessment	Learning-oriented assessment (LOA) aims to assist learners in the development of target-like second language (L2) performance over time by prioritizing the role of L2 processing and learning outcomes. It focuses not only on assessment but also on how instruction and feedback can be tailored to L2 learning (Purpura & Turner, 2013). For this reason, LOA needs to be informed not only by a theory of L2 testing, but also by a theory of L2 learning (Purpura, 2004). While previous studies have attempted to apply L2 learning theories to assessment (Chapelle et al., 2010), the present paper aims to demonstrate how L2 theories can contribute to the instruction component under the LOA framework. Specifically, the current conceptual paper provides a meta-analytic review of findings from aptitude-treatment-interaction (ATI) research to empirically support the instruction component of LOA.
		
		     Purpura (2004) tried to tie his model of grammar testing to VanPatten's (1996) L2 learning model, to conceptualize a model for LOA. VanPatten's input-processing model explains how learners process L2 input through a set of principles and corollaries that interact in complex ways with working memory (WM). The major principles of this model are predicated on a limited WM capacity for processing information, and predict that learners, especially at lower proficiency levels, will fail to make the right form-meaning/function connections during real-time comprehension due to the constrained capacity of WM (VanPatten, 2007). This highlights the importance of considering individual cognitive differences when making predictions about the effectiveness of any kind of intervention. These findings are also in line with the SLA field's growing interest in ATI research, which seeks empirical evidence to inform the tailoring of instruction based on learners' individual differences (DeKeyser, 2012; Vatz, Tare, Jackson, Doughty, 2013).
		
		     In the current paper, we will first summarize the major findings of ATI research, and then suggest several ways in which LOA might benefit from these findings. Due to the small number of studies originally conceptualized as ATI, our review includes studies that had an interaction component between cognitive aptitudes and instructional treatments. For example, Erlam (2005) showed that individual differences in WM and language analytic ability mediate the effectiveness of input-based treatments regardless of the presence or absence of metalinguistic information. At the same time, the results suggested that providing metalinguistic information may neutralize individual cognitive differences when learners are involved in production-based practice. Additional insights can be gained from a growing number of empirical studies examining how individual differences interact with the effectiveness of feedback (e.g., Goo, 2012; Révész, 2012; Sheen, 2007; Yilmaz, 2013). Individual differences in WM and language analytic ability have been found to play differential roles in mediating the effects of several types of feedback (e.g., recasts, explicit correction).
		
		     In sum, we will zero in on the effectiveness of different types of instruction and feedback based on ATI research findings. These findings have important implications for the individualization of learning and feedback in LOA, and the issue of fairness in LOA if interventions are implemented at a group level.

| Dina Tsagari University of Cyprus | **Unplanned LOA in EFL classrooms: Findings from an empirical study** | Learning-oriented assessment (LOA) refers to an assessment approach that focuses on the centrality of L2 processing and learning. It is the result of a range of procedures (planned or unplanned) used by teachers during the learning process that guides their instruction and students' learning. LOA is integrated in everyday classroom routines, assesses and aims at promoting learning (Purpura & Turner, 2013; 2014; Turner and Purpura, submitted). However, we still know little about the role that unplanned assessments play in learning a foreign language (FL) in classroom settings and the effect these types of assessment have on learning processes.

The present empirical study investigated the nature of unplanned LOA routines used by English as a FL teachers in Cyprus to assess, support and promote language learning while transferring, correcting and expanding learners' knowledge and language performance. Twenty-six lessons conducted by four EFL teachers in private language schools were observed, audio-recorded, transcribed and then analysed using a qualitative data analysis software (Atlas.ti). The analysis was informed by research literature on classroom interaction and teacher/peer feedback (Gibbs & Simpson, 2002; Hattie & Timperley, 2007; Hill & McNamara, 2012; Lyster & Ranta, 1997; Tunstall & Gipps, 1996) and by constructivist and socio-cultural theories of learning (Brown, 2007; Lightbown & Spada, 2006) and Vygotsky's work (1978) on language teaching and learning. This led to the design of an analytical framework, which identified unplanned assessment episodes that were later analysed in detail with respect to the types of feedback and assistance provided by the teachers. For the interpretation of the data, the study employed a sociocultural theory perspective (Kramsch, 2002; Lantolf, 2000; Lantolf & Thorne, 2006; Mickan, 2006a, 2006b) that has been having a significant impact on the interpretation of classroom experiences and practices.

The results demonstrate LOA in the present context of inquiry is an individual process situated in contextual and collaborative learning within unplanned assessment spaces and based on a layered set of interactions and sociocultural processes. The analysis revealed a number of unplanned assessment episodes with mainly a teacher-centred orientation to teaching and an excessive use of the IRF (Initiation-Response-Feedback) pattern where teachers made use of various types of feedback such as 'evaluative', 'descriptive' and 'corrective' and variants that scaffolded learning in various ways. The IRF pattern and its feedback variants employed by the teachers in the current educational context promoted language learning when adjusted to learners' needs. The paper will illustrate and exemplify instances of unplanned LOA instances and feedback scaffolding and discuss the complexities involved in conceptualising and applying LOA in EFL classrooms. Finally, the presentation will make research and pedagogical recommendations with the aim of enhancing teachers' LOA awareness and practices as part of their broader assessment literacy enhancement. |

| Jim Ranalli, Stephanie Link & Evgeny Chukharev-Hudilainen

Iowa State University | Automated Writing Evaluation for Formative Assessment: Investigating Accuracy and Efficiency as Part of Argument-based Validation | Increasing use of tools for automated writing evaluation (AWE) in composition classrooms suggests growing interest in the potential of these tools for learning-oriented assessment. As with all assessments, however, they should be validated in terms of their intended uses and interpretations (Kane, 2012). A recent argument-based validation framework (Chapelle, Cotos, & Lee, 2013) outlined a range of inferences that would require backing to support the use of one widely used AWE tool, Criterion, in a college-level course for improving ESL students' academic writing skills. The present research contributes to a critical review of this framework by investigating evidence for the assumptions underlying two key inferences in the argument.

In the first of two studies, we investigated backing for the evaluation inference, which is based on the claim that Criterion provides students with accurate feedback. While some published studies discuss the accuracy – in natural language processing terms, the precision — of Criterion feedback (e.g., Chodorow, Gamon, & Tetreault, 2010), the findings are not germane for our validation purposes because they address few of the types of feedback commonly received by students in our target context, which was the same as that in the study by Chapelle, Cotos, & Lee (2013). To obtain relevant accuracy data, we collected all writing submitted to Criterion by students in courses for lower- and higher-proficiency students over a semester and identified ten of their most common error types. These error types were represented in a six-hundred error sample extracted from the corpus and rated by two expert judges. The findings show high accuracy rates for some error types (e.g., ill-formed verbs) but low rates for others (e.g., those related to article usage).

The second study focused on the utilization inference, which involves the assumption that learners are able to use the AWE feedback effectively in improving their written work. The same corpus of Criterion data used in the first study was consulted to devise an error-correction task featuring multiple instances of the 10 common error types. In addition to correcting the errors, participants were also asked to rate each item for the amount of mental effort they expended in completing it. These performance and mental effort ratings were used to calculate a coefficient of instructional efficiency (van Gog & Paas, 2008). Contrary to expectations, we found that learners in the lower-level course were more efficient in making revisions based on the Criterion feedback, which may relate to how use of the AWE tool is incorporated into a course by instructors.

The findings provide neither clear support nor clear refutation of the evaluation and utilization inferences, but offer insights for how empirical research can help to validate the intended uses, interpretations, and consequences of AWE formative feedback. The refinement of validation for use of AWE in the classroom can thus expand to additional inferences to provide a unified view of AWE formative assessment. |

Ardeshir Geranpayeh & Gad Lim
 	
Cambridge English Language Assessment
	The Role of Automatic Writing Assessment in Providing Diagnostic Feedback to Learners	The field of automated assessment of text focuses on automatically analyzing and assessing the quality of writing. A number of systems have been developed, usually employing natural language processing and machine learning techniques, where language learners can submit samples of their writing and receive feedback on them. Some (e.g. Conference on College Composition and Communication, 2009) have criticized the use of these systems as being ultimately counter to good writing and instructional practices. Indeed, contemporary approaches to validity and validation require developers of assessment instruments to account not only for the construct they are measuring, but also for their use, consequences, and impact.
		
		     In this paper we discuss the development of a Write and Improve online system that can be used to provide diagnostic feedback within a learning-oriented, self-access, and/or teacher-directed assessment context, and address issues of construct validity, as well as of assessment use and impact.
		
		     The system maps learners' output to proficiency levels defined by external benchmarks and frameworks of language ability. Implicit in this mapping is the identification of positive and negative writing features related to topic relevance, organization and structure, language and style. We consider these features in terms of the degree to which they cover models and constructs of writing ability. These features can be weighted in different ways to maximize their predictive power within different language use contexts and for different L1 backgrounds. Learners can therefore receive overall and specific feedback that diagnoses the quality of their writing according to context, increasing construct validity.
		
		     Where use and impact are concerned, we have taken in to account research into how the intuitiveness, aesthetics and usability of diagnostic tools impact learning. In response, the system employs visualization techniques, highlighting parts of learners' texts according to their overall quality. This is based upon evidence from word and part-of-speech n-grams, identified grammatical constructions, and an estimated error rate. As writing quality improves, the background color gradually changes from red towards green. This displays in an intuitive way the transition from low-quality writing to high-quality writing. The system also stores all texts submitted by the learner.
		
		     Within a learning-oriented assessment context learners can, perhaps assisted by their teacher, repeatedly access the system and continuously work on improving various aspects of their writing, referencing their earlier work as desired. Using questionnaires to collect feedback from teachers and learners, we show that the system can be a useful supplement to other modes of learning that can help to promote learners' writing development.

John Norris Georgetown University	**Discussant**	
Tim McNamara University of Melbourne	**Discussant**	