BAAL Testing, Evaluation and Assessment SIG

Assessment for learning and assessment of learning: Incommensurate paradigms or complementary perspectives?

Friday 11 March 2016, 9:55 – 17:15

Meadow Suite, Park House, University of Reading, UK

Conference Programme

9:55 – 10:00	**Welcome & Introduction**
10:00 – 10:30	**Invited talk:** Learning Oriented Assessment – A systemic approach Nick Saville (Cambridge English Language Assessment, University of Cambridge)
10:30 – 11:00	Formative feedback and writing assessment: any evidence of learning? Prithvi Shrestha (Open University)
11:00– 11:20	**Coffee Break / Group Photo**
11:20 – 11:50	Does assessment facilitate learning? Students' perceptions of assessment on a pre-sessional course Qian Zhang, Cain Barriskill & Bob Fisher (University of Northampton)
11:50 – 12:20	Performance of interactional competence with and without extended pre-task planning time: the case of classroom-based speaking assessment in Hong Kong Daniel Lam (University of Edinburgh)
12:20 – 13:30	**Lunch / Poster Session**
13:30 – 14:00	**Invited talk:** Measuring second language development: Issues and challenges in instructed SLA research Andrea Révész (University College London)
14:00 – 14:30	Political, economic, social and technological principles for implementing effective learning-oriented language assessment Liz Hamp-Lyons & Tony Green (University of Bedfordshire)
14:30 – 15:00	Language policy and assessment practices: Experiences with formative language assessment for students and academic staff in EMI programmes at a Dutch University Kevin Haines (University of Groningen)
15:00 - 15:15	**Coffee Break**
15:15 – 15:45	Strategy portfolios as a means to promote L2 learner reflection and self-assessment and to improve teacher assessment of strategy use Takeshi Kamijo (Ritsumeikan University)
15:45 – 16:15	Learning from assessment: corpus insights into strategic and pragmatic competences Elaine Boyd (Trinity College London)
16:15 – 16:45	Panel Discussion
16:45 – 17:15	**Closing, then TEA SIG AGM**

Paper Abstracts

10:00 – 10:30 **Invited talk:** **Learning Oriented Assessment – A systemic approach**
Nick Saville (Cambridge English Language Assessment, University of Cambridge)

There has been considerable controversy about the negative impact of high-stakes examinations and a plea for assessment to be more learning-oriented. However, the traditional dichotomy between summative and formative assessment is not helpful in furthering this debate. In this talk, I argue for a systemic approach that puts learning at the heart of all assessment contexts.

The challenge is to combine familiar functions of language tests and examinations into a systemic relationship with assessments that result from classroom interactions. How can assessment and teaching expertise be combined to promote effective learning?

SLA researchers acknowledge the central role that assessment plays in language classrooms and the need to relate assessment principles to teaching/learning activities. While the empirical research into classroom-based assessment provides useful insights, it is still unclear how it fits within a coherent framework linking both formative and summative uses of assessment.

In taking a systemic approach, I outline key features of a model that describes how formal and informal assessments are conceptualized from a learning perspective to create a "learning ecology". I also consider the transformative role that technology can play in implementing the model and how data collection can contribute to the individualization of learning pathways within it.

10:30 – 11:00 **Formative feedback and writing assessment: any evidence of learning?**
Prithvi Shrestha (Open University)

Research on assessment and feedback in Higher Education (HE) indicates there is lower student satisfaction regarding assessment and feedback than for other aspects of HE academic support (HEFCE, 2010). It is argued that assessment and feedback is conservative and monologic in HE (Nicol, 2010). Therefore, there is a need for further investigation of how assessment could be made more responsive to students' needs and their learning. This study explored the nature of formative feedback provided in an open and distance learning undergraduate module that focused on academic literacy skills for business studies. Particularly, this study examined the relationship between the formative feedback provided by tutors and the extent to which students' academic writing development did/did not improve over two assignments. The study followed a mixed methods approach in which the data were collected through a survey (n = 22), and student (n = 8) and tutor (n = 6) interviews. Additionally, student assignments (n = 16) and associated tutor feedback (n =16) were collected from those students who were interviewed. While the survey data were statistically analysed, the interview data were thematically examined. The student assignments and the associated tutor feedback were analysed to explore any evidence of student learning (i.e., academic writing development) which may have been as a result of the tutor feedback. The findings suggest that there is some evidence of positive impacts of assessment materials and tutor feedback on student learning. These findings have implications for improving institutional assessment strategies and tutor formative feedback practices in higher education.

References
HEFCE. (2010). *National Student Survey: Findings and trends 2006 to 2009*. London: Higher Education Funding Council for England.
Nicol, David. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education, 35*(5), 501-517.

11:20 – 11:50 **Does assessment facilitate learning? Students' perceptions of assessment on a pre-sessional course**

Qian Zhang, Cain Barriskill & Bob Fisher (University of Northampton)

This study is the follow up of our previous research carried out in April 2015 and is also part of our on-going project of 'Exploring and implementing learning-oriented assessment in Pre-sessional courses at the University of Northampton'. Based on the consolidated framework of learning-oriented assessment (Carless et al, 2006, p4), there are three main areas: assessment tasks, tutors' experiences of assessment and students' experiences of assessment. Tutors' experiences of assessment were discovered through a focus group discussion (5 pre-sessional teachers) (Zhang, Fisher, Barriskill, 2015).

The aim of the current study is to identify and describe students' experiences of assessment by using a questionnaire. 239 pre-sessional students and 13 members of staff completed a 26-item questionnaire plus two open questions on their experience of assessment. The questionnaire items include the purposes and nature of assessment, self/peer assessment, marking criteria and feedback. The results indicated that students agreed that assessment is used to motivate learning and focuses on development of knowledge. In general, students found feedback improves learning and is helpful in detail. In addition, both self and peer assessments happen frequently. However, it did not seem that students were clear about the marking criteria. The results also indicated there was difference of perceptions between staff and students.

11:50 – 12:20 **Performance of interactional competence with and without extended pre-task planning time: the case of classroom-based speaking assessment in Hong Kong**

Daniel Lam (University of Edinburgh)

On the effect of pre-task planning time on subsequent task performance, the majority of the testing studies have focused on proficiency measures in the discourse output of monologic tasks, while little research has examined such effect on the interactional patterns of dialogic tasks (Nitta & Nakatsuhara, 2014). Situated within the context of the School-based Assessment (SBA) component for the HKDSE, where planning time continues to be an issue (Hamp-Lyons, 2015), this paper compares students' performance of interactional competence in a group interaction task under different task implementation conditions.

With data taken from a larger project, the performance of a same group of four students under different task implementation conditions – with and without extended pre-task planning time – was examined. The students had 4-5 hours preparation time in the real assessment, while they had only 10 minutes in a mock assessment two months later, when the students performed the same task with some changes in the discussion prompt. Micro-analysis of the assessed discourse provides preliminary evidence that the group interaction task implemented without extended planning time has a higher capacity to discriminate between stronger and weaker candidates in terms of spontaneous production of responses contingent on the previous speakers' talk. In contrast, under the extended planning time condition, where students were found to pre-plan or pre-script the assessed interaction, the distinction between the spontaneous performance of such interactional competence in real-time interaction and the use of such competence in preparing and animating scripted responses is largely obscured.

BAAL Testing, Evaluation and Assessment SIG

13:30 – 14:00 Invited talk: **Measuring second language development: Issues and challenges in instructed SLA research**

Andrea Révész (University College London)

This talk will discuss some methodological issues and challenges in measuring second language development taking a cognitive-interactionist perspective. In particular, my focus will be on exploring ways to ensure that valid elicitation instruments and scoring procedures are used in experimental and quasi-experimental studies in instructed second language acquisition research. I will highlight the importance of utilising multiple outcome measures gathered from various linguistic and communicative contexts to assess second language learning in order to decrease the risk of construct underrepresentation. Drawing on the notion of transfer-appropriate processing, I will emphasise the need for including assessment instruments in study designs that closely match the instructional treatment. I will also showcase some innovative recent instructed SLA studies, and demonstrate how behavioural measures, such as reaction time and eye-movement data, can be employed to provide evidence about improved use of second language constructions. Next, I will turn to a discussion of challenges involved in selecting valid measurement units, and argue for utilising various developmental measures that enable capturing interlanguage development that has not yet reached target-like accuracy levels. The talk will end with some methodological recommendations for the assessment of L2 development in future instructed SLA research.

14:00 – 14:30 Political, economic, social and technological principles for implementing effective learning-oriented language assessment

Liz Hamp-Lyons & Tony Green   (University of Bedfordshire)

A range of innovative assessment types have in the past ten years or so emerged in the assessment of language learning and language in use. The naming can be confusing, but Green and Hamp-Lyons (2013; forthcoming) and Hamp-Lyons & Green (forthcoming) suggest the use of the term 'learning-oriented language assessment' as a self-explanatory umbrella term.   This paper draws on several recent projects to draw out concepts and issues of principle around the central claim that effective language assessment is, or should be by its nature, learning-oriented and interactional.
Acknowledging the complexity this implies, the paper argues that a learning-oriented, interactional, approach to assessment cannot be expected to be successful in the classroom unless conditions at higher levels—the school, the education system, the state, and the society as a whole—are sympathetic to reform. We construe these conditions as forces, even principles: the PEST (political, economic, social and technological) principles that are so often forgotten or ignored in education reform. The paper also makes an increasingly familiar and urgent argument that innovative assessment types are unlikely ever to be successfully introduced until much better quality professional development for teachers in assessment literacy becomes the norm in pre-service and in-service teacher education. Evidence is building that this need is common across virtually all second/foreign language learning contexts.

14:30 – 15:00 Language policy and assessment practices: Experiences with formative language assessment for students and academic staff in EMI programmes at a Dutch University

Kevin Haines (University of Groningen)

At the University of Groningen (The Netherlands), an internationalization process has fuelled a

growth in English-medium Instruction (EMI) programs. Non-Dutch speaking students can only access higher education if Bachelor's and Master's programmes are run through English. To enable the further expansion of these programmes, the university has established an International Classroom (IC) project and a Language & Culture (L&C) policy. Over the last decade, the University Language Centre has used the CEFR as a foundation for the development of the formative assessment of the English language skills of students and academic staff in EMI settings, with a focus on the assessment of language in the academic context and the alignment of assessor interpretations. We have previously discussed our approaches to the assessment of the language of academic teaching staff in classes delivered through English (Haines et al 2013a) and of the academic writing produced in English by students following these classes (Haines et al 2013b). However, these observations were focused on the support of students and staff in specific EMI programmes, and the question now is whether we can extend these methods across the university. We believe the CEFR can provide a bridge from an EMI-focused perspective on internationalization to a discussion of language policy as it affects the entire university community, which involves the assessment of Dutch and possibly other languages as well as English. In terms of assessment, this raises complex issues, including the question of how to address the inter-relation between linguistic and intercultural skills.

References
Haines, Kevin, Estelle Meima & Marit Faber. 2013a. Formative Assessment and the Support of Lecturers in the International University. In Tsagari, Dina, S. Papadima-Sophocleous, and S. Ioannou-Georgiou (eds.), *International Experiences in Language Testing and Assessment*, 177–190. Frankfurt am Main: Peter Lang GmbH.

Haines, Kevin, Wander Lowie, Petra Jansma & Nicole Schmidt. 2013b. EMBEDding the CEFR in Academic Writing Assessment: A case study in training and standardization. *Dutch Journal of Applied Linguistics 2*(1). 77–91.

15:15 – 15:45 **Strategy portfolios as a means to promote L2 learner reflection and self-assessment and to improve teacher assessment of strategy use**
Takeshi Kamijo (Ritsumeikan University)

Although think-aloud methods and questionnaires have been extensively used to characterize individual learners' cognitive strategies, these two approaches may be insufficient for assessing metacognitive strategies in the L2 learning process. Learner reflections can be used to assess the metacognitive awareness of strategy use, such as assessment, elaboration, and development, as part of the L2 learning process. In particular, evidence-based learner reflections in the reading strategy portfolios can contribute to both assessment and learning. The present study analyzed learner reflections contained in the strategy portfolios of members of a Japanese university's L2 reading class through a mixed methods assessment. Data from participants' reflections were coded and assessed quantitatively, identifying successful and less successful learners. Furthermore, a qualitative text analysis was conducted to examine the cognitive and metacognitive strategies used by selected successful learners. The analysis indicated that these learners continued to self-assess their strategy use's efficacy and elaborated their strategies during the learning process. The study indicated that learner reflections in strategy portfolios could be a valid method for teachers to assess second language learners' strategy use in their learning process. Also, portfolios enable teachers to actively facilitate learners' self-assessment and awareness of their strategy use during the learning process. The present study is based on a framework of sociocultural theory and metacognition. The insights regarding metacognition emerging from this study have educational implications for both

practitioners and researchers.

15:45 – 16:15 Learning from assessment: corpus insights into strategic and pragmatic competences

Elaine Boyd (Trinity College London)

The build of a new spoken corpus of learner language by CASS Lancaster, based on oral assessments conducted by Trinity College London, has generated valuable insights into how learners negotiate meaning within a test. This in turn has provided a window onto a range of strategic and pragmatic competences that can be taken back to the classroom in order to enhance learners' communicative skills.

The corpus, which contains over 3 million words, was investigated for features of successful and less successful candidates at B2 level. Along with more anticipated elements, some interesting additional aspects of language emerged which were characteristic of those achieving high scores in the test. These almost all represented features of strategic or pragmatic competence.

This work in progress looks at the results so far from this project and how this data emerging from the test might offer new insights for the classroom. It suggests that some of the most effective features of communication may need to be given more overt attention by teachers and learners. It also raises the complex issue of how far tests designated for one purpose can or should overlap with multiple purposes. The study also demonstrates that, by studying what learners actually do 'around' the test, we can identify data which is genuinely helpful for classroom practice and learner progress in what can be a frustrating field for them.

****

Poster Abstracts

- **A vocabulary framework to align learning, teaching, and assessment**
  Veronica Benigno (Pearson English, UK) **& John de Jong** (Pearson Global Assessment Standards/Vrije Universiteit)

At present there is no valid framework of vocabulary which outlines what vocabulary is needed at increasing proficiency levels and therefore complements the functional guidance offered in the CEFR (Council of Europe, 2001) descriptors. This poster reports on a project to develop a meaning-based and CEFR-linked lexical inventory of general English and was set up in response to the Council of Europe recommendation to create, for each language, inventories of linguistic forms known as Reference Level Descriptions (Council of Europe, 2005). The project seeks to align the learning, teaching, and assessment practice by providing a clear description of the intended goals for vocabulary (assessment for learning) while helping assess individuals against the set standards (assessment of learning).

The inventory was built by combining frequency analysis and teacher judgements on usefulness of vocabulary. In a first step, a frequency list was extracted from an L1 corpus of about 2.5 billion words. Next, approximately 36,000 word meanings (corresponding to about 28,000 lemmas) were semantically annotated by meaning using the categorization in Specific Notions, General Notions, and Functions of the Council of Europe Vantage Specifications (2001). Each word meaning was then rated on communicative usefulness by 20 teachers. In a final step, a weighted algorithm was used to combine the frequency data and the teacher ratings and align each entry to

the CEFR.

In this presentation we propose a meaning-based approach to vocabulary learning and present a model to assess vocabulary as a unitary construct, combining quantitative and qualitative considerations about vocabulary knowledge.

● **Web-authored products, rubrics and assessment**
Neil Cowie (Okayama University)

Digital technology, especially the use of Web 2.0 tools, is being used to help language teachers provide more formative assessment of students as a classroom-based guide for future learning. Such technology has affordances that can facilitate 'constructive alignment'; that is, it can strengthen the link between class activities, assessment tasks and learning outcomes. For example, students can create and share digital products such as videos, blogs, and slideshares; these products include language mechanics but can go beyond this to showcase many different kinds of skills and abilities and can be kept in e-portfolios of work that students can use to provide evidence of their learning and achievements. However, both teachers and students need a new level of awareness in assessing such web-authored products. A possible way for both teachers and students to learn to assess such digital products is for both parties to get involved in the process of assessment, specifically in rubric construction. This poster presentation will investigate the process in which English as a Foreign Language (EFL) teachers and students in a Japanese university collaboratively negotiate the process of rubric construction and the use of such an assessment tool throughout one academic semester. The collaborative process highlights two challenges that the teachers and students face: 1) how to assess the combination of language use and digital products; and, 2) how to empower teachers and students in the digital age.

● **Distinguishing summative and formative assessment**
Simon Dawson (Bristol Universtiy)

This proposed study will focus on the importance of maintaining a distinction between formative and summative assessment. Formative assessment is typically low stakes assessment given during a course of learning with the purpose of guiding student learning in the ongoing teaching and learning context. Resource (teacher time) is generally put to providing useful guidance and time is invested in ensuring student understanding of feedback (Huhta, 2013). Summative assessment, which usually comes toward the end of a period of study and in the form of an examination, is generally higher stakes and more standardised. The outcome of summative assessment is used to inform efforts and actions outside of the course of study (Huhta, 2013). Resource for summative assessment is focused on making the measurement as accurate, reliable and fair as possible (Bloxham, 2007). Due to the ratio between student and teacher/assessor, it is unusual for students to receive more feedback on their level of performance in the exam than a single score or breakdown of scores. If greater feedback is given, the resources allocated to 'measuring accurately' may be compromised (Bloxham, 2007). This paper looks at a case where, as a result of student feedback, a summative assessment was altered to provide more feedback to test-takers (12 scores were given rather than 3). The result was more information about exam performance for test-takers but, at the same time, a significantly greater load on assessors. The paper investigates whether the change was worthwhile by gathering opinion from students and assessors through focus groups. Based on the findings, suggested revisions are made that attempt to balance the need for accurate measurement with the need for feedback from summative assessments.

- **Using Formative Assessment to Monitor Students' Progress in Writing and Reading Skills for Academic Purposes**

Saeede Haghi (University of Warwick)

In recent years, studies of assessment have witnessed a burgeoning interest in the interactions between assessment and learning and teaching practices. The use of assessment is no longer confined to measuring learners' achievement: it is also used as a means of providing information that can facilitate learning and instructional practices. Formative Assessment (William and Black, 1998) which serves such a function is regarded as a powerful means of enhancing student achievement (Irons, 2008). Considering this method of assessment and its efficacy in improving student performance, a case study project including the design, development and use of a series of formative assessment tasks for evaluating reading and writing skills for academic purposes, was conducted for a pre-sessional course in the UK. This project investigated the practicality of designing formative tasks from EAP textbooks and the effectiveness of using these tasks in relation to students' learning gains. To this end a series of narrative inquiries, including pre-sessional tutors' reflective pedagogy and students' voices in the effectiveness of using FA tasks were collected. The findings of this research project will provide insights for both EAP tutors and learners including the feasibility of designing FA tasks by adapting teaching activities in EAP text books, the realistic link between the theory and practice of FA, and the effectiveness of these tasks in finding the learning gap, monitoring learning progress, and setting and achieving learning goals.

- **The English language learning outcomes at primary school level of young EFL learners in mainland China**

Xi Lian (University of Oxford)

There is a worldwide tendency that children start learning a foreign language, especially English, from an earlier age. Governments from many countries are incorporating English as a foreign language (EFL) programs into primary school curriculum and are lowering its starting age. In such situation, attention has drawn on the effectiveness of early EFL programs at primary school level in the state education systems. However, there still has been few research into early EFL learning, especially in the formal school setting. And there still has been little known about the learning outcomes of these EFL young leaners. In specific, how much they have achieved according to the curriculum level descriptors; how they have been assessed and what task types were used by teachers/local educational authorities; and how these EFL young learners themselves be aware of their language progress.

This study therefore investigates the learning outcomes of young EFL leaners in China at primary school level, in specific, to examine whether the learning outcomes of Chinese EFL young learners have achieved the targets described by the general 'can do' statements for the language skills (listening, speaking, reading and writing) at Level 1 and Level 2 in the National English Curriculum Standards (NECS). This project first used a series of measures (large scale standardized tests and research-based assessments) to assess children's language skills and to see if their learning outcomes have achieve the NECS targets, and then investigated how teachers/local authorities actually assessed their students and what tasks they were using and how these tasks were related with the NECS targets.

● **L2 metaphoric competence: Moving from classroom task to assessment tool**
David O'Reilly (University of York)

The ability to comprehend and use metaphor in a second/foreign language (L2 metaphoric competence) has long been recognised as an important part of L2 language competence. In acquiring a second/foreign language, learners need to hone very specific metaphor related skills, which range from improving knowledge of the boundaries of conventional metaphor, multiple layering and hedges (Low, 1988) to mastery of ideational, heuristic and imaginative functions (Littlemore & Low, 2006).

For language teachers seeking to facilitate and promote students' L2 metaphoric competence, the primary concern is likely to be the development and use of metaphor based tasks for use as classroom activities rather than assessment tools. Empirical research (e.g. Azuma, 2005; Littlemore, 2001) on the other hand, has placed focus on the development of such tasks as tools for measuring (L2) metaphoric competence, a process admittedly fraught with methodological challenges.

This study reports on the development of a metaphoric competence test battery and its usefulness as an alternative form of assessment for EFL practitioners. Specifically, the rationale for the development of the test as a research and practitioner tool, the constructs measured, scoring, test administration and practical considerations are discussed.

● **The co-construction of discourse between candidates and examiners**
Lucy Passmore (King's College London)

Interactional competence (IC) has been recognised as an important aspect of both assessment and learning as it can be used to facilitate future real-life communication. A key aim of Trinity College London's Graded Examinations in Spoken English (GESE) is to promote natural interaction between the candidate and examiner. Testing a candidate's ability to interact naturally with an interlocutor could potentially have a positive impact on learning, since candidate preparation for the GESE will need to incorporate the kinds of interactional skills needed for successful communication. However, the exact nature of the interactional features elicited by the GESE has yet to be fully investigated. The aim of this exploratory study was to investigate the extent to which grade 7 of the GESE provides opportunities for the co-construction of discourse to occur between candidates and examiners and whether a case can be made that the exam promotes interactional competence. The study applied conversation analysis techniques to 10 transcribed GESE examinations in order to determine the interactional features used by examiners and candidates. The main findings were that the three phases of the GESE presented different opportunities for the co-construction of discourse and that candidates generally took advantage of these opportunities. It was concluded that grade 7 of the GESE taps several aspects of interactional competence. The findings of the study could potentially contribute to the construct validity of the GESE and promote the development of IC as part of preparation for the examination.