

Learning Oriented Assessment: A systemic view

Neil Jones and Nick Saville
Cambridge English Language Assessment

Abstract

The term Learning Oriented Assessment (LOA) is one of several used in recent years with a similar purpose: to carve out a place for forms of assessment with different priorities and values to those of traditional assessment, with its focus on reliability and validity. All such approaches have been ‘explicitly or implicitly defined in opposition to traditional externally set and assessed large scale formal examinations’ (Davison and Leung 2009:395). They are reforming initiatives aimed at curing, or at least mitigating, what are seen as fundamental problems with current approaches to assessment in schools. For an examination provider the challenge of LOA is to link familiar functions of language assessment (e.g. diagnosis, monitoring, or accreditation) into a systemic relationship with the forms of assessment resulting from planned or unplanned classroom interactions. How can quantitative and qualitative information, or assessment and teaching expertise, be best combined to promote learning?

The need for a systemic view of assessment

The “systemic view” offered in this paper presents assessment as something which operates on multiple levels in an educational context and takes many different forms. It proposes a model that encompasses both the *macro level* of framing educational goals and evaluating outcomes, and the *micro level* of individual learning interactions which take place in the classroom or outside it.

It is written from the point of view of an assessment body, more familiar with assessment’s traditional functions in education: determining curricular aims, diagnostic profiling of students at entry, monitoring progress against targets, and final evaluation of learning outcomes. Within this tradition, conceptions of reliability, validity and validation have been developed which are appropriate to these functions, as well as technical systems to support them, such as item response theory (IRT) scaling and item banking. In this trait-based conception of assessment we pursue validity by carefully defining and measuring constructs, with the primary purpose of locating candidates on a proficiency scale.

How does the expertise put into developing exams impact on learning? The historical view is a simple one: exams provide curricula which define learning objectives, and final accreditation for successful learners. Cambridge English’s assessment roots are very much in this tradition (the 100th anniversary of the Cambridge Certificate of Proficiency in English was celebrated in 2013), and most of our candidates are still found within educational settings.

Much has changed in the last twenty years: the exam suite has developed from a couple of high-level exams into a learning ladder of graded objectives, scaffolding learning from beginner level onwards. The range of support offered to teachers and learners continues to expand, as does engagement with the learning contexts in state and institutional sectors. But so far the traditional relationship between assessment and learning has not fundamentally changed: exams remain external, summative commentaries on learning. At best they have washback effects that may impact positively on outcomes. However, they may also have unintended effects (see Salamoura, Khalifa and Docherty, 2014, for two case studies

illustrating such effects), and remedial action may be required to achieve the intended outcomes (as advocated by the *impact by design* approach (Saville 2009); also discussed later in this paper).

There is a general trend for assessment bodies to engage more directly with education. Various incentives lead to closer alliances with publishers and media companies. Educational initiatives may throw up new business opportunities, such as the No Child Left Behind legislation in the US. A technological imperative leads to assessment being integrated into the delivery of learning material within learning management systems. Visions are promoted of assessment and technology-driven intelligent tutoring systems that will revolutionise education, though this has yet to be demonstrated. Thus for a range of motives assessment bodies are seeking new forms of engagement with learning. At the same time, educationists are seeking new forms of engagement with assessment.

Turner (2012) reviews the emergence of Classroom Based Assessment (CBA) as a field of activity separate from high-stakes assessment and requiring different conceptualisation of such fundamentals as reliability and validity. Turner dates this development to the 1990s, considering it to be ‘still in its infancy as an independent domain and paradigm’ (Turner 2012:66). Finding that ‘the time is ripe to begin to formulate a research agenda’ she poses a number of questions, slightly paraphrased below:

- What are the quality criteria? How can reliability and validity be reconceptualised within a socio-cultural framework?
- What are the characteristics of assessment tasks that provide a context for learning? What is the evidence for these bringing about change in learning?
- What is the nature of teacher/student feedback and reflection that influences the effort towards and outcome of further learning?
- What defines the role(s) and responsibilities of “assessor”? How do teachers and students interpret their roles? What impacts on their decision making?
- How do teachers reconcile CBA and preparation for external tests? What are the commonalities needed to create coherence across these different assessment components?
- How should pre-service and in-service teachers be supported so they can effectively engage in assessment that supports learning?

We will offer initial answers to these questions at the end of the paper.

An appropriate model of cognition

The renewed engagement with education creates pressure to change the relationship between educational assessment and the process of learning and teaching. One approach to this looks to implement far more detailed models of competence, supported by even more powerful statistical methods. This programme goes back as far as Frederiksen, Misley and Bejar (1993) who found that ‘the view of human abilities implicit in standard test theory - item response theory as well as classical true-score theory - is incompatible with the view rapidly emerging from cognitive and educational psychology.’ In this view, trait-based measures fail to capture the complexity of abilities in the way necessary to understand learning or impact positively on it. Detailed cognitive models are needed, and approaches to measurement that can deal with them. This programme has been pursued through two major projects at Educational Testing Services (ETS): work on Evidence-Centred Design (ECD), and on Cognitive Diagnosis Approaches (CDA). The focus on cognition has clear relevance to formative assessment, and is developed in this direction by Pellegrino, Chudowsky and Glaser (2001).

However, Shavelson (2008, 2009) warns against premature applications of cognitive development models, particularly for specifying learning progressions. He cites research to counter the assumption that development is neatly describable, or follows predictable paths.

Successful formative assessment is difficult to reduce to rules, and depends on a high level of teacher training.

We would add further that the cognitive approach as presented by Pellegrino et al is illustrated chiefly by examples from maths and science, which stress the process of problem-solving. It is a constructivist approach, but constructivism comes in two flavours. As Sjøberg (2007) explains, the concept of *cognitive constructivism*, developed by Jean Piaget, treats general aspects of the development of knowledge: ‘He was not so much interested in education, let alone teaching or conditions for good and effective learning’ (2007:7). Constructivism in education latterly puts more stress on social and cultural conditions for learning: *social constructivism* is associated above all with the name of Vygotsky (1986). Shepard (2000) advances a “social-constructivist” conceptual framework, stressing that ‘a singularly important idea in this new paradigm is that both development and learning are primarily social processes’. The cognitive modelling approach seems to miss this point. The goals of learning go beyond attainment of curricular objectives and understanding of concepts, to encompass a range of skills, dispositions and attitudes that inspire further learning (James and Brown 2005:10-11). Learning is chiefly about becoming a better learner, and learning cannot be divorced from its social context. And if social constructivism provides a better model for classroom learning in general then this must be all the truer for the special case of language learning, which is unique among school subjects in the range of learner attributes – cognitive, psycho-motor and affective – which it engages (Coleman 2004). There is more to it than knowledge of rules. Our own work in Cambridge has centred on developing a *socio-cognitive approach* to construct definition (presented below) which locates the development of learner cognition clearly in a social context.

A complementary relationship with teaching

We are sceptical of whether the essentially reductionist approach proposed by Pellegrino et al can actually contribute to supporting learning at the classroom level. Diagnosis is only a starting point for formative activity, which, we assert, entails *interaction*; and thus a pivotal role for teachers, as participants in or coordinators of that interaction. Jones (2012) states:

When we consider the nature of formative interactions within the classroom it becomes clear that learners’ states of understanding or mastery can hardly be analysed in isolation from the interactions themselves (Teasdale and Leung 2000). Cognition is socially constructed and begins in interaction. Model-based diagnosis of cognitive attributes requires stable observations, something hardly to be expected at the growing point where learning is happening. It is easier to see the relevance to summative than to formative assessment (Jones 2012:360).

In metaphysical terms we should see learning as an emergent power, that is, something qualitatively different from, and irreducible to, its parts (Sayer 1992:119).

In order to find an ecological and productive role in learning, we must recognise the limits of assessment-as-measurement as a means to understanding. Assessment is useful; it enables us to locate learners somewhere on a scale, and this already impacts positively on learning, by providing an orientation for learners as to their current level, and the progress that has been made. It helps to define learning goals and to measure off the distance to their achievement. It can provide a useful level of profiling regarding more- or less-developed skills, and even some more detailed diagnostics. It can link performance to interpretative frameworks such as the Common European Framework of Reference for languages (CEFR; Council of Europe 2001), adding layers of useful meaning. Assessment provides one

dimension of the learning landscape: the vertical progression (vertical if we use the metaphor of levels, stacked one above the other). Let us think of this as the quantitative dimension.

The second dimension – the horizontal, qualitative one – captures how learners who are all at the same global proficiency level differ from each other in terms of their cognition, their experience, and their learning needs. If the vertical dimension is the domain of assessment experts, the horizontal dimension is the domain of the classroom and the teacher. The model of Learning Oriented Assessment (LOA) presented in this paper thus foresees a central role for the teacher in creating an environment productive of learning, complementary to the role of formal assessment.

We also recognise that in many contexts teachers are pulled two ways, being assigned simultaneous formative and summative roles (Tattersall 2004). This complicates life for teachers and for us in defining LOA. It is matter for another paper; in the present paper we will focus on the formative role. We will have more to say below about the central importance of teacher-facilitated classroom interaction.

Impact by design

As noted above, it is becoming increasingly common for state or institutional sectors to adopt Cambridge English exams as a voluntary or integral part of certifying language learning achievements in school. Studying the impact of our exams within such new educational settings has become a research priority (see, for instance, issue 50 of *Research Notes* which includes a collection of recent Cambridge English impact studies: www.cambridgeenglish.org/research-notes). In the past impact studies have generally been conceived as a post-hoc endeavour to evaluate the wider effects, intended or unintended, positive or negative, of introducing an educational innovation such as a new exam. However, we can argue that an assessment body has an obligation as far as possible to maximise the positive and minimise or eliminate negative impact. Saville (2009; 2012) describes this as positive “impact by design”. This implies developing *a priori* models of how assessments can be introduced and integrated into a learning context, and goes beyond the design of the assessment itself: necessarily, it involves planned modifications to classroom practice (see Salamoura et al, 2014, for a discussion of the application of the impact by design model in various educational contexts). LOA, as presented below, can be seen as a theory of action aimed at achieving positive impact by design.

A theory of Learning Oriented Assessment

We accept Bennett’s (2011) view that conceptualisations of formative assessment should incorporate fundamental measurement principles, and be open to validation to the same extent (if not necessarily in the same way) as any other kind of assessment. A model of LOA must have theoretical underpinning, particularly as LOA as we present it here links all levels and dimensions of assessment into a single, coherent framework. Where large-scale assessments are linked explicitly to learning programmes and to classroom practice then they must validate each other: the large-scale assessment must validly reflect the goals of the programme, and the programme must validly set out to equip learners for the assessment.

In fact the theory comes almost ready-made. The model of validity which supports Cambridge English exam development is based on the *socio-cognitive* model of language learning proposed by the CEFR, which refers to:

...the actions performed by persons who as individuals and as social agents develop a range of *competences*, both *general* and in particular *communicative language competences*. They draw on the competences at their disposal in various contexts under various *conditions* and *constraints* to engage in *language activities* involving *language processes* to produce and/or

receive **texts** in relation to **themes** in specific **domains**, activating those **strategies** which seem most appropriate for carrying out the **tasks** to be accomplished. The monitoring of these actions by the participants leads to the reinforcement or modification of their competences. (Council of Europe 2001:9, emphasis in original).

“Socio-cognitive” indicates that language ability is seen as a mental construct exhibited in social interaction. This model has been further developed by Weir (2005b), and in collaboration with researchers at Cambridge English: four volumes in the Cambridge English Studies in Language Testing series offer detailed analyses of the skills constructs as defined in Cambridge English exams over a range of CEFR levels, for Listening (Geranpayeh and Taylor (Eds) 2013), Speaking (Taylor (Ed) 2011), Reading (Khalifa and Weir 2009), and Writing (Shaw and Weir 2007). Weir (2005a) is one of several authors who criticise the CEFR for having too little to say about underlying cognitive processes, given that the various descriptor scales intentionally focus on observable outcomes, and tend to emphasise the social dimensions of language use. Organised around Weir’s validity model, the above volumes set out to supply the useful level of detail that the CEFR itself does not.

As the CEFR excerpt above shows, at the very centre of language learning and use is *language activity*, through which a learner deals with *tasks* thrown up by daily life. This is a model of learning, and not merely use, because self-monitoring provides feedback that leads to reinforcement or modification of competences. The task is key: it is the reason for communicating, and communication is a language activity that leads to learning.

Focus on the task within the socio-cognitive model offers a straightforward conception of validity: while test tasks may not be authentic (answering multiple-choice questions is not a real-world activity), they may still demonstrate *interactional authenticity* to the extent that they engage the same cognitive processes as the real-world tasks for which they serve as proxies. It is a simple step to extend the notion of interactional authenticity to the classroom, thus putting tasks at the centre of our conception of LOA. The fundamental assessment or learning procedure thus centres on a task, which produces language activity, in conditions enabling observation, and learning. These conditions include an appropriate level of challenge, comprehensible input, and scaffolding which makes the task accessible. With these conditions satisfied learning mechanisms are enabled, above all concerning the communication of personally significant meanings. Feedback is generated, and performance is judged, through self-evaluation or evaluation by others.

The above paragraph shows assessment and learning as aspects of the same basic activity, which is adapted to particular purposes. Summative assessment, for example, requires more standardised control of contextual conditions, and formalised procedures for scoring performance. The feedback provided is more standardised and backward-looking. In the classroom, on the other hand, assessment focuses more on supporting learning (by identifying, for example, strengths and weaknesses) within particular learning contexts, and feedback is immediate, forward-looking, and more individualised.

Adaptivity and interaction

We have identified a quantitative dimension measuring vertical progression through levels, and a horizontal dimension distinguishing differences between individuals at a given level. To better illustrate the contrast we can compare the positions of two writers. Krashen (1982) takes a Chomskyan view of language acquisition as an innate capacity, proposing that learning happens by exposure to comprehensible input, which is at a level just beyond the learner’s current capacity to use (the $i+1$ level). He goes so far as to state that formal language teaching does not work. In contrast, Vygotsky (1986) views all cognition as socially constructed, so that learning happens through interaction with a more knowing other (e.g. a

teacher) and what a learner can achieve with assistance defines the *zone of proximal development*. Krashen's *i+1* level clearly relates to the vertical, proficiency dimension. Promoting learning is a matter of *adaptivity* (as in a computer-adaptive test): find the right level and everything else follows. For Vygotsky establishing the zone of proximal development is just the starting point. Learning springs from *interaction* within that zone.

Others concur. The 'interaction hypothesis' (Gass, Doughty and Long 2007, Long 1996) also sees the negotiation of meaning as the means by which learning takes place. The 'output hypothesis' (Swain 1985) argues that production and practice is necessary for the self-monitoring which enables the learner to test and modify hypotheses about the language. These positions are all consistent with the socio-cognitive model presented above, and stress the centrality for learning of purposeful language activity prompted by engagement with a task.

The validity of LOA

Turner's (2012) review of CBA identifies the issue of *alignment* between externally-measured criteria/standards and classroom-based assessment, quoting Pellegrino et al (2001: 9): 'A vision for the future is that assessments at all levels ... will work together in a system that is comprehensive, coherent, and continuous. In such a system, assessments would provide a variety of evidence to support educational decision making. Assessment at all levels would be linked back to the same underlying model of student learning and would provide indications of student growth over time.' It is clear that the LOA model presented here should deliver Pellegrino et al's vision, because all levels of assessment focus on the same criteria and construct. Validation of our LOA model would inhere in demonstrating:

- that the same socio-cognitive model of language use and learning is applied to external assessment tasks and classroom learning tasks;
- that the model successfully focuses efforts on the development of communicative language ability;
- that preparation for an appropriate external exam impacts positively on classroom practice rather than negatively;
- that the various forms of information generated by exams, mini-tests and classroom learning activities contribute useful feedback into further learning;
- that the expected positive impacts on learners are observed, e.g. development of learning-how-to-learn skills, autonomous learning, etc.;
- that teachers have a sufficient understanding of the learning process and expected outcome goals to give unplanned, spontaneous feedback which is consistent within the system as a whole.

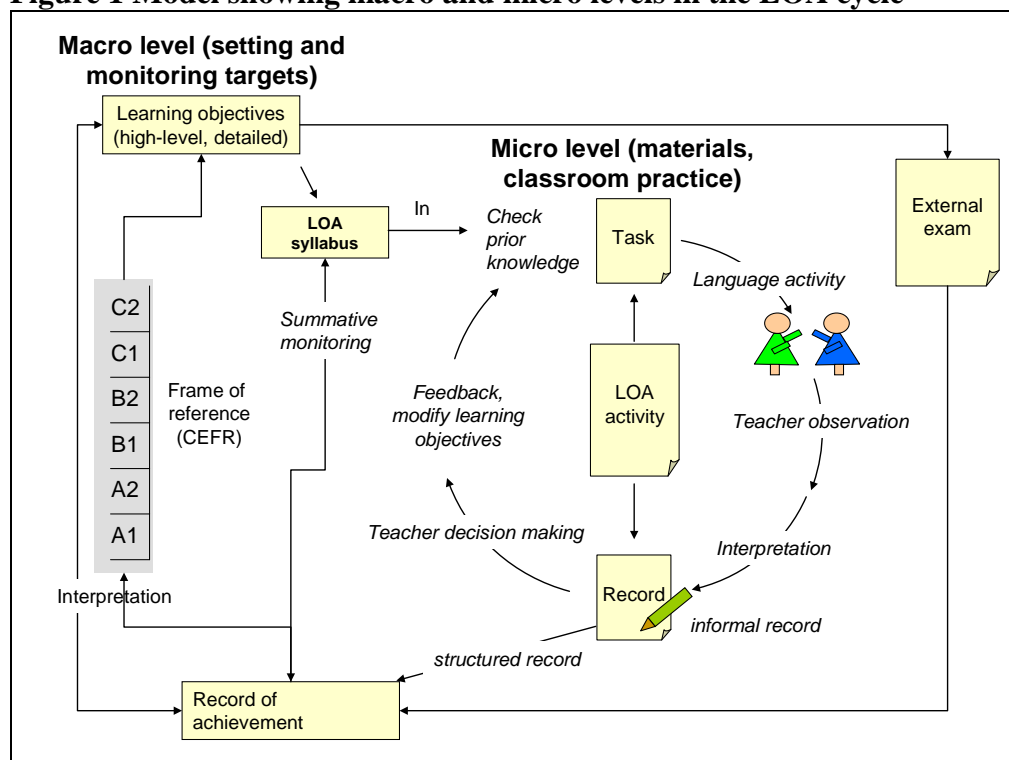
Implementing the LOA model

Moving from theoretical models to implementation we begin by identifying two levels of activity as shown in Figure 1 below. At a macro level, learning objectives are defined, both high-level and more detailed, and systems put in place to monitor progress towards their achievement. Objectives are defined within a communicative frame of reference (the CEFR here). This is simultaneously the basis for interpreting performance. From the objectives a LOA syllabus is constructed. An external exam may figure as a final or intermediate summative measure.

The micro level concerns the classroom. Figure 1 shows a cycle of stages centred on a specific LOA activity. After preparation, a task is set which generates language activity. This is observed and interpreted by the teacher, perhaps making some informal record. The teacher

decides what to do next. Feedback is given, and objectives possibly modified - repeating, extending or additionally scaffolding the activity.

Figure 1 Model showing macro and micro levels in the LOA cycle



A structured record is also captured to maintain a record of achievement for each student. This can be interpreted against the CEFR frame of reference, and contributes to periodic summative monitoring of progress against high-level objectives. External exam results also contribute to the record.

Figure 1 is a sketch to which much detail must be added: what is the exact nature of the activities that would count as LOA. Certainly, they must engage learners in purposeful language activity, have explicit learning objectives, and produce information which is relevant to tracing progress towards these learning objectives. The emphasis on record keeping, informal or formal, relates to the importance of information. Information feeds into the construction of *knowledge*, which emphasises action: knowledge is ‘not so much as a representation of the world but a way of doing things in it’ (Sayer 1992:48). The record serves several purposes, some merely administrative, but others integral to further action, and thus to learning (as implied in the very notion of feedback). The LOA model combines information from a range of activities: external assessments, internal monitoring tests, the LOA activities undertaken in the classroom. Actions taken on the basis of this information may have both summative and formative purpose.

Information is thus important, and we need practical ways of generating it. It cannot be an administrative burden imposed on the teacher. One part of a solution is to use information technology. Any activities mediated by IT systems can generate data, from which information can be extracted. This is true not only of tests but of a potentially unlimited range of collaborative or individual learning activities. With or without IT support, classroom tasks themselves might be designed to produce a record as a by-product of completing the task. Creative thinking is needed in this area. Above all, information must be co-constructed with students, who are its greatest beneficiaries. Turning information into action is critical.

Conclusion

We promised at the outset to address Turner's six questions.

Regarding quality criteria for LOA: the socio-cultural framework implies a range of high level outcomes that constitute significantly better learning. This is how its validity will be demonstrated. We also argued that in an LOA setting, large-scale assessments and learning programmes (and the way they are implemented in classroom) should be explicitly linked. This will provide added validation: do large-scale assessments validly reflect the learning programme objectives and, conversely, does the programme validly prepare the learners for the assessment?

Regarding the characteristics of assessment methods that favour learning: *interactional authenticity* is key. Tasks are contexts for purposeful interaction, not exercises to be completed.

Regarding teacher/student feedback and reflection: to be effective this must be based on serious interaction within a group. A socio-cultural constructivist approach favours this.

Regarding the role(s) of "assessor" and the accompanying responsibilities: peer assessment and self-assessment are skills that will enable learners to take responsibility for their learning. Teachers must learn to share the role of assessor, and must focus equally on immediate goals and high-level outcomes.

Regarding teachers mediating between classroom and external assessment: a focus on tasks and interactional authenticity potentially provides the link between the four worlds of the learner, the school, society and assessment. But this absolutely requires the adoption of a shared, construct-based conception of objectives (Daugherty, Black, Ecclestone, James and Newton 2008).

Regarding teacher training for LOA: We must go beyond method. Teachers should implement LOA from the inside, that is, be able to understand and work from principles (James and Pedder 2006).

A recurring theme in the literature on Learning Oriented Assessment is that of difficulty of implementation, given the new and different demands placed on all participants in the learning process: learners, teachers, assessment bodies and educational institutions. We find ourselves insisting on the importance of a few fundamental principles, and even if we acknowledge the practical issues in implementing them, we believe that current recognition of the need to improve educational practice and outcomes, supported by the properly-directed expertise of assessment bodies, can still be a powerful force for change.

References

- Bennett, R E (2011) Formative assessment: A critical review, *Assessment in Education: Principles, Policy and Practice* 18 (1), 5-25.
- Coleman, J A (2004) Modern languages in British universities: past and present, *Arts and Humanities in Higher Education* 3, 295-317.
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.
- Daugherty, R, Black, P, Ecclestone, K, James, M and Newton, P (2008) Alternative perspectives on learning outcomes: Challenges for assessment, *The Curriculum Journal* 19 (4), 243-254.
- Davison and Leung (2009) Current issues in English Language Teacher-based assessment, *TESOL Quarterly* 43 (3), 393-415.
- Frederiksen, N, Mislevy, R J and Bejar, I (Eds) (1993) *Test Theory for a New Generation of Tests*, Hillsdale: Erlbaum.
- Gass, Susan M, Doughty, C J and Long, M H (2007) *Input and Interaction: The Handbook of Second Language Acquisition*, Malden: Blackwell Publishing.

- Geranpayeh, A and Taylor, L (Eds) (2013) *Examining Listening: Research and Practice in Assessing Second Language Listening*, Studies in Language Testing volume 35, Cambridge: UCLES/Cambridge University Press.
- James, M, and Brown, S (2005) Grasping the nettle: Preliminary analysis and some enduring surrounding the improvement of learning outcomes, *The Curriculum Journal* 16 (1), 7–30.
- James, M, and Pedder, D (2006). Beyond method: assessment and learning practices and values, *The Curriculum Journal* 17(2), 109-138.
- Jones, N (2012) Reliability and dependability, in Fulcher, D and Davidson, F (Eds) *The Routledge Handbook of Language Testing*, London: Routledge, 350–262.
- Khalifa, H and Weir, C (2009) *Examining reading: Research and Practice in Assessing Second Language Reading*, Studies in Language Testing volume 29, Cambridge: UCLES/Cambridge University Press.
- Krashen, S (1982) *Principles and Practice in Second Language Acquisition*, Oxford: Pergamon.
- Long, M (1996) The role of the linguistic environment in second language acquisition, in W. Ritchie, W and Bhatia, T (Eds), *Handbook of Second Language Acquisition*, San Diego: Academic Press, 413-68.
- Pellegrino, J W, Chudowsky, N and Glaser, R (2001) *Knowing What Students Know: The Science and Design of Educational Assessment*, Washington, DC: National Academy Press.
- Salamoura, A, Khalifa, H and Docherty, C (2014) *Investigating the impact of language tests in their educational context*, in Proceedings of the 40th Annual IAEA Conference, Singapore, 2014.
- Saville, N (2009) *Developing a model for investigating the impact of language assessment within educational contexts by a public examination provider*, unpublished PhD, University of Bedfordshire.
- Saville, N (2012) Applying a model for investigating the impact of language assessment within educational contexts: The Cambridge ESOL approach, *Research Notes* 50, 4-8.
- Sayer, A (1992) *Method in Social Science: a Realist Approach*, London and New York: Routledge.
- Shavelson, R J (2008) Guest editor's introduction. *Applied Measurement in Education* 21 (4), 293-294.
- Shavelson, R (2009) *Reflections on learning progressions*, paper presented at the Learning Progressions in Science (LeaPS) Conference, June 2009.
- Shaw, S and Weir, C J (2007) *Examining Second Language Writing: Research and Practice in Assessing Second Language Writing*, Studies in Language Testing volume 26, Cambridge: UCLES/Cambridge University Press.
- Shephard, L A (2000) The role of assessment in a learning culture, *Educational Researcher* 29 (7), 4-14.
- Sjøberg, S (2007) Constructivism and learning, in Baker, E, McGaw, B and Peterson P (Eds) (2007) *International Encyclopaedia of Education* (3rd Edition), Oxford: Elsevier.
- Swain, M (1985) Communicative competence: Some roles of comprehensible input and comprehensible output in its development, in Gass, S and Madden, C (Eds), *Input in Second language Acquisition*, Rowley, Mass.: Newbury House, 232-253.
- Tattersall, K (2004) *Teacher assessment: Realistic objective or impossible dream?* paper presented at AEA Europe Conference, Budapest, November 2004.
- Taylor, L (Ed) (2011) *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing volume 30, Cambridge: UCLES/Cambridge University Press.
- Teasdale, A and Leung, C (2000) Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing* 17 (2), 163-184.
- Turner, C E (2012) Classroom assessment, in Fulcher, G and Davidson, F (Eds) *The Routledge Handbook of Language Testing*, Routledge: Oxford, 65-78.
- Vygotsky, L (1986) *Thought and Language*, Cambridge: MIT Press.
- Weir, C J (2005a) Limitations of the Council of Europe's Framework of reference (CEFR) in developing comparable examinations and tests, *Language Testing* 22 (3), 281–300.
- Weir, C J (2005b) *Language Testing and Validation: An Evidence-Based Approach*, Basingstoke: Palgrave MacMillan.