

## **Investigating the impact of language tests in their educational context**

Angeliki Salamoura, Hanan Khalifa and Coreen Docherty  
Cambridge English Language Assessment

### **Abstract**

Introducing language assessments as part of educational reform at national or institutional level is increasingly frequent with two predominant aims: increasing language proficiency and reaching a wider range of learners who can benefit from the reform. However, 'good examinations are not guaranteed to produce positive results and bad examinations do not necessarily produce bad ones' (Wall 2005:xiv). How can education providers then assess effectively whether their language education reforms work as intended?

The paper outlines the 'impact by design' approach (Saville 2009, 2012) to researching language test impact and discusses how it can help examination and education providers foster beneficial effects from their tests and avoid or mitigate harmful consequences. It is based on four maxims that collectively allow for the monitoring of positive and negative impacts in new educational assessment projects.

The approach will be illustrated with two case studies carried out in state and state-subsidised schools in Asia (Vietnam) and Europe (Spain) which implement bilingual programmes in conjunction with international external examinations. A mixed methods research design was used to collect quantitative and qualitative data from learners, teachers, parents and decision-makers. The findings will be discussed with reference to the main features of impact by design and the initiative objectives.

### **Introduction**

Goodhart's Law (1975) states that 'any purposeful action will produce some unintended consequences' as well as intended consequences. Assessments, in particular, form part of complex systems and have expected and unexpected effects, including exerting powerful influences on the individual and on educational systems and society. This paper introduces the concept of 'impact by design' (Saville 2009, 2012), the Cambridge English approach to researching language test impact, and illustrates its application with examples of projects which Cambridge English has carried out in state or state-subsidised schools in two different contexts.

### **Four maxims for investigating test impact**

Impact research in language testing investigates and seeks to understand the effects and consequences which result from the use of tests and examinations in educational contexts and throughout society. Bachman was one of the first researchers to present impact as a 'quality' of a test which should be integrated within the overarching concept of test usefulness (e.g. Bachman and Palmer 1996). Following his lead, Cambridge English Language Assessment (henceforward referred to as Cambridge English) introduced impact as one of its four essential qualities for test development and validation. The four qualities are: validity, reliability, impact and practicality (VRIP).

In 1996, Milanovic and Saville proposed a model of test impact which was explicitly designed to meet the needs of Cambridge English as an examination board. The model comprised four maxims:

Maxim 1 PLAN - Use a rational and explicit approach to test development

Maxim 2 SUPPORT - Support stakeholders in the testing process

Maxim 3 COMMUNICATE - Provide comprehensive, useful and transparent information

Maxim 4 MONITOR and EVALUATE - Collect all relevant data and analyse as required

These maxims were designed to capture key principles and to provide a basis for practical decision-making and action planning – and they still remain central to the examination board approach today.

Under Maxim 1, Cambridge English developed systems and processes and use a rational and explicit model for managing the test development processes in a cyclical and iterative way. The cyclical approach requires regular reviews and revisions to take place and for improvements to be made when necessary (Cambridge English 2013:31-32, Saville 2003:57–120).

Maxim 2 focuses supporting all the stakeholders involved in the learning, teaching and assessment processes associated with internationally standardised examinations. This is an important aspect of the approach because examination systems only function effectively if all stakeholders collaborate to achieve the intended outcomes.

Maxim 3 focuses on the importance of developing appropriate communication systems and providing essential information to the stakeholders (Cambridge English 2013:31-32).

Maxim 4 focuses on the essential research requirement to collect as much relevant data as possible and to carry out routine analyses as part of the iterative model (noted under Maxim 1). The nature of the data needed to investigate impact effectively and how it can be collected, analysed and interpreted under operational conditions has become an increasingly important part of the Cambridge English ‘impact by design’ model.

Following empirical research, Saville (2009) proposed a meta-framework designed to provide a more effective model for conducting impact research under operational conditions. By implementing this framework systematically, ‘anticipated impacts’ can be achieved more effectively and well-motivated improvements to the examination systems can be identified and put into place. Prominent aspects of this model are discussed in the next section under the concept of impact by design and are exemplified in the case studies reported in the second part of this paper. In essence, the empirical research and case studies start from and build on Maxim 4: Monitor and Evaluate, in order to feed back and inform Maxims 1-3: Plan, Support, Communicate.

### **The ‘impact by design’ approach**

Impact by design is a key feature of the expanded impact model. It starts from the premise that assessment systems should be designed from the outset with the potential to achieve positive impacts and takes an *ex ante* approach to anticipating the possible consequences of using the test in particular contexts (Saville 2009, 2010). The most important features of this model are, in turn, discussed below.

#### ***Test construct***

Impact by design builds on Messick’s (1996) idea of achieving ‘validity by design as a basis for washback’. The existence of a rational model of test development and validation is a necessary condition for creating construct valid tests and for the development of successful systems to support them (cf. Maxim 1). Adequate specification and communication of the focal constructs is crucial for ensuring that the test is appropriate for its purpose and contexts of use and to counter threats to validity: construct underrepresentation and construct irrelevant variance (Messick 1996:252).

***Test delivery systems***

While appropriate construct representation is a necessary condition for achieving the anticipated outcomes, it is not sufficient and impact by design highlights the importance of designing and implementing assessment systems which explicitly incorporate considerations related to the social and educational contexts of learning, teaching and test use. This relates to the need for effective communication and collaboration with stakeholders, as noted in the original Maxims 2 and 3 and incorporated into Cambridge English's *Principles of Good Practice*, Section 2 (Cambridge English 2013).

***Context***

Context is, arguably, the most important consideration when investigating educational processes. Understanding the nature of context within educational systems and the roles of stakeholders in those contexts has always been a priority for Cambridge English – see Taylor (2000) for a discussion of stakeholders. It is now widely recognised that educational processes take place within complex systems with dynamical interplay between many sub-systems and ‘cultures’ and so an understanding of the roles of stakeholders as participants is a critical factor in bringing about intended changes (e.g. Fullan 1993, 1999, Thelen and Smith 1994, Van Geert 2007). In conducting impact research the aim is to understand better the interplay between the macro and micro contexts within the society where the tests are being used and to determine which elements facilitate or hinder the desired outcomes. In general, diversity and variation increases as one moves from the general milieu within a country or region (the macro context) to specific schools and ultimately to the individual participants within classrooms (the multiple micro contexts at the local level involving schools, classes/groups and individual teachers and learners).

Saville (2009), therefore, claims that it is important to develop methods to understand both the general context as well as specific local cases, including dynamics which affect learning in classrooms. This leads to the need to use both quantitative and qualitative data collection methods (see Research Methods below).

***Timeline***

It is essential to know what happens when a test is introduced into its intended contexts of use and this should constitute a long-term validation plan (cf. Maxims 1 and 4). Anticipating and managing change over time – the intended and unintended impacts – within specific contexts is therefore central to this concept and it means that appropriate consideration of timescales and the timeline for implementation (often involving several phases) are central to the design of impact studies. In impact research designs there is nearly always a fundamental need to collect comparative data, and therefore to develop research designs which can be carried out in several phases over an extended period of time or replicated in several different contexts.

Impact by design is therefore not strictly about prediction; a more appropriate term might be ‘anticipation’. In working with stakeholders, possible impacts on both micro and macro levels can be anticipated as part of the design and development process, and where potentially negative consequences are anticipated, remedial actions or mitigations can be planned well in advance.

***Research methods***

Contemporary theories of knowledge and learning have played a prominent role in developing Cambridge English's model of impact, including ‘critical realism’ (e.g. Sayer 1984, 2000), contemporary views on pragmatism and constructivism. Impact by design is best situated in ‘real world’ research paradigms (Robson 2002) which are most appropriate to finding out ‘what goes on’ in contexts of test use. It also advocates the investigation of

impact from a constructivist perspective, drawing on participants' understanding and their subjective world view. The research is, therefore, shaped bottom up – from individual perspectives to broader, more general understandings. The approach is also participatory in nature because it seeks collaboration with and involvement from various stakeholders.

The current model of impact looks to 'real world' research paradigms to provide tools which can shed light on what happens in testing contexts, including mixed methods and quasi-experimental designs. Case studies, like the two examples reported in the next section, are especially useful for investigating impact at the micro level and for understanding the complexities of interaction between macro level policies and implementation in local settings. Without such methods it is difficult to find out about and understand how the interaction of differing beliefs and attitudes can lead to consensus or to divergence and diversity.

For these reasons, mixed method research designs are particularly relevant to addressing impact research questions (Creswell and Plano Clark 2011:69). Cambridge English has, therefore, created an 'impact toolkit' of methods and approaches for carrying out analyses of both large-scale aggregated data, as well as micro analyses of views, attitudes and behaviours in local settings.

Quantitative analysis of macro level group data allows us to capture overall patterns and trends, while the qualitative analysis of multiple single cases enables researchers to monitor variability in local settings and to work with the 'ecological' features of context. It is the integration of both analyses to provide an in-depth understanding of the data and insightful as well as useful interpretations which is particularly important.

### ***Participants and roles***

Finally it is important to highlight the make-up of the impact research teams. Collaboration between an international examination provider and local researchers is essential in order to capture relevant data, to bring a deeper and a more accurate understanding of the educational context which is under investigation and to shed light on both the macro and micro contextual parameters (see Context above). Many dilemmas which arise in assessment contexts can only be dealt with if a wide range of local stakeholders agree to manage them in ways which they jointly find acceptable; the challenge is to get the relevant stakeholders working together effectively to agree what needs to be done to achieve the intended outcomes.

### **Two case studies**

Having presented the impact by design approach, we will now discuss two case studies as examples of using the approach in practice. The first study was conducted in Vietnam (Khalifa, Nguyen and Walter 2012) whereas the second in Spain (Ashton, Salamoura and Diaz 2012, Salamoura, Docherty and Hamilton 2013). Both are situated in the primary or primary and early secondary education context and investigated the introduction of external assessments as part of the local educational programmes introduced in state or state-subsidised schools with the aim of raising the standards of English language knowledge. We will start by outlining the common research questions, methods and instruments shared by the studies before exploring the macro and micro contexts in which these studies are situated as well as the individual differences that, by necessity, emerge from these divergent contexts.

### ***Shared features***

*Research Questions:* The overarching research questions in both studies were the following:

1. What is the intended/unintended impact of the educational initiative?
2. What is the impact of the test/new initiative on key stakeholders, namely, teachers and learners?

The first question provides contextual information, forms a key part of the situational analysis and shapes up subsequent questions to ask and data to gather. The second question investigates the impact/effects on learning and teaching. Common constructs explored under these two questions include: attitudes towards teaching, learning and assessment; learner and teacher motivation; learner progression; teaching practice/development; parental involvement.

*Methods and Instruments:* Due to the complexity of the phenomenon of educational impact, both studies used multiple data sources in a convergent parallel mixed methods research design (Creswell and Plano Clark 2011). Quantitative and qualitative data were collected simultaneously and analysed separately but the interpretation of results drew on both strands of evidence. The triangulation of multiple sources of information built into this design resulted in a richer picture and enhanced confidence in the findings (Greene, Caracelli and Graham 1989).

Instruments used in both studies were selected from the Cambridge English 'impact toolkit' (see Research Methods above) and adapted for the specific context where necessary using expert judgement reviews prior to implementation. Quantitative instruments included surveys and test score data; qualitative instruments comprised interviews and, for young learners in particular, focus groups and (in the Spanish case study) a draw-a-picture task, a new technique which is more suited for eliciting views from young learners by asking them to draw a picture in response to a question (Harris, Harnett and Brown 2009). Finally, both studies were collaborations between Cambridge English and local researchers with an in depth knowledge of the local context.

## **Case study 1: Vietnam**

### *Context*

The context of the first study is Vietnam where the Ministry of Education and Training (MOET) has a long-standing strategic objective to raise English language learning standards by 2020 so that students are better prepared for the workplace, for studying abroad and for becoming global citizens. To meet this objective, in 1998-1999, the Department of Education and Training (DOET) in Ho Chi Minh City (HCM) initiated the provision of an Intensive English Programme (IEP) for primary students in state-funded schools whereby students are given additional English lessons. State schools typically operate on half-day basis due to high demand, insufficient classroom capacity and high class density. English is taught in two lessons per week. However, IEP schools offer a whole-day programme to students which allows for an additional eight English lessons per week.

HCM DOET's drive for additional classes was also motivated out of a sense of social responsibility. Until the introduction of IEP, only students from financially-able families had the opportunity of increasing their English proficiency through attending private language institutes. With the IEP initiative, it is hoped that students from all financial strata have the chance to increase their English proficiency at affordable fees without the need to go to private institutes or tutors.

This initiative started with one school in 1998-1999 but by 2011-2012, a total of 194 schools out of 495 state-funded primary schools had joined IEP. In 2010-2011, as a result of the high demand on IEP, HCM DOET introduced mandatory external assessment to the programme to provide a fair and reliable measure for student selection and for accountability and quality assurance purposes. *Cambridge English: Young Learners* was selected not only because of its international recognition and use in similar contexts but more importantly because it introduces children to everyday written and spoken English in a fun and motivating

way. According to a key decision maker in HCM DOET Mr Nguyen Hoai Chuong, DOET Vice Director, states that *Cambridge English: Young Learners* 'is child friendly, takes into account child psychology, is very motivational and covers all skills...if the exam is child friendly and encourages learning in a fun way, then the teaching will change accordingly, so it is a win-win situation' (cited in Khalifa et al 2012:9).

So, the first case study sought to understand the intended and unintended effects of HCM DOET's strategic decision to increase English Language provision through the IEP programme and to ensure the quality of the provision through the use of external assessment.

## Key findings

### *Positive effects*

The study revealed areas where positive effects have been achieved, involving three key groups of stakeholders (learners, teachers and parents):

- Both the qualitative and quantitative findings showed an increased focus on speaking, a direct positive effect of the introduction of *Cambridge English: Starters* (also known as *Young Learners English: Starters*), the design of which is based on a communicative approach to language learning. What is more important is that this focus did not detract from attention being paid to the other skills as evidenced by test score data. Although there is a notable language progression in terms of speaking, learners also performed well on the other skills.
- A positive change in terms of teaching practice was also noted with the adoption of certain Assessment for Learning principles (Assessment Reform Group 2002), the introduction of collaborative teaching, and the utilisation of best practices such as teacher reflection or adaptation of teaching methods to support students' learning goals and styles.
- The intervention led to increased parental involvement in their child's learning in terms of encouraging them to learn English, taking them to extra English classes as provided by IEP and providing incentives for better performance as seen from the comments made in the focus group discussions with the young learners ('I want to take the test so I can go to a Grade 3 IEP class, mom told me that'). Parental involvement and teacher encouragement were a recurring theme in the findings of this study as playing a key role in learner motivation.

### *Unintended effects*

As with most interventions, DOET's initiative led to unintended consequences too:

- The study illustrates that when decentralisation of decision making is well executed, innovative approaches that suit the local context can lead to positive effects. Although the strategic objective for improving language standards came from MOET, it was up to HCM DOET to decide on how to achieve this and it is also up to schools to decide whether to be involved in the initiative or not, which provided a sense of ownership and faith in the intervention.
- A change in decision making also took place based on discussions with policy makers during the course of this study. The initial decision to use a certain number of shields obtained in *Cambridge English: Starters* as an entry criterion to IEP was revised to create a fairer selection process.
- Another unintended effect is better utilisation of children's free time. After a half-day of schooling, children are engaged in IEP and *Cambridge English: Young Learners* study. This alleviated parents' anxiety as to how to engage their children's free time once the half day of schooling is over.

- The introduction of an internationally recognised external assessment led to a heightened sense of status among teachers. They view themselves as ‘international teachers’ because they are teaching towards international standards.
- It was mentioned earlier that with the introduction of the IEP initiative, students from financially disadvantaged backgrounds would have the opportunity to increase their English proficiency. However, 67% of the parents participating in this study have self-assessed themselves as belonging to the middle socio-economic stratum. It may be a sampling issue, but HCM DOET may want to think of how to engage more parents from the low/low-mid strata so that their children can benefit from IEP.

### *Lessons learned*

Finally, the study also revealed areas where improvements may be necessary such as better planning to disseminate information on the intervention to all stakeholders; ensuring information is provided consistently and adequately from schools to parents; further investigation as to why some principals felt that learners are not motivated to learn English; and further in-depth investigation as to why the majority of the sampled parents continue to send their children to private language institutes despite the introduction of IEP.

## **Case study 2: Spain**

### *Context*

The second study is situated in Madrid, Spain. In 2008 the Spanish government set the target that all students in Spain will speak English within 10 years (Blanco and Nicholson 2010). The promotion of English was prompted by the desire to be more economically competitive, as well as the need to work towards the 2002 Barcelona European Council Conclusions which called for European Union (EU) member states to teach ‘at least two foreign languages from a very early age’ (European Commission 2005). In addition to a central government, Spain has 17 autonomous regional communities which have devolved political powers including education. The Autonomous Community of Madrid in working towards the directives described above has, therefore, set as a future goal for the majority (80%) of secondary school learners to leave school having achieved Level B2 of the Common European Framework of Reference for languages (CEFR; Ashton et al 2012).

FERE is an association of approximately 340 primary and secondary schools in Madrid. The FERE schools are private but state funded. They follow the rules of the Autonomous Community of Madrid but are self-managing. The FERE schools must meet the minimum regulations set by the Autonomous Community of Madrid but they have complete freedom regarding the organisation of staff and teaching methodology used. It is common for the FERE schools to aim higher and to extend the minimum requirements set by the Autonomous community of Madrid.

Against the political and educational background described above and having recognised that a proficient knowledge of English as a foreign language is a necessary tool for students to reach effective and full integration within Europe, FERE set up the Bilingual English Development and Assessment (BEDA) programme in 2008. The purpose of BEDA is to implement bilingual education (the promotion of English alongside Spanish) and improve efficiency in language education within the FERE schools. The programme aims at the gradual introduction of bilingual education with growing use of English within the school, as well as in the curriculum and the school environment.

Cambridge English examinations form a key component of the BEDA programme, which also includes a teacher training programme and the promotion of a bilingual

atmosphere. As part of this bilingual atmosphere, native speaker English language assistants are provided by FERE to work in each BEDA school and schools are expected to have initiatives such as school exchanges to support the learning of English, an ‘English corner’ in the school and an English language section in the school library. At the time this study was conducted, there were over 200 FERE schools registered to take Cambridge English exams within the BEDA programme and this number is growing. *Cambridge English: Young Learners*, *Cambridge English: Key* (also known Key English Test) and *Cambridge English: Preliminary* (also known as Preliminary English Test) are the most commonly taken exams by students in BEDA. As parents pay for the exams, they have the final decision as to whether and when their child will take a particular exam, although teachers provide the recommendations and there is very close and collaborative discussion between students, teachers and parents.

The aims of the second case study was to investigate the impact of Cambridge English assessments as part of FERE’s BEDA programme as well as the overall impact of the BEDA programme in the FERE schools.

## Key findings

### *Positive effects*

- Overall, there were positive perceptions of the exams and the programme by all stakeholders.
- Learners showed high motivation for learning English and a perception of the link between assessment and learning. For example, the majority of learners reported that their test scores help them understand their English level.
- At the same time, the majority of teachers agreed that learner proficiency had improved as a result of the school joining the BEDA programme and using Cambridge English examinations, particularly in speaking.
- Teachers and parents particularly appreciated the external reference that the Cambridge English examinations provide and the fact that they are internationally recognised. As one of the BEDA co-ordinators remarked: ‘Parents are quite fond of these exams because they are from the University of Cambridge and it is not the teacher that assess [sic] them.’
- Changes in teacher practice were also noted. Training within the BEDA programme provided essential support for schools - particularly schools new to the programme. The majority of teachers agreed that the BEDA training days help them improve as a teacher and that the Cambridge English examinations have made them more motivated to teach English. The data also highlighted how teachers are using the *Cambridge English: Young Learners* test results formatively to structure their teaching practice according to their learners’ strengths and weaknesses.

### *Challenges*

Some of future challenges identified for the programme include:

- Ensuring parents and learners are given adequate info about the purpose of the exams within the BEDA programme.
- Implementing a consistent approach in integrating exam preparation & scheduling across BEDA schools. Some stakeholders felt that it is sometimes difficult to integrate the Cambridge exams into the lesson, especially when the students have different levels.
- Reviewing the teacher training courses on offer to ensure they are meeting the needs of teachers, including ‘more teaching methodology and practical things like lesson planning and running exercises in class.’



### **Impact by design in practice**

The case studies provide an exemplification of some of the key features of the impact by design approach. Both initiatives adopted external standardised exams which are based on a comprehensive test development and validation model, while the test constructs are extensively documented and publicised (Cambridge English 2013, CambridgeEnglish.org). Covering all four skills, the tests and delivery systems were set up with a positive washback in mind. In both studies, for instance, the face-to-face assessment of speaking, which is an integral part of the exam, led to a direct positive effect of increased focus on speaking in the classroom (*Test constructs* and *Test delivery systems*). The interplay between macro and micro contextual factors in bringing about desired changes was also attested, as in the example of MOET's strategic objective at a national level and DOET's implementation of it at a local context in the Vietnam study (*Context*). Impact by design is also about 'anticipating' intended and unintended impacts and planning remedial actions or ameliorations where potential negative consequences are anticipated (*Timeline*). Both studies charted a number of unintended effects (along with intended ones), such as, for example, the potential underrepresentation of learners from low or low-middle socio-economic strata in the IEP programme in Vietnam, or the restrictive entry criteria to the IEP programme and the subsequent mitigating decision to revise these criteria.

The use of mixed methods designs allowed the capture of overall patterns and trends (such as the positive perceptions of exams and the BEDA programme) through quantitative analyses of group data, as well as the identification of specific features of context that may explain some of these trends (such as the appreciation of the external reference that the internationally recognised Cambridge English exams provide) through qualitative analyses of views of smaller groups or individuals (*Research methods*). Finally, it was the make-up of the project teams comprising both Cambridge English staff with expertise in research design and analysis, as well as local researchers who were 'participants' in the teaching/learning context itself that enabled a thorough understanding of the educational context and highlighted issues that may have passed unnoticed (e.g. practices and procedures that may not have been well documented) (*Participants and roles*).

### **Closing remarks**

Change is perhaps the only constant within complex, multi-dimensional educational processes, including assessment systems. Anticipating impacts and finding out what happens in practice is critical but not enough if improvements do not occur as a result. Being prepared to manage change is therefore essential. The impact by design approach proposes a theory of action for managing change by integrating impact investigation into test development, validation and operationalisation within wider educational initiatives. The two case studies illustrate that when working closely with the stakeholders in their own contexts, this approach provides the necessary action plans and tools to determine where the initiative currently stands, what needs to be done and when/how to do it in order to achieve the desired outcomes. It is anticipated that building a body of empirical research and case studies with a range of stakeholders and from a variety of contexts will contribute significantly to refining the features and fine tuning the tools of the current model.

### **References**

- Ashton, K, Salamoura, A and Diaz, E (2012) The BEDA impact project: A preliminary investigation of a bilingual programme in Spain, *Research Notes* 50, 34-41.
- Assessment Reform Group (2002) *Assessment for Learning: 10 Principles*, available online: [www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf](http://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf)

- Bachman, L F and Palmer, A S (1996) *Language Testing in Practice*, Oxford: Oxford University Press.
- Banerjee, J (2012) Guest editorial, *Research Notes* 50, 2-3.
- Blanco, J and Nicholson, D (2010) Cambridge ESOL and Spanish school networks, *Research Notes* 40, 9-13.
- Cambridge English (2013) *Principles of Good Practice: Quality Management and Validation in Language Assessment*, Cambridge: Cambridge English Language Assessment, available online: [www.cambridgeenglish.org/gr/images/22695-principles-of-good-practice.pdf](http://www.cambridgeenglish.org/gr/images/22695-principles-of-good-practice.pdf)
- Creswell, J W and Plano Clark, V L (2011) *Designing and Conducting Mixed Methods Research* (2nd edition), Thousand Oaks: Sage.
- European Commission (2005) *Commission Communication of 1 August 2005 - The European Indicator of Language Competence, COM (2005) 356 final - Not published in the Official Journal*, available online: [europaeu/legislation\\_summaries/education\\_training\\_youth/lifelong\\_learning/c11083\\_en.htm](http://europaeu/legislation_summaries/education_training_youth/lifelong_learning/c11083_en.htm)
- Fullan, M (1993) *Change Forces: Probing the Depths of Educational Reform*, London: The Falmer Press.
- Fullan, M (1999) *Change Forces: The Sequel*, London: The Falmer Press.
- Goodhart, C A E (1975) Problems of monetary management: The UK experience, in *Papers in Monetary Economics volume I*, Reserve Bank of Australia.
- Greene, J C, Caracelli, V J and Graham, W F (1989) Toward a conceptual framework for mixed-method evaluation designs, *Educational Evaluation and Policy Analysis* 11 (3), 255-274.
- Harris L R, Harnett J A and Brown G T L (2009) 'Drawing' out student conceptions: Using pupils' pictures to examine their conceptions of assessment. In: McInerney D M, Brown G T L, Liem G A D (Eds) *Student Perspectives on Assessment: What Students Can Tell Us about Assessment for Learning*, Charlotte: Information Age Publishing, 53-83.
- Khalifa, H, Nguyen, T and Walker, C (2012) An investigation into the effect of intensive language provision and external assessment in primary education in Ho Chi Minh City, Vietnam, *Research Notes* 50, 8-19.
- Messick, S (1996) Validity and washback in language testing, *Language Testing* 13 (3), 241-256.
- Milanovic, M and Saville, N (1996) *Considering the impact of Cambridge EFL examinations*, Cambridge: UCLES, internal report.
- Robson, C (2002) *Real World Research: A Resource for Social Scientists and Practitioner-researchers*, London: Blackwell.
- Salamoura, A, Docherty, C and Hamilton, M (2013) *The use of external standardised assessment in a school context for motivational and accountability purposes*, paper presented at the 35th Annual Language Testing Research Colloquium, July 1-5 2013, Seoul, Korea.
- Saville, N (2003) The process of test development and revision within UCLES EFL, in Weir, C J and Milanovic, M (Eds) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913-2002*, Studies in Language Testing volume 15, Cambridge: UCLES/Cambridge University Press, 57-120.
- Saville, N (2009) *Developing a model for investigating the impact of language assessment within educational contexts by a public examination provider*, unpublished thesis, University of Bedfordshire.
- Saville, N (2010) Developing a model for investigating the impact of language assessment, *Research Notes* 42, 2-8.
- 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 (1984) *Method in Social Science: A Realist Approach*, London: Hutchison.
- Sayer, A (2000) *Realism and Social Science*, London: SAGE Publications.
- Taylor, L (2000) Stakeholders in language testing, *Research Notes* 2, 2-3.
- Thelen, E and Smith, L B (1994) *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge, MA: The MIT Press.
- Van Geert, P (2007) Dynamic systems in second language learning: Some general methodological reflections, *Bilingualism: Language and Cognition* 10, 47-49.
- Wall, D (2005) *The Impact of High-stakes Examinations on Classroom Teaching*, Studies in Language Testing volume 22, Cambridge: UCLES/Cambridge University Press.