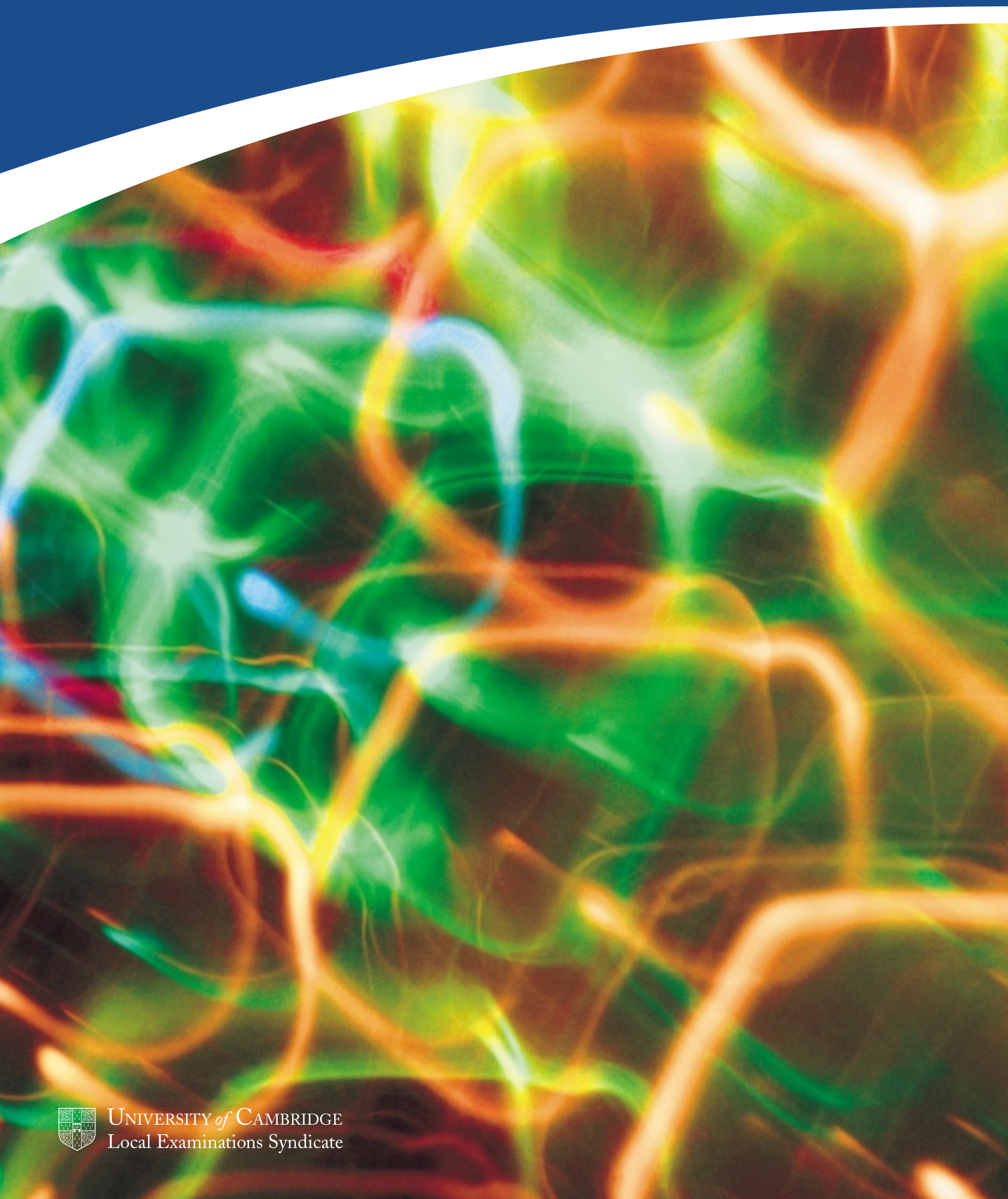


Issue 15 January 2013



CAMBRIDGE ASSESSMENT

Research Matters



UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate

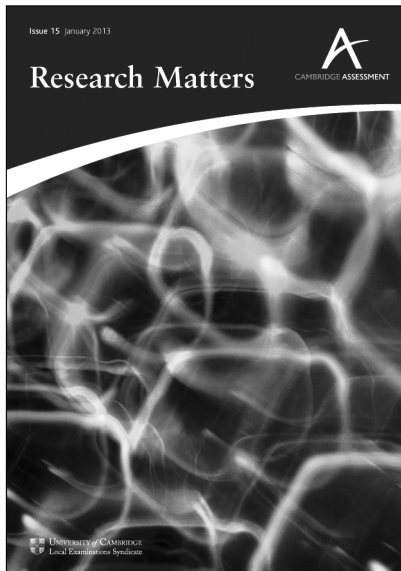


CAMBRIDGE ASSESSMENT

Cambridge Assessment is the brand name of the University of Cambridge Local Examinations Syndicate, a department of the University of Cambridge. Cambridge Assessment is a not-for-profit organisation.

Citation

Articles in this publication should be cited as:
Crisp, V. and Johnson, M. (2006). Examiners' annotations: Practice and purpose. *Research Matters: A Cambridge Assessment Publication*, 2, 10-12.



- 1 **Foreword** : Tim Oates
- 1 **Editorial** : Sylvia Green
- 2 **21st Century skills: Ancient, ubiquitous, enigmatic?** : Irenka Suto
- 9 **Independent research at A level: Students' and teachers' experiences** : Sanjana Mehta, Irenka Suto, Gill Elliott and Nicky Rushton
- 17 **Assessment for Learning in International Contexts (ALIC): understanding values and practices across diverse contexts** : Stuart Shaw, Martin Johnson and Paul Warwick
- 29 **Using scales of cognitive demand in a validation study of Cambridge International A and AS level Economics** : Jackie Greatorex, Stuart Shaw, Phineas Hodson and Jo Ireland
- 38 **Cambridge Assessment Qualitative Research Methods Reading Group** : Martin Johnson
- 38 **Statistical Reports** : The Research Division
- 39 **Research News**

If you would like to comment on any of the articles in this issue, please contact Sylvia Green.

Email: researchprogrammes@cambridgeassessment.org.uk

The full issue and previous issues are available on our website:
www.cambridgeassessment.org.uk/ca/Our_Services/Research

Research Matters : 15

A CAMBRIDGE ASSESSMENT PUBLICATION

Foreword

As this copy of *Research Matters* goes to press, with three articles focusing on A level, there is emerging a conundrum in national policy in respect of the policy on increasing the involvement of Higher Education in the design, administration and evaluation of A levels. The conundrum is this: the institutional response to the recent consultation on 'HE engagement' has been muted at best, and negative on some of the key aspects of the policy. By contrast, the response of individual academics has been overwhelmingly positive. Reflecting this contrast, the effort which Cambridge Assessment has put into commissioning HE subject forums has ensured the systematic involvement of very large numbers of academics, from across the entire HE community, not just 'elite' institutions. This endorses the approach of the Group to increasing HE engagement, which can be summarised as: 'new relations, not new institutions'. New institutions can escalate unwieldy and expensive bureaucracy, whereas carefully-constructed new relations can help positively and efficiently with re-alignment, re-design and updating of A levels. The Group's research programme is feeding systematically into these new relations, including the work contained in this volume: analytic work on '21st Century skills', examination of validity in advanced level qualifications, and elements of A level studies which constitute preparation for study at higher level.

Although contributions to enactment of evidence-driven national policy tend to assume centre stage, it is vital not to overlook Martin Johnson's brief, but highly important piece at the end of this volume. It is focused on method. Assessment is located in complex systems; for research evidence to be dependable, research design must be robust. Alongside the stream of focused research projects, the research teams in Cambridge Assessment engage in a parallel, linked programme of discourse and critique of method in educational and assessment research. I still experience a lift when I witness conference presentations based on excellent design; conversely, a deep sigh is still too often present listening to studies in important areas but with inadequate design. Design and method must remain a key area of conscious deliberation and continuous refinement if we are to convince practitioners and policy-makers of the veracity of the new knowledge we create.

Tim Oates *Group Director, Assessment Research and Development*

Editorial

This issue addresses a number of themes that are both topical and challenging. The first article from Irenka Suto opens with a quotation from Andreas Schleicher which illustrates the extent of the challenge facing us as we attempt to define 21st Century skills and to support their development in young people. As well as considering curricula and pedagogy, Suto addresses the difficulties faced by those who seek to assess such skills. Independent research and investigative skills are identified in the literature as skills that should be fostered in the 21st century and in their article Mehta, Suto, Elliott and Rushton explore the perspectives of teachers and students who engage in independent study at A level (Advanced Level General Certificate of Education). This research provides insights into the preparedness of students who progress to university where they are expected to engage in independent study and research-related tasks. It also explores how such skills could be embedded in A level courses of study.

Shaw, Johnson and Warwick discuss Assessment for Learning (AfL) practices in international contexts. The Assessment for Learning in International Contexts (ALIC) project highlights the values and practices that seem to be important to teachers across diverse national contexts. This analysis of assessment practice on the ground suggests that different factors impact on teachers' aspirations and interpretations and that research such as this is extremely valuable in identifying how the principles of initiatives such as AfL can be realised.

The final article in this issue focuses on the methodological challenges faced by Greatorex, Shaw, Hodson and Ireland as they attempted to use scales of cognitive demand in a validation study. Although there were difficulties in applying the method, the research indicated the potential for use of demands analyses and contexts in which such scales could be used in future validity studies.

At Cambridge Assessment a Qualitative Research Methods Reading Group was first convened by Martin Johnson, a research officer working in the Research Division. Over the last 5 years this group has grown both in size and remit. It is an invaluable focus for discussion and has helped to develop shared understandings which have strengthened research across the organisation as members have learned new skills and identified a range of potential methodologies.

Sylvia Green *Director of Research*

21st Century skills: Ancient, ubiquitous, enigmatic?

Irenka Suto Research Division

Introduction

Today, because of rapid economic and social change, schools have to prepare students for jobs that have not yet been created, technologies that have not yet been invented and problems that we don't yet know will arise.

Andreas Schleicher, OECD Education Directorate, 2011.

The understanding and skills needed to compete in today's global economy are arguably quite different to those upon which 19th and 20th century education systems have traditionally focussed. Life has become much more international, multicultural and inter-connected. Seismic advances have occurred in ICT and in access to it. These have enabled the economies of developed countries, including the UK's, to shift from a basis of material goods and services to one of information and knowledge (Lisbon Council, 2007; Cisco, Intel and Microsoft, 2008). Whereas the possession of detailed facts and figures was once a passport to a professional job or a university place, there is now much more emphasis on what people can do with the knowledge they can access (Silva, 2009) and on interpersonal skills. In the UK and elsewhere, aspirations towards a more meritocratic society (e.g. Aim Higher, 2011) coupled with these economic and social changes have contributed to intense competition for places at top universities and for jobs, leading applicants to seek new ways to distinguish themselves. Simultaneously, however, some employers and HE tutors complain that new recruits arrive ill-equipped, having been spoon-fed material at school or college (Tickle, 2011). It is unsurprising that references to so-called '21st Century skills' have peppered many debates over what and how today's students should learn, in order to become productive citizens.

The aim of this article is to explore some of the benefits and risks of building pedagogies and curricula around 21st Century skills. I begin by outlining some conceptualisations of 21st Century skills. I then address the question of how their development in young people can best be supported; I describe recent examples of alternative approaches used in the UK and internationally, including extended projects for sixth-form students. I also start to consider the value placed by stakeholders on the summative assessment of 21st Century skills, and finally, the feasibility of such assessment for test developers.

What are 21st Century skills?

There is no single widely-accepted definition of '21st Century skills'. Arguably, this is to be expected, given the diversity of agendas held by different educationalists, policy makers, employers, teaching unions, and higher education institutions. According to Silva (2009), there are hundreds of descriptors of the skills set, including life skills, workforce skills, interpersonal skills, applied skills, and non-cognitive skills.

One of the largest research ventures currently underway is *Assessment and Teaching of 21st Century Skills (ATC21S)*. The stated goal of this international collaboration among academics, governments and three major technology companies is to 'Change the way students are taught in the classroom, to make their education relevant for the 21st Century and beyond' (ATC21S, 2011). An initial objective of the ATC21S project was to develop clear, operational definitions of 21st Century skills. Researchers began by conducting what is probably the most thorough recent review of the literature in this field. They analysed the definitions developed and used by eleven major organisations, including the Partnership for 21st Century skills (2011) in the United States and the Lisbon Council (2007) of the European Union.

The ATC21S researchers concluded that 21st Century skills can be grouped into four broad categories: (i) ways of thinking; (ii) ways of working; (iii) tools for working; and (iv) skills for living in the world (Binkley, Erstad, Herman, Raizen, Ripley and Rumble, *in preparation*). Within these categories, they identified ten skills as encapsulating all others and accommodating all approaches. The four categories and ten skills are shown in Table 1. The table also includes the skills in the key literature reviewed by ATC21S (Binkley *et al.*, *in preparation*; Cisco, Intel and Microsoft, 2008) and outlined by the Confederation of British Industry (2007) following a consultation with its members. It can be seen that none of the organisations in Table 1 agree exactly on the skills that are needed for the 21st Century, but there is considerable overlap among them. In particular, problem-solving, and ICT operations and concepts, are listed by all organisations. Communication, collaboration, and information literacy (the ability to mine new information and interact constructively with it) are also listed frequently.

In addition to the 21st Century skills shown in Table 1, others were discussed at a recent education conference¹. A Senior Tutor at the University of Cambridge stressed the need for top university applicants to possess 'mental fluency' in their subject of interest; for example, this might be mathematical fluency, fluency in writing, or fluency in scientific thought (Partington, 2011). A closely-related skill is articulacy, an aspect of communication. Partington warned that in the UK, this is becoming the preserve of the middle classes. Other skills mentioned by Partington included high motivation, and an interest in complexity and difficulty which feeds into creativity.

Multilingualism does not feature explicitly in Table 1, but is often considered a key part of communication within a global community. The term can mean different things to different people. For some employers in other parts of the world, multilingualism boils down to employees speaking English, often in addition to a different native language. English must be mastered in order to communicate in international business meetings (virtual or face-to-face), as well as to

1. *What kind of education enables us to cope with an interconnected world?* A Cambridge Assessment event held at 1 Great George Street, Westminster, London, SW1, on 15/03/11.

Table 1: Definitions of 21st Century skills

ATC21S	21st Century skills projects reviewed by ATC21S					Confederation of British Industry (CBI) (2007)
Categories of 21st Century skills	21st Century skills	Partnership for 21st Century Skills (2011)	Lisbon Council (2007)	International Society for Technology in Education (ISTE) NETS (2011)	ETS iSkills (2011)	
Ways of thinking	1. Creativity and innovation	Creativity and innovation		Creativity and innovation	Creativity and innovation	
	2. Critical thinking, problem solving, decision-making	Critical thinking, problem solving, decision-making	Problem-solving	Critical thinking, problem solving, decision-making	Critical thinking, problem-solving	Problem-solving
	3. Learning to learn, metacognition					
Ways of working	4. Communication	Communication		Communication	Communication	Communication
	5. Collaboration (teamwork)	Collaboration	Collaboration	Collaboration		Collaboration
Tools for working	6. Information literacy (includes research on sources, evidence, biases, etc.)	Information literacy, media literacy,	Information literacy	Information literacy	Information literacy	Application of numeracy
	7. ICT literacy	ICT operations and concepts	ICT operations and concepts	Research and inquiry, Digital citizenship, ICT operations and concepts	ICT operations and concepts	ICT operations and concepts
Living in the world	8. Citizenship – local and global					
	9. Life and career	Initiative and self-direction, Flexibility and adaptability, productivity, leadership and responsibility	Flexibility and adaptability			Initiative and self-direction
	10. Personal and social responsibility – including cultural awareness and competence					Business awareness Customer care

read technical and specialised documents which cannot easily be translated into every language. For some educationalists, however, learning another language (not necessarily English) enables people to understand their own first language in a different way, building more sophisticated lifelong learning patterns (Puntis, 2011). Arguably, this latter form of multilingualism is an important contributor to the third skill construed by ATC21S: 'learning to learn'.

Another perspective is that many 21st Century skills, including creativity, problem-solving, decision-making, communication, collaboration, citizenship, and personal and social responsibility, are linked inextricably to personality characteristics and so-called 'emotional intelligence' (EI). Petrides (2001) and Petrides and Furnham (2003) have defined EI as a constellation of behavioural dispositions and self-perceptions concerning one's ability to recognise, process, and utilise emotion-laden information. EI is conceptualised as an aspect of personality, which is malleable and still developing well into a person's twenties, and is unrelated to non-verbal reasoning ability. Petrides and Furnham designed a self-report questionnaire to give comprehensive coverage of all personality traits related to emotion. The questionnaire,

which has been well validated in many countries and languages, measures self-perceived skills and abilities such as being a good negotiator and adapting well to new situations. These skills fall within the conceptualisations of 21st Century skills shown in Table 1.

A further usage of the term '21st Century' is to attach it to any subject considered or intended to be fit for purpose in the present century. '21st Century literacy', for example, can be used to describe modern literacy skills or current best practice in literacy, just as '21st Century mathematics' can describe modern approaches in mathematics. The implication is that subjects must evolve to meet new needs. Conceptual leaps in pedagogy as well as substantial changes in subject content may sometimes be evident. GCSE courses in 21st Century Science, for example, have been developed through a partnership among an educational foundation, a university, an academic publisher and an awarding body (21st Century Science, 2011). The courses focus on the knowledge and understanding needed in order to recognise the impact of science and technology upon everyday life, and contemporary contexts such as air quality and sustainable energy are used. Subject-specific uses of '21st Century' are not considered further in this article.

Approaches to developing 21st Century skills

The question of how best to support the development of 21st Century skills in young people is posed frequently. Several diverse solutions have been mooted, which are now discussed in turn.

Continue with long-standing methods of teaching 21st Century skills

One view is that the skills now termed '21st Century' skills have long been recognised by educationalists and are nothing new. Whilst a focus on their importance by the media might be novel, it is a consequence of deficits in current education systems, and the skills themselves are ubiquitous. According to Silva (2009), creative, critical and analytical thinking skills have been valued by many philosophers and educators, from Socrates 2400 years ago, to John Dewey in the twentieth century. Many long-established vocational courses can be seen to nurture 21st Century skills. For example, Rose (2011) has articulated some of the highly sophisticated analytical, problem-solving and creative skills developed on electricians' courses. Puntis (2011) has argued that traditional academic subjects such as mathematics and the sciences can also be reconceptualised in terms of the 21st Century skills they engender – a position shared by the Advisory Committee on Mathematics Education (2011). This may explain why performance in these subjects correlates highly with personal career earnings in the 21st Century (Chevalier, *in press*).

Consider the cognitive domain of Bloom *et al.*'s (1956) taxonomy of educational objectives, which comprises six levels: knowledge, comprehension, application, analysis, synthesis/creation, and evaluation. A relatively recent revision of the cognitive domain (Anderson and Krathwohl, 2001) presents the three lowest levels as ordered hierarchically, but the three higher levels as parallel. In Figure 1 I show how many of the 21st Century skills identified by ATC21S (Table 1) can be captured by the taxonomy. The exceptions are communication, collaboration, and ICT literacy. While the latter skill could not have existed widely in 1956, communication and collaboration skills certainly

did, but were conceptualised more implicitly, as necessary components of the successful usage of other cognitive abilities.

Although the positioning of 21st Century skills in Figure 1 is only approximate, it suggests that they tend to be higher order thinking skills. Arguably, the focus of present day educationalists has drifted up the hierarchy, since many routine tasks are now performed by computers and other machinery. This drift can be perceived as being part of a wider phenomenon of downwards pressure within the educational system, with higher education approaches to pedagogy and curriculum being implemented in secondary schools, and primary education approaches being applied in early years settings. The origin of such pressure is unlikely to be exclusively economical. It may also reflect recent developments in the understanding of children's cognitive capabilities. According to Oates and Grayson (2004), higher order thinking skills develop earlier in childhood than Piaget's popular theory of cognitive development would suggest. For example, infants seem able to differentiate among types of things much earlier than previously thought. Oates and Grayson argue that as development proceeds, mental representations become progressively more abstract and freed from the concrete examples and specific contexts in which they originated, and that abstraction begins before many children begin formal schooling. There is therefore a need to support the development of higher order thinking at a younger age, through the teaching of traditional school subjects or otherwise.

Develop curricula covering 21st Century skills explicitly

An alternative view on the optimal development of 21st Century skills is that curricula should be developed to cover them explicitly. Over the past decade, critical thinking, for example, has become a subject in its own right for many sixth-form students in England. According to the website of a major English awarding body offering AS and A level courses in critical thinking:

Critical thinking is a skills-based rather than content-based A Level. It develops the ability to interpret, analyse and evaluate ideas and arguments and can support thinking skills in all subject areas, from arts and humanities to sciences. (OCR, 2011)

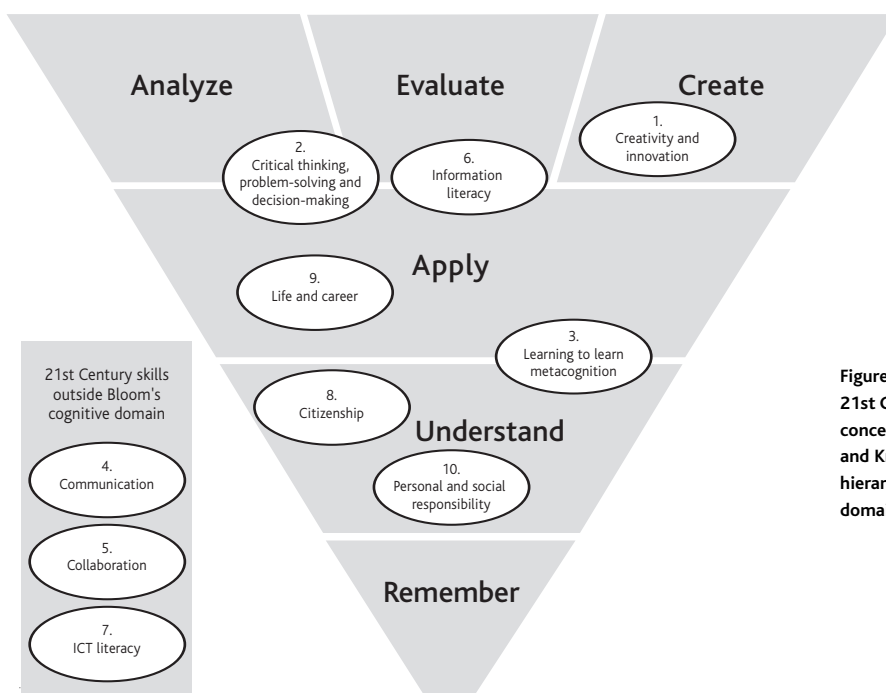


Figure 1: Approximate mapping of 21st Century skills (ATC21S conceptualisation) onto Anderson and Krathwohl's (2001) revised hierarchy of Bloom's cognitive domain

The popularity of critical thinking courses has soared over the past decade. While in 2001, just 2000 students sat AS level examinations in critical thinking, by 2009 that figure had risen to 22000 students (Black, 2010). Other thinking skills curricula have been around for longer. In 1998, for example, the implementation of three thinking skills programmes among younger secondary school students in England was reviewed by Baumfield and Oberski (1998). Findings from their case study indicated that experienced teachers faced considerable difficulties when attempting to develop the new skills needed to teach the programmes. Similar concerns over teaching capacity have been raised in the US, where thinking skills courses are well established (Sternberg, 1987). One challenge is that focussing not on a student's answer, but on the thinking underpinning that answer, is a crucial but less familiar process for many teachers. According to Black (2010), the difficulty of introducing a critical thinking course is more than that of the subject's novelty; it requires a fundamental re-orientation prior to teaching.

Critical thinking is by no means the only 21st Century skill to be taught as a separate subject. There has been considerable debate in schools and advisory services in England about whether ICT literacy (knowledge, skills and processes) should be developed through teaching ICT as a separate subject or through integrating ICT across the school curriculum (Webb, 2002). In her exploration of the positioning of ICT in the curriculum, Webb (2002) identifies three discrete approaches: (i) learning ICT as a subject; (ii) using ICT as a tool for learning (for example, using word processing software to redraft an essay or running a simulation to test a scientific prediction); and (iii) learning through ICT (situations in which the ICT facility becomes the whole learning environment by providing learning materials and acting as the tutor and assessor). Webb (2002) raises multiple concerns over ICT teaching capacity, some of which apply also to the teaching of thinking and other 21st Century skills. These concerns include a lack of skills and content knowledge among teachers, loosely defined pedagogical content knowledge, and loosely defined general pedagogical knowledge (subject specific interpretations are less well-defined for ICT than for other subjects). Borovik (2011) raises a further concern about ICT education. Whilst acknowledging the need for modern employees to be able to use ICT (that is, to be ICT *literate*), he suggests that producers and 'creators' of ICT are needed as well as users. Borovik argues that the IT industry wants schools to teach the fundamentals of actual computing, that is, the development of algorithms and their subsequent implementation in computer code. This entails sophisticated abstract thinking, not dissimilar to that used in mathematics.

Adopt a skills-centred pedagogy in schools and colleges

To encourage the development of multiple 21st Century skills, the Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA) has devised a pedagogical framework called *Opening Minds*. The framework has been created to promote 'innovative and integrated ways of thinking about education and the curriculum' (RSA, 2011). It entails five key competences: citizenship, learning to learn, managing information, relating to people, and managing situations. Teachers use the competences to develop curricula to suit their own schools, and can use whatever content they like. The idea is that they teach *through* the competences, which become the lesson objectives. According to the RSA's *Opening Minds* website:

A competence based approach enables students not just to acquire subject knowledge but to understand, use and apply it within the

context of their wider learning and life. It also offers students a more holistic and coherent way of learning which allows them to make connections and apply knowledge across different subject areas. (RSA, 2011)

Through another lens, *Opening Minds* can be viewed as a pedagogical framework with a focus on strengthening emotional intelligence. It is being used in two hundred secondary schools across the country, including the RSA Academy in Tipton, near Birmingham. The RSA Academy has an underprivileged intake, and according to James (2011), the competences help to give students a sense of being and of where they fit into the world. Affluent students arrive at other secondary schools already knowing who they are and how they can make a contribution to the global community. In contrast, students at the RSA Academy are far less sure of themselves as they do not absorb such understanding by osmosis from their home and social environments. The RSA Academy aims to develop a culture and love of learning which, James argues, is vital for becoming a global citizen but is not usually prevalent in underprivileged areas.

Students in Key Stage 3 (ages 11 to 14) at the RSA Academy spend considerable time on: global integration; creativity; critical thinking; problem-solving; collaborative approaches; ICT; and global citizenship. Each day comprises two three-hour lessons, and students work through a set of themes, where teachers attempt to show how subjects are inter-related. James (2011) claims that new students entering Key Stage 3 lack confidence and are inarticulate because they are not used to engaging in discussion. Students in Key Stage 4 (ages 14 to 16) who have successfully grasped the *Open Minds* competencies are better learners because they understand who they are, and are beginning to develop an appreciation of what education can offer them.

Nurture 21st Century skills through extra-curricular activities

Another perspective is that 21st Century skills can be nurtured successfully outside of lessons. This longstanding position has been described by Haensly, Lupkowski and Edlind (1985):

Many students seek, and seem to thrive on, activities outside of the traditional classroom setting. Such activities are variously termed the co-curriculum or the extra-curriculum, apparently depending on whether they are specific extensions of academic coursework, or are peripheral to it. Thus, students may extend and enrich previously learned academic skills through competitions (e.g. interscholastic debates) and by applying them to the real world simulations (e.g. writing skills in school publications). In the co/extra-curricular setting they may also develop and practice artistic, musical, and psychomotor talents; leadership skills, and future career and occupational skills. Interpersonal and social strategies – proficiencies not considered basic elements of the academic curriculum – may especially be constructed through participation in the extra-curriculum. (Haensly, Lupkowski and Edlind, 1985, pp.110–111)

Whilst the broad benefits of extra-curricular activities are rarely disputed, concerns over inequity and inequality have been raised. Penney and Harris (1997), for example, have drawn particular attention to the disadvantaging of girls, and of lower ability pupils, in physical education in terms of the opportunities and experiences typically available to them in extra-curricular settings. Resources can be expensive with the

consequence that both internationally and nationally, the least affluent students may struggle to compete. Bourdieu (1984) famously argued that despite the apparent freedom of choice in the arts, individuals' artistic preferences and tastes align closely with their social positions. Contributing to Bourdieu's concept of 'cultural capital', extra-curricular activities (such as playing an orchestral instrument) may play a major role in securing university places and higher-paid, higher-status jobs not through the skills engendered, but through the status conferred. More recently, Partington (2011) reports that university admissions tutors are aware of these difficulties and have an ambivalent attitude to applicants' extra-curricular achievements.

Cultivate 21st Century skills through independent research projects

In order to improve equity, some schemes and activities have been designed to be flexible enough to function both within and beyond the curriculum. The British Science Association's (2011) Crest awards offer a modern example of how 21st Century skills are being nurtured in this way. The scheme operates at three levels spanning secondary education (11 to 19 years), and awards are given for project work in science, technology, engineering and mathematics (STEM subjects), produced either individually or in teams. It is intended that Crest projects link the personal passions of students to curriculum-based learning, and at the highest level, students spend over seventy hours on their projects. The scheme is designed to enrich the curriculum but can also link into work experience placements and after-school clubs. It is difficult to ensure complete equality of opportunity, since this kind of scheme ultimately depends on teachers and resources, but scheme flexibility is an important contributor.

The Crest awards in STEM subjects resemble several more formal educational routes entailing independent project work, which lead to accredited qualifications. These routes include Project Qualifications, which have been developed for students in all subjects at three levels: Foundation (Level 1), Higher (Level 2), and Extended (Level 3). Project Qualifications are administered by several different awarding bodies in England and Wales. At Level 3 (16–19 years) Extended Project Qualifications (EPQs) can be taken alongside A levels, as well as being a compulsory element in Diplomas. Students explore a particular interest which may be unrelated to anything else they are studying. They undertake purposeful, cross-curricular study both inside and outside the classroom. This leads to the composition of a 5000-word dissertation or to a shorter report of 1000 to 3500 words accompanying a project artifact (such as a piece of furniture or a musical performance). The students' project supervisors mark their work, which may then be moderated by an assessor employed by the awarding body. The Blair government envisaged that the task of carrying out a research project of this kind requires many 21st Century skills, including creativity and imagination, problem-solving skills, independent thinking, cooperation with others, and using people as resources, while relying on solid subject disciplines (DfES, 2005).

Similar educational routes exist elsewhere, both locally and internationally. Students at The Sixth Form College, Farnborough (2011) carry out projects (without a qualification attached) which were originally a precursor to EPQs. (The college worked closely with the AQA awarding body to develop the latter.) The students identify research questions that take them beyond the confines of their A level syllabuses/specifications, and which link at least two subjects. They then

conduct their research during their holidays or alongside their other courses, generating formally-structured reports of 4000–5000 words. Assessment is formative: teachers provide their students with detailed written comments. Internationally, the International Baccalaureate's Diploma students engage in independent research through an in-depth study of a question relating to one of the subjects they are studying (International Baccalaureate Organisation [IBO], 2011). This study culminates in the production of an extended essay of up to 4000 words, which is often followed by a short concluding interview with the student's supervisor. The extended essay, which is marked by external assessors, is intended to promote 'high-level research and writing skills, intellectual discovery and creativity' (IBO, 2011).

Develop 21st Century skills in the workplace

The extent to which 21st Century skills could and should be developed in the workplace in addition to, or instead of, in schools and colleges, is a controversial issue. Apprenticeships, internships, and work experience placements provide young people with rich opportunities to develop the generic and transferable skills desired by members of the Confederation of British Industry (2007, see Table 1). Reporting on the German dual apprenticeship system, which entails alternation between training in vocational schools and training in firms, researchers have commented:

Workplace apprenticeship is the fundamental principle of vocational training in Germany and is thought to combine the most favourable conditions for developing skills. Skills include the capacity to take on the responsibilities involved in completing tasks, the ability to communicate and work in a team, the ability to think in terms of systems, and the ability of learning to learn. (Tremblay and Le Bot, 2003, p.14)

Whilst in Germany, companies provide apprenticeships for social and moral reasons and assume the costs of practical training (Tremblay and Le Bot, 2003) such obligations are felt less strongly by many companies in the UK and elsewhere. As participation and engagement in work placement schemes can be expensive, many businesses and other organisations attempt to recruit ready-trained people who already possess skills corresponding to their immediate and particular needs. In some sectors within the UK, such as journalism, fashion, and politics, unpaid internships are common. As with extra-curricular activities, however, concerns over inequity and inequality have been raised. For example, the deputy prime minister recently commented that making young people work for nothing is barring entry to those from poorer or less well-connected backgrounds and harming social mobility (BBC, 2011).

Assessing 21st Century skills

The above approaches to developing 21st Century skills are diverse, each presenting different challenges for those wanting to assess the skills cultivated. Although these challenges (discussed subsequently) may be considerable, there is nonetheless a strong desire for assessment from some quarters. ATC21S, for example, places a strong emphasis on the importance of assessing 21st Century skills. At its launch, the project's director commented:

Reforming assessment is essential to enabling any systemic change in education. And change on a global scale is required to equip students

of today with the skills they need to succeed in the workforce of tomorrow. The international education assessments in the Program for International Student Assessment (PISA), with which I was involved as Director for Education at the Organization for Economic Co-operation and Development (OECD), focus on key competencies in reading, mathematics and science but we always wanted to extend the scope to cover important new skills. In PISA 2003, we took a step by adding an assessment of problem solving, but one limited to analogical reasoning. We hoped to add information and communications technology (ICT) competence in PISA 2006 but did not succeed. We all need now to work together to advance assessment practice. (McGaw, 2008)

This comment conveys the implicit message that assessing a construct summatively adds meaning and extrinsic value to it for a greater number of people. If a 21st Century skill can be quantified, then measurements of it (that is, assessments and qualifications) can be utilised in application processes for jobs and university places, thereby making them more meritocratic. A further argument in favour of assessment is that it can drive curriculum and pedagogical developments, and for newly defined skill sets, these are often sorely needed. At a major testing organisation in the US, novel constructs have been clarified and articulated through the process of developing new tests in listening skills, people skills, and problem-solving skills (ETS, 2011).

There are mixed views on whether and how research projects for sixth-formers should be assessed. According to Guy (2011), students' projects at The Sixth Form College, Farnborough are assessed only formatively, to prevent them becoming 'high stakes'. In Guy's opinion, formal summative assessment and grading reduces highly desirable risk-taking behaviours among students. Instead of playing safe to obtain a qualification, the focus of the research process ought to be on opening up a new world of intrinsically pleasurable learning, thereby developing knowledge. On completing their projects, Farnborough students ask each other, "What did you do yours on?" rather than, "What grade did you get?".

In contrast, External Project Qualifications (EPQs) are assessed summatively by teachers, using assessment objectives provided by the awarding bodies administering them. As with Farnborough projects, however, the focus is intended to be on the investigative process that the student undertakes in order to achieve and evaluate his or her final outcome, rather than on the outcome itself. That is, the generic research and 21st Century skills displayed by the student are to be rewarded, rather than the subject content. This assessment approach is encountered infrequently in general secondary education in England and Wales, and presents considerable challenges relating both to marking task demands, and to assessors' personal expertise. Challenges relate to the complexity, variety, novelty and interdisciplinary nature of the project work and skills demonstrated. On the other hand, the EPQ's assessment framework provides an overt and tangible means of steering students and teachers towards focusing upon 21st Century skills. Moreover, recognition of the EPQ by the Universities and Colleges Admissions Service (it attracts 20 to 70 UCAS tariff points, for grades E to A* respectively), which is contingent upon assessment, may have widened uptake. UCAS recognition² is likely to have increased awareness and encouraged many schools and colleges to provide their students with new opportunities to

carry out projects, which would not otherwise have been offered.

Finally, it is worth noting that not all attempts to assess 21st Century skills are successful. A common difficulty is that some skills are too subjective and enigmatic to be measured objectively. Since creativity, for example, can involve every sense (sight, hearing, touch, smell and taste) and is almost infinite, it defies precise definition. Whilst it may be possible to assess the quality of products of creative (and additional) processes, the skill itself may not be readily assessable. Similarly, Unwin (2010) considers collaboration and initiative-taking to be highly contextual capabilities; they therefore elude comparable quantification. Whilst new conceptualisations of skills often call for innovative assessment approaches, over-ambition can be costly. In England and Wales in 2007, a £26 million project to develop computerised tests of Key Stage 3 ICT skills had to be abandoned after a novel adaptive assessment model proved unworkable and theoretically insecure (Mansell, 2007). The tests were intended to transform ICT teaching and improve ICT infrastructure in schools, but ultimately, these goals could not be realised.

Conclusions

In this article, multiple conceptualisations of 21st Century skills have been outlined. The diversity of usage of the term has been highlighted, as well as the common ground among major organisations. The inter-disciplinary skills most commonly regarded as essential for the 21st Century are problem-solving, ICT operations and concepts, communication, collaboration, and information literacy. Some would argue that these skills are as ancient as Socratic philosophy, others would argue that they are ubiquitous in education, and others would argue that they are enigmatic since they elude objective measurement.

The article also contains an exploration of the key approaches to developing 21st Century skills currently in use in England and Wales as well as in some international contexts. While some educationalists wish to continue with long-standing teaching methods, others favour the construction and adoption of new pedagogies, curricula and extra-curricular activities, which incur both risks and benefits. Several of the approaches discussed can be used in combination with one another. Assessment is a major consideration for many stakeholders, and opinions are divided over its value and feasibility. In the case of research projects for sixth-formers, summative assessment faces challenges relating to task demands and assessor expertise, and may reduce risk-taking behaviour in students. However, it provides a tangible means of widening access, as well as guiding students and teachers towards the need to focus upon skills and processes rather than outcomes. For many teachers, this entails a fundamental re-orientation in pedagogical approach which is likely to be supported by emerging communities of teaching and assessment practice.

References

- Advisory Committee on Mathematics Education (2011). *Mathematical Needs: Mathematics in the workplace and in Higher Education*. London: The Royal Society.
- Aim Higher (2011). Official website. Available online at: <http://www.aimhigher.ac.uk/candp/home/> (accessed 7 June 2011).
- Anderson, L. & Krathwohl, D. A. (2001). *Taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

2. The qualification was first considered for UCAS tariff points in 2007. Points came into effect for entry to higher education from 2010 onwards.

- Assessment and Teaching of 21st Century Skills (*in preparation*) Draft white papers. Melbourne: The University of Melbourne.
- ATC21S (2011). Official website. Available online at: <http://atc21s.org> (accessed 26 May 2011).
- Baumfield, V. M. & Oberski, I. M. (1998). *What do teachers think about teaching skills?* *Quality Assurance in Education*, **6**, 1, 44–51.
- BBC (2011). *Lib Dems under fire over unpaid interns*. Available online at: <http://www.bbc.co.uk/news/uk-politics-12983163> (accessed 19 July 2011).
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M. & Rumble, M. (*in preparation*) *Defining 21st Century skills. Draft white paper*. Part of a report to the Learning and Technology World Forum 2010, London.
- Black, B. (2010). "It's not like teaching other subjects" – the challenges of introducing Critical Thinking AS level in England. *Research Matters: A Cambridge Assessment Publication*, **10**, 2–8.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: the classification of educational goals; Handbook I: Cognitive Domain*. New York: Longmans, Green.
- Borovik, A. (2011). *The strange fate of abstract thinking*. Available online at: <http://www.maths.manchester.ac.uk/~avb/methodology.html> (accessed 4 July 2011).
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. London: Routledge.
- British Science Association (2011). *CREST awards*. Available online at: <http://www.britishecienceassociation.org/web/ccaf/CREST/index.htm> (accessed 02 June 2011).
- Chevalier, A. (*in press*). Subject choice and earnings of UK graduates. *Economics of Education Review*. (2011), doi:10.1016/j.econedurev.2011.04.007
- Cisco, Intel and Microsoft (2008). *Transforming education: Assessing and teaching 21st Century skills*. Available online at: <http://download.microsoft.com/.../6/E/9/.../Transformative%20Assessment.pdf> (accessed 6 October 2011).
- Confederation of British Industry (2007). *Shaping up for the future: The business vision for education and skills*. Available online at: www.cbi.org.uk/bookshop (accessed 6 April 2011).
- Department for Education and Skills (2005). *The 14–19 Education and Skills White Paper*. London: HMSO.
- ETS (2011). *New constructs*. Available online at: http://www.ets.org/research/capabilities/assessment_research/new_constructs (accessed 6 June 2011).
- ETS iSkills (2011). Official website. Available online at: www.ets.org/iskills (accessed 6 April 2011).
- Guy, J. (2011). Comment made at 'What kind of education enables us to cope with an interconnected world?' A Cambridge Assessment conference, March 2011, in London.
- Haensly, P. A., Lupkowski A. E. & Edlind E. P. (1985). The role of extracurricular activities in education. *The High School Journal*, **69**, 2, 110–119.
- International Baccalaureate Organisation (2011). *Diploma Programme curriculum – core requirements*. Available online at: <http://www.ibo.org/diploma/curriculum/core/> (accessed 3 June 2011).
- International Society for Technology in Education (ISTE) NETS (2011). Official website. Available online at: www.iste.org (accessed 6 April 2011).
- James, I. (2011). *Opening Minds*. Paper presented at 'What kind of education enables us to cope with an interconnected world?' A Cambridge Assessment conference, March 2011, London.
- Lisbon Council (2007). *Skills for the future* (Brussels: Lisbon Council). Available online at: www.lisboncouncil.net/component/downloads/?id=214 (accessed 6 April 2011).
- Mansell, W. (2007). Computerised testing collapses. *The Times Educational Supplement* 19/01/07, 18–19.
- McGaw, B. (2008). *News release: Cisco, Intel and Microsoft collaborate to improve education assessments*. Available online at: <http://atc21s.org/Documents/ProjectLaunch.pdf> (accessed 6 April 2011).
- Oates, J. & Grayson, A. (2004). *Cognitive and language development in children*. Oxford: Blackwell Publishing in association with The Open University.
- OCR (2011). Website for A level Critical Thinking. Available online at: http://www.ocr.org.uk/qualifications/type/gce/hss/critical_thinking/index.html (accessed 31 May 2011).
- Partington, R. (2011). *Cambridge University and education in an interconnected world*. Paper presented at 'What kind of education enables us to cope with an interconnected world?' A Cambridge Assessment conference, March 2011, London.
- Partnership for 21st Century Skills (2011). Official website. Available online at: <http://www.p21.org/> (accessed 26 May 2011).
- Penney, D. & Harris, J. (1997). Extra-curricular Physical Education: More of the same for the more able? *Sport, Education and Society*, **2**, 1 41–54.
- Petrides, K. V. (2001). *A psychometric investigation into the construct of emotional intelligence*. Doctoral dissertation. London: University College London.
- Petrides, K. V. & Furnham, A. (2003). Trait emotional intelligence: Behavioural validation in two studies of emotion recognition and reactivity to mood induction. *European Journal of Personality*, **17**, 39–57.
- Puntis, A. (2011). *An Interconnected World*. Paper presented at 'What kind of education enables us to cope with an interconnected world?' A Cambridge Assessment conference, March 2011, London.
- Rose, M. (2011). *Rethinking remedial education and the academic-vocational divide: Lessons to learn about language, cognition, and social class*. Paper presented at the American Educational Research Association annual conference, April 2011, New Orleans.
- RSA (2011). Official website. Available online at: <http://www.rsaopeningminds.org.uk/about-rsa-openingminds/> (accessed 1 June 2011).
- Schleicher, A. (2011). *The case for 21st-century learning*. Available online at: http://www.oecd.org/document/2/0,3746,en_21571361_44315115_468465_94_1_1_1_1,00.html (accessed 8 June 2011).
- Silva, E. (2009). Measuring skills for 21st-century learning. *Phi Delta Kappa*, **90**, 9, 630–634.
- Sternberg, R. J. (1987) Teaching critical thinking: eight ways to fail before you begin. *Phi Delta Kappa*, **68**, 246–249.
- The Sixth Form College, Farnborough (2011). *The Farnborough extended project*. Available online at: <http://www.farnboroughsfc2.ac.uk/Extended%20Projects> (accessed 6 June 2011).
- Tickle, L. (2011). Essay writing trips up students. *The Guardian newspaper*. Available online at: <http://www.guardian.co.uk/education/2011/apr/26/students-essay-writing> (accessed 4 May 2011).
- Tremblay, D.-G. & Le Bot, I. (2003). *The German dual apprenticeship system: Analysis of its evolution and present challenges*. Research Note No. 2003–4A. Télé-université, Université du Québec. Available online at: <http://www.telu.quebec.ca/chaireecosavoir/pdf/NRC03-04A.pdf> (accessed 19 July 2011).
- 21st Century Science (2011). Official website. Available online at: <http://www.21stcenturyscience.org/> (accessed 6 June 2011).
- Unwin, L. (2010). Learning and working from the MSC to New Labour: Young people, skills and employment. *National Institute Economic Review*, **212**, 49–60.
- Webb, M. E. (2002). Pedagogical reasoning: Issues and solutions for the teaching and learning of ICT in secondary schools. *Education and Information Technologies*, **7**, 3, 237–255.

Independent research at A level: Students' and teachers' experiences

Sanjana Mehta, Irenka Suto, Gill Elliott and Nicky Rushton Research Division

Background

The A level system in England and Wales

In this article we explore the opportunities that exist to conduct independent research at A level, from the perspectives of both teachers and students. In England and Wales, most 16 to 18 year old students wishing to take a non-vocational educational route follow courses leading to A level qualifications (Advanced Level General Certificates of Education). These courses are available in over 45 different subjects ranging from physics to philosophy (Ofqual, 2011), and are studied over two years: the Advanced Subsidiary (AS level) component in Year 12, and the A2 component in Year 13. Most students take between three and five subjects in Year 12 (at AS level) and typically drop one subject to continue with the remaining three or four subjects in Year 13 (at A2 level) (Vidal Rodeiro, 2007). The A level system provides a very common route into higher education and employment. The nature of the skills and knowledge that students develop through their A level courses therefore has important implications for the ease and success of these transitions (Wingate, 2006; Marland, 2003).

In contrast with educational systems for similar students in other European countries (for example, the German Abitur and the French Baccalaureate) and internationally (for example, the International Baccalaureate, IBO, 2011), the A level system is quite specialised (DfES, 2005). Students study relatively few subjects, but have the opportunity to do so in depth. Unlike almost any other educational system for 16 to 18 year olds in the developed world, no subject is compulsory (Porkess *et al.*, 2011); instead students may focus exclusively on the sciences, the humanities, modern foreign languages, or may choose a diverse combination of subjects if they wish. Awarding bodies administering the qualifications state that over a two-year period, students should receive 360 guided learning hours for each subject that they study (OCR, 2011; AQA, 2011), which amounts to at least five hours' teaching per week. Students are also expected to put in many hours of private study outside the classroom.

Whilst critics of the A level system have raised concerns that it forces students to specialise at too early an age and provides too narrow an education (Tomlinson, 2004), a potential benefit of A level courses may be their conduciveness to independent research and investigation. Arguably, A level students should have greater opportunities than should some of their European counterparts to engage in deep thinking and investigative work within their subjects.

Independent research and investigative skills

The research literature contains a mixture of conceptions of independent research and investigative skills. There exists no single system of classifying skills in this broad domain, and terms are often used

interchangeably. Within the cognitive domain of Bloom's (1956) taxonomy of educational objectives, analysis, evaluation, and synthesis/creation have been positioned at a high level within a hierarchy of cognitive skills (Anderson and Krathwohl, 2001). All three 'higher order' skills are used extensively in independent research and investigative activities, are valued highly at university, and relate closely to some '21st Century skills' such as critical thinking, problem-solving, decision-making, and innovation (for discussion see Silva, 2009; ACT21S, 2011). More tangible skills asserted as being important for research and investigation include: locating information using appropriate sources, determining the scope of tasks, recording and organising findings, and keeping material relevant to the original research question (O'Sullivan and Dallas, 2010). Other authors have explored the nature of independent learning. Skills considered relevant to this related concept include: planning, prioritisation, time management, getting started on assigned tasks, and organisation (Smith, 2004; Wingate, 2006; O'Shea, 2010).

Embedding opportunities within academic A level courses has not been the only approach to developing independent research and study skills. O'Sullivan and Dallas (2010) describe a collaborative project in the US, between an English teacher and the school librarian, which aimed to prepare high school students for higher education by providing them with a 'research paper' class within the general secondary curriculum. The class offered students structured guidance in selecting a topic, gathering information, and organising the information into a 10–15 page research paper. The general objective was to foster skills such as critical thinking and problem solving. In the UK, the Extended Project Qualification (EPQ) has been developed explicitly to provide students with an opportunity to develop project management skills, including planning, accessing information from a range of resources, analysing, organising and integrating findings (OCR, 2011). Students' projects culminate in 5000-word dissertations, or in shorter reports accompanying artefacts such as works of art or music. They may be on any topic, and are conducted alongside study for A levels or vocational qualifications. Other approaches include the extended essay written by students working towards the International Baccalaureate Diploma (IBO, 2011), and the British Science Association's (2011) scheme of Crest awards for projects in mathematics and science.

Whilst the above schemes have proven valuable to many students, key concerns for educationalists and policy makers relate to the equality of provision and access. Given that university tutors claim some new undergraduates are not ready for higher level study, having been spoon-fed material at school and college (Kajander and Lovric, 2005; Rees and Wilkinson, 2008), it is clear that not all students are benefitting. Although project schemes such as EPQ are funded by the state, uptake is mixed. Some individual schools and colleges go to considerable lengths to support students undertaking projects of various types, whereas others

restrict themselves much more to traditional curricula, or offer alternative forms of extracurricular activities with alternative associated benefits. Arguably, therefore, it is important for staple courses such as A levels to provide opportunities for independent research and investigation too. This may help to level the playing field at a time of intense competition for university places (Richardson, 2010, November) and employment.

It is important to note that *perceptions* of the opportunities provided by specific A level courses may vary among teachers. Although some subjects may not easily lend themselves to independent research and investigation, creative teaching and assessment approaches could bolster opportunities for students. Even in the case of the subjects that do lend themselves better to independent research, there are other factors related to the institution, the culture of the school department, teaching style and individual characteristics which determine whether students will develop such skills successfully (Lovitts, 2005). Classroom-based action research by Stohl (2010), for example, indicated that questioning and focussing on inquiry played an important role in developing important characteristics and skills among students, including: autonomy, motivation, self-confidence, independent learning and knowledge-seeking.

Further concerns over equality of opportunity relate to the availability and usage of the practical resources needed to develop research and investigative skills, as well as the advice and guidance that students receive when using them. Although many students in the 21st Century have access to extensive information via the internet, recent studies have indicated that students' ability to locate specific information and use that information appropriately is often rather weak (Owen, 2010; Salisbury and Karasmanis, 2011). McClure and Clink (2009) found that students rarely pay attention to aspects such as timeliness, authority and bias while gathering information from online sources. They found that teachers face a challenge in encouraging their students to attend to these crucial issues.

The present study

The aims of the present study were to explore teachers' and students' experiences and perspectives of independent research at A level. The study focused on three contrasting but mainstream A level courses: Economics, French and Mathematics. These three subjects were selected for investigation because they are relatively popular subjects, available for study in most types of schools and colleges for 16 to 18 year olds, and they contrast with one another. Three main issues were investigated: (i) the extent to which teachers think research and investigative skills can be developed at A level; (ii) the resources and formal guidance that students use; and (iii) whether subject-specific differences arise. The study contributed to a wider project on teaching and learning at A level, which explored related issues such as teaching beyond the syllabus (Suto *et al.*, 2011), reasons for subject choice (Mehta *et al.*, 2011), and class size (Rushton *et al.*, 2011). A questionnaire and follow-on interview methodology was used to collect data, which were analysed quantitatively and qualitatively.

Methods

Questionnaires

Two questionnaires were developed: one for teachers and one for students in Years 12 and 13 (aged 16 to 18). Ideas for themes to be

explored were generated from a series of brainstorming meetings held with subject experts and other researchers. Questions were constructed which covered all the identified themes and extended draft questionnaires were assembled. These were then refined by removing or rewording questions which seemed less likely to generate meaningful responses, and by incorporating instructions and a consent form at the start.

The draft questionnaires were piloted in a local school by a teacher and two students. The teacher was asked to complete the teacher questionnaire and then to comment on the following areas:

1. Typing mistakes or grammatical errors
2. Ambiguous questions
3. Terminology of the questions (words which were not understood or unclear phrases)
4. Multiple choice answers
 - a. Suitability of answer options
 - b. Missing answer options
 - c. Extraneous answer options

The two students were instructed to fill in independently the student questionnaire and then give feedback in response to:

1. Understanding of what each question was asking
2. Words used which were not understood
3. Ease of answering each question

The teacher was also asked to provide general feedback about the student questionnaire.

Several suggestions were received in the pilot, including: additional response options, changes to wording to clarify what was being asked, and indications of where more space was needed for the free text responses. This feedback was used to amend some questions. The overall length of the questionnaires was also reduced by removing some questions. The amended versions of the teacher and student questionnaires were reviewed by subject experts and other researchers.

Once each questionnaire had been finalised, separate versions were created for Mathematics, Economics and French. These differed only in the subject name used throughout.

In the teacher questionnaire, respondents were asked to indicate the frequency with which carrying out an investigation/research is set as: (i) classroom activity, and (ii) homework/private study. Respondents were asked to tick the most relevant option from: *twice a week or more, once a week, once a fortnight, less often, and never.*

In another question, teacher respondents were asked:

To what extent are your students required to utilise independent research skills during the course?

This was a multiple choice question and teachers were asked to tick the most relevant option (explained in the results section).

Teacher respondents were also asked to show their agreement with the following statement:

The course enables students to develop research and investigative skills that are useful in higher education

The response options to this question were: *strongly agree, agree, neither agree nor disagree, strongly disagree, and not applicable/don't know.*

In the student questionnaire, (in a question very similar to one in the teacher questionnaire,) respondents were asked to indicate the frequency

with which they were asked to carrying out an investigation/research as: (i) classroom activity, and (ii) homework/private study. Respondents were asked to tick the most relevant option from: *twice a week or more, once a week, once a fortnight, less often, and never.*

In another question, student respondents were asked:

If you are asked to carry out independent research in your subject, who or what would you consult? (e.g. the internet, TV, your parents, etc.)

This was an open-ended question. Students were provided with space for a free text response.

Three recruitment samples were identified for the survey: 200 schools and colleges following an A level Mathematics course; 100 schools and colleges following an A level Economics course, and 100 schools and colleges following an A level French course. (All three courses were administered by a major awarding body (OCR, 2011), who provided the research team with contact details.) There was no overlap between the three recruitment samples. Each sample was stratified in order to represent the full A level population for that course in terms of social deprivation, academic achievement and school/college type, as measured by indicators derived using data from the National Pupil Database.

Letters of invitation were sent to Heads of Department at the identified schools and colleges, explaining the key aims of the study and what participating would entail. Two copies of the appropriate teacher questionnaire and ten copies of the appropriate student questionnaire were enclosed, together with a pre-paid envelope in which questionnaires could be returned. The centres were encouraged to ask for additional questionnaires if needed and were given six weeks to complete and return the questionnaires.

Interviews

Following the survey, interview schedules for teachers and students were developed for probing more deeply into some issues which were considered to be in need of clarification or further exploration. A further purpose was to corroborate the questionnaire results. The schedules were piloted on one teacher and one student and a final version was produced which incorporated feedback from the pilot. All the teachers and students who had consented in their questionnaires to be contacted for a further phase of the project were identified. Two teachers (one teaching Economics and the other French) and nine students (four studying Economics and five studying French) were selected at random and were interviewed about their experiences. All interviews were carried out in interviewees' own schools and colleges by experienced interviewers. Each teacher interview took an hour to complete and each student interview lasted 30–40 minutes. Each interviewee completed a further consent form before beginning the interview. The interviews were recorded and were subsequently transcribed.

Results

Characteristics of questionnaire respondents

Forty Mathematics departments responded to the survey, returning 47 teacher questionnaires and 299 student questionnaires. Eighteen Economics departments responded, returning 24 teacher questionnaires and 228 student questionnaires. Twenty-one French departments responded to the survey, returning 15 teacher questionnaires and 136 student questionnaires. To ascertain the representativeness of the

data collected, key background characteristics of the responding schools and colleges were compared with those of the full populations of schools and colleges from which the recruitment samples were originally drawn. In terms of performance on examinations and economic deprivation (as determined by the postcodes of the responding schools and colleges), the samples of respondents were broadly representative of the full populations.

Table 1 enables comparisons to be made between the school/college types of the responding samples and the full A level populations for the three subjects. It can be seen that the sample of responding Mathematics schools/colleges seems broadly representative of the full population in this respect, although it comprises slightly smaller proportions of independent and selective state schools. The responding Economics sample comprised no Further Education colleges, proportionately more sixth-form colleges, and proportionately fewer independent schools than the full A level Economics population. The sample of responding French centres comprised no sixth-form colleges, proportionately more selective state schools, and proportionately fewer independent schools than the full A level Economics population. However, given the sample sizes (Mathematics = 40; Economics = 18; French = 21) a few differences of this kind are unsurprising. The sample was deemed sufficiently representative of the full population to be useful.

Table 1: Types of centres responding to the questionnaire

School/college type	Mathematics		Economics		French	
	% of responding schools/colleges	% of full A level Mathematics population	% of responding schools/colleges	% of full A level Economics population	% of responding schools/colleges	% of full A level French population
FE College	5	3	0	9	5	3
Sixth-form College	8	9	22	12	0	4
Independent	15	24	22	30	14	27
Selective state	15	9	11	8	29	13
Comprehensive	57	55	45	41	52	53

Data from teachers

The frequency with which students are asked to carry out investigation/research in the classroom and as homework/private study – teachers' responses

Table 2 and Figure 1 show that about half of the responding Economics and French teachers were setting investigation/research as a classroom task once a week to once a fortnight. In comparison, the Mathematics teachers reported setting this task less often for their students. While there are examples of good practice for engaging students in independent research/study in Mathematics (Hernandes-Martinez *et al.*, 2011), it appears that such activity may be relatively uncommon.

Table 3 and Figure 2 show that trends in the provision of opportunities to carry out investigation/research as homework or private study are similar to those in the classroom context. Once again, the Economics and French teachers were asking students to carry out more independent research than were the Mathematics teachers. It should be noted that

Table 2: The frequency with which students are asked to carry out investigation/research in the classroom – teachers’ responses

	Mathematics (N=47)	Economics (N=24)	French (N=15)
Number of responses	43	24	14
Twice a week or more	0	1 (4.2%)	0
Once a week	0	2 (8.3%)	4 (28.6%)
Once a fortnight	2 (4.7%)	8 (33.3%)	3 (21.4%)
Less often	24 (55.8%)	12 (50.0%)	6 (42.9%)
Never	17 (39.5%)	1 (4.2%)	1 (7.1%)

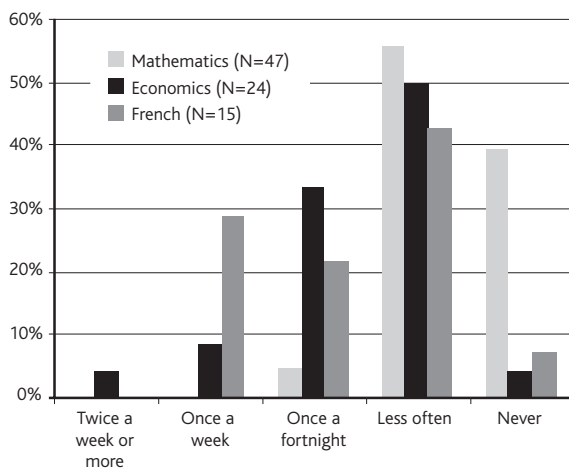


Figure 1: The frequency with which students are asked to carry out investigation/research in the classroom – teachers’ responses

although students were being asked to carry out independent research both in the class and at home, the teachers presumably had less control over how the students were carrying out assigned tasks at home and to what extent they were carrying them out entirely independently.

Teachers’ views on the extent to which students are required to utilise independent research skills during their A level courses

Table 4 and Figure 3 show that the majority of the French teachers felt that their course offered some opportunity to develop independent research skills. More than half of the Economics teachers also took this view. However, the majority of the Mathematics teachers felt either that their Mathematics course did not require an independent approach on the part of the students, or that little emphasis was put on independent research skills. These data corresponded well with the teachers’ responses to the previous question. It is also interesting to note that one Mathematics teacher and two Economics teachers believed that their course fostered independent research throughout its entirety (Table 1). Although in general, teachers felt that Mathematics does not appear to lend itself that easily to the development of independent research skills, there were a few Mathematics teachers who disagreed.

Teachers’ views about whether their courses enable students to develop research and investigative skills that are useful in higher education

The teachers’ responses in Table 5 and Figure 4 show that of the three subjects, French teachers agreed the most with the view that their course

Table 3: The frequency with which students are asked to carry out investigation/research as homework/private study – teachers’ responses

	Mathematics (N=47)	Economics (N=24)	French (N=15)
Number of responses	41	19	13
Twice a week or more	0	0	0
Once a week	1 (2.4%)	4 (21.1%)	3 (23.1%)
Once a fortnight	2 (4.9%)	6 (31.6%)	4 (30.8%)
Less often	19 (46.3%)	8 (42.1%)	6 (46.2%)
Never	19 (46.3%)	1 (5.3%)	0

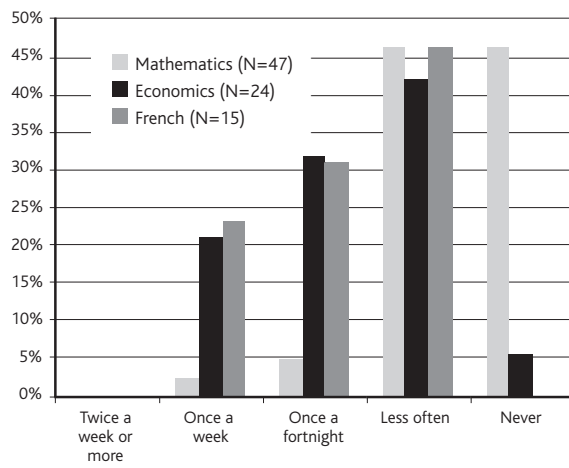


Figure 2: The frequency with which students are asked to carry out investigation/research as homework/private study – teachers’ responses

enables students to develop research and investigative skills that are useful in higher education. In contrast, the majority of Mathematics teachers disagreed with the view. Of those teachers who disagreed, it is not possible to identify whether the cause of their disagreement was

Table 4: The extent to which the students are required to utilise independent research skills during the course – teachers’ responses

	Mathematics (N=47)	Economics (N=24)	French (N=15)
Number of responses	46	23	15
A. The course fosters independent research throughout its entirety	1 (2.2%)	2 (8.7%)	0
B. The course offers some opportunity to develop independent research skills	5 (10.9%)	14 (60.9%)	14 (93.3%)
C. Much of the course does not require an independent approach on the part of the students	27 (58.7%)	6 (26.1%)	1 (6.7%)
D. Little emphasis is put on independent research skills in the course, and these skills are not significantly developed by the students	13 (28.3%)	1 (4.3%)	0

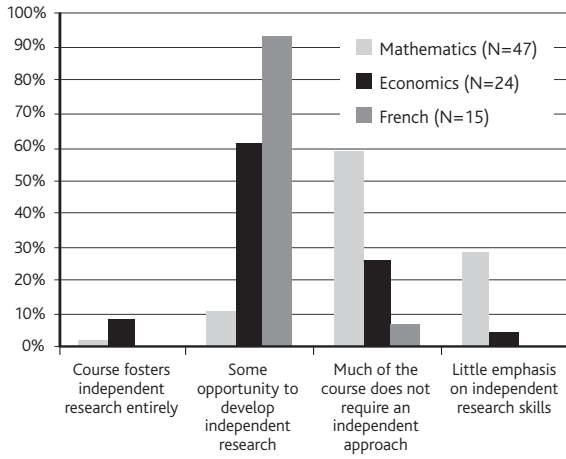


Figure 3: The extent to which the students are required to utilise independent research skills during the course – teachers' responses

Table 5: Teachers' views about whether their course enables students to develop research and investigative skills that are useful in higher education

	No. of responses	Strongly agree/Agree	Neither agree nor disagree	Strongly disagree/Disagree
Mathematics (N=47)	46	6 (13%)	12 (26.1%)	28 (60.9%)
Economics (N=24)	23	10 (43.5%)	7 (30.4%)	6 (26.1%)
French (N=15)	15	12 (80%)	1 (6.7%)	2 (13.3%)

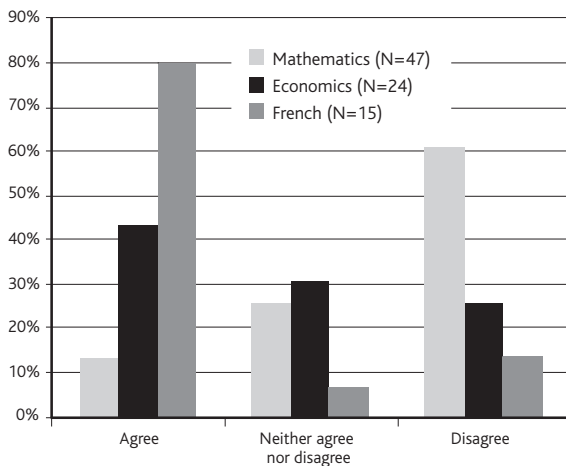


Figure 4: Teachers' views about whether their course enables students to develop research and investigative skills that are useful in higher education

about the course enabling the development of research and investigative skills or about those skills being useful in higher education. These two aspects need to be explored separately in future research.

During the interview stage of data collection, the Economics teacher who was interviewed indicated that while she did not get much chance to develop independent study skills, she tried to encourage her students to use different resources for information. Additionally, she also facilitated debates amongst her students to allow them to get used to uncertainty, that is, to become aware that many situations do not have one obvious correct answer and also to realise that different perspectives could exist in relation to the same situation.

The French teacher explained in the interview that in her school they

had a tutorial system in place to discuss study skills, organisational skills and independent learning with their students. She also explained that independent study skills were not treated as a separate entity and were an integral part of various activities, for instance, in students choosing what they wanted to read and also being encouraged to explore the latest news.

Overall, the teacher questionnaire and the interview data indicate that independent research is certainly taking place at A level, although to a greater extent in French and Economics than in Mathematics. It also suggests that students are being encouraged to develop independent research skills in a variety of contexts.

Data from students

The frequency with which students are required to carry out investigation/research in the classroom and at home – students' responses

Figures 5 and 6 show that more French and Economics students than Mathematics students responded that they carried out research, both in the classroom and as homework or private study. This finding is in line with the teachers' responses (see Figures 1 and 2). However, it is notable that more Mathematics students reported carrying out investigation and research (both in the class and as homework) than their teachers indicated. One possible explanation for this difference in students' and teachers' responses might relate to differences in teachers' and students'

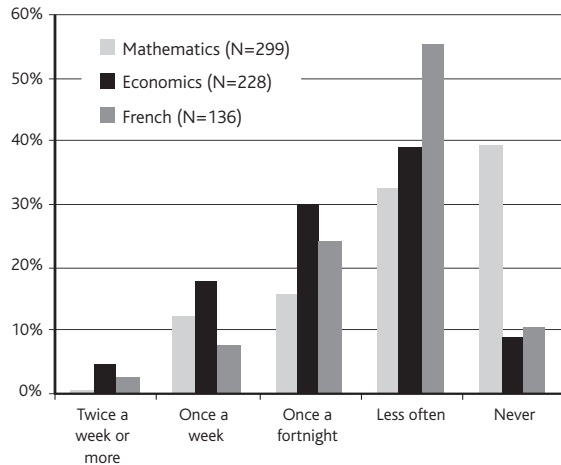


Figure 5: The frequency with which students are asked to carry out investigation/research in the classroom – students' responses

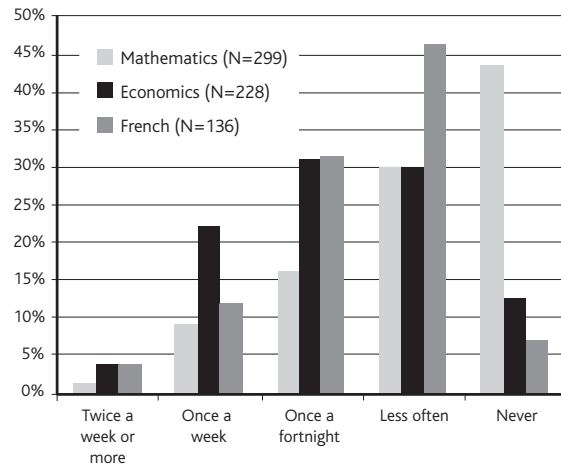


Figure 6: The frequency with which students are asked to carry out investigation/research as homework/private study – students' responses

understanding of what constitutes research and investigation. The teachers may have been more inclined to respond according to the number of times the task was set, whilst the students may have been influenced by how long it took them to complete tasks, especially if they carried out the work over a number of occasions. Since the questionnaire did not include questions exploring the differences in teachers' and students' definitions of independent research, it is not possible to provide a concrete explanation for this difference in their results. However, this strand could be explored in future work.

Sources consulted for information while carrying out independent research – students' responses

Table 6 shows that across all three subjects, the internet was the most widely used resource that students reported consulting if asked to carry out independent research. However, the internet is an extremely broad area and the students did not really explain how they were using the internet to facilitate their independent research. Table 6 also shows that a greater percentage of Mathematics students reported consulting their teacher in comparison with Economics and French students. Some students across all three subjects reported seeking help from friends or family. A greater percentage of French students indicated referring to magazines compared to Mathematics and French students. A greater percentage of Economics students than French students mentioned referring to the newspaper. Mathematics students, on the other hand, did not appear to use this source at all. A few students studying Mathematics

Table 6: The sources students consult for information while carrying out independent research

	Mathematics (N=299)	Economics (N=228)	French (N=136)
Number of responses	267	220	135
Internet	217 (81.2%)	213 (96.8%)	132 (97.7%)
Teacher/tutor	65 (24.3%)	14 (6.3%)	10 (7.4%)
Textbooks	64 (23.9%)	55 (25%)	23 (17%)
Other books	44 (16.4%)	45 (20.4%)	41 (30.3%)
Family	42 (15.7%)	28 (12.7%)	27 (20%)
Friends/peers	33 (12.3%)	15 (6.8%)	10 (7.4%)
Library	15 (5.6%)	10 (4.5%)	11 (8.1%)
Notes/Revision material	15 (5.6%)	8 (3.6%)	6 (4.4%)
Television/videos	7 (2.6%)	23 (10.4%)	17 (12.5%)
Exercise books	4 (1.4%)	0	0
Intranet	2 (0.7%)	1 (0.4%)	0
Magazines/articles	1 (0.3%)	23 (10.4%)	22 (16.2%)
Work	0	0	1 (0.7%)
Newspaper	0	47 (21.3%)	17 (12.5%)
Not asked	5 (1.8%)	0	0
Don't do	8 (2.9%)	0	1 (0.7%)
Don't know	2 (0.7%)	0	0

and one French student said that they did not do any independent research or were not asked to carry out this task.

The interview data corroborated the questionnaire results in terms of the various sources of information consulted. In the interviews the students frequently mentioned using the internet. However, a wide variety of usage from the very general to the specific was reported. For example:

I'd just Google it! And probably go on, like a bite size website or something like that. Wikipedia's always good as well. (Economics student)

The internet.... I've found the BBC website very useful for Economics – its business section. (Economics student)

When asked if the students used books to gather information while carrying out independent research, one student said:

Not really... we don't use books, we find the internet is an easier resource to use. (Economics student)

One student explained why he sought information from a relative:

I sometimes ask my cousin stuff. My older cousin, depending on what it is really. If it's more political-based, I'll just speak to my cousin about it, sometimes if I don't fully understand stuff, because he's like got a 2.1 from Cambridge in law and thinks he's God's gift. But he's (useful) sometimes. (Economics student)

In the interviews, the students were asked to define independent study skills. The following examples illustrate the variety in students' understanding of independent study:

Being able to find out information on your own but if you are struggling you should be able to ask someone else who should be able to, but not be dependant on them. Going to people and saying can you help me. (Economics student)

Just being able to work out what you need to do and to find useful information for exams. (French student)

I would say just to understand how to be in charge with your own learning. (French student)

The ability to ask further questions, gaining more knowledge further depth on a subject of your own accord, so that you deepen your understanding of a topic. (French student)

Discussion

In this study, the frequency with which Mathematics, Economics, and French teachers set independent research tasks for their students was explored. Teachers' opinions about the extent to which they thought their particular course developed students' independent research skills were also explored. Additionally, the frequency with which Mathematics, Economics, and French students thought they carried out independent research was investigated, as were the resources they consulted while doing so. Cross-subject comparisons were made.

Limitations

The study has certain limitations that are commonly associated with self-reported data. The primary weaknesses of a survey method include: multiple interpretations of the same question; limited/incomplete recall

of information; and responding to the questions based on certain preconceived expectations (Cohen *et al.* 2000). In order to corroborate the questionnaire results, a number of face-to-face interviews were conducted with teachers and students. However, the interview sample was small and may not have been representative of the sample that responded to the questionnaire. The other main limitation of this study is that it explored the development of independent research in courses offered by a single awarding body. Therefore, the sample in this study may not be representative of teachers and students of other A level courses and awarding bodies. Finally, this study is based on a small section of a wider project on teaching and learning at A level. Therefore, resources were insufficient for it to explore all the various aspects in relation to independent research.

Summary of main findings

A significant quantity of data was collected. A total of 47 Mathematics teachers, 24 Economics teachers and 15 French teachers responded to the teachers' survey. Additionally, 299 Mathematics students, 228 Economics students and 136 French students responded to the students' survey. In terms of performance in examinations, economic deprivation, and the centre type of the responding schools/colleges, the sample was sufficiently representative of the full population.

With regard to the frequency with which teachers set investigation/research tasks for their students in the classroom context, it was found that about half of the French and Economics teachers were assigning such tasks to their students from once a week to once a fortnight. On the other hand, about half of the Mathematics teachers were setting such tasks less often and about 40% of them never set investigation/research tasks. The frequency with which the teachers set investigation/research tasks as homework/private study showed the same subject-specific differences as the classroom context.

It was found that the majority of the French teachers felt that their course offered some opportunity to develop independent research skills. More than half of the Economics teachers also held the same view of their course. However, Mathematics teachers differed in that more than half of them felt that much of their course did not require an independent approach on the part of the students. The majority of the French teachers agreed with the statement that their course enabled students to develop research and investigative skills that are useful in higher education. A little less than 50% of the Economics teachers agreed with the same view, followed by only 13% of the Mathematics teachers.

This article also explored the extent to which students thought they conducted independent research. Reassuringly for the validity of the data, the same subject-specific differences emerged among the students' responses as among the teachers' responses. More French and Economics students in comparison to the Mathematics students stated that they carried out investigation/research both in class and at home. These subject-specific differences might be the result of actual difference in practice, or could be due to differences in the respondents' understanding of the survey questions.

The internet was the most frequently listed source that students across all three subjects consulted while engaging in independent research. The interview data shed further light on the general and specific usage of the internet. The survey data indicated a few subject-specific differences in the resources used for independent research. For instance, magazines and newspapers were listed by a few Economics and French students, but were not used by Mathematics students.

Conclusions and further research

To conclude, the findings from this study shed some light on the variation in preparedness of 16 to 18 year old students for independent study and research-related tasks at university. This study could help teachers to reflect upon their teaching and learning practice, and to identify whether there is scope or need for independent study and research in the A level courses they teach, and how this might be incorporated within their teaching and assessment strategies. Porkess *et al.* (2011) suggest that Mathematics teaching requires better implementation of the syllabus. This might involve getting students to think further and more in depth, for instance, through independent research. Suto *et al.* (2011) have found that significantly more teachers teaching in schools with relatively high A level Mathematics results than with relatively low results believe teaching beyond the syllabus strengthens and expands students' existing knowledge and adds interest to the course. Further work in this area with larger samples could also explore the relationship between independent research and school performance.

This study forms a small section of a wider project on teaching and learning at A level. It would be useful to carry out further research focussing exclusively on independent research; exploring, for instance, teachers' views of the effectiveness of independent research at A level, particularly when carried out at home. One Mathematics teacher and two Economics teachers claimed that their courses fostered independent research throughout their entirety. Further work could be carried out with these teachers to understand the elements of the course and their teaching that facilitate development of independent research. The interview data with teachers indicated that independent research skills were being developed in a variety of school contexts. Therefore, further work on independent research and study could be explored within a wider school context. This study included A level courses from only one awarding body; it could be replicated with courses administered by different awarding bodies to determine if the same trends emerge as in this study.

Some interesting differences emerged in the frequency of independent research reported by the Mathematics teachers and their students. Future work exploring students' and teachers' understanding and their definitions of independent research, investigation, and related terms might shed further light on such differences in self-reported data. While this study indicates that teachers across three very different subjects provided opportunity for independent research, additional work needs to be carried out to understand the practical challenges of facilitating independent research/study. If it were found that more teachers would like their students to participate in independent research, but that the syllabus or focus of the assessment objectives prevents them from engaging their students with this task, then the qualification developers would need to determine how this could be addressed.

An issue related to the effectiveness of independent research is the ability of the teachers to facilitate it. Although some teachers might be extremely competent in delivering the content of their subject, they might be less able to supervise and guide their students' research activities. This problem may be exaggerated in cases where non-specialists are responsible for teaching a particular subject. While this study gathered data about how often teachers were assigning independent research tasks to their students and whether they felt that their courses facilitated this aspect, the study information on the nature of those independent research tasks was not obtained. Additional work that looks at the strengths and limitations of independent research that is embedded in staple courses such as A levels, relative to dedicated research routes such as EPQ, would also be worthwhile.

References

- ATC21S (2011). Assessment and teaching of 21st Century skills website. Retrieved from: <http://atc21s.org>
- Anderson, L. W. & Krathwohl, D. R. (Eds.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman.
- AQA (2011). The Assessment and Qualifications Alliance website. Retrieved from: <http://web.aqa.org.uk/>
- Bloom, B. S. (1956). *Taxonomy of educational objectives: the classification of educational goals*. Cognitive Domain. London: Longman Group.
- British Science Association (2011). The British Science Association website. Retrieved from: <http://www.britishsociety.org/web/>
- Cohen, L., Manion, L. & Morrison, K. (Eds.) (2000). *Research Methods in Education*. 5th ed. London: Routledge Falmer.
- DfES (2005). 14-19 Education and Skills white paper. Retrieved from: <https://www.education.gov.uk/publications/standard/publicationDetail/Page1/CM%206476>
- Hernandez-Martinez, P., Williams, J., Black, L., Davis, P., Pampaka, M. & Wake, G. (2011). Mathematics coursework as facilitator of formative assessment, student-centred activity and understanding. *Research in Mathematics Education*, **13**, 2, 197–212.
- IBO (2011). The International Baccalaureate Organisation website. Retrieved from: <http://www.ibo.org/>
- Kajander, A. & Lovric, M. (2005). Transition from secondary to tertiary mathematics: McMaster University Experience. *International Journal of Mathematical Education in Science and Technology*, **36**, 2, 149–160.
- Lovitts, B. E. (2005). Being a good course-taker is not enough: A theoretical perspective on the transition to independent research. *Studies in Higher Education*, **30**, 2, 137–154.
- Marland, M. (2003). The transition from school to university: Who prepares whom, when, and how? *Arts and Humanities in Higher Education*, **2**, 2, 201–211.
- McClure, R. & Clink, K. (2009). How do you know that?: An investigation of student research practices in the digital age. *Libraries and the Academy*, **9**, 115–132.
- Mehta, S., Suto, I., Elliott, G. and Rushton, N. (2011, June). *Why study Economics? Perspectives from 16 to 19 year old students*. Paper presented at International Association for Citizenship, Social, and Economics Education Conference, Bath, UK.
- O'Shea, J. (2010). Helping students make the transition from school to university: The National University of Ireland-Maynooth Experience, The Higher Education Academy, Engineering Subject Centre. Retrieved from: http://www.engsc.ac.uk/downloads/helm/proceedings/session7a_oshea.doc
- O'Sullivan, M. K. & Dallas, K. B. (2010). A collaborative approach to implementing 21st Century skills in a High school senior class. *Educational Libraries*, **33**, 3–9.
- OCR (2011). *Extended Project Qualification*. Retrieved from: <http://www.ocr.org.uk/qualifications/type/projects/projects/extended/index.html>
- OCR (2011). OCR website. Retrieved from: <http://www.ocr.org.uk/>
- Ofqual (2011). Office of Qualifications and Examinations Regulation website. Retrieved from: <http://www.ofqual.gov.uk/>
- Owen, P. (2010). A transition checklist for high school seniors. *School Library Monthly*, **26**, 8, 20–23.
- Porkess, R., Vorderman, C., Budd, C., Dunne, R., & Hart, M (2011). A world-class Mathematics education for all our young people. Retrieved from: http://www.conservatives.com/News/News_stories/2011/08/-/media/Files/Downloadable%20Files/Vorderman%20maths%20report.ashx
- Rees, A. & Wilkinson, M. (2008). Scientific communication skills: The transition from Further Education to Higher Education in the UK. *Journal of College Teaching and Learning*, **5**, 9, 33–40.
- Richardson, H. (2010, November). 209,000 university hopefuls miss out on degree places. BBC News online. Retrieved from <http://www.bbc.co.uk/news/education-11765127>
- Rushton, N., Suto, I., Elliott, G. & Mehta, S. (2011, September). *Small is Beautiful? An exploration of class size at A level*. Paper presented at the British Educational Research Association annual conference, London, UK.
- Salisbury, F. & Karasmanis, S. (2011). Are they ready? Exploring student information literacy skills in the transition from secondary to tertiary education. *Australian Academic and Research Libraries*, **42**, 43–58.
- Silva, E. (2009). Measuring skills for 21st-century learning. *Phi Delta Kappa*, **90**, 9, 630–634.
- Smith, K. (2004). School to University: An investigation into the experience of first-year students of English at British Universities. *Arts and Humanities in Higher Education*, **3**, 81–93.
- Stohl, C. (2010). Inquiry in the classroom: An Action Research Report. Retrieved from: <http://www.eric.ed.gov/PDFS/ED509645.pdf>
- Suto, I., Elliott, G., Rushton, N. & Mehta, S. (2011, September). *Going beyond the syllabus: A study of A level Mathematics teachers and students*. Paper presented at British Educational Research Association annual conference, London, UK.
- Vidal Rodeiro, C. L. (2007). A level subject choice in England: Patterns of uptake and factors affecting subject preferences. Cambridge: Cambridge Assessment. Retrieved from: http://www.cambridgeassessment.org.uk/ca/digitalAssets/114189_Survey_Report_Final.pdf
- Wingate, U. (2006). Doing away with 'study skills'. *Teaching in Higher Education*, **11**, 457–469.

Assessment for Learning in International Contexts (ALIC): understanding values and practices across diverse contexts

Stuart Shaw CIE Research, Martin Johnson Research Division and Paul Warwick University of Cambridge, Faculty of Education

Conceptualising Assessment for Learning

Assessment for Learning (AfL) has been characterised as 'not a test but a process' (Popham, 2008, p.6), focused on providing qualitative insights into student understanding (Shepherd, 2008; Black and Wiliam, 1998). The Assessment Reform Group summarise AfL as: "the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (ARG, 2002). The involvement of engaged, reflective professional teachers is seen as central to the development of classroom-based assessment practices that are the foundation of AfL (Black, McCormick, James and Pedder, 2006) and these sentiments cohere with the learner centred approach found in the teacher development programmes offered by Cambridge International Examinations (<http://www.cie.org.uk/aboutcie>).

The language of AfL belongs to a ubiquitous educational discourse, being used across diverse social, economic and cultural boundaries (Swaffield, 2011). It is either seen as synonymous with formative assessment, and thus includes such practices as targeted observation or marking of work by teachers to develop students' next steps in learning (Wiliam, Lee, Harrison, and Black, 2004; James and Pedder, 2006); or it is seen as describing only those components of formative assessment that focus on students' involvement in their own learning. Here, we use the term as synonymous with formative assessment. Black and Wiliam (2009) conceptualise formative assessment as consisting of five key strategies, intended to provide contingent information upon which both teachers and students can act to progress student learning. These are:

1. *Clarifying and sharing learning intentions and criteria for success;*
2. *Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding;*
3. *Providing feedback that moves learners forward;*
4. *Activating students as instructional resources for one another; and*
5. *Activating students as the owners of their own learning.*

(Black and Wiliam, 2009, p.8)

Towards an understanding of international classroom practice

Despite the apparent clarity of pedagogic intentions and associated strategies, AfL classroom practices vary across Western educational contexts. Both Black and Wiliam (2005) and Sebba (2006) point to the differing policies, politics and cultures, both micro- and macro-, that impact upon classroom assessment practices. These influences include

the nature of national curricula, the presence or otherwise of selective education and the prominence given to the outcomes of summative assessments within a society. Black and Wiliam (2005) indicate that, even within what might superficially be seen as relatively homogenous national systems, regional and local variation in teaching and assessment practices is clear. For example, the influence of schools boards in various States in America means that a State-wide consensus on such issues as classroom-based approaches to assessment can be difficult to evidence.

When considering non-Western contexts, differing policies, politics and cultures are as likely to be important determinants of pedagogical approaches and classroom-based assessment practices as in Western contexts. With diverse national and regional educational priorities, and the different languages within which educational ideas are interpreted, comes another layer of complexity. In such international contexts, therefore, it is unsurprising to find that the development and embedding of successful assessment for learning practices seems to differ (Johnson and Burdett 2010; Akyeamong, Pryor, and Ampiah, 2006). Evidence of this comes from Johnson and Burdett's study which highlighted that, internationally, the ambitions of educators to engage with assessment for learning principles might be hindered by factors such as teacher competency levels or the promotion of conflicting theories of learning. Acknowledging the differences of AfL interpretation and practice within Western and non-Western contexts also raises the important spectre of those differences being present between Western and non-Western contexts.

Professional meaning making in relation to learning and assessment therefore seems inextricably linked to the social context in which such interpretations are based, and 'commonly used' language can be open to interpretation across different contexts (Smith, 1995). It may be, therefore, that the seemingly ubiquitous nature of the language of formative assessment within international educational discourse masks a poor shared understanding of the underlying meanings around such phraseology. Thus, if "differences between cultures are greater than those within" and "...concepts assumed to be universally understood were found to have contextually located meanings" (Andrews, 2007, p.490 and pp.495–496), then differing cultures may ascribe different levels of value to the strategies associated with AfL, and may evidence these differing values through differing classroom practices. A useful research contribution to this area of understanding would thus be to focus on eliciting the valued assessment practices held by teachers.

When examining the issue of values and practices in the UK, and possible gaps between the two, the Learning How to Learn Project surveyed 558 teachers in England (James and Pedder, 2006; Pedder, 2006). Reflecting earlier work by Torrance and Pryor (1998), James and

Pedder (2006, p.119) suggest that items in their survey relate to four themes:

- 'convergent assessment tendencies' (where there is an emphasis on linear and curriculum-oriented planning, and the use of closed questioning and summative feedback);
- 'divergent assessment approaches' (where students can take forward their own learning objectives and peer assessment practices are used);
- the promotion of guided self-assessment and opportunities for students to assess their own work and learning;
- teachers learning more about their students' learning.

James and Pedder's results revealed three underlying dimensions of assessment practice. These were:

- Making learning explicit* (defined as eliciting, clarifying and responding to evidence of learning; working with students to develop a positive learning orientation).
- Promoting learning autonomy* (defined as a widening of scope for students to take on greater independence over their learning objectives and the assessment of their own and each other's work).
- Performance orientation* (defined as a concern to help students comply with performance goals prescribed by the curriculum through closed questioning and measured by marks and grades).

James and Pedder found sizeable values-practice gaps on two dimensions that appear to be in tension (*promoting learning autonomy and performance orientation*), along with evidence that over half of the sample were unable to sustain practices across all dimensions in line with their values. Further evidence of the existence of three dimensions of assessment practice, and the presence of values-practice gaps, was found by Winterbottom *et al.* (2008a, b) when they used the Learning How to Learn survey tool with English teacher trainees. These values-practice gaps are particularly interesting and are a specific focus of this study.

Research questions

The Assessment for Learning in International Contexts (ALIC) project extends earlier understandings around AfL through using an adapted version of the James and Pedder (2006) survey tool with teachers across different international contexts. The following questions informed the ALIC project:

- Which assessment practices seem to be valued by teachers in national contexts other than the UK?
- To what extent does there seem to be congruence in the assessment values and practices of these teachers?
- Do teachers working in different national contexts evidence the importance of the same dimensions of assessment practice as those found in studies in England? (This question is not a focus of this article.)

It should be stated that there is not a presumption of a 'model classroom' sitting behind the research questions. They are broadly framed in order to be able to build a picture of the range of values and practices that seem to be important to teachers across diverse national contexts.

ALIC project methods and sample

The ALIC project gathered survey data from teachers working in national contexts that differed in terms of their linguistic and historical educational traditions. Sample selection and recruitment took into account a number of pragmatic considerations. Teacher recruitment for the project was maximised by focusing on teachers working in schools and colleges with a strong identification with Cambridge International Examinations (CIE). This focus also added a practical importance to the project outcomes. Through generating insights into the localised practices and values of teachers working in schools and colleges, the research project would enable 'feed-forward' to CIE's ongoing programme of teacher professional development provision.

The project used CIE's regional organisation structure to aid data gathering. Nations with the greatest number of schools and colleges with active CIE links were identified across each of CIE's five global regions, with a decision made not to recruit nations from the same region. This helped to maximise the geographical diversity of the sample and to potentially maximise the number of returns. This sampling approach suggested that the project should focus on teachers in Argentina, India, Indonesia, Nigeria and Saudi Arabia. An appeal for participation from at least two teachers from each approached school or college was intended to bring a sense of collegiality to the process for individual teachers, since it was anticipated that there might be a future opportunity to build a community of teachers around shared professional discussions through involvement with this project.

The ALIC project builds on the work of James and Pedder (2006), which had used a validated survey to explore the assessment values and practices of teachers in the UK. In electing to work with an existing survey instrument, the ALIC research team considered whether the James and Pedder questionnaire was sufficiently relevant to the ALIC research questions, whether it was appropriate to use in the different international contexts, and whether it facilitated collection of this information with maximal reliability and validity. Here, reliability can be understood to be the extent to which a measure – the underlying variable(s) of interest – is stable or consistent and produces similar results when administered repeatedly. This is of special value to the ALIC project, given that measurements are taken in different national contexts.

Whilst the use of validated methods (Alderson, 1992; Hawkey, 2006) should contribute positively to the validity of a research design, it is important to bear in mind that validation is context specific and has consequences if a research method is applied to a situation for which it was not designed. The ALIC project took the constructs that underpinned the original James and Pedder (2006) teacher survey and ensured that these were accessible to teachers working across a variety of national contexts. A critical review of each of the James and Pedder survey items was undertaken to ensure that the language of the survey (both the instructions accompanying the survey and the survey items themselves) was accessible to teachers for whom English may not necessarily be a first language. This involved an iterative process of discussion between the research team members. An original and a revised item are illustrated in Figure 1.

A draft of the ALIC survey was piloted with a small group of teachers in some of the sample nations in order to validate its format. Once it was complete, the survey was distributed via a dedicated website to schools and colleges in the five sample nations. 613 schools and colleges were contacted directly in three of the five study nations (Argentina: 186

Figure 1: An example of an original and a revised teacher survey item

Scale X Your assessment practices (About You)				Assessment practices	Scale Y How important are assessment practices for creating opportunities for students to learn? (About your values)			
Original Survey Item								
Never true	Rarely true	Often true	Mostly true		Not at all importance	Of limited importance	Important	Crucial
				The next lesson I teach is determined more by the prescribed curriculum than by how well my students did in the last lesson.				
Revised Survey Item								
Never true	Rarely true	Often true	Mostly true		Not at all importance	Of limited importance	Important	Crucial
				The subject curriculum I have to teach is a greater influence on what I will do in my next lesson than how well my students did in the last lesson.				

schools/colleges; India: 288 schools/colleges; Indonesia: 135 schools/colleges). Taking into consideration local arrangements in Saudi Arabia and Nigeria, indirect contacts were sent to schools and colleges through British Council offices.

The first data analysis stage involved descriptive analysis of the survey return data, and it is this phase that is reported on in this article. In order to explore comparisons between teachers' values and practices, a gap analysis compared the extent to which teachers' reported practices matched their reported values; any discrepancies were thus indicated between their professional assessment aspirations and their actual practices. The second data analysis stage of the ALIC project will be to replicate the statistical methods used by James and Pedder (2006) and Pedder (2006) in their work with teachers in the UK (not reported in this article).

Findings

Teacher demographics

Two hundred and forty two ALIC surveys were returned, with five containing no indication of teacher nationality. The data in Table 1 show that most teachers who returned the surveys were female (69%), had more than 5 years of teaching experience (83%), and were teaching 15–18 year old students (62%). There was a spread of subjects taught by teachers in the sample, although Science/Maths and English teachers made up the majority of the sample (67%). It is worth noting that the initial process of 'teacher subject' coding defined those teachers who taught multiple subjects as 'not specified', partly explaining the relatively large number of teachers who appear in this category.

The survey return rate differed for each nation (i.e. the proportion of schools and colleges from which surveys were received compared with the number of schools and colleges approached). This national difference might reflect the national variation in the methods used to approach the schools and colleges. Figure 2 shows that Indian teachers submitted the

Table 1: ALIC Survey Participant Data

Teacher Gender	N	%
Male	70	28.9
Female	166	68.6
Not specified	6	2.5
Teacher Experience		
Less than 2 Years	2	0.8
2–4 Years	20	8.3
5–10 Years	65	26.9
11–20 Years	84	34.7
21+ Years	52	21.5
Not specified	19	7.9
Experience in Current School		
Less than 2 Years	51	21.1
2–4 Years	52	21.5
5–10 Years	73	30.2
11–20 Years	30	12.4
21+ Years	11	4.5
Not specified	25	10.3
Age Taught		
10 and Under	7	2.9
11–14	60	24.8
15–18	151	62.4
18+	2	0.8
Not specified	22	9.1
Subject Taught		
Science/Maths	85	35.1
English	77	31.8
Languages	2	0.8
Social Sciences/Humanities	32	13.2
Arts	3	1.2
Not specified	43	17.8
Total	242	100

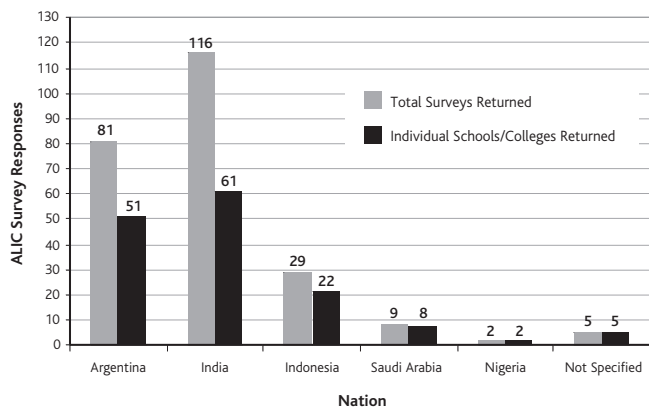


Figure 2: ALIC Returned Responses by Nation and School/College

greatest number of survey returns (a 21.2% return rate), followed by Argentina (a 27.4% return rate) and Indonesia (a 16.3% return rate).

The national survey data (Appendix 1) demonstrate variances in the profile of teacher demographics. Teachers from Argentina and Saudi Arabia were the most experienced; a majority of teachers in both nations had more than 10 years of teaching experience. The length of time that teachers had worked in their current school/college also differed across the nations. India was the only nation where the majority of teachers had worked in their current school/college for less than five years. The profile of subjects taught differed across the teachers in the different sampled nations. Teachers of English formed the largest group of respondents in Argentina, contrasting with the profile of teachers from the other nations where Science/Maths teachers formed the largest group.

Comparisons between teachers' values and practices

This article focuses largely on considering teacher values and on a values-practice gap analysis across the whole teacher cohort in general and for Indian and Argentinean teachers specifically. No attempt is made in this article to compare the data with that from the UK.

The values-practice gap analysis data show the level of match between what teachers value about their assessment work and the extent to which they feel they enact these values in practice. Data analysis looked at the comparison between those assessment practices that the teachers thought were 'important/crucial' against those that they reported using 'often/mostly'. Where there were any mismatches between assessment values and practices, a positive difference suggests that the teachers value the assessment practice more than they actually enact it. On the other hand, a negative mismatch suggests that the teachers were enacting practices that they did not value.

Findings across the data set

The data in Appendix 2 show that two-thirds of classroom assessment practices listed in the survey were highly valued by a majority of responding teachers, with 20 of the 30 survey items being considered to be 'important/crucial' for at least 88% of the surveyed teachers. The data also show that there were seven practices that were highly valued by less than a quarter of the sampled teachers.

Of the highly valued practices across the whole ALIC teacher cohort data, 10 items relate to teachers' concern with understanding more about student learning. Of these items, some relate to using evidence of learning to influence planning (Item 1), and using open questioning, encouraging discussion, clarifying learning objectives, lesson purposes and success criteria (Items 11, 18, 21, 25 and 28). These items clearly link with the first

and second of Black and Wiliam's AFL Strategies (*Clarifying and sharing learning intentions and criteria for success, and Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding*). Some items relate to providing formative feedback to respond to evidence of learning and encourage pupil involvement in learning (Items 4, 10, 15 and 20) and link with Black and Wiliam's third AFL Strategy (*Providing feedback that moves learners forward*). Item 22 ('Assessment of students' work is mainly in the form of comments') might be considered to be linked to these items, but it is not given the same value by teachers.

Teachers also placed a very high value on practices relating to the development of pupil agency in assessment and learning. This coheres with Black and Wiliam's fifth AFL Strategy (*Activating students as the owners of their own learning*). These items are connected to such things as providing opportunities for students to assess their own work and learning (Items 13, 14 and 24) and develop independence in learning (Item 9); a concern that students should engage with mistakes and problems in their work (Items 15, 16 and 25), should build on their strengths (Items 14 and 26) and should view effort as important (Item 27); and that students should be encouraged to think critically about their learning (Items 17 and 30).

Item 3 ('The main thing I look for in my assessments is whether my students know, understand or can do key sections of the curriculum') is also highly valued and is the only item that might be interpreted as sitting outside the items that can be linked with Black and Wiliam's AFL Strategies.

With respect to the least valued items, only one item fell below 50% in terms of being valued. This was Item 5 ('I tell students how well they have done compared to others in the class'), which emphasises the development of a competitive classroom ethos and a strong focus on performance orientation. Overall, items that were highly valued by a small number of teachers were those that might be linked to teacher control of assessment processes and a focus on performance goals. These included items associated with curriculum orientated planning (Items 2 and 23); closed questioning (Item 7); the provision of summative feedback, including marks and grades (Item 12); and the prioritising of teacher assessments (Item 8). A second group of items less valued by teachers were those associated with student control over assessment processes, including students taking forward their own learning objectives (Item 6) and developing peer assessment practices (Items 19 and 29).

Table 2 provides data derived from a comparison of teachers who placed a high value ('crucial'/'important') on a particular practice, against the percentage suggesting it was 'often true' or 'mostly true' in their own practice. These data only relate to items where the values-practices gap is of +/- 5 points or greater. A positive gap indicates that a practice is more valued than it is employed with students; a negative gap suggests a practice that is strongly used but is less in tune with teacher values. The data presented here focus on the most marked gaps.

The group of items (6, 19 and 29) associated with giving students more control over assessment processes was not particularly highly valued, and the largest positive gap is for Item 6 ('I give students the opportunity to determine their own learning objectives'), with Item 19 having a six point gap and Item 29 an eight point gap.

For those items that were more highly valued by teachers, there is an apparent gap between values and practices for items associated with the development of pupil agency. There are thus relatively large positive gaps between values and practices that link to promoting opportunities for students to assess their own work (Items 13 and 24), build on their

Table 2: Comparing ALIC teachers' assessment values and practices across five national contexts (only differences of +/- 5 points or greater are shown; data shown in highlight represents a negative values-practice gap)

Item		Values (%) important/ crucial	Practices (%) often/mostly	Values- Practices Gap
6	I give students the opportunity to determine their own learning objectives	73	55	+18
26	I help students to plan the next steps in their learning	88	71	+17
17	I help students to think about how they learn best	96	87	+9
29	I give students the opportunity to assess each other's work	73	65	+8
13	I give guidance to help my students assess their own work	94	86	+8
24	I give guidance to help students assess their own learning	94	86	+8
19	I give guidance to help students to assess one another's work	70	64	+6
22	Assessment of students' work is mainly in the form of comments	75	70	+5
14	I tell students about their strengths and help them to develop these strengths	98	93	+5
23	The subject curriculum determines students' learning objectives	81	88	-7
12	Assessment of students' work is mainly given as marks and grades	64	77	-13

strengths (Items 14 and 26) and think critically about their learning (Item 17).

With respect to the items with a negative gap, indicating well-used practices that are less in tune with teacher values, the largest gap occurs with respect to Item 12, the provision of feedback in the form of marks and grades.

Given that the survey return rates from Argentina and India were large enough to enable statistical manipulation, the data from these two countries can be examined in more detail.

Argentinean and Indian perspectives

The data in Appendix 3 show that 19 of the 30 classroom assessment practices listed in the survey were highly valued by a majority of responding teachers from Argentina.

The data also show that there were eight practices that were highly valued by fewer than three-quarters of the Argentinean teachers. Two items fell below 50% in terms of being valued – Item 5 ('I tell students how well they have done compared to others in the class') and Item 7 ('I use questions mainly to get factual knowledge from my students').

Table 3 shows the items where the level of congruence between the values and practices of the Argentinean teachers were least marked.

Items with the largest positive values-practice gap are associated with giving more control over assessment processes (Items 6, 19 and 29). However, practices seem to be well behind aspirations. Even where such practices are not particularly highly valued, as with Item 6 (see Table 3),

Table 3: Comparing Argentinean ALIC teachers' assessment values and practices (only differences of +/- 5 points or greater are shown; data shown in highlight represents a negative values-practice gap)

Item		Values (%) important/ crucial	Practices (%) often/mostly	Values- Practices Gap
26	I help students to plan the next steps in their learning	80	48	+32
6	I give students the opportunity to determine their own learning objectives	67	37	+30
17	I help students to think about how they learn best	96	79	+17
29	I give students the opportunity to assess each other's work	71	55	+16
24	I give guidance to help students assess their own learning	95	80	+15
13	I give guidance to help my students assess their own work	96	84	+12
19	I give guidance to help students to assess one another's work	64	53	+11
21	I help students to understand the learning purposes of each lesson or series of lessons	90	79	+11
22	Assessment of students' work is mainly in the form of comments	81	71	+10
8	My assessments are more useful than formal assessments	75	69	+6
15	I help students find ways of solving problems that they have in their learning	99	93	+6
16	I encourage students to see their mistakes as valuable learning opportunities	98	93	+5
14	I tell students about their strengths and help them to develop these strengths	99	94	+5
4	The feedback that my students get helps them improve	100	95	+5
30	I often talk to students about how they can improve their learning	100	95	+5
18	I use questions mainly so that my students give me reasons and explanations	88	93	-5
3	The main thing I look for in my assessments is whether my students know, understand or can do key sections of the curriculum	90	95	-5
23	The subject curriculum determines students' learning objectives	78	88	-10
7	I use questions mainly to get factual knowledge from my students	43	54	-11
12	Assessment of students' work is mainly given as marks and grades	56	85	-29

actual classroom practice seems to be well behind aspirations; in this case the gap is +30 points. Similarly, items associated with the development of pupil agency in assessment and learning (Items 13, 17, 24 and 26) show significant gaps between values and practices.

A number of items in Table 3 exhibit values-practice gaps worthy of some consideration (+/- 5 points or greater) including two items associated with providing formative feedback to respond to evidence of learning and to encourage pupil involvement in learning (Items 4, 15) and four additional items associated with the development of pupil agency in assessment and learning (Items 14, 15, 16 and 30).

Items 23 and 12 show evidence of a negative values-practice gap, indicating practices that are less in tune with teacher values. The gap for Item 12 ('Assessment of students' work is mainly given as marks and grades') is very large (-29 points). Other items falling into this negative gap category include Item 3, valued by 90% of Argentinean teachers and practised by 95%; Item 18, valued by 88% and practised by 93%; and Item 7, valued by 43% and practised by 54%.

The data in Appendix 4 show that 21 of the 30 classroom assessment practices listed in the survey were highly valued by a majority of responding teachers from India. Item 26 ('I help students to plan the next steps in their learning') exhibits a difference greater than +/- 5% and was highly valued by 96% of Indian teachers.

The data also show that there were seven practices that were highly valued by fewer than three-quarters of the Indian teachers. Item 5 ('I tell students how well they have done compared to others in the class') was firmly rooted at the bottom of all items in terms of the extent to which they are valued by teachers. Item 22 ('Assessment of students' work is mainly given in marks and grades') also falls into this group, being highly valued by 70% of teachers.

Table 4 shows the items where the level of congruence between the values and practices of the Indian teachers were least marked.

With respect to the items that exhibit a positive values-practice gap (indicating that a practice is more valued than it is employed with students); only 4 items are included here. The values-practice gap for all of these items is relatively small (i.e. between +5 and +8 points).

Table 4: Comparing Indian ALIC teachers' assessment values and practices (only differences of +/- 5 points or greater are shown; data shown in highlight represents a negative values-practice gap)

Item		Values (%) important/ crucial	Practices (%) often/mostly	Values- Practices Gap
2	The subject curriculum I have to teach is a greater influence on what I will do in my next lesson than how well my students did in the last lesson	64	56	+8
6	I give students the opportunity to determine their own learning objectives	78	70	+8
26	I help students to plan the next steps in their learning	96	89	+7
17	I help students to think about how they learn best	98	92	+6
12	Assessment of students' work is mainly given as marks and grades	64	71	-7

This group includes Item 2, relating to curriculum-orientated planning, and Item 6, relating to the development of student control over assessment practices. Item 2 is both practised and highly valued by relatively few teachers, whilst Item 6 is practised and highly valued by a greater percentage; in both cases the positive gap suggests an aspiration outstripping practice.

Item 12 is the only item to exhibit a negative values-practice gap for Indian teachers, indicating a practice that is not valued as much as it is employed with students (Table 4).

Discussion

It is important to say at the outset that the ALIC survey relied on self-reporting by participants. Unlike James and Pedder (2006), the ALIC team were unable to corroborate statements made in the survey through empirical sampling of teacher practices. And with respect to AfL strategies, others studies have found that teachers can be less confident than they claim to be in putting actual strategies in place (Sach, 2012). Nevertheless, if 'teachers' professional consciousness is a...fundamental determinant of teaching practices' (Yung, 2002), and if teachers' conceptions of learning are central to understanding and enacting assessment practices (Marshall and Drummond, 2006), then it is crucial to consider how they view their practices and to examine their aspirations for the future.

James and Pedder (2006) suggest that their original survey incorporated items relating to the themes of 'convergent assessment tendencies'; 'divergent assessment approaches'; the promotion of guided self-assessment and opportunities for students to assess their own work and learning; and teachers learning more about their students' learning.

In considering findings of the highly valued practices across the data set, 10 items relate to teachers' concern with learning more about student learning. Certainly, a concern with understanding student's learning, and acting upon that understanding, lies at the heart of the five key AfL strategies discussed at the start of this article. If "...formative assessment is concerned with the creation of, and capitalization upon, 'moments of contingency' in instruction for the purpose of the regulation of learning processes" (Black and William, 2009, p.10), then learning more about student's learning is vital. Yet in differing national contexts, what is considered to be an appropriate 'assessment repertoire' might include approaches that are not bounded by Black and William's (2009) key strategies. Thus Item 22 ('Assessment of students' work is mainly in the form of comments') is included, yet with relatively low value attributed to it compared to the rest of the items in the group. This may indicate that formative feedback is seen as primarily to be given in a spoken, rather than a written, form.

The very high value placed on practices related to the development of pupil agency in assessment and learning (Zimmerman, 2008) suggests a concern to develop students' metacognitive understanding of their own learning, and coheres with the sentiment of Black and William's AfL Strategy aimed at 'activating students as owners of their own learning'. Thus there is an emphasis on the learning orientation of the student, rather than on performance orientation (Dweck, 2000), together with a focus on students developing learning strategies that work best for them in a particular circumstance. Placing high value on these items suggests that teachers aspire to move students towards

self-regulated learning through appropriate scaffolding related to the contingent position of learner. Further, it again seems to suggest a clear concern amongst teachers with promoting the intentions of the five strategies of formative assessment (Black and William, 2009).

As we have indicated in the findings section, Item 3 ('The main thing I look for in my assessments is whether my students know, understand or can do key sections of the curriculum') is also highly valued and is the only item that might be interpreted as sitting outside concerns with either learning more about student learning or the development of pupil agency. Certainly it might be comfortably part of a group of items associated with curriculum-oriented concerns, and James and Pedder (2006) place it with items that suggest a performance focus. But the prescribed curriculum does not have to be a driver for a particular pedagogy and the focus on student understanding embedded in Item 3 suggested to the ALIC team that it might easily be placed with several groupings of items, not just those related to 'convergent assessment tendencies'. Thus it seems there is little contradiction amongst the highly valued items in the survey as a whole, though the meanings attributed to Item 3 deserve further investigation.

Item 5 ('I tell students how well they have done compared to others in the class') was the least valued item. It emphasises the development of a competitive classroom ethos and a strong focus on performance orientation (Dweck, 2000). Overall, the group of items that were highly valued by fewest teachers were those that might be linked to teacher control of assessment processes and a focus on performance goals. Item 8 ('My assessments are more useful than formal assessments') could be placed in this group of items, but it might be interpreted in a number of different ways; it may be seen as stressing the primacy of the individual teacher (perhaps regardless of evidence from pupils) or it might be strongly linked to the idea that considered formative assessment has more to offer than testing. Given this ambivalence, it is perhaps not surprising to see this item somewhat equivocally valued by teachers.

Items associated with student control over assessment processes were also amongst the least valued. These "divergent approaches to assessment" (Torrance and Pryor, 1998, pp.153–154) are clearly not of high value to these groups of teachers, and mirror the findings from research with teachers in the UK (James and Pedder, 2006; Winterbottom *et al.*, 2008a, b). These 'divergent' ambitions might be considered to be an end point or aspiration in terms of AfL practices, even in countries and schools where such practices are embedded, so their relatively low attributed value across nations is unsurprising.

When considering values and practices gaps a number of challenges for teachers are evident. In line with the argument made above, the group of items associated with giving students more control over assessment processes was not particularly highly valued. It nevertheless appears that for any teacher with an aspiration to achieve this level of student involvement in formative assessment there is still some way to go. The largest positive gap is thus for Item 6 ('I give students the opportunity to determine their own learning objectives').

For those items that were more highly valued by teachers, there is an apparent gap between values and practices for several items that can be broadly grouped through their association with the development of pupil agency in assessment and learning. Teachers seem much more comfortable with assessment approaches linked to developing their own understanding of students' learning than they are with promoting opportunities for students to assess their own work, build on their

strengths and think critically about their learning. This is unsurprising, as building such elements into assessment repertoires is not easy. However, the high value attributed to such practices suggests a strong aspiration to develop practice in this direction.

For items with a negative gap, it is interesting to see that the largest gap occurs with respect to the provision of feedback in the form of marks and grades. The strong drivers of accountability cultures (both on a micro-level in such things as direct accountability to parents and on a macro-level in terms of school, regional and national data comparisons) clearly have an influence here (Black and William, 2005). But it is nevertheless interesting to see how little comparative value is given to this practice compared to the level of practice itself. And though the gap is less marked, it seems clear that teachers would like some flexibility with respect to the setting of learning objectives, beyond the constraints of the prescribed curriculum.

There are some subtle differences between the perspectives of the Argentinean teachers and those reflected by other teachers in the data set, e.g. differences in the response to Item 26 ('I help students to plan the next steps in their learning') might be seen to imply rather less of a concern with children understanding how to build on their strengths and analysing areas for development in their own work. This is particularly interesting, given that other items related to the development of pupil agency in assessment and learning are given similar value ratings to those evidenced across the cohort. It might suggest a rigid curriculum structure that constrains the extent to which teachers feel that they can have an input on an individual's 'next steps for learning'. Certainly large values-practice gaps that relate to the provision of formative feedback, encouraging pupil involvement in learning and the development of pupil agency suggest a strong concern amongst Argentinean teachers to develop this area of their work.

Few Argentinean teachers highly valued those practices associated with student control over assessment practices. Of other practices highly valued by fewer than three-quarters of the Argentinean teachers, closed questioning (and the focus on performance goals with which it is often associated) and the provision of summative feedback were valued substantially less than they were by teachers in the overall study. This is interesting given what is suggested about actual classroom practices; a consideration of the values-practice gap suggests that aspirations for Argentinean teachers in many cases appear to be far ahead of current practices. For example, the gap for Item 12 ('Assessment of students' work is mainly given as marks and grades') indicates a strong, embedded practice that is considerably at odds with teachers' aspirations. It is tempting to speculate on the reasons why practices that accord with formative assessment principles lag so far behind teacher aspirations, but the data does not enable fruitful speculation in this area.

Turning to a consideration of the data from India, in general, there are small gaps between values and practices for those items linked positively to formative assessment practices. Indian teachers seem confident that their values and practices are broadly in harmony, though again we must add the caveat that interpretation without empirical evidence of practice can only be speculative. Looking in more detail at the Indian data through the lens provided by the Argentinean data, Item 26 ('I help students to plan the next steps in their learning') was much more highly valued, suggesting that Indian teachers place somewhat greater value overall on this aspect of helping children understand how to build on their strengths and analyse areas for development in their work. If Argentinean teachers are constrained by a rigid curriculum with respect

to this item, as we have speculated above, we might suggest that Indian teachers feel greater freedom in suggesting 'next steps' for individual learners.

In the Indian data, Item 8 ('My assessments are more useful than formal assessments') appears as a highly valued item, in a way that it does not in the Argentinean data. This may again be an issue of interpretation in a given national context. It might sit well with other items associated with teacher control of assessment processes and a focus on performance goals, as James and Pedder (2006) suggest. Alternatively, it might be interpreted as being associated with the idea that formative assessment has more to offer than formal testing.

Conclusions

Given the global prominence given to AfL by governments, assessment agencies, researchers and others, it is perhaps unsurprising to find that, in very broad terms, the items most valued by the teachers in this study demonstrate the considerable value placed upon practices linked positively to formative assessment principles and strategies. Certainly it seems that teachers have a particular concern with learning more about student learning and with promoting the development of pupil agency in assessment and learning. These concerns not only form the foundation of Black and Wiliam's (2009) five key strategies, they might also be seen more globally as being related to what teachers think about 'positive' pedagogy (Wiliam and Thompson, 2007). Importantly, the idea of pedagogy as we use it here includes individual and culturally-informed perspectives on communicative approaches (Mercer and Littleton, 2007), classroom participation structures (Cazden, 1986), the importance of students' metacognitive understanding of learning (Dweck, 2000; Zimmerman, 2008), the importance of student interaction and collaboration (Kutnick *et al.*, 2005), and the accountability structures that impinge on the work of the teacher (Black *et al.*, 2003).

Concern with such aspects of pedagogy, and associated assessment practices, suggests that the survey data reflect the views of professionals who are engaged, reflective and responsible. But it does seem clear that an individual teacher's response to the survey items is also likely to be considerably nuanced and strongly related to prevailing contextual imperatives. Thus, the data suggest that Indian teachers feel relatively confident that their practices match their aspirations with respect to classroom-based assessment, whilst for Argentinean teachers there are constraints that militate against their positive aspirations.

Factor analysis of the data, which will be the next step in our research, will consider how items group and this will provide further insights. The issue then will be to examine whether the underlying dimensions of assessment practice differ from those revealed by James and Pedder (2006) and why that might be the case.

The survey data as it has been interpreted thus far might suggest different ways of working with teachers in different countries. For example, Indian teachers might be helped to analyse their practice through classroom-based research, providing them with the tools to articulate good practice in their context; Argentinean teachers, on the other hand, might wish to develop peer observations that enable discussions about how best to develop their AfL aspirations. Whatever the survey data might suggest, however, it seems clear that an analysis of practices 'on the ground' is necessary if the nuances of national practices are to be fully revealed.

References

- Akyeampong, K., Pryor, J. & Ampiah, J. G. (2006). A vision of successful schooling: Ghanaian teachers' understandings of learning, teaching and assessment. *Comparative Education*, **42**, 2, 155–176.
- Alderson, J. C. (1992). *Validating questionnaires*, CRILE Working Papers 15. Lancaster: LAMEL, Lancaster University.
- Andrews, P. (2007). Negotiating meaning in cross-national studies of mathematics teaching: kissing frogs to find princes. *Comparative Education*, **43**, 4, 489–509.
- Assessment Reform Group (ARG) (2002). *Research-based principles of assessment for learning to guide classroom practice*. Available from: <http://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf>
- Black, P., Harrison, C., Lee, C., Marshall, B. & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Maidenhead: Open University Press.
- Black, P., McCormick, R., James, M. & Pedder, D. (2006). Learning how to learn and Assessment for Learning: a theoretical inquiry. *Research Papers in Education*, **21**, 119–132.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, **5**, 1, 7–74.
- Black, P. & Wiliam, D. (2005). Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal*, **16**, 2, 249–261.
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, **21**, 1, 5–31.
- Cazden, C. B. (1986). Classroom discourse. In: M. J. Wittrock (Ed.), *Handbook of research on teaching*. 3rd ed., 432–463. New York: Macmillan/AERA.
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.
- Hawkey, R. (2006). *Impact Theory and Practice: Studies of the IELTS test and Progetto Lingue 2000*. Studies in Language Testing 24. Cambridge: UCLES/Cambridge University Press.
- James, M., & Pedder, D. (2006). Beyond method: assessment and learning practices and values. *Curriculum Journal*, **17**, 2, 109–138.
- Johnson, M. & Burdett, N. (2010). Intention, interpretation and implementation: some paradoxes of Assessment for Learning across educational contexts. *Research in Comparative and International Education*, **5**, 2, 122–130.
- Kutnick, P., Sebba, J., Blatchford, P., Galton, M. & Thorpe, J. (2005). The effects of pupil grouping: *Literature review* (Literature review No. 688). London: DfES Research Report 688.
- Marshall, B. & Drummond, M. J. (2006). How teachers engage with Assessment for Learning: lessons from the classroom. *Research Papers in Education*, **21**, 2, 133–149.
- Mercer, N. & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. London: Routledge.
- Pedder, D. (2006). Organizational conditions that foster successful classroom promotion of Learning How to Learn. *Research Papers in Education*, **21**, 2, 171–200.
- Popham, W. J. (2006). Phoney formative assessments: buyer beware! *Educational Leadership*, **64**, 3, 86–87.
- Sach, E. (2012). Teachers and testing: an investigation into teachers' perceptions of formative assessment. *Educational Studies*, **38**, 3, 261–276.
- Sebba, J. (2006). Policy and practice in assessment for learning: the experience of selected OECD countries. In: J. Gardner (Ed.), *Assessment and Learning*. 185–196. London: Sage.
- Shepard, L. A. (2008). Formative assessment: caveat emptor. In: C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning*. 279–303. New York: Erlbaum.

- Smith, H. A. (1995). Introduction: cultural psychology and semiotics: Confronting meaning in educational practice. *Canadian Journal of Education*, **20**, 4, 407–414.
- Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning. *Assessment in Education: Principles, Policy and Practice*, **18**, 4, 433–449.
- Torrance, H. & Pryor, J. (1998). *Investigating Formative Assessment*. Buckingham: Open University Press.
- Wiliam, D., Lee, C., Harrison, C. & Black, P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles, Policy and Practice*, **11**, 1, 49.
- Wiliam, D. & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In: C. A. Dwyer (Ed.), *The Future of Assessment: Shaping Teaching and Learning*. 53–82. Mahwah, NJ: Erlbaum.
- Winterbottom, M., Brindley, S., Taber, K. S., Fisher, L. G., Finney, J. & Riga, F. (2008a). Conceptions of assessment: trainee teachers' practice and values. *Curriculum Journal*, **19**, 3, 193–213.
- Winterbottom, M., Taber, K. S., Brindley, S., Fisher, L. G., Finney, J. & Riga, F. (2008b). Understanding differences in trainee teachers' values and practice in relation to assessment. *Teacher Development*, **12**, 1, 15.
- Yung, B. H.-W. (2002). Same assessment, different practice: Professional consciousness as a determinant of teachers' practice in a school-based assessment scheme. *Assessment in Education: Principles, Policy and Practice*, **9**, 1, 97.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, **45**, 1, 166–183.

APPENDIX 1: ALIC Survey Participant Data by Nation

		Argentina		India		Indonesia		Nigeria		Saudi Arabia	
		N	%	N	%	N	%	N	%	N	%
Survey Returns		81	33	116	48	29	12	2	0.0	9	<0.0
Schools/ Colleges		51	35	61	42	22	15	2	0.0	8	<0.0
Teacher Gender	Male	10	12.3	38	32.8	16	55.2	2	100.0	4	44.4
	Female	70	86.4	78	67.2	13	44.8	0	0.0	5	55.6
Teacher Experience	Less than 2 Years	0	0.0	1	0.9	1	3.4	0	0.0	0	0.0
	2–4 Years	3	3.7	15	12.9	1	3.4	0	0.0	1	11.1
	5–10 Years	16	19.8	34	29.3	12	41.4	1	50.0	2	22.2
	11–20 Years	27	33.3	40	34.5	12	41.4	0	0.0	5	55.6
	21+ Years	31	38.3	17	14.7	1	3.4	1	50.0	1	11.1
	Not specified	4	4.9	9	7.8	2	6.9	0	0.0	0	0.0
Experience in Current School	Less than 2 Years	9	11.1	35	30.2	4	13.8	1	50.0	2	22.2
	2–4 Years	12	14.8	33	28.4	6	20.7	0	0.0	0	0.0
	5–10 Years	24	29.6	30	25.9	12	41.4	1	50.0	6	66.7
	11–20 Years	23	28.4	3	2.6	3	10.3	0	0.0	1	11.1
	21+ Years	8	9.9	3	2.6	0	0.0	0	0.0	0	0.0
	Not specified	5	6.2	12	10.3	4	13.8	0	0.0	0	0.0
Age Taught	10 and Under	3	3.7	2	1.7	2	6.9	0	0.0	0	0.0
	11–14	21	25.9	34	29.3	4	13.8	0	0.0	1	11.1
	15–18	53	65.4	68	58.6	19	65.5	2	100.0	8	88.9
	18+	1	1.2	1	.9	0	0.0	0	0.0	0	0.0
	Not specified	3	3.7	11	9.5	4	13.8	0	0.0	0	0.0
Subject Taught	Science/Maths	7	8.6	59	50.9	12	41.4	1	50.0	6	66.7
	English	47	58.0	22	19.0	6	20.7	1	50.0	1	11.1
	Languages	0	0.0	2	1.7	0	0.0	0	0.0	0	0.0
	Social Sciences/Humanities	13	16.0	17	14.7	1	3.4	0	0.0	1	11.1
	Arts	3	3.7	0	0.0	0	0.0	0	0.0	0	0.0
	Not specified	11	13.6	16	13.8	10	34.5	0	0.0	1	11.1

APPENDIX 2: Comparing ALIC teachers' classroom-based assessment values and practices – percentage of positive responses across five national contexts

<i>Item</i>	<i>Values (%) important/crucial</i>	<i>Practices (%) often/mostly</i>
30 I often talk to students about how they can improve their learning	100	97
4 The feedback that my students get helps them improve	99	96
15 I help students find ways of solving problems that they have in their learning	99	95
14 I tell students about their strengths and help them to develop these strengths	98	93
16 I encourage students to see their mistakes as valuable learning opportunities	98	94
1 Assessment gives me useful evidence of my students' understandings which I use to plan my next lesson	97	98
11 I talk about learning objectives with students in ways they understand	97	94
10 I tell students how well they have done compared with their own earlier performance	96	95
17 I help students to think about how they learn best	96	87
27 I think student effort is important when I assess their learning	96	99
28 I talk about assessment criteria with students in ways that they understand	96	95
20 I find students' errors are helpful because they give me information about how students are thinking	95	97
9 My classroom assessment practices help students to learn independently	94	94
13 I give guidance to help my students assess their own work	94	86
24 I give guidance to help students assess their own learning	94	86
3 The main thing I look for in my assessments is whether my students know, understand or can do key sections of the curriculum	93	95
21 I help students to understand the learning purposes of each lesson or series of lessons	92	88
25 My assessment is mainly about what students know, understand and can do	89	88
18 I use questions mainly so that my students give me reasons and explanations	88	89
26 I help students to plan the next steps in their learning	88	71
8 My assessments are more useful than formal assessments	81	77
23 The subject curriculum determines students' learning objectives	81	88
22 Assessment of students' work is mainly in the form of comments	75	70
6 I give students the opportunity to determine their own learning objectives	73	55
29 I give students the opportunity to assess each other's work	73	65
19 I give guidance to help students to assess one another's work	70	64
2 The subject curriculum I have to teach is a greater influence on what I will do in my next lesson than how well my students did in the last lesson	64	60
12 Assessment of students' work is mainly given as marks and grades	64	77
7 I use questions mainly to get factual knowledge from my students	52	54
5 I tell students how well they have done compared to others in the class	26	29

APPENDIX 3: Comparing Argentinean ALIC teachers' classroom-based assessment values and practices: percentage of positive responses

<i>Item</i>	<i>Values (%) important/crucial</i>	<i>Practices (%) often/mostly</i>
4 The feedback that my students get helps them improve	100	95
30 I often talk to students about how they can improve their learning	100	95
14 I tell students about their strengths and help them to develop these strengths	99	94
15 I help students find ways of solving problems that they have in their learning	99	93
16 I encourage students to see their mistakes as valuable learning opportunities	98	93
10 I tell students how well they have done compared with their own earlier performance	97	94
13 I give guidance to help my students assess their own work	96	84
17 I help students to think about how they learn best	96	79
27 I think student effort is important when I assess their learning	96	99
1 Assessment gives me useful evidence of my students' understandings which I use to plan my next lesson	95	99
9 My classroom assessment practices help students to learn independently	95	91
11 I talk about learning objectives with students in ways they understand	95	91
20 I find students' errors are helpful because they give me information about how students are thinking	95	98
24 I give guidance to help students assess their own learning	95	80
28 I talk about assessment criteria with students in ways that they understand	95	95
25 My assessment is mainly about what students know, understand and can do	93	91
3 The main thing I look for in my assessments is whether my students know, understand or can do key sections of the curriculum	90	95
21 I help students to understand the learning purposes of each lesson or series of lessons	90	79
18 I use questions mainly so that my students give me reasons and explanations	88	93
22 Assessment of students' work is mainly in the form of comments	81	71
26 I help students to plan the next steps in their learning	80	48
23 The subject curriculum determines students' learning objectives	78	88
8 My assessments are more useful than formal assessments	75	69
29 I give students the opportunity to assess each other's work	71	55
6 I give students the opportunity to determine their own learning objectives	67	37
2 The subject curriculum I have to teach is a greater influence on what I will do in my next lesson than how well my students did in the last lesson	64	68
19 I give guidance to help students to assess one another's work	64	53
12 Assessment of students' work is mainly given as marks and grades	56	85
7 I use questions mainly to get factual knowledge from my students	43	54
5 I tell students how well they have done compared to others in the class	14	15

APPENDIX 4: Comparing Indian ALIC teachers' classroom-based assessment values and practices: percentage of positive responses

<i>Item</i>	<i>Values (%) important/crucial</i>	<i>Practices (%) often/mostly</i>
30 I often talk to students about how they can improve their learning	100	98
4 The feedback that my students get helps them improve	99	97
15 I help students find ways of solving problems that they have in their learning	99	98
16 I encourage students to see their mistakes as valuable learning opportunities	99	96
11 I talk about learning objectives with students in ways they understand	98	96
14 I tell students about their strengths and help them to develop these strengths	98	96
17 I help students to think about how they learn best	98	92
1 Assessment gives me useful evidence of my students' understandings which I use to plan my next lesson	97	99
3 The main thing I look for in my assessments is whether my students know, understand or can do key sections of the curriculum	96	97
10 I tell students how well they have done compared with their own earlier performance	96	97
26 I help students to plan the next steps in their learning	96	89
27 I think student effort is important when I assess their learning	96	99
28 I talk about assessment criteria with students in ways that they understand	95	94
9 My classroom assessment practices help students to learn independently	94	97
20 I find students' errors are helpful because they give me information about how students are thinking	94	97
21 I help students to understand the learning purposes of each lesson or series of lessons	94	96
24 I give guidance to help students assess their own learning	94	92
13 I give guidance to help my students assess their own work	93	91
18 I use questions mainly so that my students give me reasons and explanations	89	88
8 My assessments are more useful than formal assessments	88	84
25 My assessment is mainly about what students know, understand and can do	88	88
23 The subject curriculum determines students' learning objectives	84	87
6 I give students the opportunity to determine their own learning objectives	78	70
29 I give students the opportunity to assess each other's work	71	71
19 I give guidance to help students to assess one another's work	70	71
22 Assessment of students' work is mainly in the form of comments	70	73
2 The subject curriculum I have to teach is a greater influence on what I will do in my next lesson than how well my students did in the last lesson	64	56
12 Assessment of students' work is mainly given as marks and grades	64	71
7 I use questions mainly to get factual knowledge from my students	51	51
5 I tell students how well they have done compared to others in the class	34	38

Using scales of cognitive demand in a validation study of Cambridge International A and AS level Economics

Jackie Greatorex Research Division, Stuart Shaw CIE, Phineas Hodson CIE and Jo Ireland Research Division

Introduction

The CRAS (Complexity, Resources, Abstractness and Strategy) framework is used to evaluate the cognitive demands of examination questions (Crisp and Novaković, 2009a, 2009b; Hughes, Pollitt and Ahmed, 1998; Pollitt, Ahmed and Crisp, 2007; Pollitt, Hughes, Ahmed, Fisher-Hoch and Bramley, 1998). Johnson and Mehta (2011) reviewed how CRAS was used; they endorsed some practices and made several recommendations (detailed below). This article provides an illustration of Johnson and Mehta's (2011) principles for using CRAS in the context of validating examination questions used in Cambridge International A and AS level Economics, and highlights some advantages and difficulties inherent in their methods. This article is part of our exploration of how to refine the use of CRAS, particularly with multiple choice question papers, with the aim of sharing issues and recommendations.

Development of CRAS

CRAS was developed from earlier scales of cognitive demands (Edwards and Dall'Alba, 1981) combined with examiners' views about what is more and less demanding for candidates (Hughes *et al.*, 1998). Hughes *et al.* (1998) and Pollitt *et al.* (1998) describe CRAS as having four dimensions: Complexity, Resources, Abstractness and Strategy. This was increased to five by Pollitt *et al.* (2007), who split Strategy into Task strategy and Response strategy (Figure 1).

It is Pollitt *et al.*'s (2007) conceptualisation of the dimensions that was used in this study. It has also been used by other researchers, for example, Crisp and Novaković (2009a).

Johnson and Mehta (2011) make several recommendations for using CRAS including the following:

- CRAS should only be used where the CRAS dimensions map to the constructs to be examined. CRAS is predominantly cognitive so it is only suitable for evaluating cognitive demands.
- Individual examination questions may be evaluated using CRAS but ratings on the dimensions may not be summed to give a value for the overall demand of an examination paper.
- An expert's rating of an examination question on one dimension can be compared with their rating of another examination question on the same dimension.
- Ratings on the different dimensions should not be combined to give individual questions a score for 'total demand'.

CRAS has been used to evaluate examination questions but it could also be used to evaluate the cognitive demands of text books, curricula, lesson contexts and marking criteria (Hughes *et al.*, 1998; Johnson and Mehta, 2011) and in validation studies (Shaw and Crisp, 2012; Shaw, Crisp and Johnson, 2011).

Figure 1: The CRAS Scales of demands (Pollitt *et al.*, 2007, 186)

	1	2	3	4	5
Complexity The number of components or operations or ideas and the links between them.		Mostly single ideas or simple steps. Little comprehension, except that required for natural language. Few links between operations.		Synthesis or evaluation is required. Need for technical comprehension. Make links between cognitive operations.	
Resources The use of data and information.		More or less all and only the data/information needed are given.		Student must generate or select the necessary data/information.	
Abstractness The extent to which the student deals with ideas rather than concrete objects of phenomena.		Mostly deals with concrete objects.		Mostly abstract.	
Task strategy The extent to which the student devises (or selects) and maintains a strategy for tackling the question.		Strategy is given. Little or no need to monitor strategy. Little selection of information required.		Students need to devise their own strategy. Students must monitor the application of their strategy.	
Response strategy The extent to which students have to organise their own response.		Organisation of response hardly required.		Must select answer content from a large pool of possibilities. Must organise how to communicate response.	

Table reproduced from 'Techniques for monitoring the comparability of examination standards', © Qualifications and Curriculum Authority 2007.

Cambridge International A and AS level Economics

Cambridge International Examinations (CIE) A and AS level Economics examinations are offered to students across the world in two series, November and June (CIE, 2009). This article reports on research which uses papers from the June 2011 series. The examination comprises four papers:

- AS level – multiple choice (30 questions)
- AS level – data response and structured essay (4 questions)
- A level – multiple choice (30 questions)
- A level – data response and structured essays (7 questions)

The five Assessment Objectives (AOs) for A and AS level Economics state that students are expected to:

- Demonstrate knowledge and understanding of the specified content.
- Interpret economic information presented in verbal, numerical or graphical form.
- Explain and analyse economic issues and arguments, using relevant economic concepts, theories and information.
- Evaluate economic information, arguments, proposals and policies, taking into consideration relevant information and theory, and distinguishing facts from hypothetical statements and value judgements.
- Organise, present and communicate economic ideas and informed judgements in a clear, logical and appropriate form (CIE, 2009, 5).

Validation

Educational measurement and psychological testing generally take a construct-centred approach to the validity of question papers, psychological tests and other assessments (Brown, 2010; Ertl and Stasz, 2010; Kane, 2009; Messick, 1995; Quinlan, Higgins and Wolff, 2009; Shaw *et al.*, 2011; Stobart, 2009; Threlfall, Nelson and Walker, 2007; Tran, Griffin and Nguyen, 2010; Vogt, Proctor, King, King and Vasterling, 2008). A construct-centred approach draws on the view that an underlying theoretical construct, such as mathematical aptitude, is represented by an examination mark and is the foundation on which the evaluation of an examination is built (Messick, 1989). A claim that an interpretation or use is valid must be backed by evidence that the marks from the examination adequately reflect the constructs.

To establish whether examinations elicit performances that reflect intended constructs awarding bodies must have recourse to a reasonably well-informed and coherent theoretical model underpinning the constructs of interest. The work from which this CRAS study was drawn utilised the model for validation of general qualifications proposed by Shaw *et al.* (2011), and illustrated in Shaw and Crisp (2012), which is itself situated in Kane's model of validation through argument (Kane, 1992). Based on this theoretical background, the CRAS framework was used to answer the following validation questions for CIE Economics A level:

- Do the tasks elicit performances that reflect the intended constructs?
- Do the tasks adequately sample the constructs that are set out as important in the syllabus?

The AOs are assumed to represent the intended constructs. In line with Johnson and Mehta's (2011) recommendations, the constructs were broadly mapped to the CRAS dimensions (Figure 2). Additionally, the item types were judged by the researchers to broadly map to the CRAS dimensions, with the caveat that Response strategy was less relevant to multiple choice questions than other question types. Shaw and Crisp (2012) did not report problems applying Response strategy to multiple choice questions, and on the basis of this research evidence we considered all CRAS dimensions to be suitable for use with all the question types.

Method

Six experts applied the CRAS instrument to the selected question papers, two of which contained multiple choice items and two of which contained essay and data response items. The experts were chosen on the basis of their experience as senior examiners for CIE 16–19 Economics qualifications.

Each expert was issued with the following materials:

- task instructions
- a copy of each of the question papers
- a copy of the mark scheme for each question paper
- the CRAS scales
- a response sheet for each question paper (see Appendix A for an example).

The instructions informed the experts that the exercise was about the cognitive demands of examination questions. For Task 1 they were instructed to:

- ignore the mark scheme
- familiarise themselves with CRAS and the question papers
- look at each question on each paper individually
- for each type of demand on each question, rate the level of demand of the activities the students have to do to answer the question on a scale of 1 (low demand) to 5 (high demand)
- work remotely and individually.

For Task 2 they were asked to repeat the exercise focusing on the demands rewarded by the mark scheme.

Analysis

The ratings across the five dimensions given to each question by each expert were tabulated to provide an indication of the range of demands across questions and the level of inter-rater agreement. Ratings without the mark scheme allowed inferences about construct elicitation to be made from the CRAS demands of the questions, that is, they indicated how demanding the question would appear to a candidate, teacher or other stakeholder if they did not consult the mark scheme. Ratings with the mark scheme allowed inferences about whether the mark scheme rewarded (sampled) the constructs. For the purposes of brevity the report refers to the ratings indicating the demands rewarded by the *mark scheme*, which is the primary focus of Task 2, however these ratings indicate the demands rewarded by the combination of the *question paper and the mark scheme*.

To investigate the degree to which the mark scheme rewarded the demands inherent in the question, each expert's ratings with and without the mark scheme were compared. The frequency of experts giving different ratings with and without the mark scheme was calculated for each dimension.

The scope for summary statistics was constrained by the ordinal nature of the data and the principle that ratings from different experts cannot be combined. Based on the sample size in this study most quantification was precluded unless questionable assumptions were introduced concerning equal-interval scales and common internalisation of the scales and anchor points.

Figure 2: Mapping CRAS to the assessment objectives and question types

<i>Dimension description from Pollitt et al. (2007)</i>	<i>Reason(s) for relevance to the AOs</i>	<i>Reason(s) for relevance to question types</i>
<p>Complexity The number of components or operations or ideas involved in a task and the links between them.</p>	<p>The skills in the AOs (demonstrating understanding, interpreting, explaining, analysing, evaluating, organising and communicating) can involve one or more steps, technical comprehension and synthesis or evaluation.</p>	<p>All question types:</p> <ul style="list-style-type: none"> • Can involve one or more ideas/steps • Relate to technical information • Can involve links between operations (evaluation/synthesis).
<p>Resources The use of data and information.</p>	<p>Using data and information correctly requires knowledge and understanding of economics (AO1).</p> <p>The student clearly generating information involves:</p> <ul style="list-style-type: none"> • Interpreting information presented in verbal, numerical or graphical form (AO2). • Explaining and analysing (AO3). • Evaluating (AO4). • Taking into consideration relevant information and theory and distinguishing facts (AO4). • Organising, presenting and communicating (AO5). 	<p>For multiple choice and data response students are provided with data/information such as text, graphs and statistics and can require:</p> <ul style="list-style-type: none"> • Using only the data/information provided • Generating data/information. <p>For essays students must generate much of the necessary data/information.</p>
<p>Abstractness The extent to which the students must deal with ideas rather than concrete objects.</p>	<p>All the AOs involve dealing with abstract information.</p>	<p>For all question types the content is abstract.</p>
<p>Task strategy The extent to which the students must devise (or select) and maintain a strategy for tackling the question.</p>	<p>Strategies might involve any combination of or all of the skills in the AOs.</p>	<p>Questions of all types can involve being given a strategy or devising a strategy and monitoring the application of the strategy.</p>
<p>Response strategy The extent to which the students have to organise their own response.</p>	<p>Reflects AO5 which requires the students to organise, present and communicate economic ideas and informed judgements in a clear, logical and appropriate manner.</p>	<p>The essay questions and to a lesser extent the data response questions require students to organise their own response.</p>
<p>Does CRAS map to the AOs and question types? Complexity, Resources, Abstractness and Task strategy mapped to all the AOs and question types. Response strategy reflected AO5 rather than the other AOs and was more relevant to essay and data response questions than to multiple choice questions. Therefore the constructs broadly mapped to CRAS.</p>		

Validation findings

The findings which follow are based solely on the data response and essay questions. The multiple choice papers are dealt with later under a separate heading.

There was a strong tendency for each rater to place many of the questions on a paper at the same level of demand (see Tables 1 to 4). In some contexts this could suggest a threat to validity as, if all questions are of similar demand, low ability candidates may lack sufficient low-demand questions to demonstrate their abilities and/or high ability candidates may not be stretched by sufficiently demanding questions. However, where candidates choose between optional essay questions, as is the case here, consistent demand is a desirable feature since whatever questions candidates choose, they will experience similar demands. Much greater diversity in demand levels is shown by inter-rater comparisons but, in the absence of evidence that the internalised scales of raters were similar, valid comparisons at this level are not possible.

The comparison between the demands elicited by the questions and the demands rewarded by the mark scheme, as illustrated in Figures 3 to 7, shows that there were no questions for which a consensus, or something approaching a consensus, existed on which was more demanding. As the raters as a group considered demands to be equal, or were divided on whether it was the questions or the mark schemes

which embodied greater demand, it can be inferred that demands were broadly similar across the two. This suggests that the demands rewarded by the mark scheme and those elicited by the question were similar, which provides evidence of validity.

This study alone should not be seen as providing a compelling answer to the validation questions, but can provide a valuable perspective and contribute to the body of evidence. The small number of experts used does reduce the power of this study but the number used is sufficient to warrant the conclusions and the research effort required for large sample sizes will not always be available.

Table 1: Expert ratings of questions from question paper 2 without the mark scheme

<i>Demand level (rating category)</i>		<i>Expert 1</i>	<i>Expert 2</i>	<i>Expert 3</i>	<i>Expert 4</i>	<i>Expert 5</i>	<i>Expert 6</i>
Complexity	5 (High demand)	1 2 3 4		2 3 4			1
	4		4	1	1	2 3 4	4
	3		1 2 3		2 3 4	1	
	2						2 3
	1 (Low demand)						
Resources	5 (High demand)	2 3 4				2 3 4	3
	4	1		2 3 4			1 2 4
	3		1 2 3 4		1	1	
	2			1			
	1 (Low demand)				2 3 4		
Abstractness	5 (High demand)	4				2 4	
	4	1 2 3		2 3	1 2 4	1 3	2 4
	3		2 4	4	3		1 3
	2		1 3	1			
	1 (Low demand)						
Task strategy	5 (High demand)	1 2 3 4		1 2 3 4		2 3 4	
	4						2 3 4
	3		1 2 3 4		1 3 4	1	1
	2				2		
	1 (Low demand)						
Response strategy	5 (High demand)			1 2 3 4		2 3 4	1 2 3 4
	4	2 3 4					
	3	1	1 4		1 3 4	1	
	2		2 3		2		
	1 (Low demand)						

Note: Question numbers appear in the cells in columns 3 to 8.

Table 2: Expert ratings of questions from question paper 2 with the mark scheme

<i>Demand level (rating category)</i>		<i>Expert 1</i>	<i>Expert 2</i>	<i>Expert 3</i>	<i>Expert 4</i>	<i>Expert 5</i>	<i>Expert 6</i>
Complexity	5 (High demand)	1 4					1 4
	4	2 3	4	1 2 3 4	1	2 3 4	2 3
	3		1 2		2 3 4	1	
	2		3				
	1 (Low demand)						
Resources	5 (High demand)	2 3 4		4		2 3 4	
	4		2 3	3	1	1	1 2 3 4
	3	1	1 4	2			
	2			1			
	1 (Low demand)				2 3 4		
Abstractness	5 (High demand)	4				2 4	3
	4	1 2 3	2 4	2 3		3	1 2 4
	3		3	4	1 2 3 4	1	
	2			1			
	1 (Low demand)		1				
Task strategy	5 (High demand)	2 3 4		1 2 3 4			1 2 3 4
	4	1			3 4	2 3 4	
	3		1 2 3 4			1	
	2				1 2		
	1 (Low demand)						
Response strategy	5 (High demand)			1 2 3 4		2 3 4	1 2 3 4
	4	2 3 4	1 3				
	3	1	2 4		3 4	1	
	2				1 2		
	1 (Low demand)						

Note: Question numbers appear in the cells in columns 3 to 8.

Table 3: Expert ratings of questions from question paper 4 without the mark scheme

<i>Demand level (rating category)</i>		<i>Expert 1</i>	<i>Expert 2</i>	<i>Expert 3</i>	<i>Expert 4</i>	<i>Expert 5</i>	<i>Expert 6</i>
Complexity	5 (High demand)	1 2 3 4 5 6 7		2 3 4 5 6 7		4 6 7	1
	4		1 2 5 6	1	1	2 3 5	6 7
	3		3 4 7		3 5 6 7	1	3 5
	2				2 4		2 4
	1 (Low demand)						
Resources	5 (High demand)	2 3 4 5 6 7				2 3 4 5 6 7	
	4		1 2 4 5 6 7	2 3 4 5 6 7	1	1	1 2 3 4 5 6 7
	3	1	3				
	2			1	2 3 4 5 6 7		
	1 (Low demand)						
Abstractness	5 (High demand)	2 4 6	2 6			2 6 7	1 4
	4	1 3 5 7	1 4 5	2 3 4 5 6 7	2 3 4 5 6 7	1 3 4 5	2 3 5 6
	3		3 7		1		7
	2			1			
	1 (Low demand)						
Task strategy	5 (High demand)	1 2 3 4 5 6 7		1 2 3 4 5 6 7		3 5 6	1 4
	4		1 2 4 5 6		4 7	2 4 7	2 3 5 6 7
	3		3 7		1 2 3 5 6	1	
	2						
	1 (Low demand)						
Response strategy	5 (High demand)	4 7		1 2 3 4 5 6 7		2 3 4 5 6 7	1 2 3 4 5 6 7
	4	2 3 5 6	4 7		4 7	1	
	3	1	2 3 5		1 2 3 5 6		
	2		1 6				
	1 (Low demand)						

Note: Question numbers appear in the cells in columns 3 to 8.

Table 4: Expert ratings of questions from question paper 4 with the mark scheme

<i>Demand level (rating category)</i>		<i>Expert 1</i>	<i>Expert 2</i>	<i>Expert 3</i>	<i>Expert 4</i>	<i>Expert 5</i>	<i>Expert 6</i>
Complexity	5 (High demand)	2 3 4 5 7	2 6	4 5 7			1 2 5 7
	4	1 6		1 2 3 6	1	3 4 5 7	6
	3		1 3 4 5 7		3 5 6 7	1 2 6	3 4
	2				2 4		
	1 (Low demand)						
Resources	5 (High demand)	2 3 4 5 6 7	2 5	5		2 3 4 5 6 7	1 4
	4		3 4 6 7	2 3 4 6 7	1	1	
	3	1	1				3 5 6 7
	2			1	2 3 4 5 6 7		2
	1 (Low demand)						
Abstractness	5 (High demand)	2 4 6	2 4 6			2 4 6	2 4 5 6
	4	1 3 5 7	3	2 3 4 5 6 7	2 3 6 7	1 3 5 7	1 3 7
	3		5 7		1 4 5		
	2			1			
	1 (Low demand)		1				
Task strategy	5 (High demand)	1 2 3 4 5 6 7	6	1 2 3 4 5 6 7		4	1 2 3 4 5 6 7
	4		1 2 4 5		3 4 5 7	2 3 4 5 6 7	
	3		3 7		1 2 6	1	
	2						
	1 (Low demand)						
Response strategy	5 (High demand)	4 7		1 2 3 4 5 6 7		3 4	1 2 3 4 5 6 7
	4	2 3 5 6	5		4 5 7	2 5 6 7	
	3	1	1 4 6 7		2 3 6	1	
	2		2 3		1		
	1 (Low demand)						

Note: Question numbers appear in the cells in columns 3 to 8.

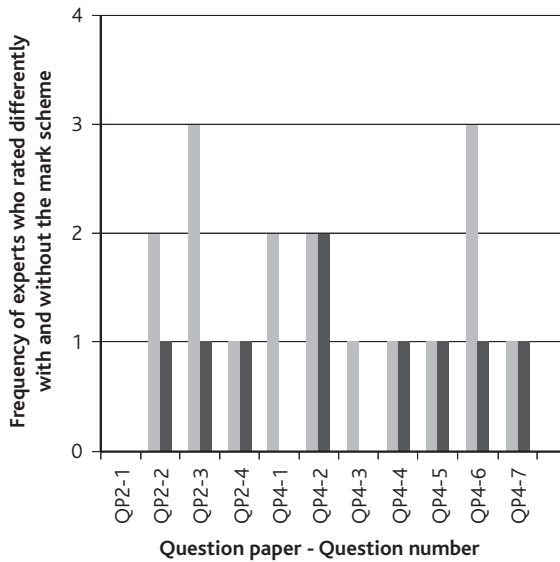


Figure 3: Frequency of experts who rated questions differently with and without the mark scheme for the Complexity dimension

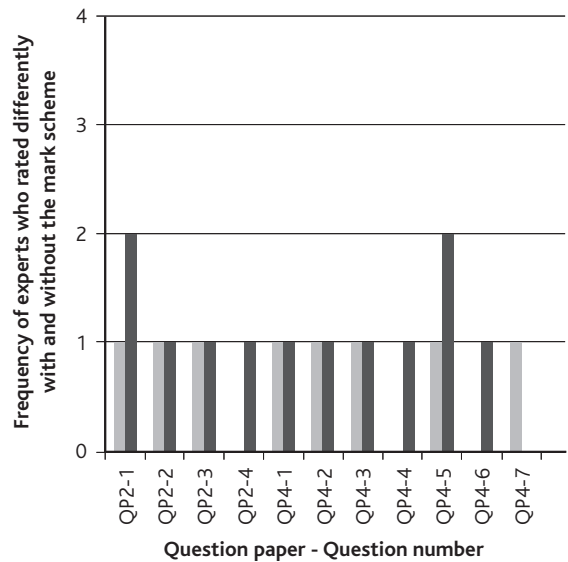


Figure 4: Frequency of experts who rated questions differently with and without the mark scheme for the Resources dimension

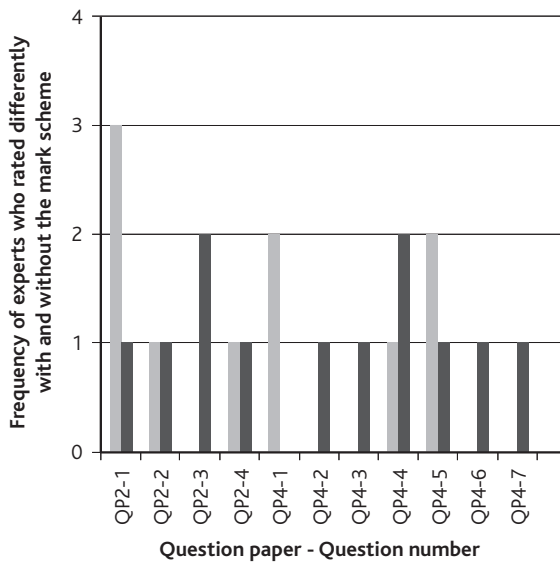


Figure 5: Frequency of experts who rated questions differently with and without the mark scheme for the Abstractness dimension

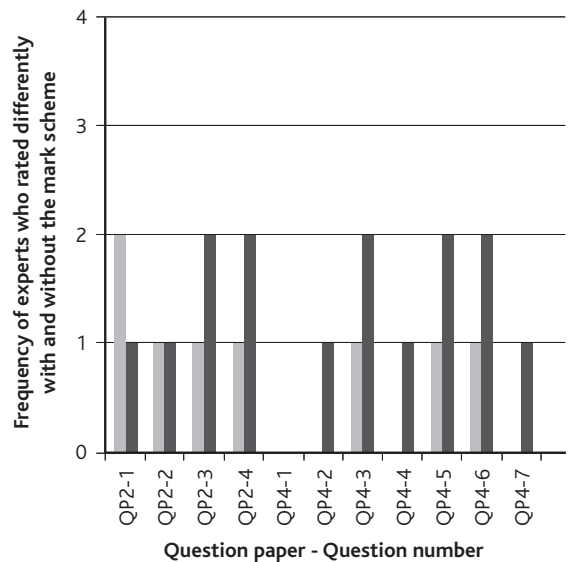


Figure 6: Frequency of experts who rated questions differently with and without the mark scheme for the Task strategy dimension

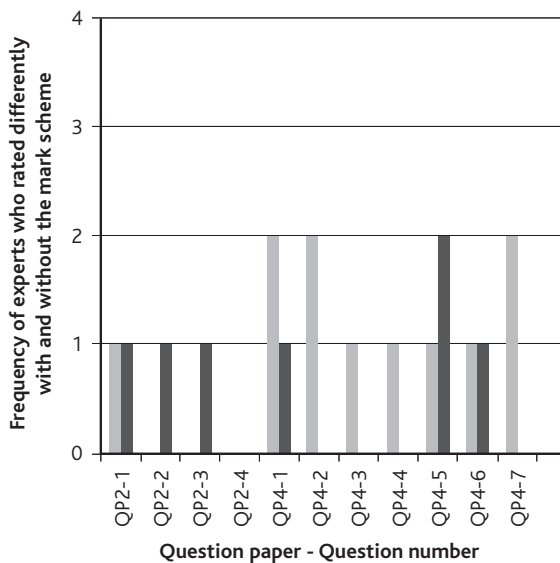


Figure 7: Frequency of experts who rated questions differently with and without the mark scheme for the Response strategy dimension

KEY TO FIGURES

- Question paper only is more demanding
- Question paper and Mark scheme is more demanding

Reflections on using CRAS with essay and data response questions

The CRAS framework allows the construction of extensive datasets of ratings but tightly circumscribes the methods available for analysing those datasets. Much of the difficulty in analysis stems from the nature of the scales used by raters. Each rater's scale is ordinal so commonly-used descriptive statistics such as means, modes and standard deviations are inapplicable. It is also not possible to compare or combine results from multiple raters on a single dimension without assuming that the raters have a common internalised scale. This assumption is difficult to support and there is no evidence that it holds for this study. Finally, there is no simple method for combining the five dimensions to give an aggregated difficulty score for an item as there is no justification for claiming that, for example, a demand of 3/5 on Resources and a demand of 3/5 on Complexity are equivalent.

One approach which can be pursued is that of making comparisons within a single demand type as rated by a single expert (Johnson and Mehta, 2011). This does allow consideration of the diversity of demand across items, though not in a strictly quantitative manner. It also allows comparison between demands elicited by the question and demands inherent in the mark scheme for a given item. These methods allow somewhat narrow conclusions given the wealth of data from which they are drawn, but the limitations on analysis inherent in the CRAS framework preclude more far-reaching analysis.

The results reported here show low levels of inter-rater agreement in terms of absolute level of demands, but interesting commonalities in terms of ranking of items. For many dimensions, the raters tended to find that all or most items mapped to the same level of demand, but differed strikingly on where that level fell on the 1–5 scale. This suggests a common understanding of how the demand of each item relates to the demand of others around it, but very different anchor points for the internalised scales. Having raters produce a rank order for the items on each dimension, rather than placing them on a scale, could allow finer distinction between items, but would not have revealed the result reported here on the homogeneity of demand across items.

The lack of consistency between raters' internal scales reported here could be related to the lack of an established community of practice. As the raters worked remotely with all information and instructions passing outward from a central hub – the research team – there was no opportunity for raters to negotiate common understandings of the CRAS framework. Though the explicit instructions were common, the tacit, internalised understandings appear significantly divergent. Wolf (1995) argued, in the context of marker reliability, that standards are conveyed through illustrations of students' work within close-knit expert networks, rather than by written criteria. Utilising an analogous process to build a shared understanding of the CRAS scales could move raters towards producing comparable ratings.

Future studies could overcome the lack of consistency between raters' internal scales through alterations to the methods. Particularly effective in avoiding the difficulties inherent in working with CRAS data are approaches based on rankings, such as:

- Collecting data using the Q sort method. Experts would initially work with one CRAS dimension. They would be given each question (or subquestion) in hard copy on a card. They would individually sort the cards, first into two piles, one for 'more demanding' and one for 'less demanding'. After completing this, the questions from the 'more

demanding' pile would be sorted into three piles, each corresponding to categories 3, 4 and 5 on the CRAS dimension. Thirdly, the questions in the 'less demanding' pile would be sorted into three piles, each corresponding with categories 1, 2 and 3 on the CRAS dimension. The number of cards in each pile would be restricted to conform to the normal distribution; with most cards in category 3, fewer in categories 2 and 4 and the least in categories 1 and 5. The experts would then rank the questions in each pile from most to least demanding. At all stages the decisions can be reviewed as necessary. Q sort data is generally analysed using cluster analysis to produce statistical summaries of similar Q sorts. The process would be repeated for each CRAS dimension for questions with and without the mark scheme. For more details about the Q sort method see van Exel and de Graaf (2005).

- Collecting data using paired comparisons. Experts would initially work with one CRAS dimension. Each expert would be presented with pairs of questions (or subquestions) and asked to indicate which question in the pair was the most demanding on a CRAS dimension. This would be repeated for all possible pairs of questions. The frequency with which each question was judged to be the most demanding would be used to produce a rank order of questions on a given CRAS dimension. The process would be repeated for each CRAS dimension for questions with and without the mark scheme. Further descriptions of paired comparison methods can be found in Vance and McCall (1934) and Crisp and Novaković (2009a, 2009b).

For both Q sort and paired comparisons each expert would, for each CRAS dimension, produce two rank orders of questions from the most to the least demanding; one with and one without the mark scheme. The frequency of experts who ranked questions differently with and without the mark scheme would be analysed. The Q sort and paired comparisons use rankings rather than ratings, and thereby overcome any leniency or severity in experts' judgements. This avoids issues of differing anchor points and internal scales but it would not standardise the experts' understanding of the CRAS scale.

Finally, the intra-rater approach used to compare the demands of the questions and the demands of the mark schemes provides a robust method for analysing CRAS data. Much of the richness of the original CRAS ratings was lost by calculating the number of experts who rated the question or mark scheme as the more demanding but the method provided useful evidence on the validity questions being addressed. The choice between rigour of analysis and maintaining the richness of the data is common to many decisions made when analysing CRAS data. The use of intra-rater comparisons in the interpretation of CRAS data is also in accordance with Johnson and Mehta (2011).

Reflections on using CRAS with multiple choice questions

The data from the multiple choice papers could not be usefully analysed with the methods used for the other question types and was therefore not presented in the results section. Comparison of the demands inherent in the questions with the demands rewarded by the mark scheme was not possible as the experts could tell what the mark scheme would contain from seeing the question paper. As the number of questions was large (30 per paper) and aggregation of data across questions was not

Table 5: Recommendations for future research in response to experts' comments

Points made by experts	Suggestions for future research which includes multiple choice questions
Response strategy is more relevant to open ended questions than closed or multiple choice questions. Experts in a study conducted concurrently with the work reported here experienced similar issues, according to discussion with the authors of Crisp and Hopkin (<i>in submission</i>).	Experts rate multiple choice questions on the first four dimensions of CRAS only, thereby omitting Response strategy. Alternatively the equivalent scale developed by Hughes <i>et al.</i> (1998) could be used as it makes no distinction between Task strategy and Response strategy.
Students have just two minutes to answer each multiple choice question and the examination setters therefore deliberately avoid complexity and unduly complex reasoning. For instance, calculations requiring more than two steps are avoided. This would lead to low Complexity ratings.	Low complexity on multiple choice papers could be viewed as a legitimate result, reflecting what candidates have to do to respond to the items. It could also be the case that two-step calculations that require "Synthesis or evaluation", "technical comprehension" or making "links between cognitive operations" (Pollitt <i>et al.</i> , 2007, 186) have high complexity ratings. If the multiple choice papers contain many questions that contain recall alone, however, CRAS might not be the best instrument to apply to them. A pre-rating meeting for experts to agree on an interpretation of CRAS could help with these issues.
Questions are designed to test one topic/problem only. Distractors (incorrect answer options) might relate to a different problem but the student is not expected to solve that problem, just to reject the distractor. This would lead to low Complexity ratings.	The act of choosing between possible responses is not necessarily cognitively undemanding and could include careful evaluation of options – a mark of higher complexity. It could also be that the complexity of items on a given multiple choice paper is generally lower than that on a matched essay paper, which would constitute a useful result.
Questions are designed to include all the necessary information in the question stem. The exception is when a question is intended to test whether the students can determine what is and is not relevant. This would give low Resources ratings.	As above, a tendency towards low ratings does not necessarily constitute an indictment of the method. Some variance in resources ratings would also be expected from the deliberate inclusion of extraneous information in some questions, requiring candidates to select resources.
Questions are designed to test higher order cognitive skills and problem solving skills and this does not seem to be reflected in the rating scale.	This criticism is a direct challenge to the validity of CRAS since if the rating scales do not reflect 'higher order cognitive skills' it is unclear what they do measure. While the CRAS framework has been found to be a useful tool in multiple studies (Hajo, 2008; QCA, 2003, 2006) this suggests the expert's internalisation of the framework was markedly divergent from that which was intended by the researchers. In future work, close attention needs to be paid to establishing with experts the nature of the instrument and good practice in its use.

possible, presentation such as that shown in Table 1 was not practical. The lack of correspondence between the Response strategy dimension and the experts' task also presented problems for analysis. As noted earlier the research evidence prior to the present study found no problems using the Response strategy dimension with multiple choice questions (Shaw and Crisp, 2012).

The following section explores these difficulties through the comments made about the multiple choice task by the raters, and proposes methods for usefully investigating multiple choice questions using CRAS.

Conclusion

In the context of the present study, the CRAS methodology provided validity evidence, though the strength of the evidence provided by the method does not justify the researcher effort required to implement it. The use of expert ratings proved very problematic as consistency of internal scales across examiners is hard to establish, and was not established here. Future work could use rank order approaches to mitigate this difficulty. The aggregation of results from multiple choice items in a manner that produced answers to the research questions was not possible here, and future uses of CRAS might best be restricted to free response items. CRAS could provide a useful tool in validity studies where both the question types and the constructs map to CRAS and either the experts produce rank orders or there is significant commonality among experts in their understanding of the scales and method.

References

- Brown, T. (2010). Construct Validity: A unitary concept for Occupational Therapy assessment and measurement. *Hong Kong Journal of Occupational Therapy*, **20**, 1, 30–42.
- CIE. (2009). Cambridge International A & AS Level Economics Syllabus code 9708. For examination in June and November 2011.
- Crisp, V. & Hopkin, R. (*in submission*). Modelling question difficulty in an A Level Physics examination. *Research Papers in Education*.
- Crisp, V. & Novaković, N. (2009a). Are all assessments equal? The comparability of demands of college based assessments in a vocationally related qualification. *Research in Post-Compulsory Education*, **14**, 1, 1–18.
- Crisp, V. & Novaković, N. (2009b). Is this year's exam as demanding as last year's? Using a pilot method to evaluate the consistency of examination demands over time. *Evaluation and Research in Education*, **22**, 1, 3–15.
- Edwards, J. & Dall'Alba, G. (1981). Development of a scale of cognitive demand for analysis of printed secondary science materials. *Research in Science Education*, **11**, 158–170.
- Ertl, H. & Stasz, C. (2010). Employing an 'employer-led' design? An evaluation of the development of Diplomas. *Journal of Education and Work*, **23**, 4, 301–317.
- Hajo, Z. (2008). *Content validity and comparability of the Lebanese national examinations in Chemistry*. Doctor of Education, Leicester, Leicester.
- Hughes, S., Pollitt, A. & Ahmed, A. (1998). *The development of a tool for gauging the demands of GCSE and A Level exam questions*. Paper presented at the BERA conference, The Queen's University, Belfast.
- Johnson, M. & Mehta, S. (2011). Evaluating the CRAS Framework: Development and recommendations. *Research Matters: A Cambridge Assessment Publication* **12**, 27–33.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, **112**, 527–535.

- Kane, M. T. (2009). Validating the interpretations and uses of test scores. In: R. W. Lissitz (Ed.), *The Concept of Validity: Revisions, New Directions, and Applications*. 39–64. USA: Information Age Publishing.
- Messick, S. (1989). Validity. In: R. Linn (Ed.), *Educational Measurement*. New York: Macmillan.
- Messick, S. (1995). Validity of Psychological Assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, **50**, 9, 741–749.
- Pollitt, A., Ahmed, A. & Crisp, V. (2007). The demands on examination syllabuses and question papers. In: P. Newton, J. Baird, H. Goldstein & H. Patrick (Eds.), *Techniques for monitoring the comparability of examination standards*. 166–206. London: Qualifications and Curriculum Authority.
- Pollitt, A., Hughes, A., Ahmed, A., Fisher-Hoch, H. & Bramley, T. (1998). *The effects of structure on the demands in GCSE and A level questions*. Report to Qualifications and Curriculum Authority.
- QCA. (2003). *Report on Comparability between GCE and International Baccalaureate Examinations*. London: QCA.
- QCA. (2006). GCSEs and IGCSEs compared: GCSE and IGCSE examinations in 2005 in English, French, mathematics and science (double award). London: QCA.
- Quinlan, T., Higgins, D. & Wolff, S. (2009). Evaluating the construct-coverage of the e-rater® Scoring Engine *ETS RR-09-01* Retrieved from <http://www.ets.org/Media/Research/pdf/RR-09-01.pdf>
- Shaw, S. D. & Crisp, V. (2012). An approach to validation: Developing and applying an approach for the validation of general qualifications. *Research Matters: A Cambridge Assessment Publication*, Special Issue **3**.
- Shaw, S. D., Crisp, V. & Johnson, N. (2011). A framework for evidencing assessment validity in large-scale, high-stakes international examinations. *Assessment in Education: Principles, Policy & Practice*, **19**, 2, 159–176. doi: 10.1080/0969594X.2011.563356
- Stobart, G. (2009). Determining validity in national curriculum assessments. *Educational Research*, **51**, 2, 161–179.
- Threlfall, J., Nelson, N. & Walker, A. (2007). *Report to QCA on an investigation of the construct relevance of sources of difficulty in the Key Stage 3 ICT tests* (S. o. E. Assessment and Evaluation Unit, The University of Leeds., Trans.). London.
- Tran, H. P., Griffin, P. & Nguyen, C. (2010). Validating the university entrance English test to the Vietnam National University: A conceptual framework and methodology. *Procedia – Social and Behavioral Sciences*, **2**, 2, 1295–1304.
- van Exel, J. & de Graaf, G. (2005). Q methodology: A sneak preview. Retrieved from <http://qmethodology.net/PDF/Q-methodology%20-%20A%20sneak%20preview.pdf>
- Vance, T. F. & McCall, L. T. (1934). Children's preferences among play materials as determined by the method of paired comparisons of pictures. *Child development*, **5**, 3, 267–277.
- Vogt, D. S., Proctor, S. P., King, D. W., King, L. A. & Vasterling, J. J. (2008). Validation of scales from the Deployment Risk and Resilience Inventory in a sample of Operation Iraqi Freedom Veterans. *Assessment*, **15**, 4, 391–403.

APPENDIX A: Example of a response sheet

Please rate the demands of the activities that the candidates have to do to answer each question where 1 represents 'low demands' and 5 represents 'high demands'. Please do NOT use the mark scheme for this exercise.

Demand type (see detailed information)

Rate from 1 to 5, where 1 represents 'low demands' and 5 represents 'high demands'

		Complexity	Resources	Abstractness	Task strategy	Response strategy
Paper 2 questions	Section A Q1					
	Section B Q2					
	Section B Q3					
	Section B Q4					

Cambridge Assessment Qualitative Research Methods Reading Group

Martin Johnson Research Division

In 2007 an invitational research seminar was convened by the Research Division to consider how qualitative research methods could inform researchers' views of assessment. That well-received seminar, which included contributions from Professors Harry Torrance and Helen Colley from the Education and Social Research Institute at Manchester Metropolitan University, suggested a broad interest in this area of thinking, and a desire to explore the issues of the use of qualitative research methods in an assessment context¹.

Since 2011 a series of Research Division-based reading groups have been organised. The remit of the group was initially to bring together researchers from across the Cambridge Assessment group to look at a variety of different qualitative research methods. The initiative was considered to be a useful way of sharing expertise amongst colleagues as well as being an important opportunity to raise awareness of the ways of using qualitative research methods in Cambridge Assessment's own research.

In its first year the reading group consisted of 17 people, and was mainly composed of researchers working in the Research Division. Since then the group has grown to 29 members and includes colleagues (both

researchers and non-researchers) from across the three different parts of the Cambridge Assessment organisation (Oxford, Cambridge and RSA Examinations, Cambridge International Examinations, and Cambridge English), as well as colleagues from other departments of the University of Cambridge (e.g. The Faculty of Education and the Judge Business Institute).

One of the main priorities of the reading group has been for its members to identify different qualitative methodologies which they feel they would like to explore further. As a result, the group has looked at the use of vignettes, focus groups, discourse analysis, stimulated recall, cognitive interviewing, mixed methods, elicitation techniques with young people, working with qualitative internet data, and is planning to look at social network analysis, and interviewing techniques.

A key objective of the reading group is to facilitate the ongoing development of its members' understanding of a diverse set of qualitative research methods. In so doing, the reading group's activity also enhances the capacity for knowledge building across Cambridge Assessment as members use newly acquired methods to gain insights into assessment.

1. More information on the ideas that were discussed at this seminar can be found in Johnson, M. (2008) 3 Rs' of assessment research: *Respect, Relationships and Responsibility* – what do they have to do with research methods? *Research Matters: A Cambridge Assessment Publication*, 6, 2–4.

Statistical Reports

The Research Division

The ongoing 'Statistics Reports Series' provides statistical summaries of various aspects of the English examination system such as trends in pupil uptake and attainment, qualifications choice, subject combinations and subject provision at school. These reports, produced using national-level examination data, are available on the Cambridge Assessment website: http://www.cambridgeassessment.org.uk/ca/Our_Services/Research/Statistical_Reports.

The most recent additions to this series are:

- Statistics Report Series No.39: Provision of level 2 science qualifications in 2011.
- Statistics Report Series No.40: Uptake of ICT and computing qualifications in schools in England 2010–2011.
- Statistics Report Series No.41: Provision of GCE A level subjects 2011.
- Statistics Report Series No.42: Uptake of GCE A level subjects 2011.

- Statistics Report Series No.43: Provision of GCSE subjects 2011.
- Statistics Report Series No.44: Uptake of GCSE subjects 2011.
- Statistics Report Series No.45: Age distribution of GCSE candidates in England 2011.
- Statistics Report Series No.46: Candidates awarded the A* grade at A level in 2011.

Additionally the following reports have been revised, to better reflect the true levels of uptake and provision of GCSEs and A levels in England:

- Statistics Report Series No.28 – revised: Uptake of GCE A level subjects in 2010.
- Statistics Report Series No.34 – revised: Provision of GCSE subjects in 2010.
- Statistics Report Series No.35 – revised: Uptake of GCSE subjects in 2010.

Research News

Conferences and seminars

Cambridge Assessment's 6th biennial conference

The 6th biennial Cambridge Assessment conference took place in Cambridge on 10 October 2012 with 'Examining risk' as the theme. It brought together more than 120 experts from within the education and assessment community with speakers from a wide range of backgrounds to consider the principles of risk and how they might apply to assessment and qualification systems.

The relationship between risk communication and regulation was discussed by keynote speakers Professor Alastair Scotland, former Director of the National Clinical Assessment Service, and Professor Ragnar Löfstedt, Director of the King's Centre for Risk Management, King's College London. Both said that a simpler, clearer, more effective and more accountable regulatory landscape, together with transparent communication, helps to build trust and minimises risk.

Isabel Nisbet, former Ofqual Chief Executive and now of University of Cambridge International Examinations, and Mick Walker, former Executive Director of Education at the Qualifications and Curriculum Development Agency, considered lessons learned from the national curriculum testing crisis of 2008.

Glenys Stacey of Ofqual and Dr Michelle Meadows of exam board AQA explored risk from the perspectives of regulator and exam board. Other panellists included: Amanda Spielman, ARK Schools and Ofqual; Russell Hobby, National Association of Head Teachers; Tim Oates, Cambridge Assessment; and David Skelton, Policy Exchange.

Further details and podcasts can be found at <http://www.cambridgeassessment.org.uk/ca/Spotlight/Detail?tag=6thCAconference>

British Educational Research Association (BERA)

The BERA Annual Conference was held from 4–6 September 2012 at the University of Manchester. Colleagues from the Research Division and CIE presented the following papers:

Irenka Suto, Rita Nádas and Lucy Chambers: *An exploration of how independent research and project management skills can be developed and assessed among 16 to 19 year olds.*

Jackie Greateorex and Sanjana Mehta: *A method for comparing the demands of specifications.*

Rita Nádas, Irenka Suto and Rebecca Grayson: *"Analyse this" – How do teachers with differing subject specialisms interpret common assessment vocabulary?*

Martin Johnson and Beth Black: *Feedback as scaffolding: Senior Examiner monitoring processes and their effects on examiner marking.*

Tim Gill and Irenka Suto: *Students' and teachers' views and experiences of A level module re-sits.*

Victoria Crisp: *The teacher as examiner: How do teachers make judgements when marking coursework?*

Victoria Crisp and Sylvia Green: *The effects of the change from coursework to controlled assessment in GCSEs.*

Jackie Greateorex and Stuart Shaw: *The validity of teacher assessed Independent Research Reports contributing to Cambridge Pre-U Global Perspectives and Research.*

European Conference on Educational Research (ECER)

In September Irenka Suto, Sanjana Mehta and Jackie Greateorex attended the ECER conference in Cadiz, Spain, and presented the following papers:

Irenka Suto: *Why is it so difficult to assess students' research reports? A model of the challenges facing teachers and external assessors.*

Sanjana Mehta: *How effective are curricula for 16 to 19 year olds as a preparation for university? An investigation of lecturers' views.*

Jackie Greateorex: *A method for comparing the demands of specifications.*

Association for Educational Assessment – Europe (AEA-Europe)

The AEA-Europe annual conference took place in Berlin in November with the theme of 'Assessment across the lifespan'. Several colleagues from Cambridge Assessment attended the conference and the following papers were presented:

Tim Oates: *"...I disagree with you at the level of principle...": contrasts and contradictions between assessment in 'academic' and 'vocational' qualifications.*

Beth Black: *The use of statistical approaches in maintaining standards in UK national examinations and the need for expert judgement.*

Rebecca Grayson: *The new A level A* grade in England: The challenge of measuring exceptional pre-university attainment.*

Jill Grimshaw: *Managing the unexpected: contingency planning and coping with crises.*

Stuart Shaw, Martin Johnson and Paul Warwick: *Understanding Assessment for Learning values and practices across diverse contexts.*

The following posters were also presented:

Nicky Rushton: *Comparing students' written performances – does an uneven performance across papers cause problems?*

Tim Gill: *Students' and teachers' views and experiences of A level module re-sits.*

Stuart Shaw and Helen Imam: *International learning and assessment through the medium of English: supporting teachers and students in multilingual educational contexts.*

EARLI SIG 1: Assessment and Evaluation

In August Gill Elliott and Carmen Vidal Rodeiro attended the sixth biennial meeting of EARLI SIG 1: Assessment and Evaluation in Brussels entitled 'Linking Multiple Perspectives on Assessment'. The main themes of the conference were formative and summative classroom assessment, large scale assessment and assessment policy.

Gill presented a paper on 'How multiple perspectives complicate comparability' and Carmen on 'Do different assessment routes (linear vs. modular) prepare students in the same way for further study?'

EARLI SIG 17: Qualitative and Quantitative Approaches to Learning and Instruction

In September Martin Johnson attended the EARLI SIG 17 meeting at the Saxion University of Applied Sciences in Deventer, the Netherlands. The theme was 'Mixed methods for analysing educational interactions'. Martin presented on 'Technologically mediated communication: methods for exploring examiners' real-time feedback interactions'.

Further information on all conference papers can be found on the Cambridge Assessment website: http://www.cambridgeassessment.org.uk/ca/Our_Services/Research/Conference_Papers

Publications

In October a Special Issue of *Research Matters* on validity was published. 'An approach to validation – Developing and applying an approach for the validation of general qualifications' describes a programme of research in which a framework for validation studies of general assessments was developed and applied to two International A level qualifications. For a copy of the Special Issue please email: researchprogrammes@cambridgeassessment.org.uk or visit the Cambridge Assessment website: http://www.cambridgeassessment.org.uk/ca/Our_Services/Research/Research_Matters

The following articles have been published since Issue 14 of *Research Matters*:

- Bramley, T. & Dhawan, V. (2012). Estimates of reliability of qualifications. In: Q. He, & D. Opposs (Eds.), *Ofqual's Reliability Compendium*. Ch.7. Coventry: Office of Qualifications and Examinations Regulation.
- Chambers, L., Galaczi, E. & Gilbert, S. (2012). Test-taker familiarity and paired speaking test performance: Does it make a difference? *Research Notes*, **49**, 33–39.
- Crisp, V. (2012). An investigation of rater cognition in the assessment of projects. *Educational Measurement: Issues and Practice*, **31**, 3, 10–20.
- Crisp, V. & Green, S. (2012). Controlled assessments in 14–19 Diplomas: Implementation and effects on learning experiences. *Evaluation and Research in Education*, **18**, 4, 333–351.
- Crisp, V. & Shaw, S. (2012). Applying methods to evaluate construct validity in the context of A level assessment. *Educational Studies*, **38**, 2, 209–222.
- Johnson, M. & Black, B. (2012). What's going on? Analysing visual data to understand context-based decision making processes. *International Journal of Research & Method in Education*, **35**, 3, 243–250.
- Johnson, M. & Nádas, R. (2012). A review of the uses of the Kelly's Repertory Grid method in educational assessment and comparability research studies. *Educational Research and Evaluation: An International Journal on Theory and Practice*, **18**, 5, 425–440.
- Johnson, M., Hopkin, R. & Shiell, H. (2012). Marking extended essays on screen: Exploring the link between marking processes and comprehension. *E-Learning and Digital Media*, **9**, 1, 50–68.
- Shaw, S. Crisp, V. & Johnson, N. (2012). A proposed framework for evidencing assessment validity in large-scale, high-stakes international examinations. *Assessment in Education: Principles, Policies and Practice*, **19**, 159–176.

CONTENTS : Issue 15 January 2013

- 2 21st Century skills: Ancient, ubiquitous, enigmatic?** : Irenka Suto
- 9 Independent research at A level: Students' and teachers' experiences** : Sanjana Mehta, Irenka Suto, Gill Elliott and Nicky Rushton
- 17 Assessment for Learning in International Contexts (ALIC): understanding values and practices across diverse contexts** : Stuart Shaw, Martin Johnson and Paul Warwick
- 29 Using scales of cognitive demand in a validation study of Cambridge International A and AS level Economics** : Jackie Creatorex, Stuart Shaw, Phineas Hodson and Jo Ireland
- 38 Cambridge Assessment Qualitative Research Methods Reading Group** : Martin Johnson
- 38 Statistical Reports** : The Research Division
- 39 Research News**

Cambridge Assessment
1 Hills Road
Cambridge
CB1 2EU

Tel: 01223 553854

Fax: 01223 552700

Email: ResearchProgrammes@cambridgeassessment.org.uk

<http://www.cambridgeassessment.org.uk>