

The meaning of curriculum-related examination standards in Scotland and England: a home-international comparison

Jo-Anne Baird^a and Lena Gray^b

^a Oxford University Centre for Educational Assessment, Oxford, UK, OX2 6PY.

Email: Jo-Anne.Baird@education.ox.ac.uk

^b Assessment and Qualifications Alliance, Devas Street, Manchester, M15 6EX

Abstract

The ways in which examination standards are conceptualised and operationalised differently across nations has not been given sufficient attention. The international literature on standard-setting has been dominated by the psychometrics tradition. Broader conceptualisations of examination standards have been discussed in the literature in England, which has curriculum-related examinations at the end of schooling. There has, however, been little analysis of conceptualisations of examination standards in Scotland. Different education systems and examinations operate in Scotland and England, and the stated value positions and processes relating to examination standards differ markedly. This paper critically examines policy positions on assessment standards in Scotland and England through the lens of recent theories of standard-setting. By analysing public statements on standards, the paper illuminates similarities and differences in conceptual bases and operational approaches, and examines the effects of these on outcomes for candidates. We conclude that both systems are operationalising attainment-referencing, but with different processes in Scotland and England and these practices do not fit within previous examination standards classifications. As such, the paper moves examination standards theory forward by concluding that there is at least one superordinate definitional category that draws upon more than one definitional stance.

Introduction

The meaning of examination standards is slippery and has been much debated in the literature in England (Baird, Cresswell and Newton, 2000; Baird, 2007; Cresswell, 1996; Christie and Forrest, 1981; Coe, 1999, 2007a, 2010a; Newton, 1997a, 1997b, 2003, 2005, 2010a). Oftentimes stakeholders discuss examination standards without realising they are talking at cross-purposes and using contradictory definitions. The international academic literature on standard-setting¹ has been dominated by the psychometrics tradition, which positions examinations standards as representing an underlying construct, a

¹ We use the term standard-setting to include the annual process, which is sometimes termed maintenance, equating or linking.

“postulated attribute of people, assumed to be reflected in test performance” (Cronbach and Meehl, 1955, 283). The psychometrics tradition is often associated with tests that are disconnected from curriculum, such as the American SATs. Broader conceptualisations of examination standards have been discussed in the literature in England, which has curriculum-related national examinations at the end of schooling and for university entrance. In this article, we continue the debate on the meaning of curriculum-related examination standards. Because this literature has been dominated by the curriculum-unrelated psychometrics tradition and latterly developed in relation to examinations in England, very little has been written that documents and conceptualises the meaning of examination standards in high stakes national examinations. Given the importance of these examinations for people’s life chances, this is a yawning gap in the literature which makes it difficult to understand examination systems internationally and to explain how they operate, how they are interpreted and why certain issues are a matter of public controversy in one country, but not in another. For example, when the results for Higher English declined in 2003, this was seen as evidence for falling educational standards in Scotland, whilst in the same year *rising* A-level pass rates were seen as declining standards in England.

Transparency of the meaning of examination standards matters to a wide range of educationalists, who are trying to enact them through teaching and other practices. Such transparency also matters to a wider group of stakeholders, including candidates, parents, politicians and the media. Each may have their own definition of what it means for standards to be “comparable”. Curriculum-related examinations, which are used for high stakes purposes such as university entrance, provide a number of comparability challenges. A key challenge is deciding which factors are legitimate to take into account when making standards decisions (Baird, 2007, 132-6).

Here, we focus upon a comparison of curriculum-related examination standards in Scotland and England for the school-leaving examinations. We contrast policy statements, standard-setting systems, controversies and cultural positioning of examination standards. Methodologically, we base our argument upon critical evaluation of published policy documents and the authors’ insider experiences of standard-setting in Scotland and England; as insider-researchers (Sikes and Potts, 2008). Both authors have had professional responsibility for standards in an English examining board and one of the authors was formerly responsible for standard-setting policy at the Scottish Qualifications Authority (SQA). Our depictions of the standards policies in Scotland and England have been constructed in part through member-checking (Creswell and Miller, 2000) with senior examining body personnel who are now or were formerly responsible for standards. In this article we present our interpretations of the stated policies

and have also discussed this with those senior practitioner-policy-makers as part of a collaborative contribution to the field (Cresswell and Miller, 2000, 126). Examination boards have a tricky, political task in managing public and stakeholder perceptions of examination standards, so we do not claim that practitioners are necessarily in a position to make the kinds of public critiques of the system that are presented here.

As these countries are part of the United Kingdom, they represent a special case for comparative research, termed a home-international comparison (Raffe, 2007). With many things in common between the countries, such as similar economic settings, home-international comparisons help researchers to understand each system better, clarify alternative policy strategies, compare the impact of alternative policy strategies, understand policy issues better by observing greater variation and understand the processes of educational change (Raffe, 2007). We briefly outline the systems used to set standards in the two countries, but our focus is more upon the meaning of standards in policy documents. Cultural contexts of current examination reforms are considered to help explain the reasons for the different approaches to the meaning of examination standards. Evidence from the outcomes of the examinations is also used to help characterise examination standards. Our focus is on the Scottish system, but we use England as a contrasting case because more has been written about the meaning of examination standards in England.

Whilst some studies have compared the level of standard of the examinations in Scotland and England, that is not our purpose: this article's central theme is how the curriculum-related, school-leaving examination standards are construed, how that differs and what the consequences of this are for the literature. There are many aspects of educational standards that are not the focus of this paper, such as the quality of education provision, the experienced curriculum and equality issues.

Defining Examination Standards

Setting and maintaining examination standards are conceptually distinct, but in practice a new examination's standards are often set with reference to a previous standard. Therefore, much of the literature on the meaning of examination standards relates to the maintenance or comparability of standards across examinations. Over the past twenty years, there have been numerous attempts to set out typologies of examination standards definitions that might be useful within the UK system of curriculum-based examinations. Most articles have argued for the adoption of a particular definition (e.g. Baird, 2007; Baird, Cresswell and Newton, 2000; Coe, 2010; Cresswell, 1996; Wiliam, 1996). Newton (2010a) synthesised this literature and criticised previous work for

confusing methods with definitions. Newton (2010a) focused upon the claims regarding attainment that justify definitions of common standards between examinations, proposing three categories of definition (Table 1), essentially categorising ways of thinking about standards according to the type of factor it is legitimate to take into account within each type of definition. Each type of definition is making a different sort of statement about what students who have achieved “comparable” standards have in common with each other. Definitions that rely upon attributes of students’ performances are *phenomenal* definitions. If a standard-setting system is based on a phenomenal definition of comparability, we would expect policy statements and operational processes characterised by a focus on particular features of student performance. The demands of the questions that students have to answer or the performances students give in examination booklets might be used as evidence for this category of definition. A second, *causal*, category, justifies standards by pointing to evidence that students with similar background characteristics (such as prior attainment) gain similar results. If a standard-setting system is based on a causal definition, then policy statements and operational processes would stress the causes of similarities and differences in attainment. Newton’s (2010a, 2010b) third, *predictive*, category, justifies common standards on the grounds that students with similar grades are able to progress to similar achievements, such as degree results or job prospects, in the future. A standard-setting system based on a predictive definition of standards would show an emphasis on progression pathways and later success, and we would expect discussion of those progression patterns (and success in them) to be the main focus of debates about comparability. We adopt this categorisation scheme due to its focus upon the *kind of attainment factor* at the basis of definitions and its explanatory power to incorporate previous definitional approaches.

Definitional category	Aspects of attainment which are maintained
Phenomenal	features, properties, or dispositions that comprise attainment: similarly graded students should share similar outcomes from learning
Causal	factors that result in attainment: similarly graded students should have had similar inputs to learning
Predictive	potential that is implicit in attainment: similarly graded student should have a similar likelihood of future success

Table 1 Categories of definitions of examination standards (Adapted from Newton, 2010a, 2010b)

Newton’s position is broader than that of Coe (2010), who argued that the fundamental view underlying all conceptions of comparable standards was

‘construct comparability’. That is, Coe (2010) conceived of all standards definitions as essentially being based upon an underlying variable that provides a measure of educational attainment in the subject of the test, for example, proficiency in chemistry. Coe’s (2010) position is consistent with a psychometric methodology and, as such, is in keeping with the philosophy in curriculum-unrelated testing. However, given the wide range of conclusions that test-users in the United Kingdom wish to draw from national examination standards and the purposes to which they put the tests, it is hard to sustain the argument that there is a single construct for each test that is used in all comparisons (Baird, 2010).

Further, some instances of standard-setting indicate that a much broader conception of comparable standards is in use. An illustrative example is the case of the General Certificate of Secondary Education (GCSE) examination in Welsh, which can be taken in two forms; for first language Welsh speakers or second language speakers. The results for 2008 showed that approximately 70% of each group was awarded at least a grade C in the two examinations. Given the different levels of construct attainment (i.e. Welsh language skill) that we would expect these groups of students to reach, it is clear that a conception of comparability other than construct comparability was in use (Newton, 2010b, 47). Psychometric ‘technology’ has arisen from psychological and educational testing. It sits most comfortably with a positivist world-view (Borsboom, 2005), in which constructs exist in people and we go out and measure them with our tests. These constructs then form a basis for conceptualising and deciding standards. Discussion of wider issues regarding the use of test scores does not sit so comfortably within the psychometric tradition. Since our aim is to investigate, present and critique the stated positions on the meanings of standards for particular examinations in Scotland and England, we need a conceptual framework that engages with the application of standards definitions in these contexts.

The categories of definitions in Table 1 are conceptually distinct, but as Newton (2010a, 2010b) pointed out, they are not necessarily tied to a particular methodology. For example, expert judgment on whether attainment is the same could be used to marshal evidence for all three categories, as could statistical techniques. Different methods can be utilised to evidence claims and these could fail to converge on whether standards are the same across examinations. Even working within a definitional category, there can be disagreements regarding whether standards have been maintained due to different nuances in meaning in use. Moreover, national examinations in Scotland and England are expected to be comparable in a range of ways: standards are expected to be comparable over time, between subjects, between qualifications and sometimes between countries, to name a few. Evidence regarding standards over time for a

qualification could indicate year on year comparability whilst data on between-subject comparability could show that standards have not been maintained. Thus, evidence regarding standards can be contradictory for reasons of definition, method or data. Our focus in this article is upon the conceptualisation of standards and we encountered methodological and data issues in applying them to policies and practices in Scotland and England, but we cannot explore those issues fully here. We return to the ontological status of standards later. Before presenting how standards are defined, operationalised and debated in Scotland, we explain aspects of the broader educational culture and give some background information about the Scottish qualifications system.

Scottish education and the Higher Examinations

Education is an aspect of Scottish culture which has long been seen as distinctively Scots by custom and right. It has been uniquely defined since the Reformation, was enshrined as unique in the 1707 Act of Union, and was made a devolved power of the Scottish Parliament in 1999. The commitment to education for all in Scotland dates back to the First Book of Discipline in 1561 (Gray, McPherson and Raffe, 1983, 40). Ideas about Scottish education and assessment are central to Scotland's national identity as a meritocratic society that values educational success for all. This conception of Scottish education has been construed by some as an anthropological myth (Gray, McPherson and Raffe, 1983, xv). Humes and Bryce (2013, 139) summed up the myth and its centrality to Scottish identity as follows.

Belief in the worth and purpose of education is linked to the sense of national identity which is regularly invoked to draw attention to the differences between Scottish and English society. This takes the form of a story or 'myth', shaped by history but not always supported by historical evidence, to the effect that Scotland is less class-conscious than England; that ability and achievement, not rank, should determine success in the world; that public (rather than private) institutions should be the means of trying to bring about the good society; and that, even where merit does justify differential rewards, there are certain basic respects – arising from the common humanity of all men and women – in which human beings deserve equal consideration and treatment.

The persistence of this national cultural positioning can be seen in the Scottish Government's (2015a, x) recent major education policy document, *A draft National Improvement Framework for Scottish Education*. It reads,

We are committed to a fairer Scotland and ensuring that every child is able to develop the knowledge, skills and attributes they will need to flourish in life, learning and work [...]

This matters to us as a nation. Scotland pioneered the idea of universal access to school education in the 17th and 18th centuries and a commitment to universal education has been part of our identity ever since. Excellence in education will be essential to our success as a nation in the future.

Scotland can be argued to have “pioneered the idea of universal access to school education” (Scottish Government, 2015a, 3), but achieving universal access is a separate matter. Some critics argue that the myth is simply untrue (H. Paterson, 1983), that it is a partial truth which masks complexity and prevents debate and change (Gray, McPherson and Raffe, 1983; McPherson, 1983), or that a more nuanced explanation is needed (Finn, 1983; L. Paterson, 2003). Comparisons of socioeconomic status effects upon educational outcomes and life chances between Scotland and England have complex findings. Whilst the *relationship* between socioeconomic status of pupils and educational outcomes at school is similar across the two countries (and stable over time), pupils at any given level of socioeconomic status gained higher educational outcomes in Scotland (L. Paterson and Iannelli, 2007).

One in five children in Scotland is living in poverty (Scottish Government, 2014), with the figure reaching one in three in the largest city, Glasgow (Child Poverty Action Group, 2014); this level of child poverty is similar to that in England (Social Mobility and Child Poverty Commission, 2014, 7). The bottom socioeconomic group had worse outcomes in school level qualifications in Scotland compared with England (L. Paterson and Iannelli, 2007). So the Scottish education system has not been better for the poorest in society’s attainment at school.

The Scottish Higher examination, which was introduced in 1888 (Philip, 1992) is the standard entrance qualification for higher education. Young people in Scotland typically take Highers in their fifth year of secondary school, at around 17 years of age. Students can choose from a wide range of subjects, and typically take four or five (although some may be at lower qualification levels). Until the introduction of *Curriculum for Excellence*, most young people would take qualifications in English, mathematics, and two or three other subjects.

The Higher has been through several reforms (McVittie, 2008), but the most recent changes have a structure that remains similar to that introduced as part of the *Higher Still* development in 1999 (Scottish Office, 1994). Most Highers contain two or three Units which are assessed by teachers on a pass/fail basis, and an externally assessed synoptic assessment which is the basis of the grading for the course (SQA, *undated*). In the new Highers for Curriculum for Excellence, this external assessment is typically split into two components: an examination of up to three hours in length, and a controlled assessment, usually a project

requiring investigation skills and allowing some choice of topic for the student (SQA, 2013). Highers are typically taken after a one-year course. In 2015, approximately 60% of school-leavers² gained at least one Higher (The Scottish Government, 2015b) and 29% gained five Highers (The Scottish Government, 2015c). The typical entry requirement for university entrance is four or five Highers at grade C or above. Grade C has traditionally been seen as a pass.³

Advanced Highers are taken at age 18, in the sixth year of secondary schooling. Whilst possession of Advanced Highers may allow some exemption or advanced standing, Advanced Higher is not itself an entry requirement for Scottish universities. Degrees are four-year courses in Scotland, reflecting the historical emphasis on breadth of study (L. Paterson, 2003; Davie, 1961). When Scottish students progress to universities in other parts of the UK, though, they may find that they must complete two or three Advanced Highers, particularly for entry to the higher status institutions.

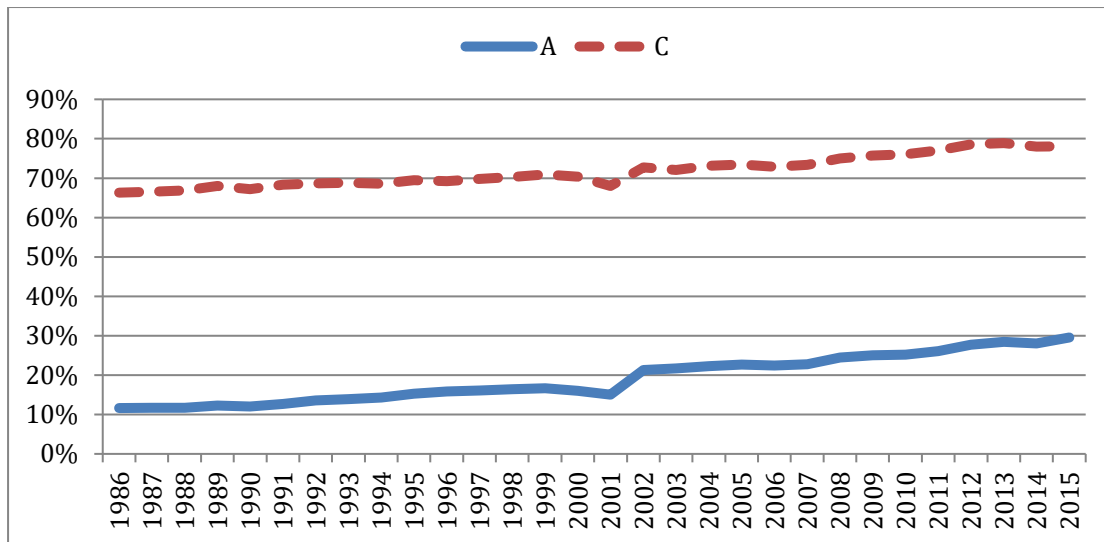
The statistics for Higher examination outcomes that gain most public attention are the outcomes at the level of entries (rather than by pupil), as published by examination boards on results day. These results are part of the context in which examination boards operate and set standards, so we briefly depict them, comparing Higher results with those for A-level and later relating results to how standards play out in each country.

There has been a steady rise in Higher attainment over time (Figure 1), with a slow rise in the Higher pass rate, from 66% in 1986 to 78% in 2015 and a sharper rise in the proportion of entries awarded grade A, from 12% in 1986 to 30% in 2015.

Figure 1 Proportion of Higher examination entries gaining grade C and above and grade A

² Children can leave school in Scotland at ages 16, 17 or 18.

³ For example, see Robin (2014).



Source: Scottish Qualifications Authority <http://www.sqa.org.uk/sqa/64717.4239.html>

Curriculum for Excellence Highers

In 2015, the SQA awarded new Highers developed to support the Government's 3-18 curriculum and assessment initiative, *A Curriculum for Excellence*. The Scottish government responded to the 2002 National Debate by establishing a Curriculum Review Group (Scottish Executive, 2003). The group's report, *A Curriculum for Excellence*, was published in 2004, arguing for transformative change in Scottish education, and emphasising the need to "Make sure that assessment and certification support learning" (Scottish Executive, 2004, 7). In 2008, *A Consultation on the Next Generation of National Qualifications in Scotland* was published, making four major proposals for qualifications reform (Scottish Executive, 2008). In contrast to recent qualifications reform in England, the Scottish reform was positioned as aiming to reduce the reliance on examinations, increasing the focus on skills rather than knowledge, and providing a qualifications system which allowed more personalised programmes to allow individual young people to reach their potential (Scottish Executive, 2008, 11).

The introduction of the new qualifications has not been unproblematic, in line with experience of previous qualifications reform in Scotland (Hayward, 2007; Hayward and Hutchinson, 2013; Hayward, 2015). Reform of examinations is risky and systemic failures have been known to happen, such as in 2000 when the Highers were last reformed (Deloitte and Touche, 2000). However, the failure in 2000 was logistical, not related to standards, as has been the case in other countries, such as in England in 2002 and New Zealand in 2003 (Baird and Coxell, 2009).

The policy context for reform of the Highers was about increasing attainment and skills for all (Scottish Executive, 2008), following an OECD review which found that too many of Scotland's young people left school with few qualifications, and too many failed to be engaged by the general academic curriculum on offer (OECD, 2007). These aims are in keeping with the view of Scottish education as inclusive. The overall pass rate, taking into account students who sat the old Highers and the new examinations (as both were available in 2015) was 74.6%; a rise of 0.4% from the previous year. As such, there was not a dramatic change to attainment for all in Scotland in 2015, but the results were still welcomed. On results day, the SQA, Scottish Government, Directors of Education and Teacher Unions separately publicly welcomed the increased pass rates, each attributing them to the success of their own policies or practices. (ADES, 2015; EIS, 2015; Scottish Government, 2015d; SQA, 2015a). Little dissention from this reaction is evident (although there is some discussion – see *The Telegraph*, 2015a).

Scotland has a single exam board, is a small country and its education system has been characterised by inter-agency partnership working (Hayward, 2007; Humes, 2013). Due to this interconnectedness, it is usually in everyone's interest to support increased pass rates as an indicator of rising standards. Bryce and Humes (2013) analysed the public reactions to pass rates in Scotland, claiming that the following pattern could be observed most years.

- The SQA highlights 'the best-ever levels' of results and asserts that they demonstrate the value of qualifications.
- Government politicians congratulate pupils and teachers on their well-earned achievements.
- Opposition politicians also congratulate pupils and teachers but claim there is public suspicion about a lowering of standards.
- Newspapers express 'fresh concerns' that qualifications are being 'dumbed down'.
- Perhaps a little too defensively, teaching unions point to the robustness of the qualifications framework and state that results inevitably fluctuate a bit from year to year.
- Local authorities claim that the results vindicate their commitments to raising standards and aspirations (with, in 2012, the directors of education in Glasgow and Edinburgh both pointing to better attendance rates in their city schools).
- Many teachers concede that increases in pass rates reflect their growing familiarity with courses and the demands of particular examinations.

(Bryce and Humes, 2013, 9685)

Periodically, overly difficult examination questions become a narrative in the Scottish press and this was the case in 2015 with the Higher Mathematics examination (e.g. *The Telegraph*, 2015b). Philip (1992) gave a historical account of the outrage at the first Higher Mathematics exam in 1888, and similar outrage at the level of demand of all subjects in 1946. These debates have a notion of inclusion at their heart.

Another enduring concern is that of social class equality in attainment. The issue featured in debates about standards in 2015, with the Conservative party education spokesperson noting on results day that “The government must not lose sight of the fact that there is still a significant attainment gap between pupils from poorer and wealthier backgrounds” (The Scotsman, 2015). The issue was raised by all opposition parties at First Minister’s Questions, and by members of the Education Committee in questioning the SQA’s Chief Executive (Scottish Parliament, 2015a, 2015b). A month after results day, the Scottish Government published the *Draft National Improvement Framework for Scottish education*, which set out four priorities for Scottish education, one of which is to close “the attainment gap between the most and least disadvantaged children” (Scottish Government, 2015c, 8).

To summarise, introduction of the Curriculum for Excellence Higher examinations went well, with no systemic failure. Controversies related to lack of attainment for all, but moreso to difficult questions on the mathematics paper. A small rise in overall attainment was welcomed by commentators, as is traditional and in keeping with the cultural position of the education and examination system in Scotland.

The meaning of standards in Scottish Highers

The Education (Scotland) Act (1996, 2), which established SQA, fore fronted the notion of ‘competence’ in educational outcomes, giving SQA the right to,

- (a) determine what it is that a person is required to do and the level of competence he is required to demonstrate in order to attain the qualification;
- (b) determine the means of assessing whether he has done what is required or demonstrated the level of competence required.

Use of the word ‘competence’ is a signal that a criterion-referenced definition (Popham, 1971) of examination standards is in use. Such a definition encompasses the writing of statements regarding students’ necessary examination performances for the award of certain grades. Standard-setting methods which accompany such a definition are typically phenomenal (Newton,

2010a, 2010b), involving the judgment of artefacts such as question papers and students' examination performances by examiners. SQA's standard-setting meetings focus upon the demands of the examination, with most emphasis being placed upon whether the question paper and marking has functioned as intended (SQA, 2015b). This is entirely consistent with a criterion-referencing approach, within a phenomenal definition of standards. Additionally, it is in keeping with the media concern regarding difficulty of questions.

The policy document describing the standard-setting procedure (SQA, 2015b) could apply in large part to practices in Scotland or England. This is typical of such documents internationally, as they are written at a broad level for a national, educational, professional audience. Scores required to be awarded particular grades are termed cut-scores (here 'cut-off scores') or grade boundaries in SQA's nomenclature. The policy indicates that SQA ensures that the cut-off scores set will reflect established standards by using the following sources of information, as appropriate to the qualification:

- performance/grade descriptions
 - exemplar material
 - the assessment and associated marking instructions
 - candidate evidence
 - statistical information (e.g. mark distributions, item statistics, cohort information)
 - other relevant qualitative and/or quantitative information
- (SQA, 2014, 10)

Statistical information is available in standard-setting meetings, but it is not fore-fronted. Of course it would be a mistake to interpret the focus of meetings as a clear indication of the prioritisation of information used in the decision-making process regarding standards and therefore of the definition of standards in use. Preparation for the meeting or the cognitive process of decision-making could prioritise different sources of evidence.

SQA's own published definition of standards is clearly a criterion-referenced, phenomenal definition.

- Qualification standards are where we define the levels of knowledge and skills required to achieve a qualification. We apply these standards to evidence of learner achievement for the purpose of gauging if learners have achieved the requisite level of knowledge and skills (otherwise known as 'level of competence').
- Assessment standards are the levels of demand on learners (how difficult the assessment is for learners taking it) and what the assessment covers.

(Retrieved from <http://www.sqa.org.uk/sqa/74496.1559.html>)

But while policy statements suggest that the definition in use is a phenomenal one, the pattern of results suggests otherwise. If Scotland were truly using such a definition based upon examiner judgments alone, then we would expect results to fluctuate, as fore fronting examiner judgement in standard-setting has this effect (Cresswell, 1997; Baird, 2007). Fluctuations have not happened in Scotland, which means that statistics which are part of the process (SQA, 2014, 10), *must* play more of a role than described in public statements.

In fact, most national examination systems use both statistics and examiner judgment in their standard-setting processes (Cizek and Bunch, 2007, 10). Lack of transparency regarding how various sources of information are utilised in decision-making is also common. Baird, Cresswell and Newton (2000) argued that examining boards had to gauge the values of stakeholders in setting standards and that the prioritisation of various sources of evidence was essentially subjective. In part, this was, they argued, due to the fact that no single source of evidence adequately captured the meaning of examination standards. Coming to an overall standards decision, utilising examiner judgment and statistical evidence, was termed 'weak criterion referencing' (Baird, et al., 2000), though this approach has been renamed a number of times since (see Newton, 2010a). A better term is Newton's (2011) 'attainment referencing', which he defined as follows,

Instead of judging students on the basis of their profile of attainment across sub-domains, in terms of clearly specified performance criteria, they are judged on the basis of their overall level of attainment in the curriculum area being examined.

(Newton, 2011, 20)

Attainment-referencing embraces qualitative information about candidates' performances and statistical information. Although not fore fronted in policy documents, statistics such as numbers sitting examinations in adjacent years, students' prior attainment, whether the schools entering students for the examinations are the same in both years, teachers' estimates of students' grades and so on, have played a role in standard-setting for Highers for some time (Elliot and Ganson, 1998; SQA, 2005). Whilst accepting that evidence, data and definitions are conceptually distinct, these sources of statistical information could not form part of a phenomenal definition of examination standards. They do not tell us about the features, dispositions or properties of students' attainments. Instead, they are related to causes of performances. We return to what this means for Newton's (2010a) categorisation scheme for examination standard definitions in the discussion section below. First, we contrast

Scotland's conceptions and operation of national examination standards with those in England.

England's Advanced Level General Certificate of Education (GCE A-level)

England's A-levels were introduced in 1951, giving a single national examination system for university-entrance selection. At the time of writing there are three examination boards in England (AQA, OCR and Pearson), one in Wales (Eduquas) and one in Northern Ireland (the Council for the Curriculum, Examinations and Assessment). Each of these offers A-level examinations in England in a market.

Students choose from a wide range of subjects at A-level and are not constrained to particular subject combinations, except insofar as universities specify entrance criteria for particular courses. The A-level itself has been through several reforms over its history and is now being changed from a modular format to an end-of-course examination. Each subject is typically assessed through three written examination papers of two hour duration. A-levels are taken after a two year course, typically at age 18. Students normally need three A-levels for entrance to university, with an AAA requirement, or even the top A* grade for the most selective courses. Approximately 40% of the cohort take at least one A-level examination and 29% gain three A-levels.⁴ Degrees are three year programmes in England. Given the differences in the number of subjects taken and the duration of the courses, the comparison between Highers and A-levels is not straightforward. Qualifications frameworks⁵ have been devised to help withdrawing correspondences, but our purpose here is to compare conceptualisations of standards, so to a large extent differences between the examinations are moot.

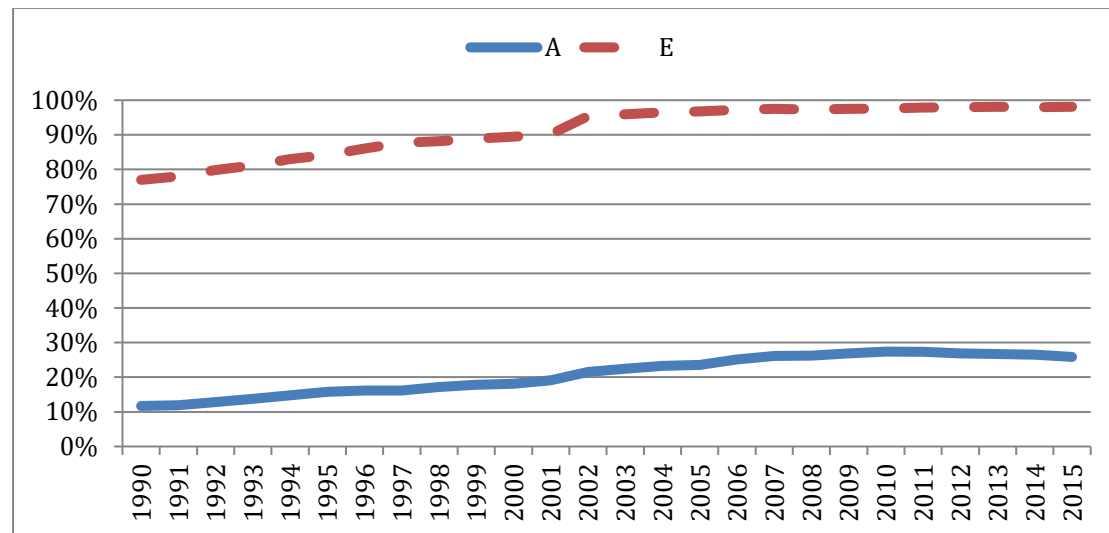
The English A-level examination system is less inclusive than Scottish Highers, with a lower proportion of the cohort entering the examinations. A-levels are graded A* to E, with an unclassified fail grade. Grade E is deemed a pass; the proportion of examination entries graded as passing has risen from 77% in 1990 to 98.1% in 2015 (Figure 2). The proportion of the cohort entered for the examination has also changed, alongside the expansion of participation in education more generally. The proportion of examination entries graded as an A or above has risen from 11.7% in 1990 to 25.9% in 2015 (2). Note that the

⁴ Calculated from Office for National Statistics figures and HEFCE (2015).

⁵ The Scottish Credit and Qualifications Framework (SCQF). <http://www.scqf.org.uk/framework-diagram/Framework.htm> Accessed 6 November 2015. The Qualifications and Credit Framework (QCF). <http://www.accreditedqualifications.org.uk/qualifications-and-credit-framework-qcf.html>. Accessed 14 December, 2015.

proportion of entries passing in A-level is higher than in Highers, due to the assignment of grade E as a pass (as opposed to grade C). However, the proportion of entries awarded a grade A is lower in England than in Scotland.

Figure 2 Proportion of A-level entries gaining grade E or above or grade A (or above from 2010)



Source: Joint Council for Qualifications <http://www.jcq.org.uk/examination-results/a-levels>

As noted above, A-level examinations are being reformed in response to a fear that standards are in decline (e.g. Coe, 2007b; Royal Society of Chemistry, 2008). The concern in England is that examination boards are competing commercially on the basis of standards, attracting customers by awarding grades more generously than competitors (House of Commons Education Committee, 2012). The prototypical media narrative in England on results day is that standards are in decline because examination results have gone up (Warmington and Murphy, 2004). Currently, culture (or at least, the media) in England seems to have an elitist stance regarding examinations, with their selective function being emphasised to a larger extent than in Scotland.

The meaning of standards in English A-levels

Examination boards set standards with reference to the regulator's *GCSE, GCE, Principal Learning and Project Code of Practice* (Ofqual, 2011). This list of sixteen information sources to be taken into account in standard-setting meetings, including qualitative and quantitative evidence and instructions from the regulators. At the time of writing, Ofqual is consulting on a replacement for this Code of Practice; the evidence that they propose to be included remains similar (Ofqual, 2015). Aside from the regulator's instructions, the list is similar to the kind of information used in Scottish standard-setting meetings. However,

standard-setting meetings in Scotland typically involve a single examiner, whilst approximately 5 - 7 examiners are involved in standard-setting in England. This is more inclusive of a wider range of examiners' views.

There was a widespread belief that A-levels were norm-referenced in the past, meaning that no matter the quality of students' performances, a certain proportion were awarded each grade. As Newton (2011, 23) pointed out, norm-referencing does not fit into any of his categories (Newton 2010a, 2010b). Norm-referencing requires a fourth category, that of 'competitive definitions', the underlying rationale being that grades are awarded on a competitive basis and testify to only a relative standard. However, the notion that A-levels were ever norm-referenced is a myth, as the outcomes data show (Newton, 2011). The misapprehension about how standards were set in the past flags the fact that it was under-conceptualised but also lacking in transparency.

Most of the time in A-level standard-setting meetings is spent on script scrutiny. This implies a phenomenal definition of examination standards. As with the Scottish system, much of the practice in England is derived from tradition, with the *Code of Practice* (Ofqual, 2011) changing incrementally. Standard-setting for A-levels in the 1990s was characterised by a predominance of the phenomenal definition and the prioritisation of evidence derived from examiners' expert judgments. However, problems with the precision of and biases in these judgments were found (Baird, 2000; Baird, 2007; Baird and Scharaschkin, 2002; Cresswell, 1996; Forster, 2005; Scharaschkin and Baird, 2000). This led to more emphasis upon statistical evidence in standard-setting.

The Joint Council for Qualifications (JCQ) is an umbrella association in which the examination boards work collaboratively to address industry-wide issues. With the added dimension of a requirement for common standards across examination boards, the JCQ and the regulator conducted a series of studies investigating inter-examination body standards (for a review see Elliott, 2011). These involved both qualitative and quantitative methods (see Newton et al., 2007 for a review). JCQ introduced a statistical comparison technique (Baird and Eason, 2004). The technique was based upon a methodology developed by Cresswell for AQA when the examination board was formed as a merger of the Associated Examining Board and the Northern Examinations and Assessment Board in 2000. AQA had to ensure that the standards were the same across the qualifications from each of the predecessor boards. Ofqual adopted and developed this technique for A-level from 2010 (Ofqual, 2014), terming it 'comparable outcomes' (following Cresswell, 2003). The method operationalises a causal definition of examination standards, as it predicts the outcomes for each syllabus on the basis of the candidates' prior attainment. The assumption is that, for syllabuses with a large number of candidates, the progress between

examinations taken at age 16 and A-level outcomes is likely to be similar from one year to the next. Thus, it is now assumed in policy terms that there will be little change in the overall attainment at A-level from year to year and that is what we have seen in the statistics since 2010.

Systemically, there is a significant issue with joined-up government thinking with the adoption of a purely causal definition in England. Schools are subjected to accountability mechanisms with an expectation that they will improve the examination outcomes of their pupils. Given that the comparable outcomes policy constrains this, Ofqual has given consideration to their operation of the policy and to its formal definition of examination standards, which it states is attainment-referencing (Ofqual, 2014). Newton (2011) argued that attainment-referencing had always been in use, but more emphasis has been given to statistics over the past twenty years, and indeed Ofqual's position is that standards decisions should be formed by weighing all of the available evidence, both quantitative and qualitative, whilst recognising that the application of the comparable outcomes policy,

In practice ... drives the final recommendations for grade boundary marks to be consistent with statistical predictions. (Ofqual, 2014)

Ofqual is grappling with the methodological issues in finding reliable indicators with sufficient precision to make confident claims of changes in outcomes between years.

Discussion

Although the literature is dominated by a curriculum-unrelated, psychometrics tradition, Newton (2010b) synthesised this with the curriculum-related literature on the meaning of examination standards published by scholars working within the English system. The current article has used Newton's (2010a, 2010b) categorisation scheme for examination standards definitions as a basis for analysing the definitions used in practice for Scottish Highers and English A-levels. By applying the scheme to these cases, we have been able to better conceptualise the stated policy positions in each country. Further, the research literature on definitions and methods indicates patterns in outcomes that can be expected from the adoption of particular stances and has allowed us to critically evaluate whether the stated policy positions pertain in practice.

SQA has a phenomenal definition in many of its public policy statements about the meaning of standards in Higher examinations; with standards representing students' competences in the examination. This is in keeping with the wider

cultural desire for inclusiveness and attainment for all in the education system. However, it is apparent from the outcomes data that the definition in use is not purely phenomenal based upon a method of examiner judgment. Indeed, the fairness of purely phenomenal approaches has been questioned in the research literature because of its inconsistency (Cresswell, 1997; Baird, 2007). We conclude therefore that SQA has been operating an attainment-referencing approach, in which they have used both phenomenal and causal definitions. Such an approach will have made defense of standards more robust to challenges mounted from either of these definitions.

England has an explicit attainment-referencing definition for its A-levels, but the use of the comparable outcomes method implies that a causal definition is fore fronted. Policy documents from the regulator discuss the tensions in finding precise and reliable methods for a more phenomenal approach (Ofqual, 2014).

Attainment-referencing might not be as publicly acceptable a standards definition in Scotland as in England. Use of statistics in the standard-setting process has its challenges in England: perhaps this would be seen as even more difficult in Scotland, but for different reasons. Scotland does not have the same accountability regime as England, but given the focus upon meritocracy, the prospect of statistical methods limiting students' attainment would be viewed negatively across the educational and political spectrum. Of course, statistical evidence does not have to constrain attainment as part of an attainment-referencing definition; it depends upon how the statistics are used in the decision-making.

Conclusion

Comparing the two systems has been enlightening, in the sense of Raffe's (2007) home-international comparison. At face value, the standard-setting processes for Scottish Highers and English A-levels appear similar, with the lists of evidence being taken into account looking the same. However, in contrasting these two cases, their different operations have been brought into better relief. In contrast to the English A-levels, the Scottish Highers have higher participation rates, there is wider cultural acceptance of rising outcomes and the standard-setting is positioned as being driven by examiner judgments of the demands of the question paper. Whilst it has been recognised that policy, politics, stakeholder values and wider culture influence standard-setting for national examinations (William, 1996; Cresswell, 1996; Baird, Cresswell and Newton, 2000; Baird, 2007), this has not been illustrated through the comparison of examination standards in two national settings previously. From our analysis, we conclude that despite the different cultural settings in these two countries in

federal Britain, both systems adopt attainment-referencing definitions of standards for their main, high-stakes, university-entrance examinations. There are good reasons for this, as attainment-referencing attempts to carry forward a standard that represents what students know and can do, as well as maintaining stability of the system by considering what similar cohorts have achieved. Not only this, considering both sources of evidence means that an argument can be established for the validity of the standards with respect to potential challenges from phenomenal or causal definitional perspectives. Given this serious advantage, it would be interesting to extend the approach taken in this article to a wider range of examination systems to see whether other national examination systems adopt attainment-referencing definitions. Extension of the comparisons to other countries might challenge our thinking regarding the definition of examination standards further.

Attainment-referencing strives for a culturally and politically acceptable definition of examination standards that is robust to challenges. Scotland and England have distinctive examination traditions, cultures and processes, yet both appear to be attainment-referencing. Different balances are struck, with Scotland's recent public policy statements appearing to be more phenomenal and England's recent policy appearing to be more causal. Acceptability of examination standards arguments differ radically in the two settings. The contexts of examination reforms in the two countries show how the interpretations of examination outcomes differs and illuminates the different environments the examination boards are operating within.

Our analysis was based upon Newton's (2010a, 2010b) categorisation scheme for examination standards definitions, but we found that the definition in use was attainment-referencing in both cases. This raises theoretical matters. First, attainment-referencing is not easily classified using Newton's scheme because, as discussed above, by Newton (2011) and Ofqual (2014), phenomenal *and* causal evidence definitions are in operation in attainment-referencing. Tensions arising in practice from this situation were described in previous work (Baird, 2007). Concentrating on what it means for theory, attainment-referencing is a super-ordinate category of examination standards definition, as it encompasses the phenomenal and causal definitional categories in the two cases presented here. Other forms of attainment-referencing might be evident in other countries and it is possible that other definitional categories and further superordinate categories would emerge from a wider consideration of countries.

References

- Association of Directors of Education in Scotland (2015) Directors Welcome Improving Examination Performance. 4 August. Retrieved from <https://adescotland.files.wordpress.com/2015/08/sqa-diet-2015.docx>
- Baird, J. (2000) Are examination standards all in the head? Experiments with examiners' judgments of standards in Alevel examinations. *Research in Education*, 64, 91–100.
- Baird, J. (2007) Alternative conceptions of comparability. In *Techniques for monitoring the comparability of examination standards*, ed. P.E. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms, 124–56. London: Qualifications and Curriculum Authority.
- Baird, J. (2010) What Constitutes Legitimate Causal Linking?, *Measurement: Interdisciplinary Research and Perspectives*, 8:4, 151–153, DOI: 10.1080/15366361003748219.
- Baird, J., & Scharaschkin, A. (2002) Is the whole worth more than the sum of the parts? Studies of examiners' grading of individual papers and candidates' whole A- level examination performances. *Educational Studies*, 28, 143–162.
- Baird, J., Cresswell, M.J. and Newton, P.E. (2000) Would the real gold standard please step forward? *Research Papers in Education*, 15, no. 2: 213–29.
- Baird, J. and Coxell, A.W. (2009) Policy, Latent Error and Systemic Examination Failures. *CADMO*, 21, 2, 105–122.
- Baird, J. & Eason, T. (2004) *Statistical screening procedures to investigate inter-awarding body comparability in GCE, VCE, GCSE, Applied GCSE and GCSE short courses*. AQA. Accessed from: P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms (Eds.) (2007), *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.
- Borsboom, D. (2005) *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press.
- Bryce, T. and Humes, W. (2013) The Scottish Qualifications Authority. In *Scottish Education*. Fourth Edition: Referendum.
- Child Poverty Action Group (2014) *Percentage of children living in poverty in Scotland*. <http://www.cpag.org.uk/content/percentage-children-living-poverty-scotland>
- Christie, T., and Forrest, G.M. (1981) *Defining public examination standards*. Schools council research studies. London: Macmillan Education.
- Cizek, G.J., & Bunch, M.B. (2007) *Standard setting. A guide to establishing and evaluating performance standards on tests*. California: Sage Publications Inc.
- Coe, R. (1999) *Changes in examination grades over time: Is the same worth less?* Paper presented at the annual conference of the British Educational Research Association, September 2–5, in University of Sussex, Brighton.
- Coe, R. (2007a) Common examinee methods. In *Techniques for monitoring the comparability of examination standards*, ed. P.E. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms, 331–67. London: Qualifications and Curriculum Authority.
- Coe, R. (2007b) Changes in standards at GCSE and A-level: evidence from ALIS and YELLIS. Report for the Office for National Statistics. April. <http://www.cem.org/attachments/ONS%20report%20on%20changes%20at%20GCSE%20and%20A-level.pdf>. Accessed 14 December 2015.

- Coe, R. (2010) Understanding comparability of examination standards. *Research Papers in Education*, 25, no. 3: 271–84.
- Cresswell, M.J. (1996) Defining, setting and maintaining standards in curriculum-embedded examinations: Judgemental and statistical approaches. In *Assessment: Problems, developments and statistical issues*, ed. H. Goldstein and T. Lewis, 57–84. Chichester: John Wiley
- Cresswell, M.J. (1997) Examining judgments: Theory and practice of awarding public examination grades. Unpublished PhD thesis, University of London Institute of Education.
- Cresswell, M.J. (2003) *Heaps, prototypes and ethics: the consequences of using judgements of student performance to set examination standards in a time of change*. London: London Institute of Education.
- Creswell, J.W. and Miller, D.L. (2000) Determining Validity in Qualitative Inquiry, *Theory Into Practice*, 39:3, 124-130, DOI: 10.1207/s15430421tip3903_2.
- Cronbach, L., and Meehl, P. E. (1955) Construct validity in psychological tests. *Psychological Bulletin*, 52, 281 – 302.
- Davie, G (1961; third edition 2013) *The Democratic Intellect, Scotland and her Universities in the Nineteenth Century*. Edinburgh: Edinburgh University Press.
- Deloitte & Touche (2000), *Scottish Executive*,
<http://www.scotland.gov.uk/library3/education/sqar-00.asp>, accessed 20 July, 2007.
- Educational Institute of Scotland (2015) Scottish Exam Results – Pupils and Teachers should be praised. Retrieved from
<http://www.eis.org.uk/public.asp?id=2715>.
- Elliott, G. (2011) 100 years of controversy over standards: an enduring problem. *Research Matters*. Special Issue 2: Comparability, 3-8.
- Elliot, D. and Ganson, H. (1999) SEB Findings on Scottish Achievements. In Scottish Education. Edinburgh: Edinburgh University Press.
- Finn, M.E. (1983) Social efficiency progressivism and secondary education in Scotland, 1885 - 1905. 175 - 196 In Humes, W. and Paterson, H.M. (Editors), *Scottish culture and Scottish Education*. John Donald Publishers Ltd.
- Forster, M. (2005) *Can Examiners Successfully Distinguish Between Scripts that Vary by only a Small Range on Marks?* Unpublished internal paper. Cambridge: Oxford Cambridge and RSA Examinations.
- Gray, J., McPherson, A.F. and Raffe, D. (1983) *Reconstructions of Secondary Education: Theory, Myth and Practice since the War*. Routledge and Kegan Paul, London.
- Hayward, L. (2007) Curriculum, Pedagogies and Assessment in Scotland: The Quest for Social Justice. 'Ah kent yir faither'. *Assessment in Education; Principles, Policy & Practice*, 14, 2, 251-68.
- Hayward, L. (2015) Assessment is Learning: the preposition vanishes. *Assessment in Education; Principles, Policy & Practice*, 22, 1, 27-43.
- Hayward, L. and Hutchison, C. (2013) 'Exactly what do you mean by consistency?' Exploring concepts of consistency and standards in Curriculum for Excellence in Scotland. *Assessment in Education; Principles, Policy & Practice*, 20, 1, 53-68.

- HEFCE (2015) *Young participation in higher education. A-levels and similar qualifications*. Higher Education Funding Council for England.
http://www.hefce.ac.uk/media/hefce/content/pubs/2015/201503/HEFE2015_03.pdf.
- House of Commons Education Committee (2012) *The administration of examinations for 15 – 19 year olds in England. First Report of Session 2012 – 2013. Volume 1*. HC 141-1.
- Humes, W. (2013) Policy Making in Scottish Education. 98-108 In *Scottish Education*. Fourth Edition: Referendum.
- Humes, W. and Bryce, T. (2013) The distinctiveness of Scottish education. 138 - 52 In *Scottish Education*. Fourth Edition: Referendum.
- Knox, J. (1561) *The First Book of Discipline*. Presbyterian Heritage Publications.
 Retrieved from: http://www.swrb.com/newslett/actualNLs/bod_ch00.htm
- McPherson, H.M. (1983) An angel on the geist: persistence and change in the Scottish educational tradition. 216 – 43 in Humes, W.M and Paterson, H.M. (Editors) *Scottish culture and Scottish Education*. John Donald Publishers Ltd.
- McVittie, J (2008) *National Qualifications: a short history. SQA Policy and New Products Research Report 3*. Retrieved from
http://www.sqa.org.uk/files_ccc/PNP_ResearchReport3_NationalQualificationsAShortHistory.pdf.
- Newton, P.E. (1997a) Examining standards over time. *Research Papers in Education*, 12, no. 3: 227–48.
- Newton, P.E. (1997b) Measuring comparability of standards between subjects: Why our statistical techniques do not make the grade. *British Educational Research Journal*, 23, no. 4: 433–49.
- Newton, P.E. (2003) *Contrasting definitions of comparability*. Paper presented at the QCA Standards and Comparability Seminar, April 3–4, in Milton Keynes, UK.
- Newton, P.E. (2005) Examination standards and the limits of linking. *Assessment in Education: Principles, Policy & Practice*, 12, no. 2: 105–23.
- Newton, Paul E. (2010a) Contrasting conceptions of comparability, *Research Papers in Education*, 25: 3, 285 — 292
- Newton, P.E. (2010b) Thinking about linking. *Measurement: Interdisciplinary Research and Perspectives*, 8, no. 1: 38–56.
- Newton, P.E. (2011) A level pass rates and the enduring myth of norm-referencing. *Research Matters*. Special Issue 2: Comparability, 20 – 23.
- Newton, P.E., Baird, J., Goldstein, H., Patrick, H. and Tymms P. (2007) *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.
- OECD (2007) *Reviews of National Policies for Education: Quality and Equity of Schooling in Scotland*. Paris. Publication: 11/12/2007. 172 pages.
<http://www.oecd.org/edu/school/reviewsofnationalpoliciesforeducation-qualityandequityofschoolinginscotland.htm>.
- Ofqual (2011) *GCSE, GCE, Principal Learning and Project Code of Practice*. May.
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/371268/2011-05-27-code-of-practice.pdf. Accessed 14 December 2015.

- Ofqual (2014) *Setting GCSE, AS and A Level Grade Standards in Summer 2014 and 2015*. Ofqual/15/5759.
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/451321/2015-08-05-summer-series-gcse-as-and-a-level-grade-standards.pdf. Accessed 14 December 2015.
- Ofqual (2015) *Consultation on: Improving Reviews and Appeals of GCSE, AS and A level Marking; Withdrawing the GCSE, GCE, Principal Learning and Project Code of Practice; New Requirement on Setting BCSE, AS and A level Grade Boundaries*. Ofqual/15/5807.
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/484076/Consultation_on_marking_reviews_appeals_grade_boundaries_and_Code_of_Practice.pdf
- Paterson, H.M. (1983) Incubus and ideology: the development of secondary schooling in Scotland, 1900 – 1939. In Humes, W. and Paterson, H.M. (Editors), *Scottish culture and Scottish Education*. John Donald Publishers Ltd.
- Paterson, L. (2003) *Scottish Education in the Twentieth Century*. Edinburgh University Press, Edinburgh.
- Paterson, L. and Iannelli, C. (2007) Social Class and Educational Attainment: A Comparative Study of England, Wales and Scotland. *Sociology of Education*, 80, 330 – 358.
- Philip, H.L. (1992) *The Higher Tradition: A history of public examinations in Scottish schools and how they influenced the development of secondary education*. Scottish Examination Board. Dalkeith.
- Popham, J. (Editor) (1971) *Criterion-referenced measurement. An introduction*. Educational Technology Publications. Englewood Cliffs, New Jersey.
- Raffe, D. (2007) *Learning from 'Home-International Comparisons: 14 – 19 policy across the United Kingdom*. 133 – 156. In Raffe, D. and Spours, K. *Policy-making and Policy Learning in 14 – 19 Education*. Bedford Way Papers, London.
- Robin, G. (2014) University of Glasgow Curriculum for Excellence Statement. University of Glasgow. Retrieved from
http://www.gla.ac.uk/media/media_273068_en.pdf.
- Royal Society of Chemistry (2008) *The Five-Decade Challenge. A wake-up call for UK Science Education?* November.
http://www.rsc.org/images/ExamReport_tcm18-139067.pdf. Accessed 14 December 2015.
- The Scotsman* (2015) Attainment gap fears over new and old Higher exams. *The Scotsman* online edition, 4 August 2015.
<http://www.scotsman.com/news/education/attainment-gap-fears-over-new-and-old-higher-exams-1-3848562>. Accessed 14 December 2015.
- Scottish Executive (2003) *Educating for Excellence: choice and opportunity. The Executive's Response to the National Debate*. Edinburgh.
<http://www.gov.scot/Publications/2003/01/16226/17177>.
- Scottish Executive (2004) *A Curriculum for Excellence*. Edinburgh.
<http://www.gov.scot/Publications/2004/11/20178/45862>.
- Scottish Executive (2008) *A Consultation on the Next Generation of National Qualifications in Scotland*. Edinburgh.
<http://www.gov.scot/Resource/Doc/226233/0061255.pdf>.

- The Scottish Government (2014) *Poverty and income inequality in Scotland: 2012/13*. National Statistics Publication.
<http://www.gov.scot/Resource/0045/00454875.pdf>.
- The Scottish Government (2015a) *A draft National Improvement Framework for Scottish Education*. June.
<http://www.gov.scot/Publications/2015/09/7802/2>. Accessed 6 November 2015.
- The Scottish Government (2015b) *High Level Summary Statistics. School Education*. June. <http://www.gov.scot/Resource/0048/00480271.pdf>
Accessed 6 November 2015.
- The Scottish Government (2015c) *Attainment and Leaver Destinations Supplementary Data*.
<http://www.gov.scot/Topics/Statistics/Browse/School-Education/leavedestla/follleavedestat/attainmentandleavers1314>
Accessed 6 November 2015.
- The Scottish Government (2015d) *Results day for students*. Retrieved from
<http://news.scotland.gov.uk/News/Results-day-for-students-1ba7.aspx>.
Accessed 6 November 2015.
- The Scottish Office (1994) *Higher Still: Opportunity for All*. Edinburgh: HMSO.
- Scottish Parliament (2015a) *Official Report: Meeting fo the Parliament 28 May 2015. Scottish Parliament*. Retrieved from
<http://www.scottish.parliament.uk/parliamentarybusiness/report.aspx?r=9973>.
- Scottish Parliament (2015b) *Education and Culture Committee 22 September 2015*. Retrieved from
<http://www.scottish.parliament.uk/parliamentarybusiness/report.aspx?r=10111>.
- Scottish Qualifications Authority (2005) *The Scottish Standard: a Guide to Pass Mark Meetings for National Courses*. Retrieved from
http://www.sqa.org.uk/files_ccc/TheScottishStandard.pdf.
- Scottish Qualifications Authority (2013) *Assessment in the new National Courses and Units*. Retrieved from
[http://www.sqa.org.uk/sqa/files_ccc/Assessment in the new National Courses and Units.pdf](http://www.sqa.org.uk/sqa/files_ccc/Assessment_in_the_new_National_Courses_and_Units.pdf)
- Scottish Qualifications Authority (2014) *SQA Awarding Body Code of Practice*.
[http://www.sqa.org.uk/files_ccc/SQA Awarding Body Code of Practice.pdf](http://www.sqa.org.uk/files_ccc/SQA_Awarding_Body_Code_of_Practice.pdf). Publication Code: FA6723.
- Scottish Qualifications Authority (2015b) *A guide to Grade Boundary Setting*. Retrieved from
[https://www.sqa.org.uk/sqa/files_ccc/A Guide to Setting Grade Boundaries.pdf](https://www.sqa.org.uk/sqa/files_ccc/A_Guide_to_Setting_Grade_Boundaries.pdf).
- Scottish Qualifications Authority (undated) *Design principles for National Courses*. Retrieved from
http://www.sqa.org.uk/files_ccc/DesignPrinciplesNationalCourses.pdf.
- Scharaschkin, A., and Baird, J. (2000) The effects of consistency of performance on A level examiners' judgements of standards. *British Educational Research Journal*, 26, 343–357.
- Sikes, P. and Potts, A. (2008) *Researching Education from the Inside. Investigations from within*. Routledge, Abingdon, UK.

- Social Mobility and Child Poverty Commission (2014) *State of the Nation 2014: Social Mobility and Child Poverty in Great Britain*. Presented to Parliament pursuant to section 8B(6) of the Child Poverty Act 2010. October.
- The Telegraph* (2015a) Questions over higher Scottish exam pass rate. The Telegraph online edition, 4 August 2015. Retrieved from <http://www.telegraph.co.uk/education/secondaryeducation/11781251/Questions-over-higher-Scottish-exam-pass-rate.html>. Accessed 14 December 2015.
- The Telegraph* (2015b) *Can you answer the tricky crocodile maths question, that left students 'in tears'*. 9 October. <http://www.telegraph.co.uk/education/educationnews/11921671/Can-you-answer-the-tricky-crocodile-maths-question-that-left-students-in-tears.html>
- Warmington, P. and Murphy, R. (20014) Could do better? Media depictions of UK educational assessment results. *Journal of Education Policy*, 19, 3, 285 – 299.
- Wiliam, D. (1996) Standards in examinations: a matter of trust? *The Curriculum Journal*, 7, 3 293–306.

Notes on contributors

Jo-Anne Baird is the Pearson Professor of Educational Assessment and Director of the Oxford University Centre for Educational Assessment, UK. Jo-Anne was formerly Professor of Education and Coordinator of the Centre for Assessment and Learning Studies at the University of Bristol, UK and has also held the positions of Head of Research at the Assessment and Qualifications Alliance, UK, Lead Editor of *Assessment in Education: Principles, Policy & Practice* and President of the Association for Educational Assessment-Europe. Her research interests include examination standards, policy and systemic aspects of assessment, e-assessment and human judgment in assessment.

Lena Gray is Head of Research in the Centre for Educational Research and Practice at AQA. Lena previously spent many years working in various roles at the Scottish Qualifications Authority (SQA), latterly as Head of Policy, Assessment, Statistics and Standards. She also has experience as a secondary teacher of English. Her research interests include examination standards, qualifications policy and design, and assessment of vocational qualifications.