

## How policy formation and implementation interacts with risks to high stakes qualifications

Stuart Cadwallader and Kate Tremain

### Abstract

The challenges associated with the formation and implementation of national qualifications policy are well documented (e.g. Baird & Lee-Kelley, 2009; Oates, 2008; Richardson, 2007). This paper elaborates on the literature by drawing on interviews with twenty-three key stakeholders from across the qualifications system. The analysis suggests that the stakeholders' major concerns centre on ensuring that new or reformed qualifications are developed to a sufficiently high quality in the context of pressurised timescales and a challenging policy environment. Issues associated with policy development and implementation have the potential to permeate the qualifications system and introduce risks to accurate administration and assessment. These risks are discussed with reference to the existing research literature and in the context of the current reforms.

*Keywords: risk, national qualifications policy, qualification reform.*

### Introduction

Richardson (2007) describes education in liberal democracies as “pregnant with values” and “inherently controversial” (p.149), arguing that it is salient in the public consciousness because it is, or has been, a part of the everyday experience of almost the entire populace. For this reason it could be argued that education policy is highly malleable – policy makers can aim it towards any number of social objectives, all of which may inspire the support of different cross sections of the electorate at different times. Given this political centrality there are frequent changes to national qualifications policy in England, both minor and sweeping, as the government attempts to influence education through qualifications<sup>1</sup>. It is necessary for the qualifications system to be adaptive if it is to successfully realise policy objectives and provide effective qualifications with efficient and reliable assessment.

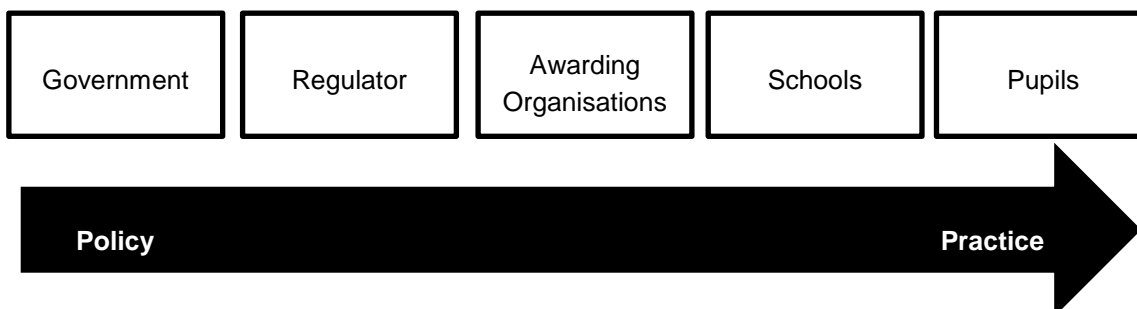
Baird & Lee-Kelley (2009) describe the mechanisms through which national qualifications policy is translated in to practice, using the specific example of how new qualifications are developed. Essentially there are three broad stages. Firstly, the government formulate a policy at a high level and send a remit letter to the regulators who inform the awarding bodies of the intended change. Next, the regulators develop qualification and subject criteria, essentially providing the rules to which new qualifications (developed by the awarding bodies) must adhere in order for the policy to be successfully realised. Finally, with these rules in place, awarding bodies develop new subject specific syllabuses (specifications) in consultation with various qualification users (e.g. teachers, employers and higher education institutions), submitting their proposal to the regulator for approval. Baird & Lee-Kelley (2009) place the key actors in the process on a

---

<sup>1</sup> See Tattersall (2003) for a discussion of how policy makers have become increasingly involved with qualifications and examinations in England since the 1970s.

continuum which ranges between 'policy' and 'practice'. Government is at the 'policy' end of the continuum, followed by the regulators, the awarding bodies, and finally schools, teachers and students are at the 'practice' end of the continuum (see Figure 1).

**Figure 1: Key stakeholders in the qualification system on a 'Policy to Practice' continuum**



Even presented in this distilled and simplified form, it is clear that the process of converting policy into practice is a complex one that involves many stakeholders, all of whom are likely to hold differing perspectives. Several authors have discussed how changes to policy can interact with the development, administration and assessment of the qualifications themselves, causing significant challenges to their successful delivery (Baird & Lee-Kelley, 2009; Oates, 2008; Richardson, 2007). An error at any stage of the development or administrative process has the potential to cause a major failure of the system. However, logic would dictate that problems with the qualification content or assessment structure that are built in from the beginning would be most problematic. Such underlying issues; introduced when policy is formulated and first implemented, are likely to introduce risks that are more difficult to mitigate and have more widespread ramifications should they be realised.

Baird & Coxell (2009) discuss the relationship between policy formation and 'latent errors'. Latent errors are flaws in the overarching structure or delivery processes of a qualification which have the *potential* to lead to tangible systemic failures. Essentially, they are the fault lines in the system. Failure along these fundamental lines could lead to problems such as candidates receiving the wrong grades, examinations not being administered correctly or perhaps more fundamental issues such as the qualification being inappropriate for the intended purpose. Baird & Coxell (2009) use case studies of three systemic examination failures<sup>2</sup> to exemplify a number of aspects of the policy development and implementation process that can produce latent errors. Their findings suggest that a key factor is that new policies can be initially unclear, making it difficult for clear pathways to implementation to be planned. This is because policy is often fluid, evolving based on public reaction and consultation with stakeholders. Such lack of clarity and stability has the potential to lead to the coexistence of differing understandings or interpretations of what a new or reformed qualification should look like. If expectations regarding content, structure and the approach to assessment are not sufficiently clear and detailed the resultant qualification may fail to meet the policy objectives or be riddled with problems regarding administration and assessment.

The introduction of new GCE A-level qualifications as part of curriculum 2000 (see HoC, 2003) provides an additional example of how latent errors, introduced through flaws in the translation of policy into practice, can result in serious failures of qualification delivery. A-levels were reformed to allow candidates to gain either an advanced level qualification (A-level) or a qualification based on half of the full content (Advanced subsidiary – AS level). This was a significant change and there were a number of issues in the first year of the new qualifications'

<sup>2</sup> Baird & Coxell's (2009) case studies were the 'Higher Still' examinations in Scotland (2000), the scholarship examinations in New Zealand (2004) and the national curriculum tests in England (2008).

availability. For example, the increase in the volume of assessment was challenging for the awarding bodies to manage and the inadequate adaptation of administrative computer systems led to some candidates receiving inaccurate grades (Richardson, 2007).

The major crisis point came in summer 2002, the first point at which candidates could certificate with a full A-level. The overall A-level pass rate increased by 4.5%, which had been anticipated due to the modular structure of the new qualifications. However, this general increase was incongruent with results at some independent fee-paying schools where teachers observed an increase in the number of pupils who had failed outright (receiving a 'U' grade). This led to major concerns from teaching professionals over the reliability of grades and whether standards, which were meant to match those of the previous incarnations of A-levels, had been adequately maintained (Richardson, 2007).

An inquiry by Tomlinson (2002) led to alterations to the grade boundary decisions that had been made by one of the awarding organisations and this resulted in the re-grading of nearly two thousand candidates. This severely affected the transition to university for a minority of pupils and widely shook public confidence in the accuracy of grading. The Tomlinson report concluded that the lack of a clear and consistent understanding about what the grading standard should be had caused the problems. This was a product of how the policy was implemented rather than any negligence or wilful manipulation from examiners, examining boards, the regulator or government (Guardian, 2002). The conclusion was that implementation had been rushed, particularly with regard to the process for identifying, evaluating and mitigating risk around the setting of standards and awarding of grades (HoC, 2003). In the dash to prepare the qualifications for ambitious deadlines it seems that latent errors were established within the assessment structure.

At the time of writing, the qualification system in England is undergoing significant changes on a scale even greater than the curriculum 2000 reforms. The key national qualifications of GCSEs and A-levels are both being reformed such that they will be first taught in either 2015 or 2016 (DfE, 2013a; Telegraph, 2013). The main thrust of the GCSE reform is to make the qualifications more demanding, such that the expectations placed on pupils in England match or exceed those that are placed on pupils in the countries or jurisdictions that perform the best in international comparisons (Gove, 2013a). Within this remit it has been stated that assessment will be linear (examinations will be taken at the end of the course rather than at various points throughout, as with a modular system) and will have a new grading scale to reflect the change in expectations. The A-level reform focusses on the need to equip pupils with the necessary skills and knowledge for higher education, with the initial intention being that the Russell group of universities will advise awarding organisations on the content of the new syllabuses (Gove, 2013b). As with GCSE reform, a linear assessment structure will be introduced.

These reforms are clearly very large in scope and provide a significant implementation challenge. Are we in danger of introducing latent error into these qualifications and repeating mistakes of the past? This paper seeks broadly to explore the high level threats to the qualifications system in the current context of policy formation and implementation. It draws on the perspectives of key stakeholders who mostly occupy the middle of the policy to practice spectrum: the regulator, the awarding bodies and teachers. The approach taken is similar to that of Baird & Lee-Kelley (2009) who use interviews to provide a useful 'at the coal face' account of how managers from awarding organisations, the regulator and government work to translate and operationalise national qualifications policy into practice. The design of this study will not facilitate sufficient depth of analysis to delve into the minutiae of the implementation process. Instead it reports a broad and contextualised overview of the significant risks, providing an informative snapshot of the challenges faced by the qualifications system as it strives to implement large scale change. It is hoped that this will help to identify areas in which to focus improvements to the development and delivery of future national qualifications in England.

## Interviews

### *Participants*

Data were collected from eighteen one-to-one interviews and one focus group with five participants. The participants were key stakeholders in the general qualification system, comprising senior managers from two major English awarding bodies and their regulator, senior examiners, teacher representatives and academics working in the field of education and assessment. Given that the targeted participants were senior within their organisations it was necessary to take an opportunistic approach to sampling. Potential interviewees were identified by the researchers and invited to participate by email. In most cases participants were acquaintances of the research team, though the researcher who predominantly carried out the interviews had only previously met two of the interviewees. The broad roles of the participants are summarised in Table 1.

**Table 1: Background of key stakeholders participating in interviews**

Stakeholder background	No. participants
Regulator	4
Awarding body A	5
Awarding body B	3
Senior examiner	2
Academic	3
Teacher <sup>3</sup>	5
Other	1
<b>Total</b>	<b>23</b>

### *Data collection*

Two interviews were used for piloting purposes, serving loosely to define the approach and use of terminology which was to be employed during subsequent data collection. The data from the pilot interviews were included in the analysis, which was considered appropriate given that the approach did not change substantially following the pilot. Participants were encouraged to shape the direction of their interviews but two main questions<sup>4</sup> were used to initiate and scaffold the discussions:

1. What do you believe to be the three biggest high level threats<sup>5</sup> to the safe delivery of exams?
2. What do you believe to be the biggest threats to public confidence in exams?

As is evident from these questions, the initial scope of the research was to capture broader notions of high level threats and risks to the qualifications system. The issues surrounding policy formation and implementation were by far the most salient in the data and therefore became the focus for this paper.

The majority of the data was collected over a period of three months and all of the interviews were carried out in face to face contexts (with one exception, a video conference). The duration

<sup>3</sup> The teachers were jointly interviewed as part of a focus group.

<sup>4</sup> The researcher also had other questions and prompts available but did not use these with all participants.

<sup>5</sup> 'Risk' and 'Threat' are used interchangeably in this paper. During the pilot interviews it became apparent that using these terms in tandem elicited a broader and richer response from our interviewees. This is likely due to the term 'risk' being deeply embedded in the discourse around specific operational issues.

of the interviews ranged approximately from twenty to sixty minutes and in all but one case they were conducted by the same researcher (a different researcher conducted one of the interviews for practical reasons). In all cases an audio recording of the interview was taken which was later fully transcribed. These transcriptions were the main source of data for the analysis but the audio files were also available during analysis for cases where the researchers sought further clarity over the context or tone of a particular excerpt.

The period during which this study was conceived and carried out was one of substantial turbulence for the English qualifications system. There can be little doubt that this had a significant bearing on the interviews. The key events were as follows:

- In **April 2012** the regulator for examining bodies in England (Ofqual) completed a report criticising teacher support seminars which were operated by the awarding organisations (Ofqual, 2012a).
- In **June 2012** a consultation on the reform of A-levels was released (Ofqual, 2012b).
- Also in **June 2012**, plans for the reform of qualifications at Key Stage 4 (KS4) were leaked from the Department for Education (DfE) to the press (Mail Online, 2012).
- In **August 2012** there was enormous controversy surrounding the results for GCSE English (Telegraph, 2012), which later resulted in a judicial review (BBC, 2013b).
- In **September 2012** the consultation into reform of KS4 qualifications was officially launched by the regulator (Ofqual, 2013a).

This context was almost certainly influential and is likely to have affected the perceptions of participants differentially depending on precisely when they were interviewed. It is arguable that if the interviews had been completed at a different time than they may have solicited different data. Our own perspective is that the context serves to enrich and enliven the data by providing contemporary perspectives that are embedded in the qualification and assessment system at a time of significant upheaval.

#### *Data analysis*

Although every effort was made by researchers to avoid prescribing their views and perspectives when interpreting the data, it is important to note that the researchers are themselves stakeholders. They are employed by an awarding organisation and are inescapably influenced by their perspective. Though this is an issue which should not be side-stepped, it is also important to point out that the approach to analysis was designed to minimise potential biases and to allow the data to speak for themselves as far as possible.

The analysis was divided into three broad stages. During the first stage, two researchers independently scrutinised each of the transcripts line by line and generated a series of descriptive or analytic categories to describe and code the content. Once this initial coding was complete, the researchers met for stage two to discuss commonalities and differences between the emerging categories, whether and how codes could be reasonably fused or split, and possible relationships between categories. It was hoped that the inclusion of two researchers in the process would allow categories to be challenged, agreed and finessed as necessary. A series of themes which the researchers felt best represented the data were agreed.

The researchers then scrutinised the transcripts for a second time, looking for examples of the themes they had identified and, crucially, for examples where the data contradicted or undermined a theme. At the end of stage three the researchers again discussed the themes and the processes by which they had been constructed. It became clear that an overarching theme, which can be divided into sub-themes, was predominant in the data and therefore forms the nucleus of this paper. The risks to the qualifications system that are associated with education policy formation and implementation were raised in some form by all participants and permeated the data throughout, linking with many elements of what the participants conveyed.

## Findings & Discussion

The overarching theme which emerged from the analysis is that the education policy development and implementation process introduces risks to the qualifications system. The findings are discussed under four headings, each representing a broad subtheme which emerged from the data. These subthemes are:

- Challenging timescales for policy implementation
- Stakeholder involvement in the shaping of qualifications
- The wider use of qualifications
- The media and public confidence in the examination system

Each of these subthemes is now unpacked, described and discussed. They are illustrated by direct quotations from the data and expanded on using the wider research literature.

### Challenging timescales for policy implementation

#### *The rapid pace of change*

Almost all of the interviewees stated that the rapid pace and high frequency of change to national qualifications policy has the potential to undermine both the quality of the qualifications and the efficiency and accuracy of their delivery. This issue is the subject of a significant body of research literature (Baird & Coxell, 2009; Baird & Lee-Kelley, 2009; Oates, 2008). Government is compelled to implement policy rapidly because elections take place every five years. They must demonstrate to the electorate that their policies have a positive impact within this timeframe. Participants in this study stated that politically prescribed schedules can compromise important phases of development such as planning and piloting:

*The political imperative to try to get things done quickly... if you push ahead too strongly on that then you can see that safe delivery might be quite difficult and, rather than ending up with a whole load of better A-levels, we might end up with some that aren't half as good because we've rushed them. (Regulator)*

Note that this is distinct from suggesting that the threat stems from low quality policy. Specific governments, politicians and initiatives were discussed by participants in terms of their strengths and weaknesses but judgements about their relative value were mixed. For example, the decision to revert to a linear assessment system was viewed positively by some participants but was a cause of concern to others. A need to review and develop policy was accepted and there was an acknowledgement that changes to policy could have either a positive or negative impact on the qualifications system and education more widely.

*So if you take the current Government, I absolutely believe that they believe (if that makes sense) that what they're doing is to improve the chances for students coming out of the UK educational system. (Awarding Body A)*

The universal issue was that the *process* of implementation, particularly with regard to the timescales, was problematic. The excerpt below captures this nuance:

*There's a risk, yes, that it might not be a very good policy but that's usually negotiated quite well between the people who put it in place and the regulator so the awarding bodies usually find a way of turning what might at first glance appear to be a duff policy into something that's at least half workable... there's kind of a caveat there which is, in order for that buffer to work effectively between policymakers and*

*policy implementers, there's got to be a period during which ideas are tested and that's a big problem with rapid policy implementation in that it doesn't allow sufficient time. (Academic)*

Part of the notion captured in the quotation above is that the regulator and the awarding bodies mediate the impact of changes to policy by essentially making the policy practical. They sculpt the policy into something that can be practically implemented. Baird & Lee-Kelley (2009) found that the managers interviewed in their study were highly pragmatic about implementing policy. Their participants suggested that there was not a clear enough distinction between policy and implementation and that this contributed to managers having difficulty in adequately defining what their actions should be. This contradicts one of the assumptions of many academic models of policy implementation: that those managing the transition know what to do in order to operationalise the policy objectives (Schofield, 2004). In reality, policy *formation* appears to be an on-going process which coincides with *implementation*. As a result the approach is largely one of containment – to adhere to the general overall direction of the policy objectives but to make them feasible by avoiding unnecessary change and minimising disruption to existing structures and processes (Baird & Lee-Kelley, 2009). In other words a fairly conservative approach is used in order to successfully deliver the changes within the imposed timescales.

#### *Implementing policy: qualification quality*

It could be argued that such enforced pragmatism can be damaging to the quality of qualifications. Some of the interviewees suggested that working in such an environment undermines the capability for stakeholders to improve qualifications:

*If you could remove that political interference, I think that the changes would be very slow and I think they would be natural evolution that would lead to improvements. (Awarding Body B)*

The issue may be that there is incongruity between the time allotted for policy development relative to that allocated for development of the qualification itself. Oates (2008) reasons that the increased risk of system failure associated with reform is not caused exclusively by tight deadlines but also by a lack of synchronisation with regard to the various phases of policy implementation (Oates used the term 'temporal discontinuity' to describe this). He suggests that most failures stem from a breakdown in the sequencing of design, piloting, evaluation and implementation. In other words, some parts of the process are allotted a reasonable period of time while others are not.

A-level reform in 2010 provides a good example of this. Three months were available for the crucial process of syllabus and question paper design, a period which was only five *per cent* of the total time between the publication of the original White paper to outline the policy change in 2005 and the first A-level examination in 2010 (Baird & Lee-Kelley, 2009). The problem was not a lack of time in itself but an inappropriate allocation of time between phases that placed strain on a key stage of implementation. It is important to add that both Oates (2008) and Baird & Coxell (2009) stress that failures associated with policy implementation are not caused purely by the demanding deadlines imposed by politicians and civil servants. Assessment professionals are part of the political network that instigates reform and how they manage the process and negotiate timescales accordingly is crucial. This paper will discuss how stakeholders who operate towards the practice end of the continuum can influence policy and its implementation in further detail in a later section.

#### *Implementing policy: qualification delivery*

Distinct from concerns that the challenging deadlines can threaten the *quality* of the qualification is that they can also threaten the *delivery* of the qualification by undermining processes of

administration and assessment. During the policy development and implementation process the focus is on the development of new qualifications and planning how they will be delivered. This is incredibly demanding on personnel and resource given the complexity of the process. The quote below illustrates the level of resource that is ploughed into managing change:

*What I think is soul-destroying is the effort that goes into managing change in our system. So if you think of all of that policy process, how many man hours that will soak up, all the discussions that take place, the documents that need to be written, then they need to be re-written into actual practice documents, the changes to computer systems, the retraining of everybody, the rolling out of different exam papers, the regulation of all of this that goes on. (Academic)*

The risk is that experienced senior personnel and key resources are directed away from the general operational processes involved in running and improving the examination system (for example, developing assessment and optimising marking procedures). This introduces the potential for either failure in the existing system or the introduction of latent error into the new one because attention is divided and resources are thinly distributed:

*Well I mean you need that basic resource there looking after the delivery to ensure that it doesn't go wrong. When your eye is off the ball, that's when things can go wrong. (Awarding Body A)*

Several of the participants suggested that the negotiation of change creates an environment which increases the likelihood of certain systemic failures in this way. This observation hints at an indirect risk: by focussing so intently on the future we may end up neglecting the present.

### **Stakeholder involvement in the shaping of qualifications**

#### *Stakeholder influence over policy*

The analysis so far paints a picture of a very hierarchical system in which stakeholders responsible for implementing policy are highly reactive to the actions of policy makers but have little influence on the policy itself. This is a simplification as in reality there is interdependency between stakeholders at different levels of the process (Baird & Coxell, 2009). Much of the qualifications and assessment expertise resides with awarding bodies and schools and it is therefore important that these stakeholders have an active role in policy formation:

*We can help policymakers because we've got policy memory and we've got evidence from the past and we can use that evidence to help policymakers make better policy. (Awarding Body A)*

The data suggest that the mechanisms by which implementers influence policy are imperfect. The expertise of policy implementers needs to be better used to actively assist in the formation of policy rather than just for the practicalities of implementation. Policy formation can be described as a 'murky process' (Baird & Lee-Kelley, 2009, p. 63) - it is difficult to delineate because there are myriad ways in which policy can be influenced and a large number of factors, groups and individuals that may be directly or indirectly involved. The Ofqual and DfE consultations provide a sense of the scale with regard to the sheer number of people who hold a stake in qualifications and have something to contribute to policy (see Ofqual, 2013a). Some of the interviewees suggested that the relationship between awarding bodies and policy makers had become less passive than had been the case in the past, with the policy implementers becoming more influential during policy formation and planning.

*I think the exam boards...have obviously recognised this and really lobby Government (as) a big part of their activity now. (Academic)*



The role of Ofqual in brokering the relationship between government and the awarding organisations is considered crucial if policy implementers such as awarding organisations and teaching professionals are to engage successfully with policy makers and influence the development of policy:

*If it is going to be a regulated market, and clearly our society says it should be, you actually need a strong regulator that's prepared to be independent of Government and that's quite a hard thing for them to do... They have to have people with enough expertise... and they have to have leadership that is brave and strong enough to actually stand up against Government policy if Government policy is wrong. Now for someone running an independent but Government-funded body, that is a test. (Awarding Body A)*

Since this data were collected it would appear that Ofqual have had some success in effectively communicating risks to the Department of Education. For example, the proposed timescale for A-level reform originally suggested that the new qualifications would be first taught in September 2014 but this was postponed until September 2015 under the advice of Ofqual (Gove, 2013b; Telegraph, 2013). There is also evidence that the regulator, under advice from awarding bodies and other stakeholders, have influenced government policy regarding assessment. With regard to tiered assessment, for example, the original policy steer was to eradicate the use of tiering but this was adjusted to allow some exceptions following mediation from the regulator (Ofqual, 2013a). The mechanism by which stakeholders communicate with government to shape policy development seems to have functioned effectively in these cases, at least to some degree.

However, there are also cases where the regulator and other stakeholders have been unsuccessful in influencing policy. For example, stakeholders robustly challenged the decoupling of AS qualifications from A-levels on the grounds that the current system functions well and promotes fairer access to universities (BBC, 2013a). This advice, dispensed via a consultation which was carried out by the regulator, appears to have been unheeded (Gove, 2013b). The Chief Regulator's letter to the Secretary of State for Education appears to suggest that the structure of a qualification is a matter which is beyond Ofqual's remit, unless the proposed structure presents a risk to the maintenance of standards:

*We take the view that decisions about structure are curriculum policy decisions for Government, provided they do not affect our ability to meet our objectives. We have considered – as we should – your policy intent, the responses to the 2012 A level consultation and other relevant factors including our equality impact assessment, and have concluded that we should act consistently with the policy you have set out. Your decisions do not jeopardise our ability to maintain standards and meet our wider objectives. However, they may affect the volume of assessment and the costs of qualifications. We will keep you informed as we explore this, and discuss options. (Stacey, 2013, p.2)*

The extent to which the regulator tried to influence policy regarding the decoupling of AS and A-levels is unclear from these letters but this excerpt raises an interesting question regarding the role of the regulator. To what extent should the regulator mediate apparent disagreements between policy makers and other stakeholders? It could be argued that Ofqual should have provided a clearer message to policy makers that the decoupling of the qualifications was strongly contested within the stakeholder community and may be associated with risks beyond those regarding examination standards and the other factors that are contained within their scope as regulator.

### *Stakeholder influence over curriculum development*

In contrast to the more positive view that there were avenues through which policy implementers could influence policy makers, some participants suggested that the increasingly political approach to curriculum design has slowly stripped expertise away from awarding organisations and teachers. This demonstrates a tension within the data, with some interviewees believing that policy implementers were becoming better at engaging with policy makers but others stating that a gradual sapping of expertise had reduced their capacity to shape policy. The latter group suggest that as curriculum has become an increasingly political arena the awarding bodies have shifted their focus towards the processes and logistics of exam delivery. Tattersall (2003) describes how the 1988 Education Act, which essentially made the centrally controlled national curriculum compulsory, had a profound impact on the awarding organisations. Its introduction had essentially fractured the 'seamless process from learning (the design of syllabuses) to evaluation (examination and grading)' (Tattersall, 2003, p.17).

The effects of this dislocation may still be being felt today. Some interviewees identified this as a significant risk as expertise was being lost and awarding organisations were becoming increasingly disempowered to challenge and influence policy. This was perceived as stifling for innovation in qualification and assessment design and there was concern that the increasingly process driven nature of awarding organisations was eroding their link with schools. In other words the system may be becoming more reactive and fatalistic.

*I think the exam boards have, because curriculum was centralised, lost a lot of power in terms of deciding what the curriculum would be and so that affected the kind of staff they employed and so on... So that they all got pretty good at delivery with their electronic marking systems and all the rest of it but actually that consideration of what the changes in the exams were doing to learning was really... had become something external to the exam boards themselves.*  
(Academic)

An increase in expertise in the administration of qualifications may have reduced the likelihood of risks associated with delivery occurring but it has arguably embellished a more impactful risk: the risk that qualifications will not be fit for purpose. Awarding bodies may have become efficient at producing and administering qualifications and assessment but they have reduced their capacity to engage with the teaching community and to innovate in a way which is educationally beneficial for learners. There seems to be a threat that awarding bodies are prioritising reliability over validity by focussing on developing efficient methods of administration while allowing policy makers to define the content. They are not helping to mediate the relationship between teachers and policy makers in the way that they used to.

The risk that curriculum has become too hierarchically defined by policy makers and is therefore not being adequately informed by practice was discussed by interviewees from awarding organisations but also by the teachers. The perception is that less agency is being given to teachers; those who are engaging with pupils on a daily basis:

*...teachers will feel alienated, they won't feel the need to, you know, innovate because it's being told 'this is the exam, this is what's important, you get on with it', so they don't become proper partners or innovators in the system and I think that does have an impact on teaching and learning.* (Teacher)

## The wider use of qualifications

### *The purpose of examinations*

National qualifications policy does not occur in a vacuum and threats to the delivery of qualifications were also perceived to stem from adjacent arenas of education policy. The data suggests that qualifications are being used to try to appease a range of stakeholders and are therefore attempting to fulfil too many purposes. As a result they may not be adequately optimised for any one of their intended purposes. This could threaten the validity of a qualification and undermine its educational value:

*I think if anyone could clearly establish what it was meant to be doing, we'd be a lot more successful at doing it because, you know, if you take, I don't know, GCSEs for example, for some students a GCSE is an entry to college, for some students it's an entry to work, for some students it's an entry to A-level... so whilst we're not very clear about what the qualifications are for then I think it's going to be very difficult to say we delivered them. (Awarding Body B)*

The concern here is that unless the purposes of a given qualification are clearly defined and given a realistic scope then they cannot deliver on the intended policy objectives. Newton (2007) points to the vast array of purposes to which education assessment is put, identifying eighteen different types (while also specifying that the list is not definitive!). Isaacs (2013) provides a case study which serves as an example of why multiple purposes may be problematic. She suggests that the failure of the 14-19 Diploma may have in part been a result of conflicting purposes at its core. The government saw it as a panacea for a variety of educational issues and expected it to deliver a host of benefits, ranging from boosting basic numeracy and literacy through to supplying skills and knowledge necessary for higher education and employment. Arguably a single qualification is inadequate for successfully fulfilling so many challenging, and sometimes contradictory, policy objectives (Isaacs, 2013).

### *The influence of accountability measures*

The school accountability system was seen as a key source of latent error given that general qualifications such as GCSEs are used as a measure of school success and are therefore high stakes for schools as well as pupils. Schools face potential closure should they be deemed as failing based on these accountability criteria and this is perceived to be having a negative impact on the examination system:

*A key risk in this area is the role that, particularly GCSEs, play in accountability of school performance and league tables and how that is potentially influencing, well more than potentially, probably is influencing the way in which schools approach exams and preparing students for them... (Regulator)*

This reflects a concern that the pressurised environment in which schools operate may have realigned priorities in a subtle way, shifting attention towards measures of accountability rather than directly on to the educational requirements of pupils. Though one could argue that these are designed to be in concert, there is a significant body of literature around concerns that accountability drivers may not necessarily benefit learners (e.g. Mansell, 2007). There is some limited evidence that the accountability system encourages strategic approaches to assessment in schools such as early and multiple entries to exams (see Acquah, 2013) and that much of this is aimed at the candidates who tend to perform more poorly (Taylor, 2012). In other words, the risk is that the use of qualifications for accountability purposes is undermining their primary educational purpose.

The fact that general qualifications are so central to the evaluation of schools strains their relationship with awarding organisations and this may introduce risk into the system by placing pressure on resources:

*...we know the pressures that are out there on the students themselves, on their teachers to continually improve. So if the students, the teachers don't see any improvement or don't necessarily see what they were expecting, immediately they're back knocking on the awarding bodies door and they're quite clear it's the awarding body that's got it wrong so managing the volume of the communication that is needed with schools and colleges post the issue with results potentially does threaten the operation of the system because, if you are focussing your staffing on that, rather than some of the staffing that might be moving on to preparing for the next series, you have got a tension in the demands there which can have a knock-on effect. (Awarding Body A)*

Though these concerns came across very strongly, several of the participants were keen to point out that they considered a reform to the current accountability system to be preferable to the removal of school accountability. It was recognised that some form of accountability was necessary and that the government would always want to keep this in place. In fact, the Department for Education are planning to review the accountability system following a consultation with the education community (DfE, 2013b). There were concerns about the direction which may be taken with the accountability system:

*I think the big, big worry at the moment (is) that the political view of education is that the aim is to somehow rank or separate young people rather than deliver a system that helps all young people get on and give them a chance at achieving and doing well, that's what I'm very worried about at the moment. (Teacher)*

### **The media and public confidence in the examination system**

The manner in which the government, the media and public opinion interact was considered important to policy formation and implementation. The qualifications system is driven by market forces and it is arguable that the qualifications themselves are a form of currency (Jones, 2011). If this is the case then public confidence is absolutely crucial – if qualifications users do not believe in the validity and reliability of a qualification then it ceases to have value and the market for it dissipates. This is true even if, in reality, the qualification is generally fit for purpose and is administered and assessed with relatively few issues:

*...if you get the Secretary of State standing up and saying the exam system's discredited, that's a pretty big headline, people remember that. What they don't remember is the 95% of things that are right... (Awarding Body A)*

When confidence wavers the public is justified in expecting corrective action from government and the regulator. In fact, maintaining public confidence is one of Ofqual's primary objectives (Ofqual, 2013b). The level of esteem in which the public holds the qualifications system is inescapably tied to the information and perspectives conveyed by the media, which tend to be negative (Newton, 2005; Warmington & Murphy, 2004). Interviewees perceived the type of stories that the press tend to write as unjustly negative and hyperbolic. Stories about examination errors were the most widely cited example, with errors which may only have impacted upon a very small proportion of candidates often being made to appear much more widespread and damning:

*People don't know very much about the examination system and what they learn about it, they generally get through the media and through the media they get both politicians' views of the system and the media's view itself. Those are related but slightly different. The media's interest in the system is selling newspapers and for making good radio broadcasts and good television broadcasts and increasingly good social networking and that thrives on bad news, ok? And obviously the public's confidence in the system is undermined by the fact that the media concentrates on bad news for, in its own thinking, a very good reason because they rely on selling newspapers and that's what sells newspapers. Because the public also learns about what politicians views are of the system through the media, any nuance just completely goes by the by. (Regulator)*

Warmington and Murphy (2004) examined the education news coverage of assessment results and suggested that it was polarised and ritualistic, focussing on the same issues (e.g. grade inflation) at various points of the assessment cycle. They go on to suggest that those who are working in areas of education which attract media interest need to become more aware of the needs and drives of the media in order to break such cycles and generate stories which are more constructive and nuanced. Some of the interviewees echoed this, highlighting the need for a strong and aligned industry position when dealing with unjustly negative press coverage. This type of communication may prove challenging as there is some evidence that stakeholders who are not closely involved with the examination system tend not think in terms of validity and grading reliability, focussing instead on *errors* of administration or marking (Chamberlain, 2011). The complexity of the issues may make informed debate within the media challenging.

Bad press alone may be uncomfortable for some stakeholders but it would not necessarily have a direct negative impact on the safe delivery of qualifications. However, the extra pressure may stifle innovation or misdirect the attention of awarding bodies. They suggested that the focus of awarding organisations could be inadvertently drawn towards avoiding or responding to bad press coverage when in fact efforts should be concentrated exclusively on delivering valid and reliable assessments:

*The Press I think is the biggest risk... In the first respect that, if something goes wrong, they will pick up on it and they will magnify it in order to tell a story but also it predisposes key stakeholders to the massively risk adverse within the system so things that could actually make the system less risky are less likely to be implemented because of fear about what the Press would say if they got hold of that thing that you were planning to do and spun it in a particular way or if you tried it and it went wrong. (Awarding Body A)*

Despite this, several stakeholders were keen to point out that errors were unacceptable and perhaps the pressure of press scrutiny could reduce their frequency. In other words the press helps to keep those operating the examination system accountable and therefore drives positive activity in terms of the mitigation of risks regarding administration and assessment.

*There does tend to be a media feeding frenzy on these things and, you know, maybe in the past there were a great number of errors but because there wasn't this spotlight put on etc., that doesn't really address the issue does it? You don't want any errors in examination papers. (Regulator)*

The impact that the media can have on public confidence and therefore the instigation of new policy can be further exemplified by the A-level grading fiasco in 2002. When the difficulties in

setting boundaries emerged, the extent of the issue was substantially miscalculated by the media who suggested that *tens of thousands* of pupils may have received the wrong grade. In fact, following the inquiry by Tomlinson (2002) and subsequent alterations of grade boundaries, fewer than 2,000 candidates were re-graded. Though this is obviously a substantial figure and the impact on these individual pupils should not be understated, this figure is actually only 0.7% of the 285,000 candidates who sat examinations with the affected awarding body. The media had, intentionally or not, substantially exaggerated the impact of the error, damaging public confidence and ultimately contributing to the resignation of the Secretary of State for Education at that time, Estelle Morris (Richardson, 2007). Through this example we can see how the reform process can be reactive and cyclical. Following a chain of events such as this, an incoming Secretary of State would likely have sufficient public support to repair a system which was publically and politically 'broken' by instigating further reform.

## Conclusions

The significant threats to the qualification system which stem from the policy development and implementation process, as perceived by the stakeholders interviewed in this study, can be summarised by four points:

1. The implementation process does not allocate adequate time for the specification of qualification content and the design of assessment. Key policy decisions need to be finalised earlier such that there is ample time to pilot new or reformed qualifications and their delivery systems.
2. The mechanisms by which stakeholders can influence policy formation are not sufficiently effective. Those stakeholders involved in the practice of delivering and administering qualifications can offer a great deal of expertise which would be beneficial to the process of policy formation.
3. Government policy requires qualifications to perform too many functions and this is a significant challenge for content specification and assessment design. This has the potential to introduce latent error into the qualification structure.
4. The relationship between the media, government and the qualifications system is a complicated and often damaging one. Though it can drive improvement regarding qualification delivery by causing stakeholders at all levels of the system to be accountable for errors, it can also cause instability which can threaten qualifications by encouraging a focus on revolution rather than evolution.

It is important to acknowledge the limitations of these findings. The analysis is explorative and focuses on threats at a fairly macro level. The data do not provide sufficient depth to unpick the underlying processes and mechanisms or truly to explore the interactions between the various stakeholder groups. Presenting an overview which combines new data with the existing literature is a useful exercise but the data cannot be used to plan a solution to the four issues listed above, only make them explicit as issues that are worthy of discussion, closer scrutiny and perhaps further research.

The findings reflect the views of a relatively small number of stakeholders at a specific point in time. It was only possible to gather the perspectives of stakeholders who occupy what can be loosely described as the middle ground between policy and practice – those who mediate the transition. It would have been useful and informative also to gather the perspectives of policy makers who are responsible for generating and developing national qualifications policy and to understand their position on its implementation. Equally, though data was gathered from a group of teachers, the sample is not extensive and does not fully capture the position of those who are 'on the ground' - those in classrooms trying to educate pupils and prepare them for assessment despite regular change in qualification content and structure. There would also be a

great deal of value in gathering the perspectives of pupils who are studying within the framework of the qualifications system.

There is a significant need for lessons to be learned from the collective experience of developing and implementing national qualification policy. With reference to the diplomas, Isaacs (2013) states that they carried with them the 'seeds of their own destruction' (p.10). Their multiple core purposes were in conflict with one another and their assessment structures were designed in a manner which made them too complex. In order to avoid building such fundamental flaws into the new A-levels and GCSEs it is necessary to learn from the mistakes of the past and ensure that policy makers, policy implementers and qualification users are able to communicate adequately and develop qualifications that are fit for purpose.

We need to start to exhibit "policy memory" (Higham & Yeomans, 2007) in order for policy to lead to incremental improvement in our qualifications. The process for specifying content and designing assessment needs to be better planned and allotted more time, the policy objectives need to be clarified collectively, and the relationship that policy makers have with the media and public opinion needs to be less reactive. Hopefully the desire to improve the process in these ways is realistic as well as idealistic and policy makers and policy implementers can refine the way in which they work together.

## References

- Acquah, D. (2013). *School Accountability in England: Past, Present and Future*. Manchester: Assessment and Qualifications Alliance.
- Baird, J. A., & Coxell, A. (2009). Policy, latent error and systematic examination failures. *CADMO: An international journal for educational research*, XVII(2).
- Baird, J. A., & Lee-Kelley, L. (2009). The dearth of managerialism in implementation of national examinations policy. *Journal of Education Policy*, 24(1), 55–81.
- BBC (2013a, January 23). A-level plans challenged by school and university heads. *BBC*. Retrieved from <http://www.bbc.co.uk/news/education-21156370>
- BBC (2013b, February 13). Court rejects bid to overturn GCSE grades. *BBC*. Retrieved from <http://www.bbc.co.uk/news/education-21426396>
- Chamberlain, S. (2011). *Qualification users' perceptions and experiences of assessment reliability*. Manchester: Assessment and Qualifications Alliance.
- DfE. (2013a). *Reforming Key Stage 4 qualifications consultation - Government response*. Department for Education. Retrieved from <http://www.education.gov.uk/schools/teachingandlearning/qualifications/gcses/a00221366/gcse-reform>
- DfE. (2013b). *16-19 Accountability Consultation*. Retrieved from <https://www.education.gov.uk/consultations/index.cfm?action=consultationDetails&consultationId=1924&external=no&menu=1>
- Gove, M. (2013a). *Ofqual policy steer letter: reforming Key Stage 4 qualifications*. Retrieved from <http://www.education.gov.uk/schools/teachingandlearning/qualifications/gcses/a00221366/gcse-reform>
- Gove, M. (2013b). *Ofqual policy steer letter: Reform of GCE A Levels*. Retrieved from <http://media.education.gov.uk/assets/files/pdf/l/ofqual%20letter%20alevels%20v2.pdf>

- Guardian. (2002, September 28). No one to blame: but what went wrong? *The Guardian*. Retrieved from <http://www.theguardian.com/uk/2002/sep/28/politics.alevels2002>
- Higham, J. & Yeomans, D. (2007). Policy memory and policy amnesia in 14–19 education: Learning from the past? In *Policy-making and policy learning in 14–19 education*, ed. D. Raffe, and K. Spours, 33–60. London: Bedford Way Papers, Institute of Education, University of London.
- HoC. (2003). *A level standards* (No. HC 153). House of Commons Education and skills committee. Retrieved from <http://www.publications.parliament.uk/pa/cm200203/cmselect/cmmeduski/153/153.pdf>
- Isaacs, T. (2013). The diploma qualification in England: an avoidable failure? *Journal of Vocational Education & Training*, 65(2), 277–290.
- Jones, B. (2011). *Regulation and the qualifications market*. Manchester: Assessment and Qualifications Alliance.
- Mail Online. (2012). Return of the O-level. *Mail Online*. Retrieved from <http://www.dailymail.co.uk/news/article-2162369/Michael-Gove-plans-scrap-dumbed-GCSEs-bring-O-Levels.html>
- Mansell, W. (2007). *Education by numbers: the tyranny of testing*. London: Politico's.
- Newton, P. E. (2005). The Public Understanding of Measurement Inaccuracy. *British Educational Research Journal*, 31(4), 419–442.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170.
- Oates, T. (2008). Going round in circles: temporal discontinuity as a gross impediment to effective innovation in education and training. *Cambridge Journal of Education*, 38(1), 105–120.
- Ofqual. (2012a). Exam board seminars - Final report. Ofqual. Retrieved from <http://www2.ofqual.gov.uk/news-and-announcements/83-news-and-announcements-news/885-changes-to-seminars>
- Ofqual. (2012b). A Level Reform Consultation. Ofqual. Retrieved from <http://comment.ofqual.gov.uk/a-level-reform/home/a-level-reform-consultation/>
- Ofqual. (2013a). GCSE Reform Consultation. Ofqual. Retrieved from <http://comment.ofqual.gov.uk/gcse-reform-june-2013/>
- Ofqual. (2013b). Ofqual corporate plan 2013-16. Ofqual. Retrieved from <http://ofqual.gov.uk/files/2013-08-09-corporate-plan-2013-16.pdf>
- Richardson, W. (2007). Public policy failure and fiasco in education: perspectives on the British examinations crises of 2000–2002 and other episodes since 1975. *Oxford Review of Education*, 33(2), 143–160.
- Schofield, J. (2004). A model of learned implementation. *Public Administration*, 82(2), 283–308.
- Stacey, G. (2013). Letter to Secretary of State for Education: Reform of A Level qualifications. Retrieved from <http://ofqual.gov.uk/news/a-level-reform-letter-from-michael-gove-14th-march-2013/>



- Tattersall, K. (2003). Ringing the changes: education and assessment policies, 1900 to the present. In *Assessment and Qualifications Alliance, Setting the standard: a century of public examining by AQA and its parent boards* (pp.7-28). Manchester: Assessment and Qualifications Alliance.
- Taylor, R. (2012) Chapter 3 – Analysis of the trends in early entry at GCSE. Manchester: Assessment and Qualifications Alliance.
- Telegraph. (2012). GCSE results 2012: row erupts over marking of English exam papers. *Telegraph*. Retrieved from <http://www.telegraph.co.uk/education/secondaryeducation/9494605/GCSE-results-2012-row-erupts-over-marking-of-English-exam-papers.html>
- Telegraph. (2013, September 6). Michael Gove's GCSE and A-level reforms delayed. *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/education/educationnews/10292690/Michael-Goves-GCSE-and-A-level-reforms-delayed.html>
- Tomlinson, M. (2002). Inquiry into A-level standards: final report. Department for Education. Retrieved from <http://webarchive.nationalarchives.gov.uk/20130401151715/https://www.education.gov.uk/publications/standard/publicationDetail/Page1/TMN2#downloadablepart>
- Warmington, P., & Murphy, R. (2004). Could do better? Media depictions of UK educational assessment results. *Journal of Education Policy*, 19(3), 285–299.