



Who's to produce and who's to choose?

Assessing the future of the qualifications and assessment market

Gabriel Heller Sahlgren

Research report 10



CENTRE FOR THE STUDY OF MARKET REFORM OF EDUCATION

23 Great Smith Street, Westminster, London, SW1P 3BL • www.cmre.org.uk

Who's to produce and who's to choose?

Assessing the future of the qualifications
and assessment market

Gabriel Heller Sahlgren

About the author

Gabriel Heller Sahlgren is Director of Research at the Centre for the Study of Market Reform of Education (CMRE), an affiliated research fellow at the Research Institute of Industrial Economics in Stockholm, Sweden, and a PhD student at the London School of Economics. He is the author of numerous publications on issues relating to applied microeconomics, including *Incentivising Excellence: School Choice and Education Quality* (CMRE and IEA 2013).

Acknowledgements

The author thanks Henrik Jordahl and Tim Oates for valuable comments and suggestions, and James Croft for rigorous editorial input.



First edition published 2016

by The Centre for the Study of Market Reform of Education Ltd

ISBN 978-1-5136-1601-8

The moral right of the author has been asserted.

All rights reserved. No part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form, or by any means (electronic, mechanical, photocopy, recording, or otherwise) without the prior permission of the publisher. Any person who does any unauthorised act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

A CIP catalogue record for this book is available from the British Library.

Every care has been taken that all information was correct at the time of going to press. The publisher accepts no responsibility for any error in detail, inaccuracy or judgement whatsoever.

Typeset by Great White Designs, Stroud, Gloucestershire

Printed and bound by Lightning Source Ltd.

CONTENTS

Foreword	v
Executive Summary	1
1 Introduction	7
2 The current English qualifications market	11
3 Concerns, allegations, and proposed solutions	15
4 Monopoly, franchising, and user choice: a framework	17
4.1 Who should pay for qualifications and assessment?	17
4.2 Who should provide qualifications and assessment?	18
4.2.1 The first question: government versus private provision	18
4.2.1.1 How strong is the case in favour of government service provision?	22
4.2.1.2 Implications for qualifications and assessment	25
4.2.1.3 Empirical evidence from other related services	28
4.2.1.3.1 Case 1: schooling	29
4.2.1.3.2 Case 2: health and elderly care	30
4.2.1.4 Verdict: private provision is preferable to a single government board	31
4.2.2 The second question: franchising or user choice?	31
4.2.2.1 A difference in substance or degree?	31
4.2.2.2 The general case for user choice	32
4.2.2.3 Making things a bit more complicated	33
4.2.2.4 Implications for qualifications and assessment	36
4.2.2.5 The current market dynamic	37
4.2.2.6 Empirical evidence from other relevant services	46
4.2.2.6.1 Case 1: schooling	47
4.2.2.6.2 Case 2: health care	48
4.2.2.7 Spreading the risk	49
4.2.2.8 Verdict: user choice is preferable – but reform is still desirable	51
5 Suggestions for a better-performing system	53
5.1 Focus on minimum standards rather than strict equivalency	53

5.2	Use an empirical output approach to assess comparability between qualifications and specifications	54
5.3	Consider introducing a national cohort-referenced comparability test	55
5.4	Change league tables in line with the new remit	56
5.5	Provide information for schools and universities on the differences between qualifications and specifications	58
5.6	Maximise the possibility of market entry	59
5.7	Monitor potential problems with price competition – and consider some regulation as a measure of last resort	60
5.8	Trial the system rigorously	61
6	If the government still opts for franchising...	63
6.1	Allow a measure of discretion in the procurement process	63
6.2	Use a model that values quality at least as highly as price	64
6.3	Contract out by subject rather than by function	65
6.4	Consider allowing more than one provider per subject to spread the risk	66
6.5	Measure schools' satisfaction with exam boards	66
6.6	Trial the system rigorously	67
7	Conclusion	69
	References	73

FOREWORD

HOW SHOULD THE QUALIFICATIONS of students be assessed? This is one of the most defining and important aspects of any education system. England, Wales, and Northern Ireland have a unique system of qualifications and assessment distinguished by choice and diversity. There is much to be said about this model, and critics have recently questioned whether standards can be upheld under competition between multiple providers. If incentives are maladjusted, competition might lead to a ‘race to the bottom’ instead of promoting quality. In this monograph, Gabriel Heller Sahlgren addresses such concerns in a fresh, competent and appealing way.

Although there is a lack of research to rely upon, Heller Sahlgren’s treatment of the topic is systematic – both grounded in theory and evaluated by evidence. He achieves this by applying the framework of transactions cost economics. A central insight of this theory is that quality expectations are not met automatically, but require costly efforts of specification and verification. Private provision may be superior to government monopoly, but only as long as private providers cannot shirk on non-contractible dimensions of service.

Private provision could take on different forms. The first main model is characterised by procurement or franchising. The other main model is distinguished by user choice. By carefully comparing the two private models with government provision, the monograph combines a comprehensive treatment of the subject with a discussion of the most policy relevant issues. There emerges a categorical ranking of the models such that the present user choice model is preferable to franchising, which in turn is preferable to government provision. In addition, there are instructive discussions about how each of the private models could be improved to get the most information value at the lowest cost.

This is a seminal work. In the future, informed discussion about the proper model of qualifications and assessment shall not be able to sidestep the issues examined here.

Henrik Jordahl

Associate Professor and Program Director,

The Research Institute of Industrial Economics, Stockholm

October 2016

EXECUTIVE SUMMARY

HISTORICALLY, SCHOOLS IN ENGLAND, Wales, and Northern Ireland have had the right to decide which qualifications their pupils take from a range of options offered by multiple independent providers. This allows an element of diversity in assessment and qualifications, and stimulates competition between different exam boards – features that are unique in an international perspective.

Yet in the last couple of years, the merits of choice and competition in this area have faced increasing criticism in the media. Competition, it is said, introduces perverse incentives, inducing exam boards to dumb down their qualifications and inflate grades. Instead of raising quality, critics argue, competition leads to a ‘race to the bottom’.

Some have proposed that franchising might resolve this issue, others that we should transition to provision via a single government exam board. Under a franchising system, exam boards would compete in a tendering process to be able to deliver examinations in a specific subject for a set number of years.

Theoretically, whether or not a service should be delivered by the government or externally by independent organisations depends on who is able to offer the service most efficiently – which in turn depends on the transaction costs involved in ensuring services are delivered according to expectations. If the price for buying a service, including the external transaction costs, is lower than the internal production and transaction costs, there is a clear economic rationale for doing the former rather than the latter.

The key issue for assessing transaction costs is the level of contract incompleteness. If quality is difficult to contract, providers may engage in quality shirking on non-contractible dimensions of the service. While private

providers have stronger incentives to cut costs and improve quality than the government, the impact on overall quality could be positive, zero, or negative depending on how strong the relationship is between costs and quality and whether or not there are opportunities for innovations.

In practice, there are few public services that fulfil the conditions that would warrant direct government provision. We argue this also holds true in qualifications and assessment. This is because there are relatively few non-contractible elements involved; reputational mechanisms appear strong; and the scope for private sector innovation is likely to be considerable. The case for government provision becomes even less persuasive considering the possibility of it giving preference to non-profit providers, which have weaker incentives than for-profit providers. However, we note that the case for giving preference to non-profit providers is often weak given the lack of empirical evidence suggesting they outperform for-profit providers in similar public services.

Furthermore, empirical evidence from related services indicates that franchising via procurement can improve efficiency and non-contractible quality in complex public services that include more non-contractible elements than qualifications and assessment, suggesting that incentives to invest in quality overall often outweigh any incentives to engage in harmful cost cutting.

We therefore argue that there is little reason to support the option of a single government examination board. The approach is likely to increase costs without any short-term gains in relation to quality improvement – while at the same time decreasing the potential for innovation and thus quality in the future.

The principal choice, then, is between a franchising model and a user choice model. Under the stylised franchising model, government picks winners in a tendering process; under the user choice model, users pick winners that meet stipulated criteria. As far as choice is concerned, in theory, it could improve matching between pupils and qualifications, while at the same time generating efficiency gains and innovations on an on-going basis as a result of competitive pressures. In some situations, however, choice might instead lead to lower quality. For example, if price competition is allowed, effects on quality will depend on how users weight quality relative to price.

There is, however, no evidence that choice and competition have led to a decline in the standards of national qualifications. This is unsurprising given

the strict regulatory framework in place. Incentives for schools to choose what they perceive to be easier qualifications are mostly a product of the equivalency framework and the way the value of qualifications are weighted in school league tables. These would largely remain in a procurement system. Moreover, perceived quality appears more important than price in schools' decision-making processes. There is no evidence of excessive price competition undermining quality in the present choice-based system.

In addition, empirical evidence from other complex public service markets suggests that choice and competition can work well even in cases where price competition is allowed, provided there is proper attention to system design. The regulatory structures in the current qualifications and assessment market make it difficult for exam boards to compete by raising standards (although competition appears to function somewhat better on this score for alternative qualifications). While the reputation of qualifications and board reputation seem important in the market for alternative qualifications, the accountability framework is constructed in such a way as to make schools more sensitive to their league table performance than to reputational mechanisms, which might otherwise be more effective drivers of quality.

Overall, therefore, there is little evidence that the qualifications and assessment market functions as poorly as its critics claim. In fact, current market dynamics suggest that choice of alternative qualifications has generated some improvements. In addition, the presence of multiple providers has effectively decreased the risk of system failure in ways that neither a single government board nor the suggested franchising model could have achieved. Accordingly, rather than abolishing the market, we argue instead that the government should focus on optimising mechanisms geared to quality improvement.

In this paper, we set out a reform agenda for how the government can achieve such quality improvements. First, the accreditation framework should be less prescriptive in its attempts to ensure comparability between different qualifications, subjects, and specifications, and instead be based on meeting specified minimum standards. This would mean that boards could offer alternatives based on higher, but not lower, quality and standards. This would generate better matching, and allow exam boards to brand themselves on the standards and quality provided. End-users would then be in a better position

and have stronger incentives to openly differentiate between different providers and their products.

To assess the qualitative difference between them, we argue in favour of a more reliably empirical approach than stipulating what students have to know and accomplish to achieve different grades. Establishing post-hoc outcomes at university or in the labour market among pupils taking different qualifications and specifications would be a far more effective way of assessing their value.

In addition, it would also be beneficial to introduce a general cohort-referenced competency test in order to provide a comparability metric with which to judge pupil performance across different qualifications and specifications. Similar to the national reference test at year 11 to be introduced from 2017, the results would not be included in league tables, but would be offered as a way of helping schools and end-users differentiate between different qualifications and specifications. Possibly, the competency test could be taken by representative samples of pupils taking different qualifications and specifications, rather than entire cohorts, thereby decreasing the costs of administrating it.

Ofqual should provide information to schools and end-users regarding differences between qualifications, subjects, and versions to ameliorate the potential for information asymmetries in the market. It should also commission, or undertake, research to better establish their relative value.

Reflecting this alteration in the remit of the regulator, school league tables should be reformed. One option is to simply publish results separately for different qualifications in the same subject area, and different versions of the same, for all subject qualifications, which would decrease incentives for schools to take easy options. Another alternative would be to use the empirical equivalence framework based post-hoc university and/or labour market outcomes as outlined above. The national cohort-referenced competency test could also be used to calibrate results based on performance on the test. A third alternative would be to create a system for end-users to assess comparability themselves based a combination of the latter two options.

It is important to maximise the possibility of market entry to increase competition among existing suppliers and generate incentives to innovate. The regulatory burden currently appears extensive, acting as a hurdle for new entrants. Reforming

the accreditation framework to focus on minimum requirements would make it possible to lower some of these barriers.

While there is little to suggest that excessive price competition has been a barrier to quality competition to date, we believe it is important to monitor this closely. If problems should arise in the future, the government should consider some regulation.

Finally, if possible, reforms should be subject first to rigorous trial. Any unintended consequences could thereby be addressed ahead of national implementation.

In the event that the government decides to move towards a franchise system without user choice in spite of our recommendation, we also provide some principles upon which we believe such a system should be built.

First, it is desirable to allow a measure of discretion in the procurement process to ensure that past performance, broadly speaking, may determine the awarding of future contracts. The reputation mechanism is useful in the procurement of complex services because it minimises the risk of suppliers engaging in harmful cost reductions.

In order to minimise the risk of harmful price competition, we also recommended that the evaluation model employed values quality at least as highly as price. In addition, it would be preferable in this scenario for government to contract out qualifications and assessment by subject, rather than by function, in order to ensure that incentives are aligned to improve the overall product.

While unorthodox, we also believe consideration should be given to allowing more than one exam board to supply the same qualification and subject in order to minimise system risk in the event of service failure. This is possible even if user choice is removed.

Since user choice would be abolished in a franchising model, it would be useful to carry out customer satisfaction surveys from randomly selected schools that have been assigned to different exam boards. The idea would be to measure subjective interpretation of the quality across different boards.

And, again, since unintended consequences are common in policymaking, we believe any potential new procurement regime should be trialled prior to subjecting the qualifications and assessment system to radical reform.

I INTRODUCTION

THE PRIMARY PURPOSE OF qualifications and assessment in education is to provide information about pupils' knowledge and ability. Assessment informs future employers as well as higher education institutions (HEIs) about candidates' suitability and preparedness for their chosen career or next steps in education and training.

Historically, schools in England, Wales, and Northern Ireland have had the right to decide which qualifications their pupils take from a range of options offered by multiple independent providers. This allows an element of diversity in assessment and qualifications, and stimulates competition between different exam boards – features that are unique in an international perspective.

Yet in the last couple of years, choice, and the competition it stimulates, have faced increasing criticism from politicians and in the media. Competition, it is argued, introduces perverse incentives that induce exam boards to dumb down their qualifications and inflate grades. Instead of raising quality, competition, it is alleged, leads to a 'race to the bottom'.

Some have proposed that franchising might resolve this issue, others that we should transition to provision via a single government exam board. Under a franchising system, exam boards would compete in a tendering process to be able to deliver examinations in a specific subject for a set number of years.

As a result of this criticism, the Coalition government proposed to abolish exam board competition, and instead introduce franchising via procurement, whereby boards would compete for exclusive contracts to deliver specific qualifications in specific subject areas for a set number of years. In this franchising model, the present system, whereby schools choose the exam board offering with which they want to work, would thus be dismantled. Instead of

continuous competition for schools' business, exam boards would compete for politicians' business every time each subject was up for tendering.

While the proposal was abandoned, it resurfaced during the Conservative majority government that followed, and is still favoured by some in the new regime – as is the idea of abolishing independent exam boards altogether and nationalising provision under a single government board.

Theoretically, whether or not a service should be delivered by the government – as, for example, qualifications through a single government exam board – or externally by independent organisations, depends on who is able to offer the service most efficiently. This, in turn, depends on the transaction costs involved in ensuring services are delivered according to expectations. If the price for buying a service, including the external transaction costs, is lower than the internal production and transaction costs, there is a clear economic rationale for doing the former rather than the latter.

However, even if we accept that independent provision is desirable, the question of how to arrange such provision remains. There are two stylised models under consideration: (1) a franchising model through procurement or (2) a user choice model, such as the one in operation. Under the franchising model, the government picks winners in a tendering process; under the choice model, users pick winners according to a standards and regulatory framework determined by the government.

Certainly, as discussed in more detail in Section 4, the distinction between user choice and procurement is fluid rather than sharp. Indeed, procurement could technically be combined with user choice among several suppliers following the conclusion of a tendering process. However, since the whole rationale for the reforms under consideration is to use procurement as a way to abolish choice altogether, the monograph principally discusses the case for and against franchising without choice vis-à-vis the user choice model.

In theory, choice could help improve matching between pupils and qualifications, while at the same time generating efficiency gains and innovations on an on-going basis as a result of competitive pressures. In some situations, however, choice might instead lead to lower quality. For example, if price competition is allowed, effects on quality will depend on how users weigh quality relative to price.

The purpose of this monograph is to evaluate whether monopoly, franchising, or user choice is preferable in qualifications and assessments. Drawing upon economic theory and empirical evidence from different sectors, it discusses the advantages and disadvantages of each approach applied specifically to the English qualifications and assessment market. It further suggests which model is likely to be preferable – and which reforms should be implemented to make sure this model works as well as possible. In addition, in the event that the government opts not to follow the paper’s recommendations, we also offer some broad principles upon which the alternative system should be built, in order to maximise its potential to work well.

The monograph proceeds as follows. Section 2 describes the current qualifications market in England; Section 3 discusses allegations and concerns that have arisen in recent years; Section 4 provides an in-depth analysis of the relative virtues of monopoly, franchising through procurement, and user choice in service provision using an economic framework and empirical evidence; Section 5 presents a broad reform agenda to help us move towards a better-performing user-choice system; Section 6 discusses broad principles around which a potential franchising system could be built; and Section 7 concludes.

2 THE CURRENT ENGLISH QUALIFICATIONS MARKET

HISTORICALLY, ENGLAND, WALES, AND Northern Ireland have a long tradition of choice and diversity in qualifications and assessment for pupils aged 16 and older.¹ For many years, schools in these nations have been able to choose different qualifications for lower-secondary and upper-secondary school pupils – such as GCSEs and A levels – from multiple independent providers. In this system, schools pay exam boards for each qualification entry, with the fees determined by the exam boards. These providers design qualifications, administer, and mark examinations, and award grades to pupils. Having grown in number over the course of the twentieth century, when they were affiliated to various universities, the number of boards has decreased significantly in the past decades, due in part to a widespread perception that economies of scale could be achieved through rationalisation. Eventually, therefore, most of them also lost their connections to universities (Education Committee 2012; Steinberg 2002).

Today, all exam boards must be approved by the central regulator Ofqual and abide by its quite extensive regulation, as stipulated in its various rulebooks. These include the General Conditions of Recognition and booklets describing its qualification-specific requirements (see Ofqual 2016a, 2016b). For example, the marking procedure is under close supervision and is generally mostly determined by statistical predications based on prior cohorts' achievement across all candidates and exam boards (Bramley and Vidal Rodeiro 2014). Ofqual also mandates how content should be assessed, for example via exams

1 For a discussion of the history of the English qualifications market, see Shackleton (2014).

or teacher-based assessment (Bassett 2014). Far from being a 'free market', as is often asserted (e.g. Gilbert 2011), the qualifications and assessment market is in fact highly regulated.

Today, there are seven providers offering academic qualifications in the three nations: the Assessment and Qualifications Alliance (AQA); the Council for Curriculum Examinations and Assessments (CCEA); the University of Cambridge International Examinations (CIE); Pearson Edexcel; the International Curriculum and Assessment Agency Examinations (ICAAE); the Oxford, Cambridge and Royal Society of Arts (OCR); and the Welsh Joint Education Committee (WJEC). Only one of these boards (Pearson Edexcel) is a profit-seeking organisation, while CCEA is the only board directly owned and operated by a government, as part of the Northern Irish Department of Education. The other five providers operate as non-profit charities.² The relevance of this dominance of non-profit providers is explored in Section 4.2.1.1 in relation to the case for government provision.

As Figures 1 and 2 (over) show, there was little change in providers' respective aggregate market shares for GCSEs and A levels – the most common qualifications in secondary school – across the three nations in the period 2010/11 to 2014/15, although Pearson and WJEC have raised their market shares slightly at OCR's expense in the GCSE market. Overall, AQA is currently the leader in both types of qualifications in a market that is entirely dominated by five of the seven boards.³

Available data also indicate that market shares in core GCSE subjects have generally been similarly stable over the past couple of years. The exception is in GCSE science where AQA and Pearson raised their market shares at OCR's expense (see Ofqual 2013, 2014, 2015a, 2016c). The Education Committee (2012) noticed broadly the same conclusion based on previous years. Certainly, the aggregate market shares mask the fact that there is considerable churn between the boards across years, as schools move to and from different boards (see Education Committee 2012). Nevertheless, there is little evidence that such moving has a considerable impact on market shares.

2 Counting all types of qualifications on offer – including vocational and language courses – there are currently about 160 providers (Ofqual 2016c).

3 Data are obtained from Ofqual (2016c).

Figure 1
GCSE market shares

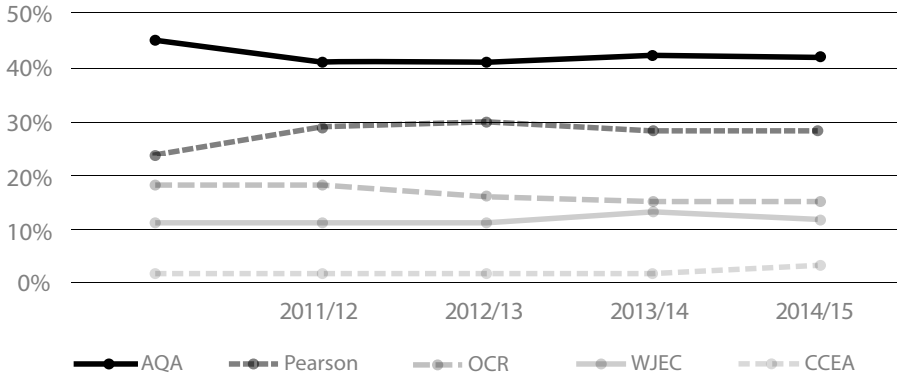
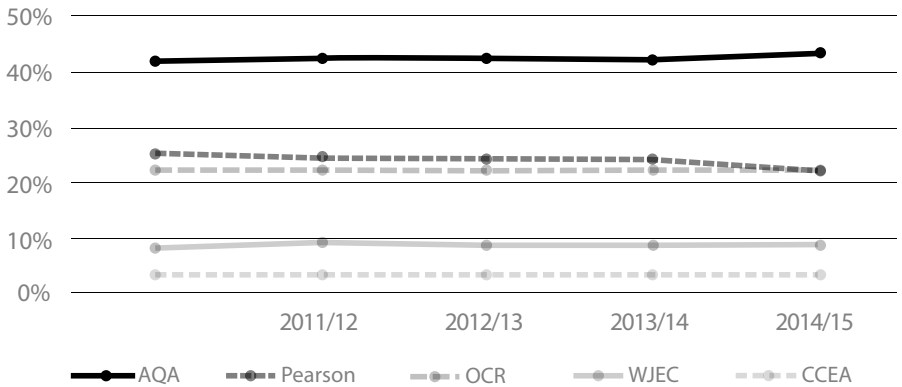


Figure 2
A level market shares



Moreover, schools can also opt for some approved alternative qualifications, such as the IGCSE and the Cambridge Pre-U, which are deemed equivalent to GCSEs and A levels respectively. In other words, schools may choose not only alternative specifications of the same national qualification, but also entirely different qualifications that are approved by Ofqual for state funding. These qualifications are also regulated, albeit to a less extent than the official national ones.

Overall, therefore, in international perspective, the English, Welsh, and Northern Irish education systems offer a unique level of diversity in provision of qualifications and assessment. Indeed, few countries, if any, have similar systems and most have single exam boards in the form of government monopolies (Education Committee 2012).

3 CONCERNS, ALLEGATIONS, AND PROPOSED SOLUTIONS

THE CURRENT QUALIFICATION AND assessment market clearly satisfies the criteria of a ‘quasi-market’, combining public funding with user choice among independently-operated providers (Le Grand 1991, 2007). Whether it can be said to be a functioning quasi-market, however, is a question that has been the subject of increasing debate in recent years. Over a decade, critics argued that exam boards’ interest in maximising market share induces them to dumb down standards and inflate grades. Instead of working to raise quality, it is argued, competition leads to a ‘race to the bottom’.

Consequently, in 2010, Cambridge Assessment (2010) held its ‘Standards Debate’, which was the first public airing of research on the issue of sliding standards. It pointed to a number of factors driving reductions in the standards of public examinations; structural changes in qualifications; patterns of incentives in government, schools, and exam boards; repeated reforms; and subtle mechanisms embedded in awarding practices. In other words, an organisation in the industry itself took the first step to deal with the perceived problems in the examination system.

A year later, a row broke out after revelations by *The Telegraph* offered evidence that informational seminars provided by exam boards included an element of coaching by which teachers were taught how to ensure their pupils performed well in exams. Video was released of examiners boasting about how easy their exams were. It was further disclosed that some teachers clearly engage in rent-seeking behaviour through open discussion of their choices of exam board and which have the least demanding curricula and grade boundaries (Winnet,

Watt, and Newell 2011). These revelations induced an inquiry by the Education Committee (2012), the report of which highlighted the danger of competition on syllabus content as especially pernicious. It argued that 'substantial improvements are needed to change the incentives in the system', should the current model of choice and competition be retained (Education Committee 2012, p. 5).

In response to these allegations of malpractice and concerns about perverse incentives, the government floated the idea of (effectively) abolishing choice in core subjects in favour of a procurement model. According to this model, exam boards would compete for exclusive five-year franchises for different subjects (see DfE 2012).

Although the idea was dropped, due to opposition from Ofqual and concerns about whether it would have fallen foul of EU competition law (see Ofqual 2012a; Grice and Garner 2013), it resurfaced again in 2015 following Ofqual's concern that one board had come dangerously close to missing the deadline for completing all its marking.⁴ In response, Schools Minister Nick Gibb declared that the incident was evidence of a need for 'long-term, fundamental reform'.⁵ Two alternatives were said to be under consideration: (1) a single government board, entailing in effect wholesale nationalisation of provision, or (2) franchising via procurement along the lines of the model proposed two years previously (Duncan 2015; Weale 2015).

In other words, the quasi-market in qualifications and assessment is currently under considerable pressure – and it appears likely that the government will attempt some form of reform to curtail user choice at some point in the future. Before embarking on such a mission, however, it is crucial to understand the costs and benefits of the three different models – monopoly, procurement, and user choice – and consider the evidence supporting the idea that alternative arrangements would generate improvements.

⁴ Note that there is little evidence that EU law was in fact an obstacle to the reforms. As procurement would have been achieved through competitive tendering, 'classification as a public services contract (which, being education, would be for Part B services) or even a public services concession ought not to have been such a major concern. Such a process, if followed correctly and carefully, would be consistent with any EU procurement or EU Treaty obligations applicable to Part B contracts or concessions' (Smith 2013).

⁵ Quoted in Weale (2015).

4 MONOPOLY, FRANCHISING, AND USER CHOICE: A FRAMEWORK

FROM AN ECONOMIC PERSPECTIVE, establishing a well-functioning model for provision and delivery of welfare services, such as qualifications and assessment, is far from straightforward. Here, we discuss the three principal alternatives – monopoly, procurement, and user choice – and their relative strengths and weaknesses from a theoretical and empirical standpoint.

4.1 Who should pay for qualifications and assessment?

What is the economic rationale for government involvement in the funding of qualifications and assessment? One reason is the potential for externalities. Certain private goods involve spill-over effects by which gains or losses to consumers pass to other members of society. The financing and provision of qualifications and assessment are part and parcel of the education process, which can have positive spill-over effects on health, earnings, and living standards (e.g. Bratti and Leombruni 2014; Ricci and Zachariadis 2013; Wantchekon, Klačnja, and Novta 2015).⁶ Education investments may also generate higher tax revenues and thus benefit society as a whole through the provision of public goods and other services. Normally, it is better to be the direct beneficiary rather than merely benefit from externalities, although the latter may also mean that people could free ride and in effect gain larger net benefits. Ideally, public financing of education, including the provision of qualifications and assessment, should align public and private interests to ensure that under-investment does not occur (see Bergman 2013; Heller Sahlgren 2013).⁷

6 See McMahon (2010) for a review.

7 Note, however, that this does not necessarily justify full public funding for all children. In a review article

Another important justification for government involvement in the provision of qualifications and assessment is that capital markets for children's education do not function particularly well. Children cannot commit to taking loans and later paying them back when they are adults, and there is also uncertainty about how much education parents should invest in. Investments in education are generally undiversified and insuring people against the risk that their investments may not pay off would generate moral hazard by protecting parents from the consequences of poor decision-making. Solving the problem through 'internal family financing' – that is, by letting older generations pay for the younger – is not sufficient if some cannot afford or will not pay for it. As Hoxby (2006:9) argues: 'We need only consider the problem of a poor family with a very gifted child who could benefit from a world-class education to see that underinvestment could occur.' As assessment and qualification are core functions of the educational process, this also applies to these features more specifically.

4.2 Who should provide qualifications and assessment?

While the above indicates the government should be involved in financing qualifications and assessment, it does not necessarily imply that it should provide them in house.⁸ Here, we analyse the economic arguments in favour and against government provision vis-à-vis private provision, with and without choice, by considering a more general theoretical framework, which we then apply to the qualifications and assessment sector.

4.2.1 *The first question: government versus private provision*

We begin by considering the case for and against a government monopoly in public services. As a starting point for the analysis, it is useful to consider the

surveying the evidence, McMahon (2010) suggests the research indicates that a little over 50 per cent of the total costs should be publicly provided.

8 Note that government provision could technically be combined through market competition, similar to the 'market socialism' that was prevalent in former Yugoslavia (see Estrin 1991). Examples in the welfare sector include school choice programmes that exclude private providers (e.g. Deming et al. 2014; Lavy 2015). However, since this option is not under consideration for the qualifications market, we do not discuss it further.

question of in-house production vis-à-vis outsourcing from the perspective of a for-profit company. Most companies choose to produce some services in house and buy others on the market, and the choice depends on which model maximises profits in the long term (Bergman 2013). Normally, other providers may deliver similar services to lower costs, due to economies of scale or the advantages associated with specialisation. For example, it is unlikely to be cost efficient for a company to hire its own cleaning staff instead of purchasing cleaning services from an external organisation (Elinder and Jordahl 2013). Cleaning is a non-complex service, and it is fairly easy to determine whether or not cleaners have done a good job.

Yet producing more complex services in house may be justified in order to ensure appropriate quality assurance. Through vertical integration, companies may have oversight over the entire production process, allowing them to monitor quality more effectively (Carlton and Perloff 2003). Such monitoring may be difficult to ensure when purchasing services externally, especially when the level of quality is difficult to identify and contract. Since external providers have lower incentives to maximise efficiency in the production of services, and are less informed about the buyers' exact needs and intentions, information asymmetries favouring the external provider give rise to principal-agent problems – when the provider's incentives do not align with the interests of the buyer. In such circumstances, buyers must also spend time and money on making sure the service is delivered as intended.

A key question, then, is whether higher transaction costs outweigh lower production costs that may arise by purchasing services externally. As services get more complex and specialised, and as quality specifications get more difficult to identify and contract, transaction costs increase. For very complex services, the external transaction costs may be so high that it is more efficient to produce these in house despite the fact that this induces higher direct production costs. This is the case if the total costs for the buyer – including the costs associated with ensuring that the service is produced to a satisfactory standard – become higher than the gains from outsourcing the production to external organisations.

Generally, the above also applies to the public sector.⁹ Like private production, public production necessitates clear instructions – via employment contracts

⁹ Simply put, both types of production require contracts. In the case of public production, there are employment contracts between the government and civil servants, which specify what the latter are

– regarding what (and how) services should be produced, as well as monitoring to ensure standards are upheld, although these transaction costs are normally assumed to be lower than those associated with outsourcing to external organisations.¹⁰ Meanwhile, private organisations often have lower production costs than public organisations, for example because of stronger incentives to work hard due to the profit motive, fewer stakeholders, more flexible use of employees, a clear division of labour, stronger focus on results, and the lack of soft-budget constraints (e.g. Dixit 1997; Donahue 1991; Kornai, Masking, and Roland 2003; Shleifer 1998).¹¹

In other words, whether or not a service should be delivered by the government or externally by independent organisations depends on who is able to offer the service most efficiently – which in turn depends on the transaction costs involved in ensuring services are delivered according to expectations. If the price for buying a service, including the external transaction costs, is lower than the internal production and transaction costs, there is a clear economic rationale for doing the former rather than the latter.

How can we determine whether transaction costs in the external production of services are too difficult to overcome? As indicated above, a key issue is whether or not it is possible to formulate contracts that can guarantee a certain level of quality, which is a difficult task in the case of complex services. For example, in education, it is difficult to stipulate precisely everything that schools should produce in detail. Education quality is multidimensional and some would argue that it is impossible to reduce such quality to easily measurable yardsticks, such as test scores or other quantifiable data. There is also a myriad of ways in which providers can manipulate the data to present themselves in a favourable light (see Figlio and Loeb 2011).

supposed to accomplish. In the case of private production, there are contracts between the government and independent providers, which in turn sign employment contracts with their employees. If it were possible to sign all-encompassing contracts, the government could achieve the same results with public or private production (Hart, Shleifer, and Vishny 1997).

10 It is not clear, however, that this assumption is valid. Indeed, due to difficulties in ensuring effective public sector management, the costs involved in monitoring performance within the public sector may be higher than external transaction costs (Elinder and Jordahl 2013).

11 As noted in Section 4.2.1.1, there are also some differences in this respect between for-profit and non-profit organisations.

But even if it were possible to contract all types of quality, it may not be desirable to impose the strict top-down accountability and regulation that would be required. Indeed, the Coalition government's education reforms in 2010-15 were predicated on the idea that school autonomy is beneficial for pupil outcomes. At the same time, increasing accountability to central governments via funding agreements and other control mechanisms may stifle any real increase in autonomy (Mansell 2016). In other words, if contracts are too strict this may undermine the very purpose of outsourcing service provision to independent providers in the first place.

Yet, at the same time, if contracts are not detailed enough, in the right respects, this may make it easier for providers to exploit information asymmetries and engage in quality shirking via cost cutting non-contracted aspects. External providers of a service may invest time in improving quality or reducing costs, and have the discretion to decide, unilaterally, which, if any, uncontracted aspects of a service to deliver. They have, in other words, what are known as 'residual rights of control' (Hart and Moore 1990), and thereby the option of reaping the rewards of keeping down costs and taking higher profits. If they instead choose to invest in quality, by contrast, they must seek to obtain a higher price from the government. Assuming that providers ignore adverse effects on quality as a result of cost reductions, they may thus face stronger incentives to increase profits by cutting costs rather than to improve quality (see Hart, Shleifer, and Vishny 1997). As a result, costs may decrease, but the same applies to quality – and the overall implications for efficiency, quality per pound spent, are ambiguous.

On the other hand, if the government instead chooses in-house production, it retains the residual rights of control over aspects of the services that are not covered in contracts. This means that employees involved in the service production need the government's approval to pursue either cost reductions or quality improvements in regard to these aspects. It also means they receive only a relatively small share of the returns to either strategy, since theirs is not an owner's stake – unlike that enjoyed by owners of the external organisation. As a result, government in-house production of a service generates lower incentives for both cost reduction and quality improvement compared to outsourcing to an external provider (Hart, Shleifer, and Vishny 1997). The question, then, is

whether or not the external providers' incentives help stimulate higher, equal, or lower efficiency given these additional incentives.

Overall, the above framework indicates that private production should lower costs compared with public production, while the impact on quality is more ambiguous. The effect may be positive, zero, or negative depending on how strong the relationship is between costs and quality – and whether or not there are opportunities for innovations that could improve quality (Hart, Shleifer, and Vishny 1997). In other words, whether or not procurement from the private sector is preferable to government production is likely to depend strongly on the service in question.

4.2.1.1 How strong is the case in favour of government service provision?

In practice, however, conditions favouring monopoly provision by the government are rarely applicable (see Shleifer 1998). Indeed, even when cost reductions have negative effects on non-contractible quality, private organisations' incentives to innovate may still ensure that they provide higher efficiency overall. In cases where innovations are important, the case for government monopoly provision appears especially difficult to justify.

In addition, just because independent providers may choose to ignore adverse effects on service quality, it is not clear that they will in fact do so. If providers generate sub-par non-contractible quality, this may cause reputational damage – which in turn may lead the government to decide in favour of another provider in the future, or even take the service in house and produce it itself.¹² While discretion is often feared to induce corruption, it may actually be advantageous in procurement for services with incomplete contracts, since it is generally supported by more dynamic, informal governance mechanisms (such as ethos, culture, brand values and reputation) of the kind that characterise the private sector (see

12 Historically, EU competition law has put limits on the ability of contracting authorities to make use of suppliers' past performance when rewarding future contracts. However, education has been considered 'light touch' for regulatory purposes, which means that member states have wide discretion about how to formulate criteria and rules for awards (CSS 2015a). In addition, recent changes to the general framework mean that: 'Poor performance under previous contracts is explicitly permitted as a grounds for exclusion [of suppliers from the selection stage]' (CSS 2015b, p. 4). Of course, given the result of the Brexit referendum in June 2016, it is not clear whether these regulations will apply in future in the UK context.

Malcomson 2012; Spagnolo 2012). Indeed, research suggests some discretion may be important for producing higher quality in outsourcing (see Butler et al. 2013; Coviello, Guglielmo, and Spagnolo 2016; Decarolis, Pacini, and Spagnolo forthcoming; Kachour, Mamavi, and Nagati 2016; Spagnolo 2012). Depending on how difficult it is to observe quality, the reputation control mechanism is likely to be more or less well functioning – but is highly unlikely to play no part whatsoever.¹³ This suggests that allowing a measure of discretion to assess the performance of private providers on non-contractible quality elements would be sensible. Those that have underperformed in the past may be suspended, penalised in other ways, or even barred entirely, thus giving them strong incentives not to squander their reputation by decreasing quality. In this way, it would be possible to ensure that desired levels of quality will be delivered.¹⁴

In general, *ex ante* competition at the point of tender also helps to ensure that potential providers are the ones most suitable for delivering the service in question, at the lowest possible costs to the government (Andersson and Jordahl 2011). The addition of user choice and *ex post* competition can add a further level of insurance, as a reduction in quality is likely to lower future demand. User choice should incline providers to cost reductions that do not have adverse effects for quality, while at the same time incentivising them to maximise innovation (Shelifer 1998). The relative merits of allowing *ex post* competition via user choice in qualifications and assessments, rather than solely *ex ante* competition via franchising, are discussed in detail in Section 4.2.2.

In relation to the case for government welfare provision of qualifications and assessments, a final point to note is that there are important differences

13 The reputational mechanism hinges on the assumption that the government can observe and understand the fact that lower quality is being produced in the first place. If this is not the case, a provider's reputation may not necessarily match the overall quality it delivers (MacLeod and Urquiola 2012). Yet if the government has difficulties in observing and understanding quality, it hardly makes the case for government provision any stronger either.

14 Discretion can also be exercised more generally to select the highest-quality supplier in the first place. For example, buyers could ask all competing suppliers to provide them with detailed plans for future provision and a discussion of how and why it would stimulate quality in the sector. This gives buyers considerable opportunities to take into account expected levels of quality that is more difficult to measure. While buyers could inspect these plans themselves, they may instead allow an external expert panel to judge the different bids, thereby decreasing discretion slightly and therefore also the risk of corruption (see Bergman 2013).

between market players with different ownership structures. The prevalence of not-for-profit providers weakens the case for government monopoly provision as a solution to quality shirking even further. Whereas for-profit firms distribute profits to owners, non-profit organisations can distribute surpluses in other ways, such as providing better working conditions for staff. Nevertheless, such perquisites are not as valuable as cash. Non-profit status, therefore, weakens incentives to maximise profits (Andersson 2009; Glaeser and Shleifer 2001).¹⁵ This, in turn, provides a quality signal in the provision of complex services where transaction costs are especially high (or perceived to be high). In such situations, non-profit companies may use their status as a way to reassure customers and other actors of their intentions (Glaeser and Shleifer 2001). Since non-profit status translates into weaker incentives to pursue cost reductions that negatively affect quality, the case for a government monopoly becomes even weaker. Of course, non-profits lack the advantage of generating taxable profit, which overall may make for-profit provision preferable, but they do offer a useful alternative to a mixed market for governments that may be seriously worried about high-powered incentives, and their consequences, in private firms.¹⁶ In such circumstances, they may even choose only to allow non-profit actors to operate.¹⁷

Overall, therefore, sole government provision of a service appears to be the best option economically only in cases when all of the following apply: (1) there are considerable opportunities for engaging in cost cutting at the expense of non-contractible quality; (2) the importance of and opportunities for innovation are low; (3) ex post competition is weak and user choice is ineffective; and (4) reputational mechanisms are weak (Shleifer 1998). In practice, there are

15 See Hoxby (2003) for a discussion applied to the schools market.

16 Because profits can be taxed, they may thereby be offset against costs, representing a saving for the government even if for-profit providers' prices are not lower. This advantage is not present when the external organisation is a non-profit organisation. Of course, increasing salaries in non-profit organisations could also increase taxes, but is dependent on their deciding to distribute surpluses in this way. Furthermore, since profit-maximising incentives are lower, one may assume their surpluses may be lower in general. In other words, as long as efficiency is not negatively affected, allowing for-profit providers to deliver services should be preferable.

17 Historically, EU law may have constrained policymakers' discretion in this case, but procurement of educational services can be restricted to non-profit organisations (CSS 2015a) as, for example, in the case of academies and free schools, which, in terms of their legal personality, are by definition non-for-profit.

only a few such cases within the realm of welfare services that satisfy all these conditions.¹⁸

4.2.1.2 Implications for qualifications and assessment

Are there any reasons to believe that the qualifications and assessment sector would be one of the few cases where outright government provision is preferable? Based on the above framework, the short answer is: no.

First, qualifications and assessment include relatively few non-contractible elements compared with other welfare services. This is partly because qualifications and assessment are supposed to form the basis of reliable metrics for aspects of the educational process that are by definition measurable and quantifiable. At a general level, their most essential purpose is precisely to provide and summarise information about pupils' ability and knowledge via such metrics (Croft and Howes 2012). Unlike most welfare services, regulators of qualifications and assessment have the advantage of being able to scrutinise most crucial aspects of the service – such as design of qualifications and syllabuses, modes of assessment, marking criteria, and marking procedures – prior to delivery.¹⁹ The problem of non-measurable quality is thus unlikely to be insurmountable in the procurement of qualifications and assessment.²⁰

This remains the case even where the information provided through qualifications and assessments is used for purposes other than intended (Newton 2007). Yardsticks and benchmarks are soon required to measure quality with reference to these purposes. New assessments may even be required to provide

18 As we discuss in Section 4.3, choice and ex post competition may work quite differently in qualifications and assessment, making their effects more ambiguous.

19 The level of effort put into the examination process by examiners is an important exception. Yet this would also be the situation under a single examination board. Furthermore, conditions and pay for examiners – which may affect their level of effort – could form part of any procurement decision and are thus contractible. For example, it would be possible to require suppliers to use piece rate models, as is common in England, or release compensation models where examiners obtain permission to use time during their regular employment to mark scripts (see SQA 2014).

20 Certainly, the fact that non-contractibility does not appear to be a major problem in qualifications and assessment does not mean there are no non-contractible elements at all. Some quality elements that are measurable may still be unforeseeable prior to the award of contracts or the ability to operate (Bergman 2013). Yet there is little to suggest these problems are especially severe in qualifications and assessment – and they should also be ameliorated by the reputational mechanism discussed below.

better measurement. Indeed, qualifications and assessment with too many purposes will inevitably compromise on fitness (Pellegrino, Chudowski, and Glaser 2001), making it important that all purposes and yardsticks are made explicit and empirically verifiable (Coe and Heller Sahlgren 2014) – irrespective of whether they are provided by the government or independent suppliers. In other words, high-quality qualifications and assessment to some extent *require* their purposes to be contractible and verifiable.²¹

While this does not mean there are no non-contractible elements at all in qualifications and assessment, there is little reason to assume that non-contractibility is an insurmountable problem that would warrant full government provision. For this reason, it is also difficult to envisage considerable opportunities for providers to engage in cost cutting at the expense of non-contractible quality.

Whatever wiggle room is left to reduce quality one would expect to be further ameliorated by the reputation mechanism, which should function well in the qualifications and assessment system. Indeed, as discussed in Section 3, the fallout from the revelations of malpractice of a couple of years ago is evidence that poor performance can cause considerable reputational damage and, indeed, even potentially jeopardise the future of the entire industry. While in the cases referenced, most of the allegations suggested activities that would be in breach of any regulation or procurement contract, this is clearly relevant for potential breaches of any unforeseen non-contractible elements. Such ‘fire alarms’, which allow whistle-blowers from different sectors of society to bring suspected malpractice to public notice, should be regarded as important market accountability mechanisms (McCubbins and Schwartz 1984). In combination with some discretion in the tendering (or approval/monitoring) process to ensure that the ‘shadow of the future’ hangs over providers, such mechanisms could play an important role in reducing the effects of any information asymmetries which favour providers over buyers, in terms of both contractible and any non-contractible quality elements.

Furthermore, the scope for innovation in qualifications and assessment is likely to be considerable and important, especially in times of rapid economic and social transformation. A considerable advantage with competition – *ex ante*

21 We note that the idea of explicitly stating and justifying the intended purposes of qualifications and assessments is a clear recommendation of the authoritative Standards for Educational and Psychological Testing (AEARA, APA, and NCME 1999).

and/or ex post – among independent providers is its shaping of a discovery process, which helps ensure that the content and delivery of qualifications and assessment continue to meet the needs of employers and HEIs.²² Independent providers have strong incentives and ability to involve such institutions and other end-users in product and service design to an extent that is unlikely to be matched by the government alone. Indeed, history also seems to justify this concern in qualifications and assessment specifically.²³

While there is a case to be made that a single government exam board would concentrate expertise, improve the sharing of best practices, and generate more economies of scale (e.g. SCORE 2012), these advantages may be of little benefit if incentives are weak and generate subpar quality, as the history of government monopolies in other sectors has convincingly shown (see Le Grand 2007; Shleifer 1998). Indeed, the argument that government provision is important for ensuring economies of scale could be made for *any* service in *any* sector, which highlights its weakness. The main question is whether or not any benefits achieved through monopolisation via economies of scale outweigh a lack of incentives. In most circumstances, this is rarely the case.

We further note that most exam boards currently operating on the market are non-profit organisations, which, as highlighted above, have weaker incentives to engage in cost cutting at the expense of non-contractible quality. If fears of high-powered incentives are considerable, the government could restrict tendering to non-profit organisations (CSS 2015a). While there is little reason to believe that for-profit organisations would provide worse qualifications and assessment quality as such, we merely note that the existence of non-profit independent exam boards makes the argument in favour of direct government provision even weaker.

Overall, therefore, there is little reason to believe that government provision of qualifications and assessment is preferable to independent provision. Far from

22 For a discussion of competition as a discovery process in general terms, see Hayek (2002).

23 Historically, most successful innovations in qualifications and assessment have come from independent organisations, which have balanced societal needs, as articulated by the government, with the interests, requirements, and concerns of universities, employers, and educators. Examples include the Nuffield and Salters' science suites, Ridgeway History, and MEI Maths. The government, on the other hand, has hardly been involved in such innovations at all – and when it has been involved, its innovations have not been particularly successful. Examples include General National Vocational Qualifications and National Vocational Qualifications (see Cambridge Assessment 2012).

all of Shleifer's (1998) conditions are fulfilled.²⁴ Comparing the properties of qualifications and assessment with the general economic theoretical framework outlined in Section 4.2.1, we find few reasons to support outright government provision over independent provision.

4.2.1.3 Empirical evidence from other related services

Whether independent provision by itself is more effective in practice than government monopoly provision is likely to depend on the properties of the service in question as well as the model for independent provision. To the best of our knowledge, there are no studies evaluating the effects of private provision of qualifications and assessment on costs and quality. We thus briefly review the evidence from other sectors. Because franchising without user choice has been floated as a key alternative to direct government provision of qualifications and assessment, we focus on studies analysing outsourcing only rather than those that mix features of procurement with choice. Private provision with and without user choice is considered separately in Section 4.2.2.

In regard to procurement more broadly, the literature as a whole indicates that outsourcing reduces costs without hurting quality. This is especially the case in services with a low level of contracting difficulty, but apparently also often in services with more difficult contracting problems. The exceptions are in some 'credence services' where arguably only the provider has information on the optimal level of quality and quantity, and this information cannot be verified either *ex ante* or *ex post*, which allows producers to overstate or understate the need for a service. Even then, however, the evidence appears mixed (see Andersson and Jordahl 2011).²⁵ Nevertheless, we do not believe most

24 The level of choice and *ex post* competition will be determined by the choice of model for independent provision. We discuss this in Section 4.2.2. For the purposes of argument, here we assume that it will play no role whatsoever, as private provision without choice is presently considered an attractive alternative to direct government provision. In respect of the model of independent provision, as is always the case, how well it works in practice will depend on system design.

25 The prime example highlighted in the literature is residential youth care, where many quality aspects may be very difficult to contract or be subject to secrecy, as in the case of therapy. However, even in the case of residential care, design of outsourcing appears crucial. In Sweden, for example, there is clear evidence that social services do not always take the care they should in the writing of contracts and in monitoring the quality of the service provided (see Lindqvist 2008). Clearly, the way the outsourcing process is handled will affect the quality and cost of the service provided.

of this research, which focuses on unrelated services, is particularly informative for the purpose of understanding the effects in qualifications and assessment. Instead, we focus on more specific cases that we believe offer some lessons for the viability of procurement in qualifications and assessment.

4.2.1.3.1 Case 1: schooling

First, we consider the case of schooling, which we believe offers interesting lessons for the subject under investigation. After all, the provision of schooling, like qualifications and assessment, is part of the education production function, although it is likely to involve more non-contractible features than qualifications and assessment.²⁶ Research indicates that handing over operating responsibility for existing American and English schools to independent organisations often appears to raise achievement among children already attending these schools (see Abdulkadiroğlu et al. 2016; Eyles and Machin 2015; Eyles, Hupkau, and Machin forthcoming). Intriguingly, research indicates the positive effects found in America are generally achieved with considerably lower public funding (see Batdorff et al. 2014), suggesting the efficiency gains are considerable indeed. Related research also indicates that independent organisations do not achieve these gains via strategic coaching but rather via bona fide improvements in human capital (Cohodes 2016). Other research finds that contracting out schools to for-profit providers had positive effects on mathematics test scores after six years, while the impact of non-profit providers was zero (Chingos and Peterson 2009), while another study finds positive effects of contracting out schools to non-profit providers (Ruble 2015). Interestingly, field experiments seeking to replicate best practices from the independently managed American charter-school sector in low-performing state schools had similar effects on pupils attending those schools (Fryer 2014). Thus, although we cannot entirely rule out manipulation or a decline in non-observable quality, this research indicates that independent providers, in some settings at least, appear to improve cognitive performance, the measurement of which is the primary purpose of qualifications and assessment.

26 For example, as noted in Section 4.2.1, many argue that school quality cannot be reduced to quantifiable metrics, whereas such quantifiable metrics are part and parcel of qualifications and assessment.

4.2.1.3.2 Case 2: health and elderly care

Another interesting case concerns health and elderly care, which are welfare services that are likely to involve a considerable degree of non-contractible quality (e.g. Bergman and Jordahl 2014). Some research indicates that independent provision of publicly funded health care can improve efficiency. In Germany, Tiemann and Schreyögg (2012) find that privatisation of public hospitals generates higher efficiency, both when considering the number of inpatient cases per year and in-hospital mortality rates per year as output variables. The effects are especially pronounced in the case of for-profit provision, but apparent also with non-profit provision. The fact that efficiency gains are apparent also when considering mortality indicates that non-contractible quality did not suffer as a result of privatisation.²⁷

Similarly, in a recent study, Bergman et al. (2016) analyse the impact of elderly care procurement in Sweden, where about 10 per cent of nursing homes have been contracted out to private operators – mostly for-profit corporations, often owned by private-equity firms – during the period of study. The authors focus specifically on mortality as a measure of non-contractible quality to understand whether or not procurement of a highly complex welfare service makes providers cut corners.²⁸ They find that both mortality rates and costs per resident fall in the municipality as a whole as a result of privatisation. This indicates that efficiency gains as a result of innovation and competition were more important than incentives to cut costs, generating higher non-contractible quality at lower cost compared with in-house production.²⁹

These cases suggest that independent provision of complex welfare services may under certain conditions improve efficiency, although the precise effects

27 Generally, most research focuses on differences between different types of private hospitals, rather than differences between private and state hospitals, and are often not sophisticated enough to unveil causal relationships (see Herrera et al. 2014).

28 Mortality is difficult to contract because it is likely to entail screening residents and because it is a noisy quality measure with which to judge the performance of care homes. Consequently, mortality was not included in the contracts (Bergman et al. 2016). This, in turn, means that care homes would not necessarily be penalised for cutting costs in ways that increased mortality.

29 Again, most research focuses on differences between different types of private elderly care homes, rather than differences between private and state care homes. Similarly, most are not sophisticated enough to unveil causal relationships (see Commondore et al. 2009).

are likely to be dependent on system design (e.g. Bergman 2013; Bergman and Jordahl 2014). Nevertheless, the research indicates that related welfare services – which arguably are more complex and include more non-contractible quality elements than qualifications and assessment – can be successfully procured.

4.2.1.4 Verdict: private provision is preferable to a single government board

Overall, neither economic theory nor empirical evidence provides a good justification for government provision of qualifications and assessment. Furthermore, we note that nationalisation itself would be a gargantuan and costly exercise, as examination boards would have to be compensated to a cost of over a billion pounds (Husbands 2015). We therefore conclude that private provision is preferable to government provision in qualifications and assessment in the English context.³⁰

4.2.2 *The second question: franchising or user choice?*

Having ruled out government provision via a single exam board as a good option, we now turn our attention to the question of whether or not an ideal model of independent provision in qualifications and assessment should include user choice. We begin by considering the general case for the two different models and then apply this model to the specific service under consideration.

4.2.2.1 A difference in substance or degree?

In welfare services, user choice is sometimes incorporated into the service design with a view to generating more efficient and equitable provision. Combining public funding with independent provision and user choice gives rise to ‘quasi-markets’ (see Le Grand 1991; 2007).³¹ It is useful to think about quasi-markets as ‘procurement plus’, since procurement includes the other two elements too.

30 Outsourcing out each function would certainly create logistical problems and poor connections between providers should different functions for the same subject be outsourced among different providers (e.g. Education Committee 2012), however procurement by subject does offer a way of avoiding these problems.

31 Technically, independent provision is not necessary for quasi-markets, since users can be allowed to choose between different public providers. Indeed, the English quasi-market in schooling was long characterised by relatively little independent provision, especially in the primary sector. Nevertheless, this option is not on the table for the qualifications and assessment system and so we refrain from discussing it in detail.

In fact, procurement models could technically include choice and ex-post competition as well, therefore making the distinction less clear.³²

To distinguish between the procurement and user choice models conceptually, we consider the principal difference to be that procurement involves time-limited contracts, which the government then awards to one winning provider, or a given number of winning providers, whereas a full-fledged choice model involves in principle an indeterminate number of suppliers. These must meet the regulatory standards necessary to enter the market, as determined by an authorising government agency. The authorising agency usually then has the power to revoke authorisation depending on performance. In general, the choice model thus involves a greater degree of user empowerment and more dynamic ex post competition – with greater scope for new competitors to enter the market as demand changes – but no ex ante competition whereby the government attempts to pick winners, as in the procurement model.³³

4.2.2.2 The general case for user choice

Why would the choice model be preferable? One reason is that user choice and ex post competition may minimise opportunities for suppliers to engage in cost cutting with negative consequences for non-contractible levels of quality. This is because demand can respond directly to lower quality (Shleifer 1998). If users can assess the quality of a service they buy directly from producers on their own, and the market is perfectly competitive, suppliers should face optimal incentives since they get a lower price for producing lower quality as a result of engaging in cost reductions – and a higher price for generating higher quality by innovation (Hart, Shleifer, and Vishny 1997). In other words, the choice model may decrease the importance of the problem of incomplete contracts, as discussed in Section 4.2.1.

32 For example, in the tendering process, the government could decide that several organisations fulfil the requirements to supply desirable services and thus allow users to choose between them (see CSS 2015a). However, in the case of qualifications and assessment, we note that the rationale behind reforms under consideration is to abolish user choice altogether via franchising. Moving to a procurement system that allows user choice would appear relatively pointless from this perspective.

33 Certainly, the distinction between the two models is still fluid rather than distinct, with the difference depending on whether or not there is an open or restricted tendering process and the extent to which the government attempts to select winners or merely focus on minimum requirements.

Of course, more generally, the choice model may be preferred because it sharpens incentives among suppliers to focus on users' preferences. In other words, choice may improve efficiency through better matching between users and producers than can be achieved via procurement. Second, competition may improve efficiency by forcing lower-performing suppliers to improve, or ensuring that they exit the market when user demand shifts decisively to higher-performing producers. These features often make choice models theoretically preferable to procurement models (at least those without any element of choice), and indeed are the basis of most arguments in favour of choice and ex-post competition in welfare services such as education (e.g. Friedman 1962; Hoxby 2003). In theory, choice should usher in a more competitive market that should generate more efficient and responsive welfare systems in general.

4.2.2.3 Making things a bit more complicated

Yet the story is more complicated in most markets for welfare services than the general appeal to the textbook case of a well-functioning free market suggests. In a completely unfettered market, incomplete contracts may also mean that choice and ex post competition merely aggravate the problem (see Hart 1975). The outcome of competition is thus likely to depend on whether users are able to observe non-contractible aspects of quality and whether they are responsive to changes in such quality.

In general, incomplete markets are often characterised by considerable product differentiation, as oligopolistic producers seek to distinguish themselves in order to raise profits (e.g. Tirole 1988). Most services are heterogeneous by nature – since one delivery is never exactly the same as another – and so are user preferences for those services. This is certainly the case in public services. For example, in education, some parents may prefer progressive education, while others prefer more traditional instruction. The combination of heterogeneity in services and preferences tends to give rise to suppliers with market power that increases inversely with the level of substitutability between them. The result is either monopolistically competitive markets, characterised by the existence of many suppliers that users can access with relative ease, such as markets for physicians, or differentiated product oligopolies, if the market is dominated

by a few actors that continuously interact over extended periods of time, as is the case on hospital markets (see Gaynor and Vogt 2000). Monopolistic competition is also generally characterised by ease of entry and exit as well as extensive supply and variety of products. Because demand is relatively elastic, suppliers have incentives to keep costs low and outcompete each other by advertising and strong differentiation. Oligopolies, on the other hand, are often characterised by quite significant entry barriers and strong incentives to keep price competition low.

We note that oligopolistic market structures may be competitive as long as entry barriers are kept low. In this dynamic view, firms compete against both existing and potential producers – through innovation and quality rather than price alone. The incentives among entering and existing firms to innovate come from the prospects of capturing monopoly rents. Innovations that change the dynamics of the game, through ‘creative destruction’, give rise to firms that can thus come to temporarily dominate the market (Schumpeter 1994). In fact, oligopolistic competition may under certain situations be the market structure most conducive for innovation – precisely because price competition is replaced by competition on the basis of innovation, which in turn spurs product differentiation (Baumol 2002).

Empirical evidence supports a nuanced version of this intuition. The empirical relationship between competition and aggregate levels of innovation appears to follow an inverted U-curve: at low levels, increasing competition spurs total levels of innovation, but at high levels this impact levels off or even becomes negative. Research further indicates that firms in sectors close to the technology frontier innovate as a response to increased entry threats – in order to capture new rents – whereas firms in sectors farther from that frontier tend to innovate less with more entry threats, since they know they have little chance of winning against a potential entrant (see Aghion and Akcigit 2016; Aghion, Akcigit, and Howitt 2014). This goes to show that oligopolistic market structures are not necessarily uncompetitive, but competition is not always visible and hinges on ensuring free entry and exit to the market.

In markets with considerable product differentiation, economic theory predicts that competition should generate higher quality when prices are regulated. But when prices are unregulated, it is more ambiguous in regard to

the level of quality that will be produced. This will depend on users' relative responsiveness to price and quality – which in turn depends on the extent to which these properties can be measured and observed. In cases where price is measured inaccurately, but quality is measured accurately, demand is likely to be less responsive to price than quality and thus allow suppliers to increase their rates – but also generate higher quality. On the other hand, if quality is measured inaccurately, but price is measured accurately, prices will decrease but the level of quality supplied will be too low since demand will then be less responsive to quality.³⁴ In other words, if users are not sensitive to price, they should be responsive to differences in quality between producers. Consequently, in such markets, both prices and quality should be high. But in markets where buyers are sensitive to price, producers may instead compete on this dimension – and generate both lower prices and quality (see Gaynor 2006; Gaynor and Town 2011; Propper and Dixon 2011).³⁵ Thus, theory predicts that choice and competition will affect quality differently depending on the market in question and how the rules of the game are designed.

Yet another important question is whether or not users' understanding of quality is in line with or differs from what is socially desirable. For example, patients are probably equally interested in making sure they get healthy as the rest of society. But in some cases, the congruence of interest is less clear. An important case is antibiotics consumption. While society has a strong interest in preventing antibiotics resistance, patients are perhaps (understandably) mostly interested in getting healthy as quickly as possible. Indeed, some evidence indicates that competition between healthcare providers increases antibiotics prescriptions (Fogelberg 2013). In such cases, competition may raise quality in the eyes of users but not in the eyes of society.

The above framework indicates strongly that characteristics of markets, and of market design, are key for the outcome of competition. Providing welfare services via choice and competition in quasi-markets may thus be a more

34 In addition, with several dimensions of quality, there may be overproduction of the easily measurable dimensions and underproduction of the other.

35 Gaynor (2006) and Propper and Dixon (2011) focus their discussions on health care, but the theoretical predictions are general.

complex endeavour than standard economic arguments about the virtues of markets typically suggest.

4.2.2.4 Implications for qualifications and assessment

Based on the above intuitions, how and why would we expect user choice to affect quality and price in the qualifications market? Applying the general case in favour of choice and competition, we first note that allowing schools to choose the qualification provider may facilitate better matching between pupils and the specific curricula on offer. Such matching would be impossible to ensure within a franchising system without choice.

Second, we note that competition between exam boards may force them to provide better quality, through various forms of innovations, for example. Whereas procurement without choice only provides *ex ante* competition in the tendering process, user choice forces boards to continuously compete for schools' business. Schools must choose their qualifications and assessment for boards to survive.

However, unlike most schooling markets, prices offered by exam boards are not fixed, which means that predictions about competition on quality depend on the extent to which schools value quality in qualifications and assessment – and the extent to which they can discern quality levels between different boards. However, as discussed in Section 4.2.2.3, if schools are not responsive to quality, we may see considerable price competition instead.

In addition, even if they are sensitive to quality, the effects depend on how schools (parents and pupils) perceive quality. If they consider boards that generate higher grades to be offering a better service, this indicates that competitive incentives may merely result in grade inflation and a reduction of standards rather than higher-quality qualifications and assessment. Indeed, as discussed in Section 3, this has been precisely the concern with the existing market. If this is the case, there may be a mismatch between what users want and what is in the wider interests of society.

Certainly, schools may take into account the needs of the specific end-users for whom they cater, including HEIs and employers. Such end-users are not required to accept grades from different exam boards as equal and could

discriminate between them. They might in theory evaluate them differently and give different offers to different pupils – depending on the board by which they are assessed. In such circumstances, boards still have relatively strong incentives to compete by raising quality.

Moreover, if universities and employers feel that qualifications and grades do not offer enough information, they can, and do, opt for alternative selection measures, such as their own entrance tests, to assess candidates independently of their external qualifications. An extension of this practice would decrease the incentives on schools to pick what they perceive to be the easiest qualifications, since they may not adequately prepare pupils for the alternative selection measures taken by universities.

In regard to the reputation mechanism, this probably functions less well among non-selective HEIs, since these have lower incentives to ensure that standards are maintained. This is not to say that they do not care about standards. Indeed, non-selective universities have an interest in ensuring that incoming students are reasonably prepared for their courses, especially given the time and costs associated with the remedial courses now commonplace at English universities (Suto 2012). Still, their financial interest in maximising the number of students enrolled are likely to take precedence.

Overall, therefore, the effects of user choice and ex post competition in qualifications and assessment are theoretically ambiguous. While there could be a virtuous cycle in which choice maximises matching and competition spurs improvements and innovation, there could also be a race to the bottom due to price competition or because the perceptions of quality among schools and those of wider society are not necessarily aligned.

4.2.2.5 The current market dynamic

While there is no direct evidence of the impact of user choice and competition on the quality of qualifications and assessment, we offer a number of observations of current market dynamics to aid understanding of how these features interact.

First, in regard to fears that price competition would decrease quality, we note that the current market does not appear to be characterised by price

competition at all (Ofqual 2012b). According to schools, price is not generally a factor in the choice of boards – mainly because of the costs of switching boards, and the risk that results might be compromised in the process. Instead, the quality of qualifications, and the desire to match them with pupils' needs, appears to be the most important factor (Ofqual 2015b).³⁶ While subjective preferences should be interpreted with caution, we note that they correspond to the regulator's assessment and more generally with how oligopolistic markets work – as described in Section 4.2.2.3. The threat of price competition driving down standards therefore does not seem to square up with how this particular market functions at the moment.

Second, we note that the regulatory framework essentially prevents competition on standards. The government underwrites national qualifications through an equivalency framework designed to assure stakeholders of the comparability of different qualifications and versions of the same qualification. This effectively makes it more or less impossible for boards to brand themselves as producing higher-standard specifications in the same qualification. The equivalency framework forces different types of national qualifications – academic and vocational – into one measure of school performance for the purposes of computing aggregate league table scores (see Bassett 2014; Croft and Howes 2012). Such equivalence could increase perverse incentives among schools to choose qualifications that are perceived to be the easiest. The framework makes it more difficult for schools and end-users to observe any remaining quality differences between boards also.³⁷

While there is no direct evidence of the impact of choice and competition on quality in qualifications and assessment, some research analyses whether or not schools switch to boards with more candidates obtaining at least a C grade in GCSE English in the period 2000-03. There is no strong evidence of this. When schools change boards – quite a rare event in general – they switch to ones with higher and lower previous proportions of candidates with at least a C

36 Prices between boards are very similar – with the difference in the average price amounting to about £2 per entry (Ofqual 2015b).

37 Indeed, we note HEIs never specify entrance requirements separately for different specifications, produced by different exam boards, indicating that they at least officially treat them as equal. This is not surprising since they are all considered the same in the eyes of the government, due to the perceived need of equivalency.

grade to almost exactly the same extent (Malacova and Bell 2006).³⁸ Of course, other things, including attempts to improve matching, may also explain the switches. Although it is difficult to draw strong conclusions from this research alone, it does not support the notion that schools choose exam boards based on their perceptions of which ones offer the easiest specifications – at the very least not on this basis alone.³⁹

Available evidence therefore does not support fears that exam board competition drives down standards. At the same time, historically, it appears that competition has helped to ensure a measure of innovation in the sector (see Croft and Howes 2012). End-users and educators were able to take their ideas of new syllabuses to different exam boards. Some of their ideas were rejected; others were rejected before later being accepted by another board. The existence of several boards ensured that one negative response would not necessarily kill the innovations before they were even tried. As discussed in Section 4.1.2.1, successful innovations by one board have then often spurred others to improve their own products (see Cambridge Assessment 2012). In addition, exam boards in today's market have invested heavily in developing more robust and secure technology to ensure a secure and effective system overall, spurred by strong reputational concerns (Husbands 2015). While this is merely suggestive evidence, it is consistent with the idea that choice and competition may help spur innovations in qualifications and assessment.

That the evidence does not support fears that exam board competition drives down standards is unsurprising given the equivalency framework and Ofqual regulation. If there is any whiff of a specification being too easy, the regulator acts swiftly. For example, in 2015, AQA was ordered to make its GCSE mathematics paper more challenging following mock tests taken by a sample

38 To be exact, 47.22/52.77 per cent of all switches across all years – calculated from absolute numbers in Malacova and Bell's (2006, p. 31) Table 3 and columns 1-3 – mean that schools move to a board with lower/higher shares of pupils receiving at least a C grade compared with the incumbent board. Only when considering switches over the period, rather than between individual years, is there a considerably higher proportion (37.6 percentage points) of boards that switch to boards with higher pass rates than those that switch to boards with lower pass rates.

39 Similarly, only when considering switches across the whole period rather than across individual years is there evidence that schools improved their average grades by switching. In some years, schools actually appear to lose out from switching. Of course, we also note that any improvements may be the result of better matching or changes in the ability of pupils rather than successful attempts to game the system.

of pupils (Richardson 2015). And in 2016, while a petition demanded that Pearson should lower the grade boundaries for the GCSE mathematics paper, Ofqual instead forced the board to increase them to ensure the specification was not easier to pass than those offered by other boards. This judgement was entirely based on statistical predictions (Wiggins 2016a). Given the strict regulation, it is essentially impossible to compete by lowering standards in national qualifications.⁴⁰

We note that the data provided in Section 2 supports this picture. Aggregate market shares, overall and in core subjects, have been more or less stable from year to year. If one board had provided objectively easier specifications in one year, and schools do indeed habitually search for this specification, one would expect a large shift towards that board in any given year. Yet this is not generally the case. If schools indeed do search for the objectively easiest specification, a large proportion of them are apparently not successful. Instead, it could mean that schools switching boards are primarily motivated by other factors than how easy a specification is in an objective sense, such as their perceptions of how different specifications fit their pupils specifically – or, in other words, improved matching between specification and pupils. As discussed below, just because teachers may believe a qualification is objectively easier than another does not mean that is the case.

Certainly, some gaming does occur. For example, it is clear that schools have historically been incentivised to offer, and encourage take-up of, vocational qualifications deemed to be equivalent to the national academic qualifications in league tables (see Bassett 2014; Croft and Howes 2012), and indeed the number of vocational qualifications taken in the period to 2010 grew, to their clear advantage. In another example, quite recently, it was revealed that a considerable number of schools have enrolled their pupils in the European Computer Driving Licence qualification, which requires about three days of teaching but is deemed equivalent to a regular GCSE in the league tables (Adams 2016). It is not hard to imagine that this incentive is playing a part in rising enrolment for this qualification.

40 For more in-depth evidence on how inter-board comparability in grading is maintained, see Ofqual (2015c).

Of course, the issue here is not exam board competition per se, but rather how league tables are designed. The incentives to opt for easier vocational qualifications would be the same even if there were only one examination board offering a particular subject, as long as the equivalency framework was in operation. If the government stipulates that vocational qualifications should be deemed equivalent to academic qualifications, schools will naturally react to those incentives – irrespective of whether or not choice and ex post competition exist.

In fact, just as the equivalency framework makes it difficult to compete by making exams easier, it makes it equally difficult to compete by making them harder. For example, while AQA was recently ordered to make the paper it offered in GCSE mathematics more difficult, OCR, Pearson, and WJEC were all ordered to make theirs easier following research suggesting they were not ‘accessible’ enough (Richardson 2015; Swinford 2015). Again, the equivalency framework is prescriptive and helps ensure that boards are effectively unable to compete on standards, whether upwards or downwards.⁴¹

To some extent, however, we recognise that in the alternative qualifications market there may be a degree of competition on standards. For example, following reforms to scrap what was effectively a state-school ban on offering alternative qualifications, there was a marked increase in the number of state schools offering IGCSEs between 2010 and 2015. In 2010, 846 schools, most in the independent sector, taught IGCSEs. A year later, the figure had increased to 1,402 schools, two years later to 1,842 schools, and by 2013 to 2,677 schools. Historically, the IGCSEs – bearing some resemblance to traditional O levels – were often perceived as more demanding than regular GCSEs, partly because they involved less coursework and were linear rather than modular, and therefore became the preferred choice among independent schools (see Paton 2013a, 2013b; Stewart 2014, 2015). If so, the changes could be evidence of a virtuous cycle in which more demanding qualifications had earned a better reputation than less demanding qualifications and thus gained market share. This is especially the case since the changes took place during times of considerable reform uncertainty for regular GCSEs; switching to a more stable qualification would seem a rational response.

41 Again, for more in-depth evidence on how inter-board comparability in grading is maintained, see Ofqual (2015c).

Yet the government has since made a U-turn, admonishing both state and independent schools to switch back to the regular GCSEs by excluding IGCSEs from league table scores. Now, in sharp contrast, the argument is that IGCSEs are in fact easier than regular GCSEs, following reforms to improve the latter – which no longer include any coursework or oral examinations at all (Adams 2014; Morgan 2015). Some viewed the switch in favour of IGCSEs as evidence that schools were seeking out easier qualifications (Paton 2013b; Stewart 2013). This would then again highlight the difficulties involved in ensuring that competition between qualifications generates a race to the top.

We believe this interpretation is too simplistic. While some schools may have viewed the qualification as easier, the dynamic indicates a much more nuanced picture overall. First, the flight began before the recent changes in the GCSEs, such as the scrapping of coursework, had been finalised and introduced (see Gurney-Read 2014; Paton 2013a; 2013b). Indeed, elements of these reforms rather appeared to have been modelled on the IGCSE, which would suggest that a private qualification had spurred improvement in the national brand.

Second, there is little evidence that IGCSEs are in fact objectively easier than regular GCSEs. A recent benchmarking exercise by NARIC (2016), the national agency devoted to comparisons of international qualifications, found that the IGCSE is a rigorous qualification with a standard comparable to that of the new GCSE. Moreover, end-users in fact appear to view the former more favourably than the latter. Whereas 47% of responding HEIs and 66% of employers agree that regular GCSEs are good preparation for further study, 72% of both HEIs and employers view IGCSEs in this light (Ofqual 2016d).

Finally, far from seeing improved grades following the switch to IGCSEs, schools often saw their grades fall. This is not too surprising because of changes in cohorts, and Ofqual's research also found that grade boundaries were suitable (Wiggins 2016b).

It is important to note that a mere shift in itself does not indicate attempts to game the system. Indeed, such an interpretation ignores the fact that different types of qualifications and assessment may fit some pupils better than others, as highlighted in Section 4.2.2.4. For example, girls appear to benefit from continuous assessment, whereas boys fare worse on average (see Bramley, Vidal Rodeiro, and Vitello 2015). Switching to qualifications offering different types

of syllabus or assessment is therefore always likely to affect different pupils differently – even if the qualifications are equally difficult. The same applies to oral versus written assessment: while some pupils may be able to better present the knowledge they have acquired orally, which may induce their teachers to switch to qualifications that include more such assessment, other pupils are going to be relatively stronger in presenting their work in a written format, which may induce their teachers to switch to qualifications that include more written assessment. This does not necessarily mean that one or the other form of assessment is objectively easier, but could instead indicate that the matching mechanism is functioning as it should.

Certainly, teachers and schools may confuse a qualification that suits their particular pupils for a qualification that is easier in an objective sense. In other words, schools that act on local rumours that a qualification is objectively easier by switching may in reality unwittingly act on correct information indicating that the qualification is especially suitable for their pupils. If so, the switch would still indicate positive incentives in the system. What schools and teachers believe is irrelevant, as long as their actions lead to better outcomes.

In sum, overall, there is in fact little proper evidence to support the argument that schools only opted for IGCSEs because these were in fact easier than regular GCSEs in an objective sense, and more to suggest that IGCSEs, at least to some extent, brought healthy competition to the system.

We also note that reputation – in terms of end-user assessment of qualifications quality – appears to have a role in the alternative qualifications market. First, end-users clearly distinguish between vocational and academic qualifications. Survey results included in Ofqual’s 2016 annual report show 17% and 21% of responding HEIs and employers respectively valuing Level 3 BTECs as highly as A levels. And while 53% of HEIs believed BTECs to offer good preparation for further study, the figures for academic qualifications were between 62 and 95% (Ofqual 2016d). Such differences are likely to lead to different entry requirements. Indeed, for example, whereas the University of Oxford recognise BTECs, the University of Cambridge normally does not. Other HEIs, such as London College of Fashion, believe BTECs may even provide better preparation for their specific courses (see Reidy 2015). In other words, end-users appear to form their own view of academic and vocational qualifications, regardless of

what the accountability framework stipulates. This is a case of the reputation mechanism in action.

Furthermore, similar differences appear when alternative academic qualifications are compared with traditional qualifications. Indeed, while 62% of HEIs view A levels as good preparation for further study, the figures are 85% for international A levels and the Cambridge Pre-U, and 95% for the IB programme.⁴² At the same time, 65% and 64% of HEIs and employers respectively value International A levels 'as highly as' regular A levels, while 75% and 60% of HEIs and employers respectively say the same thing about the IB programme. Meanwhile, 61% of HEIs value the Cambridge Pre-U 'as highly as' A levels (Ofqual 2016d).⁴³ This suggests that end-users of qualifications and assessment appear to discriminate between national and alternative qualifications in ways that should improve the functioning of the reputation mechanism in the market as a whole.

That HEIs discriminate between certain qualifications is evident in their entrance requirements. Perhaps the best example is the case of international qualifications. In this regard, the London School of Economics (LSE) provides a guide to how different international qualifications correspond to traditional A levels, and how it, as an institution, distinguishes between different countries' systems in terms of standards. Pupils applying to undergraduate degrees from some countries, such as Austria, Finland, and France, can expect to receive a specified offer solely based on their performance in their national qualification. However, pupils from other countries, such as Albania, Russia, and Ukraine, cannot apply to the university based on their national qualifications alone. Applicants from other countries, such as Montenegro, Latvia, and Sweden, are informed about minimum offers based on their qualifications – but are also warned that they may be asked to sit the three-hour UG Admissions Assessment (see LSE 2016a). This test is also used to assess candidates with BTEC qualifications, whereas applicants sitting Scottish Advanced Highers, the Cambridge Pre-U, and the IB programme can expect a specified offer solely based on their examination performance (LSE 2016b). While the LSE's own

42 The Cambridge Pre-U was produced as a more demanding alternative to traditional A levels in order to better prepare students for university studies (see CIE 2011).

43 No employers responded to the question in regard to the Cambridge Pre-U.

judgements are based on information from NARIC, it draws more nuanced, more specific interpretations to separate them further.

Other universities have similar, but not exactly the same, guidelines, indicating that they are often capable of discriminating, and do discriminate, between considerable numbers of different qualifications, providing their own interpretations of which qualifications and grades are equivalent to the English ones.⁴⁴ Indeed, research shows that universities are more likely to judge grade equivalency of A levels and the IB programme based on the future degree performance of candidates taking the different qualifications than on the official UCAS tariff (Green and Vignoles 2012). In other words, end-users appear to act on the reputation and their own experience of different types of qualifications.

Furthermore, an increasing number of universities are now setting their own entrance examinations, often regardless of the qualifications held by applicants (see Swain 2016). These tests should decrease the value of efforts to seek out qualifications, and particular versions of such, that are perceived to be easier. The existence of university entrance examinations relativises the importance of national qualifications. If anything, such tests should make it even more important for schools to choose qualifications that will prepare their pupils effectively for these entrance examinations. To some extent, therefore, they have a quality safeguarding function in the system.

Certainly, the reputation mechanism is less likely to function well for schools whose pupils apply in the main to universities and courses that are not oversubscribed. While recruiting universities do have an interest in ensuring entering students have been well prepared, they are likely to have stronger financial incentives to accept students on minimum course requirements. In the case of recruiting HEIs, then, the reputation mechanism is unlikely to help encourage schools to choose more difficult qualifications.

Furthermore, and most importantly, because of how the accountability framework is constructed, schools in general are likely to primarily strive toward higher performance in league tables based on the qualifications their pupils take, rather than acceptance rates to specific universities or employers.

⁴⁴ Again, while the broad judgements appear to be based on information from NARIC, institutions normally make their own subjective interpretations of what they consider equivalent in terms of grades necessary for entry.

Unfortunately, this applies to both national and alternative qualifications. This further decreases the reputation mechanism's power in the competitive process to raise the standard of qualifications – and further highlights how the equivalency framework, and the apparatus that supports it, prevents competition that could induce improvements in quality. If schools mostly care about their league table scores, rather than which HEIs or employers their pupils may wish to apply to afterwards, the reputation mechanism will not work for quality-improvement purposes.

Overall, therefore, we conclude that the current qualifications and assessment market is not characterised by strong competition on either price or standards when it comes to different versions of the same qualifications. The principal problem is that the system is essentially set up to ensure that boards cannot compete on quality in terms of standards – whether by lowering or increasing them. The only partial exception is the dynamic in the market for alternative qualifications, which in combination with the reputational mechanism among end-users may induce some upward competition on quality in some qualifications. Still, the equivalency framework in combination with the fact that many (or most) schools are more likely to be dependent on their league table results than university enrolment rates or employment figures means that the reputation mechanism is severely constrained.

Of course, the overall quality of qualifications and assessment is naturally broader than just the academic standard of qualifications. In this respect, we note that the current market dynamic in fact appears to have generated a more secure and effective system overall via investments and innovations in technology. This aspect is often neglected, but is an important quality aspect linked to private providers reputational concerns on the existing market.

4.2.2.6 Empirical evidence from other relevant services

While the above discussion provides 'soft' evidence on the functioning of the current qualifications market, it does not provide any research analysing the effects of choice and competition on bona fide quality in qualifications and assessment. In the absence of such research, however, we consider other relevant research analysing the effects of choice and competition in other types of publicly

funded welfare services, as in Section 4.2.1.3. This helps us judge the viability of choice and competition in qualifications and assessment more reliably.

4.2.2.6.1 Case 1: schooling

As in Section 4.2.1.3, we first consider the evidence of choice and ex post competition among independent suppliers in schooling. Overall, the research indicates that parental choice and school competition generate small-to-moderate positive effects on pupil outcomes (see Heller Sahlgren 2013). Certainly, the effects differ across countries, but the general picture indicates that choice and ex post competition can raise achievement to some extent.

For example, the Swedish voucher system appears to have improved short- and longer-term national outcomes as well as scores in the international test TIMSS, while at the same time having no or even negative effects on costs. Importantly, the effects on national outcomes are driven mostly by free school competition with state schools (Böhlmark and Lindahl 2015). Analysing the longer-term effects of competition from independently operated schools on PISA scores, West and Woessmann (2010) display similar findings: competition from independently operated schools produces better results and lower costs, thus raising productivity overall. Again, the effects on pupils in independently operated schools are only slightly higher than the impact on pupils in state schools.

Since user choice of independent providers – which is the mechanism generating ex post competition between them – appears to lead to improvements also for pupils not attending such providers, this suggests the same effects would not necessarily be achieved by merely contracting out schools to private providers without allowing any choice whatsoever. Indeed, in Florida, the mere threat of allowing independent-school choice led to improvements in state schools (Figlio and Hart 2014), suggesting the competitive mechanism – induced by the potential for choice – can improve school productivity by itself. While we note that it is difficult to draw too strong conclusions in this respect, this indicates that choice and ex post competition may be preferable to procurement without choice in schooling.

Nevertheless, there is also research suggesting that choice and competition is not always helpful. In Chile, the evidence does not indicate much benefit from the voucher reform enacted in the 1980s (Hsieh and Urquiola 2006). Unlike Sweden,

Chilean schools have been able to select pupils with academic tests, which probably allowed them to compete by other means than generating higher quality (MacLeod and Urquiola 2012). However, this merely indicates that all choice and competition are not equal, highlighting the importance of ensuring healthy incentives in any system that allows these features.

While schooling is an important case to consider for understanding the possible effects of choice and competition in qualifications and assessment, the most important limitation is that it rarely involves full price competition.⁴⁵ We thus also consider welfare services where both quality and price competition have been allowed.

4.2.2.6.2 Case 2: health care

An important case study involves the health care markets, where price competition tends to play an important role (see Gaynor and Vogt 2000). As we noted above, it is theoretically unclear whether or not quality increases as a result of competition in such markets.

Overall, empirical research suggests that health care competition under fixed prices generates improvements in the quality of care, while competition with market-determined prices has more mixed effects (see Gaynor and Town 2011). For example, NHS reforms that stimulated the entry of Independent Sector Treatment Centres, as well as patient choice under a fixed price, private regime, improved efficiency in publicly operated hospitals without negatively affecting their clinical quality (Cooper, Gibbons, and Skellern 2016).⁴⁶ On the other hand, American research indicates that competition with market-determined prices and independent providers can be both positive and negative for quality (e.g. Gowrisankaran and Town 2003; Rogowski, Jain, and Escarce 2007; Volpp et al. 2003).⁴⁷ Interestingly, however, research indicates that market entry in oligopolistic

45 Sometimes, as in Chile and Australia, top-up fees have been allowed, which have ensured that prices can adjust upward to increased demand – but they cannot adjust downward farther than the price floor determined by the voucher amount.

46 Efficiency is measured by pre-surgery length of stay for hip and knee replacements, which is less influenced by patient characteristics compared with length of stay until discharge. This is important since the authors also find that the new entrants took on healthier patients as well, thus leaving patients who were more ill for the publicly-operated hospitals.

47 Examples without private providers also include the internal NHS market in the early 1990s, where

markets with price competition improves quality (Cutler, Huckmand, and Kolstad 2010), suggesting it is important to minimise entry barriers to the market. Indeed, overall, the research indicates that design of the system is important for whether competition improves or decreases quality.

It is likely that the mixed findings in markets where providers also compete on price are due to differences in how sensitive users are to differences in quality vis-à-vis price, which, as argued above, in turn is likely to depend on the extent to which quality and prices are measured and observed. An important lesson for other welfare markets with market-determined prices is that desirable quality indicators should be made as easily available as possible.

Overall, evidence suggests that choice and competition can work in welfare services that appear more complex than the provision of qualifications and assessment, both in services of high relevance to such provision (such as schooling) and even when prices are not fixed and choice is indirect (as in health care). While design certainly matters, we believe the evidence base in other welfare services lends further support to the assessment of the current market in Section 4.2.2.5: that choice and competition in qualifications and assessment may function reasonably well, provided there is the right amount of regulation.

4.2.2.7 Spreading the risk

Apart from the straightforward economic theory and evidence discussed above, we also note that choice between multiple providers spreads the risk in the system in ways that neither a government monopoly nor the suggested franchising model could achieve. Indeed, the latter would be associated with considerable risks of system failure, since it would put the delivery of all examinations – whether by subject or (temporarily) by process/function – in the hands of one

purchasers bought health services from hospitals and elective care on behalf of patients, whom had little choice themselves. Research indicates this market lowered costs and waiting times (outcomes that were published), while the impact on unobservable quality, measured by death rates after heart attacks, deteriorated (Propper, Burgess, and Gossage 2008). On the other hand, research analysing the effects of the choice and competition reforms to the NHS under New Labour – where price was fixed – indicates that hospital competition generated higher quality, again measured by death rates after heart attacks, without raising costs (Cooper et al. 2011; Gaynor, Moreno-Serra, and Propper 2013). However, apart from price competition, we note that the lack of patient choice under the NHS internal market in the early 1990s is a key difference compared with the New Labour reforms that also sought to stimulate patient choice of GPs.

board. If problems were to arise in the process, they could potentially have considerable negative consequences (e.g. Croft and Howes 2012; Husbands 2015). Allowing schools to choose an exam board does not mean there will never be any failures, but it does mean (1) that any such failures are unlikely to affect more than a certain share of pupils, and (2) that properly functioning exam boards can quickly take over.

One example of how a single board could induce more risk is the case of what happened with the Scottish Qualifications Agency (SQA), the government-run monopoly board in Scotland, in 2000. In this instance, administrative and computer errors led to inaccurate results being sent out to schools and pupils. The SQA was forced to re-check about 150,000 exams, and process 120,000 appeals. The whole process took months to resolve (see Britten 2000; Scott 2000). Another example is what happened when exam papers in the French Baccalaureate, administered and carried out by a French government agency, were leaked (Bamat 2011). In both these instances, having qualifications and assessment provision in the hands of one organisation magnified the problems so that the entire system was affected, rather than just part of it.

Similarly, in England, the government body responsible for the national curriculum assessment tests up to KS4 has recently in effect contracted out different aspects of the examination process to different operators. As in government monopolies, there have been problems with this system in the past. For example, in 2008, the non-profit American educational organisation ETS was responsible for marking – but failed to deliver due to problems with computer systems and administration (see BBC 2008). In fact, problems occurred this year as well, when tests and answers were leaked in advance (Adams and Weale 2016). Again, these problems could occur in any system, but they become magnified when they occur in systems where all provision is concentrated in one organisation.

In a system with multiple examination boards, risk is spread across different organisations, which means that any problems are more likely to be manageable. For example, in 2011, mistakes in examination papers were found in different subjects in different boards (Richardson 2011). Problems affecting relatively few pupils tend to be more manageable than problems affecting many pupils. In fact, as noted, competition between boards is likely to have driven some

innovations in this respect (see Husbands 2015). The incentives to do so are considerable given that mistakes are frequently spotted and exposed.

Certainly, system risk could technically also be reduced via government procurement in which the same subject is outsourced to several different providers, while at the same time abolishing schools' choices between them. Schools could then be randomly allocated to different providers. However, this is fundamentally different from the franchising model under consideration. We also note that the arbitrariness of randomly allocating schools and pupils to different products is unlikely to be attractive to either schools or boards. And, of course, it would also mean that the potential advantages of choice would still disappear. Nevertheless, we do acknowledge that it would theoretically be possible to both abolish choice and spread the risk in the system.

4.2.2.8 Verdict: user choice is preferable – but reform is still desirable

Overall, therefore, we believe there is a stronger case for allowing user choice and ex post competition in the qualifications and assessment system, rather than just ex ante competition and no choice in a procurement framework. The current system essentially allows for no competition on standards, which has been the principal case against it. While there is little evidence to suggest that competition could raise standards in national qualifications within the confines of existing regulation, there are indications that choice and competition among alternative qualifications at least to some extent has spurred higher quality. More rigorous research from other relevant services does indicate that choice and competition can improve quality, although the effects appear to partly hinge on system design. We also believe the fact that a system of multiple exam boards considerably decreases the risk of system failure is an important reason why it should not be replaced with a franchising model that concentrates the supply of qualifications and assessment in each subject.

Nevertheless, this does not mean that the current system functions well. Indeed, we have highlighted considerable obstacles to the market producing higher quality overall. System design is crucial for the effects of any quasi-market; in the next section, we outline a reform agenda in this respect.

5 SUGGESTIONS FOR A BETTER-PERFORMING SYSTEM

SINCE WE BELIEVE THAT choice and competition are preferable to a franchising model based on procurement without user choice, this section provides a reform agenda for the former option. Regardless of the system, it is crucial that it is designed properly to maximise its potential – and that policymakers allow any reforms to first be trialled in a rigorous framework before they are scaled up at the national level. Here, we provide general principles on which more detailed reform proposals could be built and later rigorously tested prior to national implementation.

5.1 Focus on minimum standards rather than strict equivalency

The principal hurdle preventing choice and competition from functioning properly is that the equivalency framework – within and between qualifications – makes it difficult for exam boards to differentiate themselves on quality in terms of standards. This is especially true in the light of the fact that schools are primarily judged by their league table performance, not by statistics relating to where pupils continue further study or enter employment.

In order to allow for innovations and spur stronger competition on quality, we therefore believe the equivalency framework should be revised. Rather than seeking to ensure precise equivalence, the regulator should ensure that all qualifications, and different versions of the same qualifications, meet specified minimum standards, without underwriting that they are equal in every respect. Ofqual would still approve all qualifications that could be offered, but would no longer underwrite their equivalence. Rather, it would be restricted to ensuring that all meet the stipulated requirements.

This would enable boards to offer specifications and alternative qualifications of higher, but not lower, standards than the specified minimum, and would be similar to how school voucher programmes normally work (see Heller Sahlgren 2013). It would also allow stronger specialisation, which could enable better matching between pupils and different qualifications and assessment – and allow exam boards to brand themselves on standards and quality provided, within and between different qualifications, to both schools and end-users. End-users, in turn, would be in a better position and have stronger incentives to judge the differential quality of qualifications and assessments, as is already the case with international qualifications, as highlighted in Section 4.2.2.

Overall, changing Ofqual's remit so that it inspects minimum standards rather than seeking to enforce strict equivalency is important to reform of the incentive framework in the current system. Allowing market actors to judge whether or not different qualifications and assessments are equal would be likely to lead to more active product differentiation based on quality, as demonstrated in other markets characterised by oligopolistic competition.

5.2 Use an empirical output approach to assess comparability between qualifications and specifications

There are two different ways for end-users to judge the quality of qualifications and assessment. First, one may, as is currently the case, use an input-based model to examine what pupils have to know and accomplish to achieve different grades in different qualifications and specifications. Second, one may empirically study post-hoc outcomes at HEIs or employers among students with different qualifications and grades.

In an interesting illustration, Green and Vignoles (2012) use the latter method to determine how A levels and the IB programme should be judged in university admissions. By comparing pupils taking these different qualifications – holding constant university attended, subject studied, and background variables – they determine the equivalence based on degree performance. They find that the official UCAS tariff score vis-à-vis the one for A levels underestimates pupil performance at the top end of the IB grade distribution, but overestimates pupil performance at the lower end of the IB grade distribution. Interestingly, they

also find that universities' actual pattern of recruitment follows the empirical equivalence scale to a higher extent than the official UCAS tariff – suggesting that universities act rationally and appear more (but not perfectly) informed about the relative predictive power of A levels and the IB programme for future degree performance than the government agency.

We believe similar 'empirical equivalence scales' could be used for most qualifications and specifications. By statistically establishing what grades indicate comparable achievement based on future performance – either in the education system or, for vocational qualifications, in the labour market – it would be possible to determine more rigorously how different qualifications and specifications compare for future performance purposes.

One purpose of Ofqual could thus be to establish such scales of empirical equivalency in place of the ad hoc input-based model, which fails to capture important differences between different qualifications. This could then be offered as a general guide – on a voluntary basis – for HEIs and employers when trying to judge the comparability of different candidates.

5.3 Consider introducing a national cohort-referenced comparability test

As a further tool for judging the relative performance of different candidates taking different qualifications and specifications, it could also be beneficial to introduce a general competence test at the end of secondary school. Such a test would only be used for the purposes of allowing schools and end-users to compare different candidates taking different qualifications. Possibly, the test could be taken by representative samples of pupils taking different qualifications and specifications, rather than entire cohorts, thereby decreasing the costs of administering it.

The idea would be that the test would focus on general competencies rather than subject knowledge. It would therefore by no means be a substitute for regular qualifications, which provide more in-depth assessment of knowledge across a vast number of subjects. It would also not be included in any calculations of league tables, apart from possibly being used to calculate comparability between different qualifications and specifications.

The lack of visibility would itself make it less of a high-stakes test than regular qualifications, although the fact that the national distribution could be used to determine the calibration of different qualifications and specifications means that it is not entirely free of incentives. Since incentives are important for motivating pupils to perform, we believe this is a good thing. Yet, crucially, schools could not by themselves produce a more favourable position for their pupils in this distribution, since the calibration would depend on the average performance of all pupils sitting the same qualification or specification. Furthermore, since the test would be cohort referenced, it would be possible to abolish any remaining perverse incentives among schools to teach to the test, if it is designed properly (see Barlevy and Neal 2012; Neal 2010).

Since the test would be cohort referenced, results would only tell schools and end-users how pupils taking different qualifications perform in relation to each other on average. But this would be the main purpose of the exercise: by comparing pupils in the national distribution of this test, it would be possible to benchmark and calibrate grades from different qualifications and specifications. For example, if pupils receiving an A* grade in one A-level specification on average perform lower than pupils with a B grade in another specification on the cohort-referenced test, it would provide end-users and schools with information regarding how the different grades compare across the different specifications.

Overall, therefore, the empirical comparability approach combined with a national norm-referenced competency test would allow schools and end-users to better judge the standards of different qualifications and specifications. In this sense, it would produce better information and incentives to improve standards in the system than are presently in evidence.

5.4 Change league tables in line with the new remit

These alterations to Ofqual's remit suggest a number of changes to league tables as currently constructed.⁴⁸ We believe that league tables do fulfil a useful function,

48 Currently, the DfE headline metrics for secondary school scores are based on absolute pass rates in (1) GCSEs and equivalents, (2) the English Baccalaureate (EBacc), and (3) English and mathematics separately. They also include (4) progress measures since KS2 in English and mathematics. In 2017, the headline metrics will be revised to include the new Progress 8 measure, which is a value-added measure that includes five EBacc qualifications and three optional GCSEs or equivalents. For post-16 institutions, performance is instead

as research indicates they both improve efficiency and equity in the English education system (Burgess, Wilson, and Worth 2013). Merely abolishing them outright would therefore not be a good option.

A better alternative would be to publish results separately for different types of qualifications and specifications. These scores would be based on the type of qualifications rather than pupils. Rather than mixing pupils taking different types of qualifications, school performance would be judged by its average performance in each qualification type.

This means that average grades and any value-added metrics could still be calculated based on the performance of all pupils taking a certain type of approved qualification or specification by a specific exam board. In other words, instead of reporting average pupil performance across different types of qualifications – by forcing different qualifications and specifications into one score – league tables would instead convey average school performance in different qualifications.

Such a move would considerably decrease any incentives for schools to choose easier qualifications – such as some vocational options – that historically have been merged with more difficult ones in league tables. It would in all likelihood also have the positive side-effect of forcing both schools and end-users to consider the quality of different types of qualifications – while introducing stronger incentives among exam boards to highlight why their products and services should be preferred over others.

Of course, this proposal is not without its drawbacks, chief among which is its relative complexity, potentially making it more difficult for schools and end-users to discern differences in standards.

Fortunately, there are other alternatives. One option would be to retain the simpler framework, but instead use the empirical equivalence framework and/or the national cohort-referenced test for calibration purposes. This would ensure that different qualifications could be merged to generate a less complicated scoring framework.

reported as (1) 'A level performance', (2) total 'academic performance' – which includes performance in A levels, the IB programme, Cambridge Pre-U, and AQA Baccalaureate – and (3) 'vocational performance'. These metrics are based on pupils rather than qualification. For example, the overall performance of pupils is included in the performance category as long as they took one qualification in that category.

Another alternative would be to use the first approach of publishing results separately for different types of qualifications and specifications together with the results of the calibration exercise, providing important information to actors regarding how different qualifications and specifications compare. This would retain the more complicated structure of league tables, but also provide the key to understanding them. There may be other alternatives.

Overall, therefore, while league tables should be retained, they should be brought into line with Ofqual's new remit, as outlined above. On either model outlined: (1) using a decentralised framework, where performance measures are calculated separately for different qualifications and specifications, or (2) using an empirical, output-based method to determine equivalency, we believe the overall qualifications and assessment market would function better.

5.5 Provide information for schools and universities on the differences between qualifications and specifications

In the reformed system, we believe an important aspect of Ofqual's remit should be to provide reliable information to schools and end-users. It could therefore undertake research – or outsource such research to academic institutions – that establishes the relative value of different qualifications for different purposes. The institution could then disseminate this information to schools, HEIs and employers.

To take into account schools' subjective experiences of the level of quality provided in different aspects of qualifications and assessment, it may also be useful to carry out user satisfaction surveys among randomly selected schools using products provided by the boards. Ofqual could then publish the results from such surveys as well.

By giving better information to schools and end-users of qualifications, Ofqual would be able to decrease potential information asymmetries that could benefit producers over users. Including this within Ofqual's remit, alongside the other reform measures suggested, is therefore likely to generate a better-performing qualifications and assessment market.

5.6 Maximise the possibility of market entry

As discussed in Section 4.2.2, maintaining free entry to the qualifications and assessment market is important to maximising innovation and keeping providers sharp to the possibility of fresh competition. New firms are often disruptive innovators. They may even be so disruptive as to change the rules of the game on the market, forcing existing providers to alter their behaviour to keep up. It is therefore important to reduce regulatory entry barriers and retain only those that are necessary to ensure that all qualifications meet the stipulated minimum requirements.

Currently, the regulatory burden appears quite extensive, which acts to deter new entrants. Ofqual requires that new boards prove they can develop and deliver high-quality qualifications with a robust assessment and awarding framework; that their operations are scalable; and that their finances are secure. For boards entering the market for the first time, without track record, it is difficult to show they meet these criteria (see Bassett 2014). While some regulation is inevitable, reforming the equivalency framework to focus on minimum requirements rather than strict equivalency, as discussed in Section 5.1, should enable some of these regulatory entry barriers to be lowered. This would stimulate competitive threats among existing exam boards.

Here, we refrain from getting into the specific details of such deregulation, but rather suggest that the government invites suggestions from universities, think tanks, and other relevant groups as to how the current regulatory barriers to entry could be reduced without generating excessive risks in the system.

While some regulatory barriers to entry are necessary, it is crucial that these are kept to a minimum. This is especially important given that natural entry barriers appear to be quite high (Frontier Economics 2015).⁴⁹ Certainly, innovation always carries some risks, but it increases in importance in times of rapid economic and social transformations. Maximising the possibility of

⁴⁹ For example, there are non-negligible fixed sunk costs, such as the costs of developing specifications and marketing, which require larger volumes for exam boards to recover. This clearly favours already large producers. Second, while costs of marking are variable, they are also in some measure inflexible because of the need to maintain rosters of examiners well in advance of assessment. Other problems may include high reputational entry costs: any mistakes are scrutinised in detail by the media and government, while schools may be too risk averse to choose a new provider (see Frontier Economics 2015).

market entry, and therefore the sense of competitive threat, is therefore likely to improve the functioning of the qualifications and assessment system.

5.7 Monitor potential problems with price competition – and consider some regulation as a measure of last resort

As discussed in Section 4.2.2, economic theory and evidence from markets where prices are not fixed provide an ambiguous picture regarding the effects of competition on quality. It could increase but may also decrease, depending on how sensitive schools are to price relative to quality. Existing evidence indicates that schools are not price sensitive, indicating there is little competition on this aspect of service delivery in the qualifications and assessment market at the present time.

Yet the existing evidence is far from comprehensive, and the picture may in any case change following the reforms suggested here. It would therefore be valuable to carry out research analysing the extent to which price competition is a feature of the present market – and the impact it may have on the quality of qualifications.

The price mechanism normally fulfils a crucial information function in regular markets (Hayek 1945), which should be mimicked as far as possible on quasi-markets. We therefore believe there is a good case to retain the possibility of competition on price in a reformed qualification and assessment market. If it works well, we may both save money and obtain better quality in the long run.

However, in education, many are also likely to believe that quality should be prioritised over price. If the system appears to produce more price competition than quality competition, it may therefore be important to tweak it. One way of doing so would be to designate a minimum allocation of per-pupil funding to qualifications and assessment, while allowing schools to spend more if they wish. This would essentially produce a voucher system with top-up fees, ensuring that exam boards could always charge more for higher standards – but never lower than the determined minimum level.

Again, it is important to emphasise that there is little to suggest that excessive price competition pushes quality downwards in the current system. But given the importance attached to quality in welfare services, including qualifications

and assessment, it would be worth monitoring the system closely in this respect – and considering safeguards against the possibility that the reforms proposed might usher in excessive price competition at the expense of quality.

5.8 Trial the system rigorously

We believe the reforms considered here would improve the functioning of the qualifications and assessment market. However, there are always unintended consequences involved in policymaking. If possible, therefore, the reforms should be trialled first to ensure that any such unintended consequences are identified and addressed prior to national implementation. As Oates (2014) points out, English education policies have too often lacked proper trial phases, which allow evaluation of what has and hasn't worked before policy designs are finalised. This is in contrast to, for example, Singapore, which tends to test most policies prior to national implementation.

There are various ways to conduct experimental trials. For example, the government could randomly choose one region where the new regulatory framework would be tried across all schools for a certain number of years. Another alternative would be to apply it to randomly chosen schools across the country. There are other possible options to consider.

In general, we support proper evidence-based policy and therefore argue in favour of larger-scale trials of the reforms suggested before scaling them up to national policy. There are different ways such trials could be constructed, and presenting exact details is beyond the scope of this monograph. Future research and policy advice should focus more closely on the question of trial design specifically.

6 IF THE GOVERNMENT STILL OPTS FOR FRANCHISING...

IN THE EVENT THAT the government decides to move towards a franchise system without user choice in spite of our recommendation, we also offer here some principles around which we believe a successful franchising model should be built. There is a range of different ways in which a procurement system could be designed, and it is crucial the government gets it right. Here, we discuss some of the issues that are likely to be important.

6.1 Allow a measure of discretion in the procurement process

In a franchising model without user choice, users would no longer be able to hold providers to account in the market. This would make it even more important to ensure that the ex ante competitive process for contracts ensures that suppliers are incentivised to raise quality. As noted in Section 2, it is preferable to allow a certain amount of discretion in the procurement process so that suppliers have strong reputational incentives to ensure that they do not cut corners on whatever aspects of quality may not be explicitly specified in contracts. While discretion is often feared to induce corruption, it is often advantageous in procurement for complex services because it ensures that the ‘shadow of the future’ constantly hangs over producers.

We therefore believe that allowing a measure of discretion to assess the performance of private providers on non-contractible quality elements would be sensible. Those that have underperformed in the past may be suspended, penalised in other ways, or even barred entirely, thus giving them strong incentives not to squander their reputation by decreasing quality. In this way, it would be possible to ensure that desired levels of quality would be delivered.

Discretion can also be exercised more generally to select the highest-quality suppliers in the first place. For example, the government could ask all competing suppliers to provide it with detailed plans for future provision and a discussion of how and why it would stimulate quality in the sector. This gives buyers considerable opportunities to take into account quality that may be more difficult to measure. While the government could inspect these plans itself, it may instead allow an external expert panel to judge the different bids, thereby decreasing discretion slightly and therefore also the risk of corruption (see Bergman 2013). By allowing a measure of discretion in the procurement process, it may be easier to select a high-quality supplier in the first place.

Historically, EU competition law has generally put limits on the ability of contracting authorities to make use of suppliers' past performance when rewarding future contracts. However, education has been considered 'light touch' for regulatory purposes, which means that member states have relatively wide discretion about how to formulate criteria and rules for awards (CSS 2015a). In addition, recent changes to the general framework mean that: 'Poor performance under previous contracts is explicitly permitted as a grounds for exclusion [of suppliers from the selection stage]' (CSS 2015b, p. 4). Of course, given the result of the Brexit referendum in June 2016, it is not clear whether or not these regulations will apply in future in the UK context. Regardless, if the government decides to move towards a procurement system without user choice, we believe it is important to give allow a measure of discretion in the tendering process.

6.2 Use a model that values quality at least as highly as price

A key issue in the procurement process is how much the government should value lower prices relative to higher quality. The goal is to optimise the level of quality in relation to cost. Offers with different quality and price have to be compared. The buyer could use the best price–quality ratio or simply the lowest price offered by competing suppliers to determine who should be rewarded a contract.

In this case, it is clearly important that the government does not solely focus on price. Qualifications and assessment are complex services of crucial importance to the functioning of the education system, where stimulating higher standards and quality is key. Focusing solely on price would induce bidders to compete solely on

this aspect rather than on quality. The result could be excessive price competition and lower standards and overall quality than would otherwise be the case.

Using the best price–quality ratio would be a better option since it allows bids to be equally weighed on the basis of quality and price. However, quality may be quantified or assessed with considerably more noise than price, especially when using expert panels as suggested in Section 6.1. If this is the case, there may be stronger incentives to invest in winning contracts by lowering price rather than raising quality.

Since the aim behind franchising would be to secure standards and overall quality, the government may instead want to explicitly prioritise quality over price in the tendering process. The most straightforward way to do so would be to set the price in advance and then evaluate bids solely on the basis of quality. Such ‘fixed-price auctions’ prevent price competition altogether and therefore sharpen incentives among suppliers to develop high-quality bids. Of course, since quality is multidimensional, it would also be necessary to decide on how different aspects of quality should be valued in the process.

Alternatively, it may be possible to use a price–quality ratio model in which quality is weighted more heavily than price. This would ensure that suppliers have incentives to compete with high-quality bids, while at the same time ensuring that price is also considered to some extent.

To compare the price–quality ratios for different bids, it is useful to either convert price into quality points or different quality aspects into a specific price. It is preferable to convert quality into price. This is because approaches attempting to convert price into abstract quality scales often entail that different aspects of quality are not optimised in relation to each other or to the price offered. Converting quality into price is more transparent since it means that both quality and price are measured on a common and easily interpreted scale (Bergman 2013).

6.3 Contract out by subject rather than by function

There are different ways to contract out qualifications and assessment. Two key alternatives are: (1) contract out the entire operation for a subject; or (2) contract out across different functions (such as marking).

The advantages of contracting out by subject are that it ensures a more joined-up process and incentives to ensure that the entire operation works well. The risk

with the latter model is that the different functions would not link up well when handled by different suppliers, who would also be competitor. The experience in national curriculum tests in the years prior to KS4 does not suggest this type of procurement would work well partly for these reasons (see Baird 2015; Baird, Elwood, and Isaacs 2012). We thus believe there is a stronger case for contracting out the entire operation of qualifications and assessment by subject rather than by function.

6.4 Consider allowing more than one provider per subject to spread the risk

As discussed in Section 4, an important disadvantage with a franchise model that abolishes choice is that it decreases the ability to spread the risk in the system. For this reason, it may be useful to allow more than one exam board to supply qualifications and assessment in the same subject, provided that several boards' are deemed to fulfil the requirements stipulated by the government. Schools could then be assigned to different exam boards randomly. While such a system is likely to be perceived arbitrary to schools – which would be assigned different products without having chosen them – it would be useful to help decrease the risk of system failure. In addition, with different exam boards providing the same qualification in the same subject, it would provide even stronger incentives to provide high standards and overall quality to maximise the possibility of winning future contracts in the same subject. In other words, it would increase the possibility that the 'shadow of the future' works to sharpen incentives for the better.

6.5 Measure schools' satisfaction with exam boards

In a user choice model, schools judge exam boards by purchasing or rejecting their products and services. Since user choice would be abolished in a franchising model, it would be useful to carry out customer satisfaction surveys from randomly selected schools that have been assigned to different exam boards. The idea would be to measure subjective interpretation of the quality provided in different areas by different boards. The surveys could then also be considered in future tendering processes.

6.6 Trial the system rigorously

As in the case of our reform proposals for the existing user-choice model, we would urge the government to trial the system rigorously prior to implementation. This should be done irrespective of the design specifics of any new procurement system; there are always unintended consequences involved in policymaking. Regardless of what a potential new procurement regime would look like, we believe it is important to trial the arrangements to ensure that any unintended consequences are identified and addressed prior to subjecting the entire qualifications and assessment system to such a radical reform.

7 CONCLUSION

FOR DECADES, SCHOOLS IN England, Wales, and Northern Ireland have had the right to decide which qualifications their pupils take from a range of options offered by multiple independent providers. This system is essentially unique; no other country allows choice and competition to the same extent in this particular area of public policy.

However, in the past couple of years, subjecting qualifications and assessment to market forces has become increasingly contentious. Critics argue that competition between boards gives rise to perverse incentives to lower standards and inflate grades. Because of these concerns, the pressure to abolish the market in favour of either a single government exam board or a franchising model has increased.

Drawing on economic theory and evidence, this monograph has evaluated which of these models – monopoly, franchising, or user choice – is likely to be best for the provision of qualifications and assessment, and sketched a reform agenda for ensuring that that model functions optimally.

We noted that government monopolies could be justified in service provision with considerable opportunities for producers to engage in cost cutting at the expense of non-contractible quality. However, compared with many other complex welfare services, where the evidence shows private provision has often been successful, there are relatively few non-contractible quality elements of qualifications and assessment. Since innovations are likely to be increasingly important, and because exam board reputation functions as an important accountability mechanism, there is little to suggest that full monopolisation would be a good option.

We then went on to consider whether the independent organisations delivering qualifications and assessment should be chosen through franchising – which effectively

means that the government picks winners in a tendering process – or through a user choice model in which schools have the right to choose between multiple providers.

In the end, we have concluded that the user choice model is a better option. The current regulatory framework does not in fact allow for downward competition on standards, which has historically been the principal case against it. But it also ensures that competition is unable to raise standards in national qualifications. The market for alternative qualifications, however, indicates that quality competition is possible. In addition, research from other welfare services indicates that choice and competition can improve quality in the system, although the effects appear to hinge on system design. We also believe the fact that the current system considerably decreases the risk of system failure is an important reason why it should not be replaced with a franchising model that concentrates the supply of qualifications and assessment in each subject.

However, we have also identified considerable obstacles for choice and competition to act as a lever for higher quality in qualifications and assessment. To improve the system, we have therefore suggested the following reform agenda:

1. Revise the equivalency framework so that the regulator ensures minimum standards rather than strict equivalency.
2. Use a statistical, empirical approach to assess comparability between qualifications and specifications based on future pupil outcomes.
3. Consider introducing a national cohort-referenced competency test as a tool with which to judge the relative performance of candidates taking different qualifications and specifications.
4. Change league tables in line with Ofqual's new remit, either by publishing results separately for different qualifications and specifications, or by using the empirical method and/or the national cohort-referenced test to calibrate the different qualifications and specifications.
5. Provide information for schools and end-users to minimise potential information asymmetries in the market.

6. Maximise the possibility of market entry by decreasing regulatory barriers, which should be possible given reforms to the equivalency framework.
7. Monitor potential problems with price competition, and consider some regulation as a measure of last resort if problems emerge.
8. Trial the system rigorously before national implementation to avoid unintended consequences as far as possible.

Certainly, there may be other worthwhile changes to consider and we would welcome any suggestions in this respect. The monograph's main conclusion is merely that such reforms should seek to improve the market rather than abolishing it outright. The market's current workings, let alone its potential, simply do not warrant such a radical solution.

Nevertheless, if the government still decides to abolish the user-choice model currently in place, we have also presented some principles upon which a franchising system without user choice should be built:

1. Allow a certain amount of discretion in the procurement process so that suppliers have strong reputational incentives to ensure that they do not cut corners on any non-contractible quality measures.
2. Use a model that values quality at least as highly as price in the procurement process to minimise the risk of excessive price competition.
3. Contract out qualifications and assessment by subject rather than by function to ensure healthy incentives in the system.
4. Consider allowing more than one provider per subject to spread the risk of failure in the system.
5. Measure schools' satisfaction with exam boards as a gauge of subjective quality in the absence of user choice.
6. Trial the system rigorously before national implementation to avoid unintended consequences as far as possible.

There is a range of different ways in which a franchising system could be designed, and it is crucial the government gets it right. While we first and foremost favour reforms to improve the existing choice-based system, following these principles may at least improve the chances of success for a procurement regime in which such choice has ultimately been discarded.

REFERENCES

- Abdulkadiroğlu, Atila, Joshua D. Angrist, Peter D. Hull, and Parag A. Pathak. 2016. "Charters Without Lotteries: Testing Takeovers in New Orleans and Boston." *American Economic Review* 106(7):1878-1920.
- Adams, Richard. 2014. "Schools complain over IGCSEs U-turn." *The Guardian*, 1st July: <https://www.theguardian.com/education/2014/jul/01/schools-igcse-league-tables>.
- Adams, Richard. 2016. "Schools under scrutiny in crackdown on league table 'gaming'." *The Guardian*, 8th April: <https://www.theguardian.com/education/2016/apr/08/gcse-league-table-gaming-schools-crackdown>.
- Adams, Richard and Sally Weale. 2016. "Fresh humiliation as Sats answers published online for second time." *The Guardian*, 10th May: <https://www.theguardian.com/education/2016/may/10/fresh-humiliation-as-sats-answers-published-online-for-second-time>.
- AERA, APA, and NCME. 1999. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association. American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME).
- Aghion, Philippe and Ufuk Akcigit 2016. "Innovation and Growth: The Schumpeterian Perspective" in *Economics Without Borders: Economic Research for European Policy Challenges*, edited by Richard Blundell, Estelle Cantillon, Barbara Chizzolini, Marc Ivaldi, Wolfgang Leininger, Ramon Marimon, Laszlo Matyas, and Frode Steen. Cambridge: Cambridge University Press.
- Aghion, Philippe, Ufuk Akcigit, and Peter Howitt 2014. "What Do We Learn From Schumpeterian Growth Theory?" in *Handbook of Economic Growth*. Amsterdam: North-Holland, pp. 515–563.
- Andersson, Fredrik. 2009. "A Trickle-Down Theory of Incentives with Applications to Privatization and Outsourcing." IFN Working Paper No. 784, Research Institute of Industrial Economics, Stockholm.
- Andersson, Fredrik and Henrik Jordahl. 2011. "Outsourcing Public Services: Ownership, Competition, Quality and Contracting." IFN Working Paper No. 874, Research

Institute of Industrial Economics, Stockholm.

- Baird, Jo-Anne. 2015. "GCSE results remain stable – so why are ministers concerned exams are dumbed down?" *The Conversation*, 20th August: <https://theconversation.com/gcse-results-remain-stable-so-why-are-ministers-concerned-exams-are-dumbed-down-46288>.
- Baird, Jo-Anne, Jannette Elwood, and Tina Isaacs. 2012. "Written evidence." Written evidence submitted to the Education Select Committee Inquiry on 'The administration of examination for 15-19 year olds in England'. Education Committee, House of Commons, London.
- Bamat, Joseph. 2011. "Exam cheats send education system into turmoil." *France 24*, June 25: <http://www.france24.com/en/20110623-france-hunt-math-examination-cheating-education-baccalaureate-chatel>.
- Barlevy, Gadi and Derek Neal. 2012. "Pay for Percentile." *American Economic Review* 102(5):1805-1831.
- Bassett, Dale 2014. "Regulatory overkill: school accountability, qualifications, and the future" in *Tests worth teaching to: incentivising quality in qualifications and accountability*, edited by Gabriel Heller Sahlgren. London: CMRE, pp. 46-57.
- Batdorff, Meagan, Larry Maloney, Jay F. May, Sheree T. Speakman, Patrick J. Wolf, and Albert Cheng. 2014. "Charter School Funding: Inequity Expands." Report, Department of Education Reform, University of Arkansas, Fayetteville, AR.
- Baumol, William J. 2002. *The Free-Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*. Princeton, NJ: Princeton University Press.
- BBC. 2008. "Sats marking contract is scrapped." *BBC*, 15th August: <http://news.bbc.co.uk/1/hi/education/7562835.stm>.
- Bergman, Mats. 2013. *Upphandling och kundval av välfärdstjänster – en teoribakgrund*. Stockholm: Fores.
- Bergman, Mats A., Per Johansson, Sofia Lundberg, and Giancarlo Spagnolo. 2016. "Privatization and quality: Evidence from elderly care in Sweden." *Journal of Health Economics* 49:109-119.
- Bergman, Mats and Henrik Jordahl. 2014. "Goda år på ålderns höst? En ESO-rapport om konkurrens i äldreomsorgen." Rapport 2014:1, Expertgruppen för studier i offentlig ekonomi, Finansdepartementet, Stockholm.
- Böhlmark, Anders and Mikael Lindahl. 2015. "Independent Schools and Long-run Educational Outcomes: Evidence from Sweden's Large-scale Voucher Reform." *Economica* 82(327):508–551.
- Bramley, Tom and Carmen Vidal Rodeiro. 2014. "Using statistical equating for standard

- maintaining in GCSEs and A levels.” Cambridge Assessment Research Report, Cambridge Assessment, Cambridge.
- Bramley, Tom, Carmen Vidal Rodeiro, and Sylvia Vitello. 2015. “Gender differences in GCSE.” Cambridge Assessment Research Report, Cambridge.
- Bratti, Massimiliano and Roberto Leombruni. 2014. “Local human capital externalities and wages at the firm level: Evidence from Italian manufacturing.” *Economics of Education Review* 41:161–175.
- Britten, Nick. 2000. “SQA board face sack after exams fiasco.” *The Telegraph*, 5th October: <http://www.telegraph.co.uk/news/uknews/1368984/SQA-board-face-sack-after-exams-fiasco.html>.
- Burgess, Simon, Deborah Wilson, and Jack Worth. 2013. “A natural experiment in school accountability: The impact of school performance information on pupil progress.” *Journal of Public Economics* 106:57-67.
- Butler, Jeffrey V., Enrica Carbone, Pierluigi Conzo, and Giancarlo Spagnolo. 2013. “Reputation and Entry.” Working Paper 2013:3, Konkurrensverket, Stockholm.
- Cambridge Assessment. 2010. “Exam standards: the big debate.” Report and recommendations, Cambridge.
- Cambridge Assessment. 2012. “Written evidence submitted by Cambridge Assessment.” Written evidence submitted to the Education Select Committee Inquiry on ‘The administration of examination for 15-19 year olds in England’. Education Committee, House of Commons, London.
- Carlton, Dennis W. and Jeffrey M. Perloff. 2003. *Modern Industrial Organization*. 4th edn. Boston: Pearson.
- Chingos, Matthew and Paul E. Peterson. 2009. “Impact of For-Profit and Nonprofit Management on Student Achievement: The Philadelphia Intervention, 2002-2008.” PEPG 09-02, Program on Education Policy and Governance, Harvard University, Cambridge, MA.
- Chowdry, Haroon and Luke Sibieta. 2011. “School funding reform: an empirical analysis of options for a national funding formula.” IFS Briefing Note BN123, London.
- CIE. 2011. “Cambridge Pre-U: Guide for framing offers.” Guide, University of Cambridge International Examinations, Cambridge. <http://www.cie.org.uk/images/121798-cambridge-pre-u-guide-for-framing-offers.pdf>.
- Coe, Robert and Gabriel Heller Sahlgren. 2014. “Incentives and ignorance in qualifications, assessment, and accountability” in *Tests worth teaching to: incentivising quality in qualifications and accountability*, edited by Gabriel Heller Sahlgren. London: CMRE, pp. 74-99.

- Cohodes, Sarah. 2016. "Teaching to the Student: Charter School Effectiveness in Spite of Perverse Incentives." *Education Finance and Policy* 11(1):1-42.
- Comondore, Vikram R., P J. Devereaux, Qui Zhou, Samuel B. Stone, Jason W. Busse, Nikila C. Ravindran, Karen E. Burns, Ted Haines, Bernadette Stringer, Deborah J. Cook, Stephen D. Walter, Terrence Sullivan, Otavio Berwanger, Mohit Bhandari, Sarfaraz Banglawala, John N. Lavis, Brad Petrisor, Holger Schünemann, Katie Walsh, Neera Bhatnagar, and Gordon H. Guyatt. 2009. "Quality of care in for-profit and not-for-profit nursing homes: systematic review and meta-analysis." *BMJ: British Medical Journal* 339(7717):381-384.
- Cooper, Zack, Stephen Gibbons, Simon Jones, and Alistair McGuire. 2011. "Does Hospital Competition Save Lives? Evidence From The English NHS Patient Choice Reforms." *Economic Journal* 121(554):F228-F260.
- Cooper, Zack, Stephen Gibbons, and Matthew Skellern. 2016. "Does Competition from Private Surgical Centres Improve Public Hospitals' Performance? Evidence from the English National Health Service." CEP Discussion Paper No. 1434. London: Centre for Economic Performance, London School of Economics.
- Coviello, Decio, Andrea Guglielmo, and Giancarlo Spagnolo. 2016. "The Effect of Discretion on Procurement Performance." CEP Discussion Paper No. 1427. London: Centre for Economic Performance, London School of Economics.
- Croft, James and Anton Howes. 2012. "When qualifications fail: reforming 14-19 assessment." London: CMRE.
- CSS. 2015a. "The public contracts regulations 2015: guidance on the new light touch regime for health, social, educational and certain other service contracts." Report, Crown Commercial Service, London.
- CSS. 2015b. "A brief guide to the EU public contracts directive (2014)." Report, Crown Commercial Service, London.
- Cutler, David M., Robert S. Huckman, and Jonathan T. Kolstad. 2010. "Input Constraints and the Efficiency of Entry: Lessons from Cardiac Surgery." *American Economic Journal: Economic Policy* 2(1):51-76.
- Decarolis, Francesco, Riccardo Pacini, and Giancarlo Spagnolo. Forthcoming. "The Role of Reputation when Awarding Contracts: Empirical Evidence from a Vendor Rating System." Unpublished manuscript, Boston University.
- Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2014. "School Choice, School Quality, and Postsecondary Attainment." *American Economic Review* 104(3):991-1013.
- DfE. 2012. "The Evidence Base for Proposed Reform of the Examination System at Key Stage 4." Memorandum submitted by the Department for Education as part of the

- Education Committee's report 'From GCSEs to EBCs: the Government's proposals for reform', London.
- Dixit, Avinash. 1997. "Power of Incentives in Private versus Public Organizations." *American Economic Review* 87(2):378-382.
- Donahue, John D. 1989. *The Privatization Decision: Public Ends, Private Means*. New York: Basic Books.
- Duncan, Emma. 2015. "Exam boards are failing. Let's nationalise them." *The Times*, 9th August: <http://www.thetimes.co.uk/tto/opinion/columnists/article4531186.ece>.
- Education Committee. 2012. "The administration of examinations for 15-19 year olds in England." London.
- Elinder, Mikael and Henrik Jordahl 2013. "Kontrakt, kostnader och kvalitet" in *Välfärdstjänster i privat regi: framväxt och drivkrafter*, edited by Henrik Jordahl. Stockholm: SNS Förlag, pp. 89-120.
- Estrin, Saul. 1991. "Yugoslavia: The Case of Self-Managing Market Socialism." *Journal of Economic Perspectives* 5(4):187-194.
- Eyles, Andrew, Claudia Hupkau, and Stephen Machin. Forthcoming. "School reforms and pupil performance." *Labour Economics*.
- Eyles, Andrew and Stephen Machin. 2015. "The Introduction of Academy Schools to England's Education." CEP Discussion Paper No 1368. London: Centre for Economic Performance, London School of Economics.
- Figlio, David and Cassandra M. D. Hart. 2014. "Competitive Effects of Means-Tested School Vouchers." *American Economic Journal: Applied Economics* 6(1):133-156.
- Figlio, David and Susanna Loeb 2011. "School Accountability" in *Handbook of the Economics of Education*, edited by Eric A Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North-Holland, pp. 383-421.
- Fogelberg, Sara. 2013. "Effects of Competition between Healthcare Providers on Prescription of Antibiotics." IFN Working Paper No. 949, Research Institute of Industrial Economics, Stockholm.
- Friedman, Milton 1962. "The Role of Government in Education" in *Capitalism and Freedom*. Chicago, IL: University of Chicago Press, pp. 85-107.
- Frontier Economics. 2015. "Understanding awarding organisations' commercial behaviour before and after the GCSE and A-level reforms." Ofqual/15/5596, Ofqual, London.
- Fryer, Roland G. 2014. "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments." *Quarterly Journal of Economics* 129(3):1355-1407.
- Gaynor, Martin. 2006. "What Do We Know About Competition and Quality in Health

- Care Markets?" NBER Working Paper No. 12301, National Bureau of Economic Research, Cambridge, MA.
- Gaynor, Martin, Rodrigo Moreno-Serra, and Carol Propper. 2013. "Death by Market Power: Reform, Competition, and Patient Outcomes in the National Health Service." *American Economic Journal: Economic Policy* 5(4):134-166.
- Gaynor, Martin and Robert J. Town. 2011. "Competition in Health Care Markets." NBER Working Paper No. 17208, National Bureau of Economic Research, Cambridge, MA.
- Gaynor, Martin and William B. Vogt 2000. "Antitrust and competition in health care markets" in *Handbook of Health Economics*, edited by Anthony J. Culyer and Joseph P. Newhouse. Amsterdam: North-Holland, pp. 1405-1487.
- Gilbert, Francis. 2011. "I blame the free market for exam board cheats." *The Guardian*, 9th December: <https://www.theguardian.com/commentisfree/2011/dec/09/free-market-exam-board-cheats>.
- Glaeser, Edward L. and Andrei Shleifer. 2001. "Not-for-profit entrepreneurs." *Journal of Public Economics* 81:99-115.
- Gowrisankaran, Gautam and Robert J. Town. 2003. "Competition, Payers, and Hospital Quality." *Health Services Research* 38(6):1403-1422.
- Green, Francis and Anna Vignoles. 2012. "An empirical method for deriving grade equivalence for university entrance qualifications: an application to A levels and the International Baccalaureate." *Oxford Review of Education* 38(4):473-491.
- Grice, Andrew and Richard Garner. 2013. "Michael Gove forced into humiliating U-turn over EBacc." *The Independent*, 7th February: <http://www.independent.co.uk/news/education/education-news/michael-gove-forced-into-humiliating-u-turn-over-EBacc-8484074.html>.
- Gurney-Read, Josie. 2014. "GCSE reforms: explained." *The Telegraph*, 21st August: <http://www.telegraph.co.uk/education/educationnews/11046160/GCSE-reforms-explained.html>.
- Hart, Oliver. 1975. "On the optimality of equilibrium when the market structure is incomplete." *Journal of Economic Theory* 11(3):418-443.
- Hart, Oliver and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98(6):1119-1158.
- Hart, Oliver, Andrei Shleifer, and Robert W. Vishny. 1997. "The Proper Scope of Government: Theory and Application to Prisons." *Quarterly Journal of Economics* 112(4):1127-1161.

- Hayek, Friedrich A. 1945. "The Use of Knowledge in Society." *American Economic Review* 35(4):519-530.
- Hayek, Friedrich A. 2002. "Competition as a discovery procedure." *Quarterly Journal of Austrian Economics* 5(3):9-23. Translated by Marcellus S. Snow from Hayek's lecture 'Der Wettbewerb als Entdeckungsverfahren' in 1968.
- Heller Sahlgren, Gabriel. 2013. *Incentivising Excellence: School Choice and Education Quality*. London: CMRE and IEA.
- Herrera, Christian A., Gabriel Rada, Lucy Kuhn-Barrientos, and Ximena Barrios. 2014. "Does Ownership Matter? An Overview of Systematic Reviews of the Performance of Private For-Profit, Private Not-For-Profit and Public Healthcare Providers." *PLoS One* 9(12):e93456. doi:10.1371/journal.pone.
- Hoxby, Caroline M. 2003. "School Choice and School Productivity. Could School Choice Be a Tide that Lifts All Boats?" Pp. 287-342 in *The Economics of School Choice*, edited by Caroline M Hoxby. Chicago, IL: University of Chicago Press.
- Hoxby, Caroline M. 2006. "School Choice: Three Essential Elements and Several Policy Options." Keynote speech delivered 30th June 2005 to the New Zealand Association of Economists 46th Annual Conference 2005, Christchurch, New Zealand.
- Hsieh, Chang-Tai and Miguel Urquiola. 2006. "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program." *Journal of Public Economics* 90:1477-1503.
- Husbands, Chris. 2015. "Exam boards: ripe for nationalisation?," Blog post. UCL Institute of Education, 12th August: <https://ioelondonblog.wordpress.com/2015/08/12/exam-boards-ripe-for-nationalisation/>
- Kachour, Maher, Olivier Mamavi, and Haithem Nagati. 2016. "The Role Of Reputation In Market Entry: Evidence From French Public Procurement." *Journal of Applied Business Research* 32(3):805-814.
- Kornai, Janos, Eric Maskin, and Gerard Roland. 2003. "Understanding the Soft Budget Constraint." *Journal of Economic Literature* 41(4):1095-1136.
- Lavy, Victor. 2015. "Long Run Effects of Free School Choice: College Attainment, Employment, Earnings, and Social Outcomes at Adulthood." NBER Working Paper No. 20843, Cambridge, MA.
- Le Grand, Julian. 1991. "Quasi-Markets and Social Policy." *Economic Journal* 101(408):1256-1267.
- Le Grand, Julian. 2007. *The Other Invisible Hand: Delivering Public Services through Choice and Competition*. Princeton, NJ: Princeton University Press.
- Lindqvist, Erik. 2008. "Privatization of Credence Goods: Theory and an Application

- to Residential Youth Care.” IFN Working Paper No. 750, Research Institute of Industrial Economics, Stockholm.
- LSE. 2016a. “Country/regional specific information.” Admissions information, London School of Economics, London. <http://www.lse.ac.uk/study/informationForInternationalStudents/countryRegion/Home.aspx>.
- LSE. 2016b. “Information for applicants studying for other UK qualifications.” Admissions information, London School of Economics, London. <http://www.lse.ac.uk/study/undergraduate/howtoapply/lseentryrequirements.aspx#InformationforapplicantsstudyingforotherUKqualifications>.
- MacLeod, W. B. and Miguel Urquiola. 2012. “Competition and Educational Productivity: Incentives Writ Large.” Discussion Paper No. 7063, Bonn.
- McMahon, Walter W. “The External Benefits of Education” in Brewer, Dominic and Patrick J. McEwan (eds), *The Economics of Education*. Oxford: Elsevier, pp. 68–79.
- Malacova, Eva and John F. Bell. 2006. “Changing boards: investigating the effects of centres changing their specifications for English GCSE.” *The Curriculum Journal* 17(1):27-35.
- Malcomson, James M. 2012. “Relational Incentive Contracts” in *The Handbook of Organizational Economics*, edited by Robert Gibbons and John Roberts. Princeton, NJ: Princeton University Press, pp. 1014-1065.
- Mansell, Warwick. 2016. “Academies: autonomy, accountability, quality and evidence.” Research Report, York.
- McCubbins, Matthew D. and Thomas Schwartz. 1984. “Congressional Oversight Overlooked: Police Patrols versus Fire Alarms.” *American Journal of Political Science* 28(1):165-179.
- Morgan, Nicky. 2015. “Standards in schools matter more than league table positions.” *The Telegraph*, 29th January: <http://www.telegraph.co.uk/education/educationopinion/11377852/Standards-in-schools-matter-more-than-league-table-positions.html>.
- NARIC. 2016. “Benchmarking Cambridge IGCSE in the Context of the UK Education System.” Report, Cheltenham.
- Neal, Derek. 2010. “Aiming for Efficiency Rather than Proficiency.” *Journal of Economic Perspectives* 24(3):119-131.
- Newton, Paul E. 2007. “Clarifying the purposes of educational assessment.” *Assessment in Education: Principles, Policy & Practice* 14(2):149-170.
- Oates, Tim 2014. “The ‘qualifications sledgehammer’: why assessment-led reform has dominated the education landscape” in *Tests worth teaching to: incentivising quality*

- in qualifications and accountability*, edited by Gabriel Heller Sahlgren. London: CMRE, pp. 28-45.
- Ofqual. 2012a. "Ofqual's Response to the Education Select Committee Report (HC141-1)." Report, London.
- Ofqual. 2012b. "Written evidence submitted by Ofqual (Annex E)." Education Committee, House of Commons, London. Written evidence as part of the Education Committee's inquiry 'The administration of examinations for 15-19 year olds in England'.
- Ofqual. 2013. "Annual qualifications market report 2013." Ofqual/13/5325, London.
- Ofqual. 2014. "Annual qualifications market report – England, Wales and Northern Ireland 2012/13." Ofqual/14/5492, London.
- Ofqual. 2015a. "Annual qualifications market report – England, Wales and Northern Ireland 2013/14." Ofqual/15/5778, London.
- Ofqual. 2015b. "School and College Purchasing Behaviours." Ofqual/15/5594. London.
- Ofqual. 2015c. "Inter-board comparability of grade standards in GCSEs and A levels." Ofqual/15/5819. London.
- Ofqual. 2016a. "General Conditions of Recognition." London.
- Ofqual. 2016b. "GCSE (A* to G) Qualification Level Conditions and Requirements." London.
- Ofqual. 2016c. "Annual Qualifications Market Report – England, Wales and Northern Ireland 2014/15." London.
- Ofqual. 2016d. "Perceptions of A levels, GCSE and Other Qualifications in England – Wave 14." Ofqual/16/6040. An Ofqual commissioned report with a survey carried out by YouGov, London.
- Paton, Graeme. 2013a. "Three-fold rise in schools offering alternative GCSE exam." *The Telegraph*, 18th August: <http://www.telegraph.co.uk/education/educationnews/10248601/Three-fold-rise-in-schools-offering-alternative-GCSE-exam.html>.
- Paton, Graeme. 2013b. "Schools moving to 'easy' IGCSE exams to boost grades." *The Telegraph*, 21st August.
- Pellegrino, James W., Naomi Chudowsky, and Robert Glaser, eds. 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.
- Propper, Carol, Simon Burgess, and Denise Gossage. 2008. "Competition and Quality: Evidence from the NHS Internal Market 1991–9." *Economic Journal* 118(525):138-170.
- Propper, Carol and Jennifer Dixon 2011. "Competition between hospitals" in

- Understanding New Labour's market reforms of the English NHS*, edited by Nicholas Mays, Anna Dixon, and Lorelei Jones. London: The King's Fund, pp. 78-88.
- Reidy, Tess. 2015. "Will taking a BTEC help or hinder your university application?" *The Guardian*, 21st July: <https://www.theguardian.com/education/2015/jul/21/will-taking-a-btec-help-or-hinder-your-university-application>.
- Ricci, Francesco and Marios Zachariadis. 2013. "Education Externalities on Longevity." *Economica* 80(319):404-440.
- Richardson, Hannah. 2011. "Three more errors found in A-level and GCSE exam papers." *BBC*, 21st June.
- Richardson, Hannah. 2015. "New maths GCSEs too hard, says exams regulator." *BBC*, 21st May: <http://www.bbc.co.uk/news/education-32831905>.
- Rogowski, Jeannette, Arvind K. Jain, and José J. Escarce. 2007. "Hospital Competition, Managed Care, and Mortality after Hospitalization for Medical Conditions in California." *Health Services Research* 42(2):682-705.
- Ruble, Whitney. 2015. "The Effect of Contracting Out Low Performing Schools on Student Performance." Working Paper 1521, Department of Economics, Tulane University, New Orleans.
- Schumpeter, Joseph A. 1994. *Capitalism, Socialism and Democracy*. New York: Routledge.
- SCORE. 2012. "The administration of examinations for 15 to 19 year olds in England." Written evidence submitted by SCORE, Education Committee, House of Commons, London.
- Scott, Kirsty. 2000. "Scottish exam authority 'riddled with ineptitude'." *The Guardian*, 7th October: <https://www.theguardian.com/uk/2000/oct/07/kirstyscott1>.
- Shackleton, J. R. 2014. "The history of qualifications and the role of competition" in *Tests worth teaching to: incentivising quality in qualifications and accountability*, edited by Gabriel Heller Sahlgren. London: CMRE, pp. 6-27.
- Shleifer, Andrei. 1998. "State versus Private Ownership." *Journal of Economic Perspectives* 12(4):133-15.
- Smith, Ruth. 2013. "Government U-turn on EBC examinations – was EU procurement law really to blame?" *Procurement Portal*, 8th February: <http://www.procurementportal.com/blog/blog.aspx?entry=225>.
- Spagnolo, Giancarlo. 2012. "Reputation, competition, and entry in procurement." *International Journal of Industrial Organization* 30(3):291-296.
- SQA. 2014. "Terms and Conditions of Appointment." Scottish Qualifications Agency, Glasgow. http://www.sqa.org.uk/sqa/files_ccc/TandCs2015Markers.pdf.

- Steinberg, Bene't. 2002. "Examination boards: regulatory overkill or pattern for the future." *Economic Affairs* 22(1):44-50.
- Stewart, William. 2013. "Schools turn to IGCSEs because they are "way easier"." *Times Educational Supplement*, 21st August: <https://www.tes.com/news/school-news/breaking-news/schools-turn-igcse-because-they-are-way-easier>.
- Stewart, William. 2014. "IndependentschoolsabandonGCSEsforinternationalalternative." *Times Educational Supplement*, 30th August: <https://www.tes.com/news/school-news/breaking-news/independent-schools-abandon-gcse-international-alternative>.
- Stewart, William. 2015. "More schools switching to IGCSE." *Times Educational Supplement*, 22nd May: https://www.tes.com/news/school-news/breaking-news/more-schools-switching-igcse?dm_i=14DE,3F2ST,63BA7A,C8CLX,1.
- Suto, Irenka. 2012. "What are the impacts of qualifications for 16 to 19 year olds on higher education? A survey of 633 university lecturers." Report, Cambridge Assessment, Cambridge.
- Swain, Harriet. 2016. "Will new university admissions procedures boost social mobility?" *The Guardian*, 18th April: <https://www.theguardian.com/education/2016/apr/18/university-admissions-social-mobility-tests-students>.
- Swinford, Steven. 2015. "Exams boards ordered to re-write papers which are 'too difficult'." *The Telegraph*, 21st May: <http://www.telegraph.co.uk/education/educationnews/11621191/Exam-boards-ordered-to-dumb-down-GCSE-maths.html>.
- Tiemann, Oliver and Jonas Schreyögg. 2012. "Changes in hospital efficiency after privatization." *Health Care Management Science* 15(4):310-326.
- Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.
- Volpp, Kevin G. M., Sankey V. Williams, Joel Waldfogel, Jeffrey H. Silber, J S. Schwarz, and Mark V. Pauly. 2003. "Market Reform in New Jersey and the Effect on Mortality from Acute Myocardial Infarction." *Health Services Research* 38(2):515-533.
- Wantchekon, Leonard, Marko Klačnja, and Natalija Novta. 2015. "Education and Human Capital Externalities: Evidence from Colonial Benin." *Quarterly Journal of Economics* 130(2):703-757.
- Weale, Sally. 2015. "Ministers consider scrapping exam boards, as students await A-level results." *The Guardian*, 9th August: <https://www.theguardian.com/education/2015/aug/09/exam-boards-students-a-level-results-nick-gibb-reform>.
- West, Martin R. and Ludger Woessmann. 2010. "Every Catholic Child in a Catholic School': Historical Resistance to State Schooling, Contemporary Private Competition and Student Achievement across Countries." *Economic Journal* 120(546):F229-F255.

- Wiggins, Kaye. 2016a. "Ofqual forces exam board to raise C-grade boundary on maths GCSE." *Times Educational Supplement*, 25th August: <https://www.tes.com/news/school-news/breaking-news/ofqual-forces-exam-board-raise-c-grade-boundary-maths-gcse>.
- Wiggins, Kaye. 2016b. "IGCSE results cannot be trusted, report claims after board compensates for 'lenient' grading." *Times Educational Supplement*, 15th April: <https://www.tes.com/news/school-news/breaking-news/igcse-results-cannot-be-trusted-report-claims-after-board-compensates>.
- Winnet, Robert, Holly Watt, and Claire Newell. 2011. "Exam boards investigation: Michael Gove will reform system to prevent boards competing." *The Telegraph*, 9th December: <http://www.telegraph.co.uk/education/educationnews/8947391/Exam-boards-investigation-Michael-Gove-will-reform-system-to-prevent-boards-competing.html>.

The question how should the qualifications of students be assessed is one of the most defining and important aspects of any education system.

England, Wales, and Northern Ireland have a unique system of qualifications and assessment distinguished by choice and diversity. While there is much to commend this model, critics have recently questioned whether standards can be upheld under competition between multiple providers. If incentives are not properly aligned to outcomes, competition might lead to a 'race to the bottom' instead of promoting quality.

In this monograph, Gabriel Heller Sahlgren applies transaction cost economics and such empirical evidence as is available on the subject to determine which of several different models of provision ought in theory to be most advantageous. A central insight of transaction cost theory is that quality expectations are not met automatically, but require costly efforts of specification and verification. Private provision may be superior to government monopoly, but only as long as private providers cannot shirk on non-contractible dimensions of service.

Private provision could take on different forms. The first main model is characterised by procurement or franchising. The other main model is distinguished by user choice. By carefully comparing these two private models with government provision, this monograph combines a comprehensive treatment of the subject with a discussion of the most policy relevant issues. What emerges is a categorical ranking of the models, elucidated by instructive discussion of how each of the private models could be improved to get the most information value at the lowest cost.



23 Great Smith Street, Westminster, London, SW1P 3BL

www.cmre.org.uk
info@cmre.org.uk

ISBN 978-1-5136-1601-8

