What is meant by "rigour" in examinations?

Isabel Nisbet[1]
(nisbet.isabel@gmail.com)

"If thou, Lord, wilt be extreme to mark what is done amiss: Lord, who may abide it?" (Psalm 130, verse 3 (Book of Common Prayer))

Abstract

The term "rigour" ["US "rigor"] is often used to justify educational reforms, including changes to public examinations. The paper considers the development of the concept of "rigour", noting in particular a recent shift from more negative to more positive connotations. It makes some observations about the application of "rigour" to curriculum, but deals mainly with its application to different aspects of assessment. The conclusion is that no one structure or task type can be assumed, prima facie to be more rigorous than its opposite, or more educationally desirable. It is important that debates about examination reform do not confuse the different concepts associated with "rigour". [Key words: assessment; rigour; examination reform]

……

The term "rigour" ("rigor") is often used, normally as a term of approval, by advocates of educational reform. One international assessment organisation has established a distinguished panel charged with developing "a blueprint for assessments which are internationally benchmarked and rigorous"[2] This paper will consider the development of the concept of "rigour", particularly in educational contexts. It will then briefly address its application to curriculum and pedagogy, but mainly ask what is meant by "rigour" in assessment. This is particularly topical in countries, such as the UK, in which changes to public examinations have been justified as increasing rigour. "We will make GCSEs more rigorous by stripping out modules," the UK Secretary of State told Parliament in 2010[3], and in 2012 he referred with approval to the "rigorous and respected exams taken by Singapore's students"[4]. Her Majesty's Chief Inspector of Schools, Sir Michael Wilshaw, told a UK Parliamentary Committee that in the past, "useless vocational qualifications" had been introduced, "which lack rigour – and are seen to lack rigour by both students and employers"[5]. The examinations regulator for England, Ofqual, states "We regulate because … the public needs to be assured that standards and rigour are being maintained." (Ofqual, 2013) What does this mean? What are rigorous examinations? And is rigour always a good thing in assessment?

[1] At the time of writing, Isabel Nisbet was Director of Education, South East Asia, for Cambridge International Examinations. Since April 2014 she has been Executive Director of the A Level Content Advisory Board (ALCAB), advising UK Ministers on behalf of some universities on the content of A levels for offer in England. The views expressed in this paper are the author's own and not necessarily those of Cambridge International Examinations or ALCAB.

[2] "Pearson convene international expert panel to define new gold standard in assessment", announcement published on 13 August 2012, obtained from http://uk.pearson.com/home/news/2012/August/pearson-convene-international-expert-panel-to-define-new-gold-standard-in-assessment.html

[3] Rt Hon Michael Gove, Oral Statement on the Schools White Paper, House of Commons Hansard, 24 November 2010, col 267

[4] Rt Hon Michael Gove, Oral Statement, House of Commons Hansard, 21 June 2012, col 1025.

[5] Sir Michael Wilshaw, HM Chief Inspector of Schools, in oral evidence to the House of Commons Education Select Committee, 12 February 2014

The concept

Early uses of "rigour" in English are close to its Latin etymological root ("rigor", meaning "stiffness"), conveying notions of strictness and severity (Nelson, 2011). Rigour implied punctiliousness – punishing every wrong-doing – and Chaucer contrasted it, disapprovingly, with the more noble virtues of "pacience" and "temperaunce"[6]. In the Old Testament, the stern Egyptians forced the exiled Children of Israel to "serve with rigour"[7], but the Isrealites were later instructed not to be so severe in governing themselves ("Ye shall not rule one another with rigour"[8]). This root concept of "rigour" is largely negative, particularly in its early usages, although some words associated with it have more positive connotations – thoroughness, perfectionism, carefulness, not allowing looseness or mistakes – and even reliability and "high standards". Many of these words are found in contemporary debates about examinations.

A development from the root concept of "strictness" has been the use of "rigour" to describe intellectual features – "rigorous arguments", "rigorous analysis" and "academic rigour". These normally imply thoroughness and carefulness, with no step missed out, every statement based on evidence and conclusions following logically from the premises of the arguments used. In this sense, "rigorous" is normally a term of approval: academic publications which were full of unjustified polemic, illogical or sloppily referenced would be criticised as lacking rigour.

But none of these senses of "rigour" appears to fit much of its use in educational writing in the 20th century, particularly in the USA. Nelson (2011) talks of a recent "paradigm shift" in the concept, moving "from harshness to excellence". Today the flag of "rigour" ("rigor") is commonly flown by educationists advocating aiming for the highest academic standards for all students. Barbara Blackburn emphasises that "rigor" is "not about severity or hardship" and is "not a measure of the quantity of content to be covered". Her definition is:

> "Rigor is creating an environment in which each student is expected to learn at high levels, and each is supported so he or she can learn at high levels, and each student demonstrates learning at high levels" (Blackburn, 2008)

Writers advocating "rigor" in this sense use the term positively, to support the highest ambitions for all, encouraging all students to aim for the highest standards and supporting them in each step of their progress. Rigo[u]r is contrasted with "low ambition", "dumbing down" and "second best". Although such language is often associated with the right wing of educational politics in the USA and the UK, its champions range wider. Some of Blackburn's thinking would be shared by Michelle Obama, who took a group of school children from inner London (mainly black girls) to Christ Church in Oxford. "Look around you," she said to them. "A renowned university that has trained so many of the world's brightest minds and greatest leaders. All of us believe that *you belong here*."[9]

In Box 1 I have set out some of the negative and positive words associated with "rigour", in its root sense ("strictness"), the later intellectual sense ("thoroughness", "carefulness") and the modern sense of "excellence", "high ambition". Many of these words are frequently found in discussions of curriculum and assessment. These discussions can be confused by

[6] The Frankiln's Tale, lines 773-786, Geoffrey Chancer, *h*
[7] Exodus 1, 11-14
[8] Leviticus 25,53)
[9] *The Guardian*, Wednesday 25 May 2011

using arguments associated with one meaning of "rigour" to justify or criticise actions which are associated with another – for example, by using the language of high ambitions for all to justify strict mark schemes in examinations, or by ridiculing advocates of the "high ambitions" sense of "rigour" by citing the negative connotations of the root concept. I shall try to avoid such lack of rigour in the remainder of this paper.

Box 1 **RIGOUR/rigorous**

NEGATIVE	POSITIVE
strict, severe, harsh, unbending, tough, "no excuses" every mistake picked up, … stressful,	thorough, perfectionist, careful, well-argued, based on evidence, no short cuts, structure, purpose "the real thing" complex, rich scholarly, aiming for excellence

Curriculum and pedagogy

As the focus of this paper is principally on assessment, I shall not do justice here to the rich discussion in the literature of "rigour" as applied to curriculum and pedagogy. The bibliography cites a range of sources, with an interesting selection published by the Heschinger Institute (Heschinger, 2009). I shall confine myself to three observations which also have relevance to assessment:

1. Many modern accounts of rigour in curriculum and pedagogy argue for thoroughness and structure , which hearkens back to the "academic rigour" concepts. A mathematics academic[10] is quoted in the Heschinger volume as saying "A rigorous course… examines details, insists on diligent and scrupulous study and performance, and doesn't settle for a mild or informal contact with the key ideas". Advocates of this kind of rigour tend to criticise textbooks which contain "gobbits" or extracts, and courses which encourage students to apply academic concepts to real-life situations in what they think is a superficial way, before they have properly understood the concepts.

2. Academic courses are not, by definition, more "rigorous" than vocational/practical courses. Courses in each category can differ in their degree of rigour. This point has been obscured by the fact that some of the champions of "rigour" in the sense

[10] Robert Talbot, Associate Professor of Mathematics and Computing Science, Franklin College, Franklin, Indiana

of high ambitions for all students have criticised systems which encourage students who are lower academic achievers to "make do" with vocational options at what the critics think is too young an age. It seems to me that most of the words in Box 1 associated with "rigour" could be applied to the learning of many kinds of knowledge and skill, including practical skills. The exception is perhaps the group of concepts associated with intellectual argument, which is more suited to the construction of arguments than to the construction of buildings, but most of the other connotations of "rigour" could be applied to demanding practical skills training as well as to academic programmes. In the 1990s, the content of GCSE Dance, for offer in England and Wales, was changed to include more dancing and less essay-writing. I do not believe that it can be concluded from that sentence alone that the exam became less rigorous, in any of the senses I have outlined (with the one exception, perhaps, of "requiring rigorous argument").

3. Can rigour stifle creativity? This argument can be applied both at school level and at more advanced levels. Michael Morpurgo, the writer and former Children's Laureate in the UK, argued that "rigour", in the root sense of strictness, when applied to the teaching of language in schools (for example, penalising spelling mistakes) could make children afraid to experiment with the use of words[11]. Robert Nelson argues that concepts of rigour in university settings have favoured analytic thinking rather than creative thinking (Nelson, 2011). I agree with Nelson that most of the concepts associated with "rigour" – and particularly intellectual/academic rigour – tend to favour structure and accuracy rather than flair and vision. However, the more positive version of "rigour", in the sense of "high ambitions for all" or "exposure to excellence" might be said to inspire children and extend their thoughts and ideals. Probably, the concept of rigour is now too wide for generalisation on the relation to creativity. However, the questions raised by the critics have sufficient resonance to encourage us to look critically on any educational innovation justified as increasing "rigour" by asking the question "What about creativity?".

Assessment

What does it mean to apply these varying senses of "rigour" to assessment and examinations? And are rigorous assessments necessarily better than less rigorous ones? It may help to consider how the concept(s) may be applied to different aspects of assessment[12].

Construct

Subject-content: One link between the discussion of rigour in assessment and its application to curriculum concerns the subject-matter of the assessment – for example, the syllabus for a public examination. It could be argued that a rigorous assessment is a valid assessment of a rigorous curriculum. If rigour in the curriculum denotes thoroughness or structured progressive learning, arguably a valid assessment should measure whether the candidate has travelled all the stages of that journey. However, on that understanding the assessment borrows its label of "rigour" from the content being assessed.

[11] "Michael Morpurgo: What Michael Gove calls 'rigour' I call 'rigor mortis'", *Daily Telegraph*, 7 December 2012

[12] I am indebted to my Cambridge colleagues Stuart Shaw and Tom Bramley for parts of the argument which follows

From first principles, there seems little link between most of the concepts of "rigour" in Box 1 and the sheer volume of content in a syllabus. For example, it is not clear why a syllabus which required students to study three Shakespeare plays would be more "rigorous" (in most senses) than a syllabus which required them to study only one. The only two possible links I can suggest are with "toughness" or "stress" (associated with the root meaning) – because the students would have to do more reading - or with "richness" and "quality" at the positive end of the spectrum, with the students of three plays exposed to more of the richness of Shakespeare's works. However, the most telling factor would be the kinds of knowledge and understanding being developed and then tested, however many plays were studied. There may be subject-specific arguments for more content in a syllabus – for example, that studying a longer period in history can increase the structure and depth of the student's historical understanding – but the case for that would need to be made, subject by subject. Prima facie it cannot, I suggest be assumed that more volume of content necessarily means greater rigour.

External acceptability/appropriateness: Another argument is that "rigorous" examinations or qualifications demonstrate knowledge and skills that are recognised as important and respected by the external users of the outcomes – for example, by employers or colleges and universities. Some of the recent critiques of certain vocational qualifications in the UK (eg Wolf, 2011) have argued that they lack rigour in that respect – that employers and colleges do not respect them – rather than that they are less rigorous just because they are vocational rather than academic.

Level and type of knowledge or skill tested: Hess and others (Hess et al, 2009) have argued for a measure of "cognitive rigor" which is linked to a combination of Bloom's taxonomy (in the revised form, which incorporates creativity) and Webb's "Depth of Knowledge". They link the different levels with different task words ("list, describe"… "compare.." "What if…") and argue that the better the assessment tasks fit the higher/deeper levels of both frameworks, the more they display "cognitive rigour". In this context, Hess and his colleagues tend to equate "cognitive rigour" with "higher order thinking skills".

In response, I would observe, first, that "cognitive rigour" in this sense cannot be equated with "difficulty". Some of the world's most difficult tests are tests of memorisation (see Box 2) and some of the language assessments which are normally rated as the most demanding require memorisation of thousands of quasi-pictorial characters.

Box 2
Dubai's Quran recital competition

"[The Quran has 6,000 verses, spread across some 600 pages in 114 chapters.] The judge recites a random verse from the Quran and asks the candidate to continue reciting where he left off. 13-year-old Tajikistani wunderkind Lutfillo Kholikov … is the only competitor thus far who, in addition to reciting the correct verse, previewed his answer by stating .. which page the verse was on, its corresponding chapter and in which city the chapter was revealed, as well as the first and last word on the page where the verse was located." (The National (UAE) 11 August 2012).

Secondly, it is an open question how best to assess the higher orders of Bloom's and Webb's classifications. For example, level 4 of Webb's Depth of Knowledge ("extended thinking") might arguably be best demonstrated in tasks which took more time than the normal length of an examination paper and which required the student to plan and carry out research. This might suggest an extended project of some kind, rather than a time-limited examination.

Thirdly, a test of higher-order thinking may exhibit some of the positive feature of "rigour", while avoiding the pejorative overtones of the negative root meaning. For example, in Singapore, some of the science questions in the Primary School Leaving Examination were designed to be "creative", asking pupils to think about new situations which they had not learned before. Pupils were reported as describing the paper as "interesting", "new" and "tricky", or even as "more fun"[13]

Validity, or lack of certain types of invalidity: Some readers of this paper may take the view that all the aspects of assessment addressed are really sub-divisions of validity. I shall leave that discussion for another day. However, one aspect of validity/invalidity which seems particularly close to many of the ideas associated with "rigour" is relevance of the assessment tasks to the knowledge and skills being measured, or (with apologies for the triple negative) lack of construct-irrelevant invalidity.

An assessment might lack rigour, in both the root, negative, sense and the more positive senses if candidates could do well without demonstrating the prescribed knowledge or skill. As Ofqual argued in its study of predictability, lack of rigour in this sense could be a consequence of question papers that were either over-predictable or unpredictable. With over-predictable assessments, for example, recall of memorised material and drills could replace the demonstration of higher-order skills. And excessively unpredictable assessments would lack reliability (Ofqual, 2008).

Because of the roots of "rigour" in "strictness" we tend to use it to exclude assessments where it is possible to get by without having the required knowledge. Other types of construct-irrelevant invalidity (for example, culture bias in the wording of questions) may have the opposite effect – of denying appropriate results to candidates who deserved them. We would not normally use the language of "rigour" to criticise such failures, but that is not to deny their importance.

Difficulty – "real" and perceived: Discussions of "difficulty" in assessment normally contrast the "judgemental" approach, where the test score indicates the candidate's position on an notional abstract spectrum of levels of achievement, with the "statistical" approach, where the standard for a particular score relates to the proportion of the defined population expected to achieve it (see Bramley, 2005 and a range of publications by Cresswell (eg Cresswell, 1996)). However, the notion of "rigour", at least in some of its earlier uses, seems to prey in aid the psychological effect of the rigour on its recipient. In that context, a "rigorous" assessment is expected to feel difficult. We are told, with approval, that students feel that they have been "put through their paces" or "given a hard time".

In response to this, I would make two observations. First, as we have seen from the example from Singapore, it is possible to increase rigour in some senses (eg by testing deeper knowledge) without increasing psychological stress on the candidates. Second, perceptions of difficulty have been shown to be culturally determined. Research by Shen and Pedulla showed that, *within* countries, students who said they found mathematics easy did better in

[13] *The Straits Times*, 13 October 2013

mathematics tests than their compatriots who said they found mathematics hard. However, *inter*nationally, the mean scores of students from countries where students, on average, said they found mathematics difficult was higher than that of countries where students said they found the subject easy (Shen and Pedulla, 2000). The researchers speculate about the reasons for this interesting finding, but for my purposes it is sufficient to note that perceived difficulty is not the same as "real" difficulty (in either the judgmental or the statistical sense) and that a more "rigorous" assessment may not necessarily feel more difficult.

Structure and type of assessment. We have seen that in the UK, the Secretary of State said that a move from a modular structure to terminal examinations made GCSEs "more rigorous". As national policies on modularisation of exams have changed, researchers have found that those structural changes produce "winners" and "losers" among students, depending on age, the nature of the subject being studies and, sometimes, gender (McClune, 2001). Taverner and Wright found that A level mathematics students with comparable GCSE scores got better results by half a grade in a modular A level, compared with a terminal examination (Taverner and Wright, 1997). Does that mean that terminal examinations are more difficult? Or more severe/stressful (as in some of the root uses of "rigorous")?

There seem to me to be three possible reasons why modular examinations might be thought to lack rigour, compared with terminal examinations. The first is that most modular structures offer opportunities to employ "gaming" strategies and the availability of opportunities to resit modules to get the best outcome. Hence the outcomes may be determined by construct-irrelevant factors.

Secondly, modular assessments may require shorter-term memory as the knowledge tested will have been studied recently and there will be less to remember at the one time – and a shorter period to remember it - than is needed of terminal examinations covering the whole syllabus. On this argument, terminal examinations are more "rigorous" as they require more to be remembered at the one time. If that is the case, the question of whether that is a good thing or not depends on the relevance of the more extended memory required to the construct being assessed, and to the use of the test results. For those looking for "cognitive rigour" measured by Bloom's taxonomy and Webb's Depth of Knowledge, a move to requiring greater memorisation be irrelevant,

A more persuasive argument is that linear assessment may allow for more extended questions which require students to apply concepts developed in one part of the curriculum to others, or to new situations. This would fit with the notions of "rigour" linked to the opportunity to display higher-order thinking skills. However, it will only hold if the tasks in the terminal examination really do test higher-order thinking. Also, the relevance of this argument may depend on the nature of the subject – for example, whether the syllabus involves a linear development of understanding or skill, with the highest level achieved at the end of the course, or whether it involves a sequence of more disparate topics. It may also depend on the age of the students and the extent to which they mature intellectually during the course.

There are, therefore, some prima facie reasons why terminal examinations might be more "rigorous" than modules. But much depends on the context, the subject studied and the students. The total amount of time available to demonstrate the required knowledge and skill may be less with a terminal examination than with modules, and syllabus writers of some subjects may find it difficult to "cram" the content into the limited examination window. Also, as we have seen, some kinds of extended knowledge are arguably better demonstrated through extended project work.

Test items: The pleas for more "rigour" in examinations have often been accompanied by demand for examinations to have more extended tasks, such as essays, rather than multiple choice questions or one-word answers. Arguably, the "extended thinking" featured in Webb's higher-scoring depth of knowledge categories can be better demonstrated in tasks where the candidate has to extend his or her thinking and develop an argument. However, the validity of such tasks depends on the construct being assessed. For example, diagnostic decisions in medicine may be more validly tested by multiple choice, as in real life the doctor has to choose among a finite range of options. Clearly poorly-constructed multiple choice questions, where the candidate could pass without demonstrating the knowledge required, for example by gaming strategies such as opting for the third option every time, would not be rigorous. But good-quality multiple choice may be the most appropriate form of assessment for some constructs. That is over and above the arguments based on reliability, which led several of the medical Royal Colleges in the UK to make greater use of objective testing and less of extended essays and interviews[14].

It has also been argued that "rigorous" assessments have more "holistic" tasks, rather than breaking up the task into parts which the student can tackle sequentially. Tasks set out in steps have been derided as "bite-sized". It seems plausible that tasks which are combined in a holistic question may be more rigorous in the root sense of "daunting" or "severe", and that they might provide more opportunitites for extended thinking. But the validity of such a change depends on the purpose of the assessment. For example, if a test is aiming to rank-order candidates for selection to schools or colleges, then it will be important to be able to distinguish among middle-achievers, who may be able to do some of the tasks but not all of them, from those who can do little or none of any of the tasks. Also, students may be demotivated by the appearance of the holistic tasks and not do themselves justice. For these reasons, some of the questions in Singapore's Primary School Leaving Examination were restructured in 2013, "to guide pupils to the answers. Even if their answers were only partially right, pupils would still be able to earn part of the marks"[15] In the words of the Minister, this was done "to bring everyone's focus back from chasing points to really learning". Was this approach less rigorous? Perhaps, in some senses. But most commentators would argue that it was educationally justified, and that the outcomes gave a more valid measure of the achievement of the middle group of candidates.

Marking – standards for pass/grades: One possible way of making examinations more "rigorous" could be simply to raise the standard of achievement required for "passing", or for the award of particular grades. This could be done without changing the construct examined – for example by raising cut scores above the level required for comparability with previous years. There are policy arguments for and against making such changes, but it seems to me to be difficult to link them to the concept of "rigour" except in the root, negative, sense of "severe" or "stressful". If a candidate needs a Grade C to get to college and if it becomes more difficult to achieve a grade C – with no other changes to the curriculum or the structure of the assessment – then his anxiety will be increased, but it is difficult to describe the change as an increase in "rigour".

Marking – strictness: one aspect of marking which appears close to the root, negative, sense of "rigour" is strictness. In England, changes have been made to mark schemes for some public examinations to apply more penalties for poor spelling, punctuation, grammar and mental arithmetic. In that sense, "rigorous" mark schemes can be described by some of the

[14] I owe this point to Professor Dame Lesley Southgate

[15] *The Straits Times*, 23 November 2013

words at the top of Box 1 – picking up all mistakes, punctilious… It seems to me that strictness in this sense does relate to the root, more negative, sense of "rigour". But is rigour in this sense desirable? The answer depends, I believe, on the purpose of the assessment. As the psalmist observed in the quote at the beginning of this paper, an "extreme" approach, picking up everything that is "done amiss" will catch us all out. Thus if the purpose of the

assessment is to distinguish across the range of candidates being tested, that approach will be self-defeating. It may also invite Michael Molpurgo's charge that fear of mistakes may stifle creativity. On the other hand, if the primary purpose of the assessment is to police a standard of competence so that no-one below it will be allowed to practise, then the negative words associated with the root concept of "rigour" become positives (see Box 4). Protecting the safety of the patient is more important than generous assessment of the surgeon and our lives may depend on our surgeon's assessment having been carried out with rigour, in its root sense. .

Box 3: Not a real case, thank goodness

"The inquest was told that neurosurgeon Mr X had made an incision two inches away from the correct spot in the brain of the deceased. During his professional training, Mr X had regularly made inaccurate incisions, but an examiner told the coroner that Mr X – and others with similar levels of competence – were usually placed above the line in their examinations in order not to discourage students from taking up neurosurgery, and to give them positive reinforcement, as it was a difficult specialty. Also, Mr X's grandmother had passed away the day before his final exam, and the examiners took that into account."

Conclusion

I have described how the concepts associated with "rigour" have developed in educational contexts, noting in particular the "paradigm shift" from the initial concepts of strictness and harshness to excellence and high ambitions. It is essential that arguments for or against reforms aiming to increase rigour are focussed on the sense of "rigour" intended by the reformers. I have looked briefly at issues about curriculum and pedagogy raised by concepts of "rigour" and the bibliography includes further sources on that. I have then considered what is meant by "rigour" in assessment and the aspect of assessments to which the adjective "rigorous" might be applied. It is a varied picture, and no one form of assessment or structure is prima facie more rigorous than its opposite in all contexts. The questions of whether reforms to examinations increase rigour - and whether they are desirable – need to be considered in context, selecting from the wide spectrum of concpets that has become associated with the concept of "rigour".

Bibliography

Belshaw, Doug (2012), "What constitutes 'rigour' in our 21st-century educational systems?" Downloaded from http://dmlcentral.net/blog/doug-belshaw/what-constitutes-rigour-our-21st-century-educational-systems

Blackburn, Barbara R (2008), *Rigor is not a four-letter word*, Routledge, Abingdon/New York (2nd edition 2013)

Bramley, Tom (2005), "Accessibility, easiness and standards", *Educational Research*, 47:2, 251-261

Cresswell, MJ (1996), "defining, setting and maintaining standards in curriculum-embedded examinations: judgmental and statistical approaches", in H Goldstein & T Lewis (eds) *Assessment: problems, developments and statistical issues*, New York, Wiley

Hechinger Institute (2009): The Hechinger Institute on Education and the Media, Teachers College, Columbia University, *Understanding and Reporting on Academic Rigor*, June 2009, can be downloaded from http://hechinger.tc.columbia.edu/what-we-do/publications

Hess, Karin K, Jones, Ben S, Carlock, Dennis and Walkup, John R (2009), "Cognitive Rigor: Blending the Strengths of Bloom's Taxonomy and Webb's Depth of Knowledge to Enhance Classroom-level Processes", Education Resources Information Center (ERIC) publication ED517804, 7 March 2009

Jago, Carol (2001, 2nd ed 2011), *With rigor for all*, available as PDF from heineman.com

McClune, Billy (2001): "Modular A levels – who are the winners and losers? A comparison of lower-sith and upper-sixth students' performance in linear and modular A level physics examinations", *Educational Research*, 43:1, 79-89

Nelson, Robert (2011), "Toward a history of rigour: An examination of the nasty side of scholarship", *Arts and Humanities in Higher Education*, 10, 374-387

Ofqual (2008), *Predictability studies report: A study of GCSE and GCE level examinations*, August 2008, Ofqual/08/3866

Ofqual (2013), *General Conditions of Recognition – September 2013*, Ofqual/13/5306

Shen, Ce and Pedulla ,Joseph J (2000), "The relationship between students' achievement and their self-perception of competence and rigour of mathematics and science: A cross-national analysis", *Assessment in Education: Principles, Policy & Practice*, 7:2, 237-253

Taverner, Sally and Wright, Martin (1007), "Why go modular? A review of modular A level mathematics", *Educational Research*, 39:1, 104-112

Wolf, Alison (2011), *Review of Vocational Education – the Wolf Report*, DfE, London, March 2011, available at https://www.gov.uk/government/publications/review-of-vocational-education-the-wolf-report