ASSESSMENT and QUALIFICATIONS ALLIANCE

Public Trust in High-Stakes Assessment and its Measurement

> *"To say we trust is to say we believe that individuals and institutions will act appropriately and perform competently, responsibly, and in a manner considerate to our interests."*
> (Barber, 1983; cited in Mechanic, 1996, p. 173).

ABSTRACT

Sociologists (and journalists) have increasingly claimed that society is less trusting of public institutions (O'Neill, 2002). Public trust in the examination boards responsible for setting, marking and grading national examinations is important, given the uses to which assessment results are put; examination results impact upon the life chances of students and the careers of those who teach them. This paper discusses possible indicators of public trust in examination boards. In particular, it explores the findings from an analysis of enquiries about results data. The analysis shows despite apparent improvements in quality of marking, the number of re-mark requests continues to increase yearly. To what extent, however, are such trends an indication of a lack of public confidence in assessment or a result of other societal pressures on students and teachers to perform? How else could public trust in assessment be objectively measured?

INTRODUCTION

In the UK, the information provided by examination boards responsible for high-stakes assessment is put to a multitude of purposes. Examination results impact upon the educational and employment choices of students, parents and employers. They provide a basis for selecting individuals for higher education courses or particular jobs and qualify individuals to perform certain vocational or professional activities. Assessment outputs are not only used to judge students, but also those responsible for teaching them – teachers and schools may be rewarded or reproved on the basis of their students' performance. Furthermore, examination results can be used in the development of policy arguments for alternative modes of education and training and to compare the educational accomplishments of one society with those of another (O'Neill, 2005). For these reasons, it is important that those who use examination results can trust and have confidence in them. Indeed, trust has been acknowledged as central to the credibility of the examination system in the UK. William (1996) argued that the maintenance of examination standards is partly dependent upon those responsible for standard setting being trusted by users of examination results:

> *"Examination results are 'social facts'; like bank notes they depend for their value on the status that is accorded to them within a social system".*
> (Wiliam, 1996, p.304).

Over recent years, it has increasingly been claimed that contemporary society is characterised by a 'crisis of trust' or 'culture of suspicion' (O'Neill, 2002). Opinion polls support such claims, showing a marked decline in trust in a number of public institutions and professions (Cabinet Office, 2004; cited in Kelly 2005; Duffy *et al* 2003; Ryan, 2000; cited in Kelly, 2005). According to theorists, such as Giddens (1990), this current state of affairs is a consequence of an overall decline in deference to authority. In a post-modern society, trust must be earned by governments, organisations and groups from autonomous, reflexive individuals.

The extent to which this 'culture of suspicion' has permeated the area of assessment in the UK is yet to be assessed. Mechanic (1996) identified current events and media images as being key in shaping trust in public institutions. Examinations and the agencies responsible for administering and grading them have been subject to a number of scandals e.g. the Scottish Qualifications Authority crisis in 2000, the Edexcel crisis in 2001, and the A-level results crisis in England in 2002. These scandals, coupled with the increased scrutiny of examination standards - common headlines include assertions that educational standards are falling, examinations are getting easier and so on (Murphy, 2004) - and the technicalities of grading examinations (Warmington & Murphy, 2004) by the British press seem likely to have impacted upon the public's trust. Worryingly, there is also evidence to suggest that negative events are more visible, carry greater psychological weight, are perceived as more credible, and hinder the kind of experiences needed to overcome distrust in the future (Slovac, 1993).

Trust is clearly central to the examination system in the UK – its very legitimacy depends upon it. Furthermore, the cost of distrust for those agencies responsible for administering examinations is likely to be high. It could result in excessive amounts of time being devoted to grievance procedures and even a loss of entries as schools and colleges seek more 'trustworthy' alternatives. So, how could examination boards best measure and monitor public trust in their activities and outputs? One approach would be to treat observable trends, such as the increase in uptake of the International GCSE (which are not recognised by either the Department for Children, Schools and Families or the Qualifications and Curriculum Authority (QCA)) by Independent Schools ((Marley, 2007), or the popularity of the International Baccalaureate in place of traditional A-levels, as indictors of mistrust. Alternatively, fluctuations in enquiries from schools/ colleges following the publication of examination results may suggest changes in trust levels. The purpose of this paper is to explore the value of the latter as an indicator of public trust in high-stakes assessment.

ENQUIRIES ABOUT RESULTS

The examination boards offer a number of Enquiries about Results (EAR) services for high-stakes examinations, one of which is a post-results review of the marking. Schools/colleges or students can request that an examination script is re-marked when they are dissatisfied with the original mark received. Following a re-mark a candidate's mark may increase, decrease or remain the same.

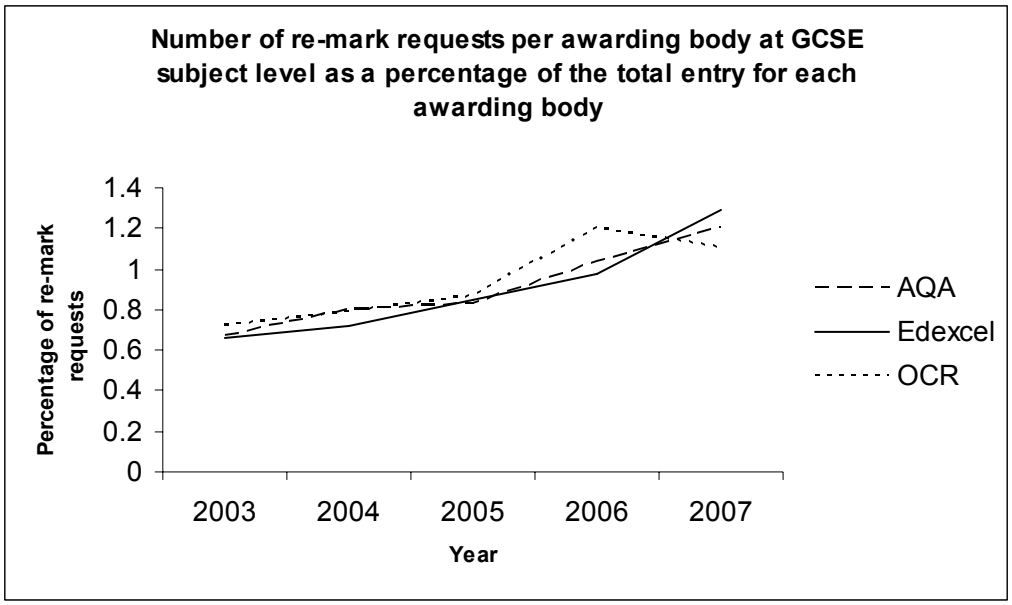
The QCA define a re-mark as *"A process in which the second examiner reviews the marking of the first examiner to ensure that the authorised mark scheme has been applied reliably"* (QCA, 2007, p.67). In line with this, re-marking examiners are advised that the aim of the re-mark is to determine whether the mark awarded is a fair reflection of the script's merit; if it is the mark should not be changed. Examiners are also reminded that a re-mark does not imply that the original marking, which will have been monitored during the marking period, is in anyway inappropriate. It is imperative that no undue generosity is given to students whose scripts are being re-marked as to do so unfairly disadvantages students who have not applied for re-

marking. Equally, however, re-marking exists to provide a check on reliability, and examiners must make mark adjustments if they consider the original assessment to be unreliable.

Rates of re-mark requests over time

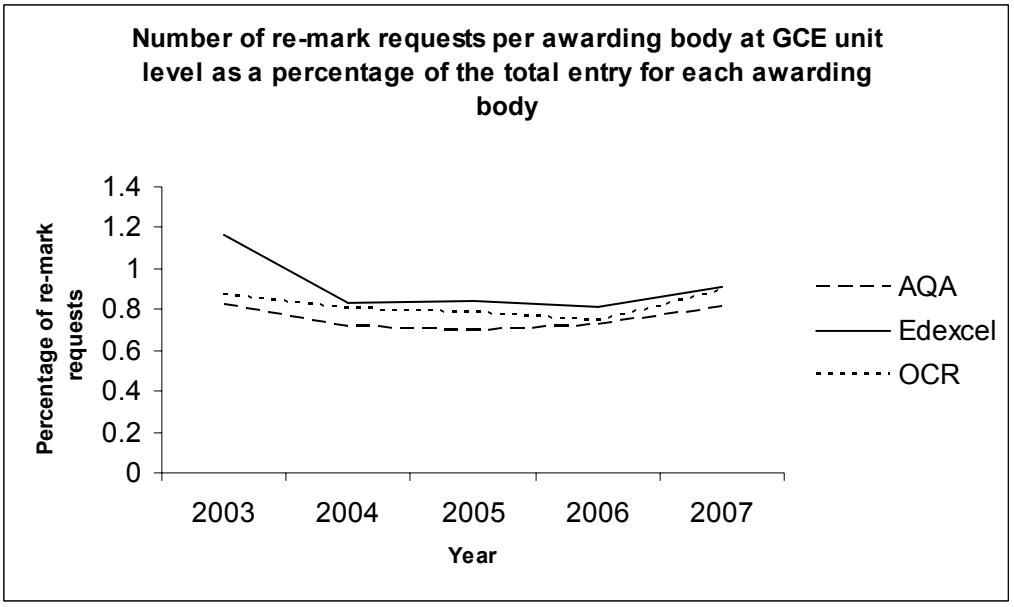
Meadows and Taylor (2008) explored the pattern of re-mark requests and grade changes over a five-year period (2003-2007). They found that for GCSE, there had been a steady increase in the rate of re-marks requests as a proportion of the total entry (Figure 1). Furthermore, the pattern of increase was similar across the three examination boards in England. For all examination boards, the rate of enquiries as a proportion of the total GCSE entry was very small, between 0.6 and 1.3 *per cen*t.

Figure 1



The pattern of GCE re-mark requests as a proportion of the entry was less linear (Figure 2). Generally, the number of requests decreased between 2003 and 2005 and then began to increase in 2006 returning to approximately the initial rate in 2007. The rate of enquiries as a proportion of GCE entry was again very small for the three boards.

Figure 2



In 2002, there was a highly publicised A-level results crisis in England. Reforms to the curriculum had resulted in 94.3 *per cent* of candidates passing, an increase of 5 *per cent* from the previous year. Allegations that the government had manipulated grade boundaries to prevent an even higher pass rate were rebuffed, and a number of high-profile resignations ensued. Similarly, Edexcel results came under particular scrutiny in 2001 when it became clear that some scripts had been wrongly graded. It seems likely that the relatively high rate of GCE re-mark requests in 2003 was a consequence of previous year's bad press. Indeed, the findings of a survey conducted by Ipsos MORI (2006) on behalf of QCA showed that the events of summer 2002 had damaged public confidence in the A-level. There is some evidence to suggest, therefore, that EAR data provide an indication of the fluctuations in trust brought about by educational scandals.

Proportion of grade changes

Meadows and Taylor (2008) also considered the proportion of re-marks that lead to grade changes. They found that the proportion of re-marks that led to grade changes was low (e.g. 23 *per cent* for AQA's GCSEs in 2007 and 14 *per cent* for AQA's GCEs in 2007). Moreover, they noted a tendency for the proportion to steadily decrease for GCSE (Figure 3) and GCE (Figure 4) between 2003 and 2007.

Figure 3

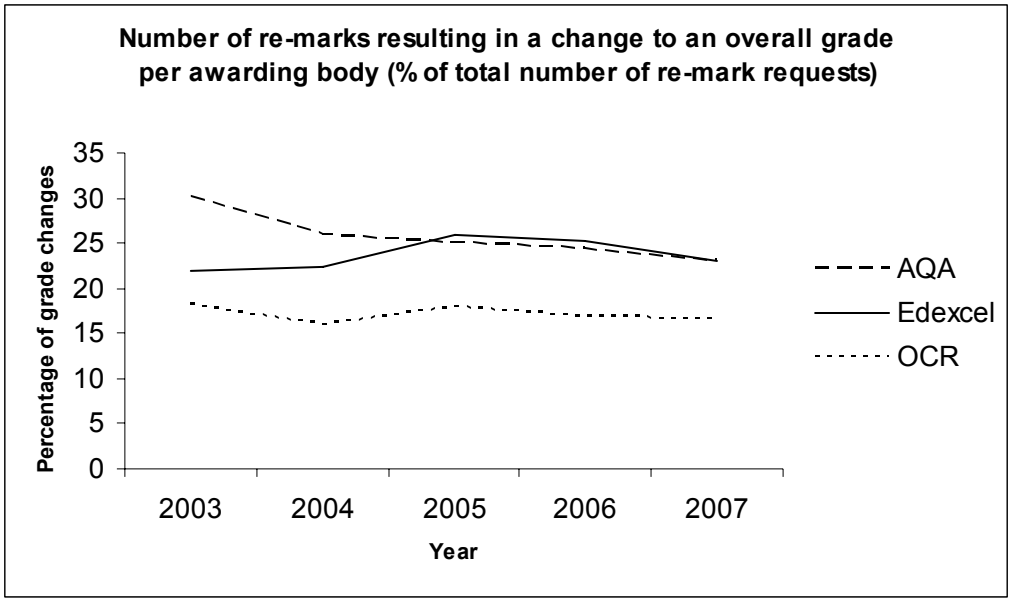
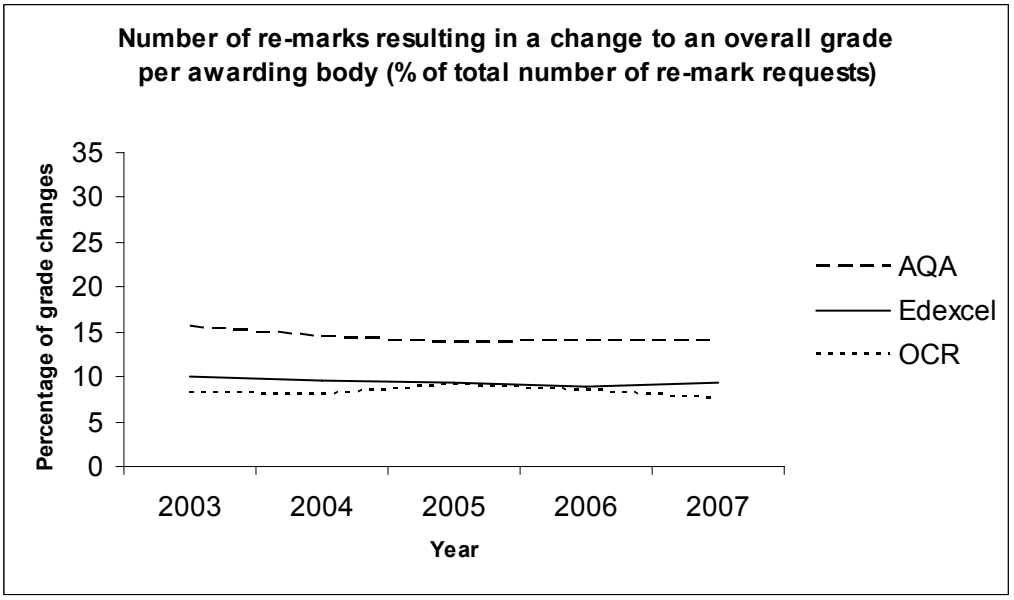


Figure 4



The decrease in the proportion of enquiries leading to grade changes may be a product of improvements in quality of marking. Research conducted into the impact of electronic marking, for example, has shown that it has had a positive impact upon marking reliability (Taylor, 2007). This finding, however, seems at odds with the reported increase in enquiries shown in Figures 1 and 2. If the quality of marking is improving (as shown by a reduction in the proportion of re-marks that result in a grade change) then why have enquiries steadily increased over the past 5 years? Meadows and Taylor (2008) propose a number of reasons for the increase in the number of enquiries, of which a decline in trust is only one:

> *"…it could also reflect an increasing understanding of the inevitability of some level of assessment error, or the increased pressure on teachers and students to achieve."*
> (Meadows & Taylor, 2008, p. 3)

Teachers and schools are increasingly judged on the basis of examination results. This is perhaps why enquiries tend to be targeted at those students whose marks are towards the top of the grade boundary (Pinot de Moira, 2003) – maximising the likelihood of an upgrade. It appears that the increase in enquiries observed over recent years may be ill-founded; not being based on a genuine increase in erroneous marking. The public's trust in the examination system may well have been damaged by educational scandals leading them to query results, but there is no clear and easy way to disentangle trust from other motivational factors e.g. to get further up the ever important school league tables. To truly get at the crux of public trust in the examination system an objective measure is needed.

HOW ELSE COULD PUBLIC TRUST IN ASSESSMENT BE MEASURED?

Examining how the issue of trust has been dealt with in other areas of social life highlights strategies for the measurement of trust in the examination system. Like education, the medical profession in the UK has been subject to a number of scandals, such as the conviction of the GP Harold Shipman, the enquiry into paediatric cardiac surgery in Bristol and the removal of organs from children at Alder Hey Hospital. Intense media scrutiny over medical incompetence is thought to have eroded public trust in health care institutions and its providers (Calnan & Rowe, 2004). Calnan and Rowe (2004) also link a decline in public trust in health care to structural factors, such as the way that the NHS is run and financed and the increased pressure placed on NHS budgets by an ageing population, the rising cost of technology, and increases in public sector pay. This loss of faith in the medical profession is by no means unique to the UK – Schlesinger (2002; cited in Simone, 2007), for example, traced the decline in public confidence in the medical profession in the United States.

Within the medical field there is a burgeoning body of literature about the causes, correlates and consequences of trust. Trust is recognised as a *"coherent psychological construct that can be reliably measured and differentiated from related concepts such as satisfaction"* (Hall, Dugan, Zheng, & Mishra, 2001). Many research teams have developed multi-item scales to measure trust in physicians (Anderson & Dedrick, 1990) and in hospitals or the medical system (LaVeist, Nickerson, and Bowie, 2000; Van der Schee, Braun, Calnan, Schnee, & Groenewegen, 2007). Mechanic (1996) importantly notes, however, that the distinction between interpersonal and institutional trust is a simple way of capturing a more complex reality. The two types of trust are correlated and mutually supportive (Parker & Parker, 2003) e.g. public trust is in part influenced by an individual's experiences in contacts with representatives of an institution or system. Illustrations of how medical researchers have sought to measure 1) interpersonal and 2) institutional trust are given below:

Patient-physician trust (interpersonal)

The Trust in Physician Scale (developed by Anderson and Dedrick, 1990) provides an example of a trust measure that operates at the level of interpersonal relationships. The Trust in Physician Scale is an 11 item, single score scale (see Table 1), selected from 25 items developed from patient interviews or adapted from other measures. Responses are scored on a five-point Likert scale. The scale was originally developed and tested on non-insulin dependent diabetic men (n=266) seen at a Veterans Administration Medical centre. The scale demonstrated high internal consistency with a reported Cronbach's alpha of .90 and .85 in two separate studies conducted in the same population. Construct validity was supported by positive associations with locus of control, with desire for clinician control, and with patient satisfaction with the meeting (Thom, Ribsil, Steward, Luke, & The Stanford Trust Study

Physicians, 1999). However, since all data were cross-sectional, there was no information on the test-re-test reliability of the scale or the predictive validity of the scale.

Table 1: Items featuring on the Trust in Physician Scale

Item
1. I doubt that my doctor really cares about me as a person.
2. My doctor is usually considerate of my needs and puts them first.
3. I trust my doctor so much I always try to follow his/her advice
4. If my doctor tells me something is so, then it must be true.
5. I sometimes distrust my doctor's opinions and would like a second one.
6. I trust my doctor's judgements about my medical care.
7. I feel my doctor does not do everything he/she should about my medical care.
8. I trust my doctor to put my medical needs above all other considerations when treating my medical problems.
9. My doctor is well qualified to manage (diagnose and treat or make an appropriate referral) medical problems like mine.
10. I trust my doctor to tell me if a mistake was made about my treatment.
11. I sometimes worry that my doctor may not keep the information we discuss totally private.

Thom *et al*. (1999) tested the validity and reliability of the Trust in Physician Scale using a more general, primary care population of male and female adult patients. Adult patients (n=414) from twenty practices were recruited to participate in the prospective, six-month study. At the beginning of the study, participants completed the 11 item Trust in Physician Scale, as well as measures of demographics, preferences for care, and satisfaction with care received from the physician. Continuity, satisfaction with care, and reported adherence to treatment were measured at six-months. Reliability, construct validity, and predictive validity were evaluated using correlation coefficient and analysis of variance techniques.

The findings revealed that the Trust in Physician Scale had essentially the same high internal consistency as in the original study (Anderson & Dedrick, 1990). Trust, as measured by the Trust in Physician Scale, was also found to be a significant predictor of patients' satisfaction with care received from their physician, continuity with the same physician, and self-reported adherence to medical advice assessed at six-months. Moreover, trust remained a significant predictor of satisfaction, continuity and adherence after controlling for baseline satisfaction, suggesting that trust is conceptually distinct from satisfaction. Such findings have implications for the development of a scale to measure trust in the examination system; one might find that trust levels have predictive power in terms of continuity with an examination board, the number of enquiries about results and so on.

Trust in health care (institutional)

Van der Schee *et al* 2007 studied public trust in health care in three European countries. Data were collected in Germany, The Netherlands, England and Wales using a postal questionnaire. Items on trust in health care (as opposed to the general questions on overall confidence in today's health care system) were taken form a validated scale (Straten, Friele, & Groenewegen, 2002). Factor analysis of the public trust in health care scale revealed six dimensions: 1) patient-centred focus of health care providers; 2) macro-level policies concerning health care; 3) professional expertise of health care providers; 4) quality of care; 5) communication and provision of information, and 6) quality of cooperation between health care providers. The items pertaining to each dimension are given in Table 2:

Table 2: Dimensions of the public-trust-in-heath-care scale

How much do you trust that…	
Patient-centred focus	Patients are taken seriously Patients get enough attention Patients are listened to Doctors spend enough time on their patients
Macro-level policies	Cost-cutting does not disadvantage patients Patients will be able to pay for their own health care if they have to Medical help and patient care will not be compromised by the shortening of waiting lists Patients won't be the victim of rising costs of health care Waiting times are never too long
Professional expertise	Doctors can do everything Doctors know everything about all sorts of diseases New treatments are put into practice in the health care system The education and training of doctors in this country is one of the world's best
Quality of care	Patients always get the right dose of their medicine Doctors don't prescribe medicines too late Patients always get the right medicine A lot of care is taken to keep patients' medical information confidential in the health service Doctors always do enough tests Patients will always get the best treatment Doctors always make the right diagnosis
Communication and provision of information	Patients get sufficient information about the effects of their treatment Patients get sufficient information about the various treatments that are available The information given to patients is clear and understandable Patients get sufficient information about the cause of their problem Doctors discuss things fully with their patients
Quality of cooperation	Health care providers are good at co-operating with each other Patients aren't given conflicting information High levels of specialisation do not cause problems in the health care system

The public-trust-in-health-care scale provides a framework for the development of an instrument to measure public trust in the examination system. What's more, findings from studies that have used the public-trust-in-health-care scale have implications for the uses that a trust measure concerned with the examination system could be put. Statistical analysis conducted exclusively on the data from England and Wales, revealed that the key determinants of public trust were patient-centred care and levels of professional expertise (Calnan & Sanford, 2004). Consequently, it was suggested that policy makers concerned with the erosion of public trust in health care in the UK target these aspects of the system. If a similar measure were developed for the examination system, it could be used to identify where levels of distrust are high and the key determinants of public trust. Resources could then be targeted to increase public trust in examinations.

Conclusion

Theorists and opinion polls suggest that public trust in many institutions and professions is declining. It is imperative that examination boards responsible for setting, marking and grading examinations in the UK make efforts to measure and monitor public trust. Trust is fundamental to the credibility of the examination system. The cost of distrust could be high in both symbolic and monetary terms for examination boards. On the surface, observable trends such as increases in enquiries may seem likely indicators of change in public trust in the examination system. Indeed, there are empirical findings to suggest that enquiries data are sensitive to education scandals, which are thought to undermine public trust.

Much more notable, however, is the fact that whilst the number of re-mark requests has steadily increased over time, the proportion of re-marks leading to a mark change remained more or less stable (and if anything declined). Such findings suggest that many re-mark requests may be ill- founded and not based solely on concerns over quality of marking. It seems likely that public trust is a contributing factor in trends such as the increase in enquiries, but that other factors such as pressure on schools and colleges to perform also have an important part to play.

What is needed is an objective measure of public trust in the examination system. An erosion of trust in the medical profession has been acknowledged for some time, with researchers developing psychometric scales to measure both interpersonal and institutional trust. Such scales pose a way forward for the development of a scale to measure trust in the examination system and hint at its utility. As in the field of medicine, a trust scale in the examination system may have predictive powers and provide insight into how resources could be targeted to best maintain and enhance pubic trust. It is with this in mind, that the authors have scheduled qualitative research for early 2009 to lay the foundations for the construction and eventual validation of a public-trust-in-the-examination-system scale.

Lucy Billington & Rachel Taylor
September, 2008
e-mail: lbillington@aqa.org.uk
rtaylor@aqa.org.uk

References

Anderson, A.D., & Dedrick, R.F. (1990) Development of the Trust in Physician Scale: A measure to assess interpersonal trust in patient-physician relationships. *Psychological Reports*, v67 pp.1091-1100.

Calnan, M., & Rowe, R. (2004, November) Trust in health care: An agenda for future research. Paper presented at the Nuffield trust seminar.

Calnan, M. W., & Sanford, E. (2004) Public trust in health care: The system or the doctor? *Qual Saf Health Care*, v13 pp. 92-97.

Duff, D., Downing, P., & Skinner, G. (2003) *Trust in public institutions*. MORI Social Research Institute.

Giddens, A. (1990) *The consequences of modernity*. Stanford, CA: Stanford University Press.

Hall, M.A., Dugan, E., Zheng, B., & Mishra, A.K. (2001) Trust in physicians and medical institutions: What is it, can it be measured, and does it matter? *The Milbank Quarterly*, v79 n4 pp. 613-639.

Ipsos MORI. (2006) *GCSEs and A level: The experiences of teachers, students, parents and the general public.* Report commissioned by the Qualifications and Curriculum Authority.

Kelly, M. (2005) Public confidence in British official statistics. Retrieved 14 September 2007 from http://www.statistics.gov.uk/about/data/public_confidence/downloads/BristishOfficialStat.pdf

LaVeist, T.A., Nickerson, K.J. & Bowie, J.V. (2000) Attitudes about racism, medical mistrust, and satisfaction with care among African American and White cardiac patients. Medical Care Research and Review, v57 n1 pp. 146-61.

Marley, D. (2007) Independent schools rush to international GCSEs. *The Times Educational Supplement* 28 September, p. 15.

Meadows, M., & Taylor, R. (2008) Enquiries about results – an analysis of marking reviews. AQA report?

Mechanic, D. (1996) Changing medical organisation and the erosion of trust. *The Milbank Quarterly*, v74 n2 pp. 171-189.

Murphy, R. (2004) *Grades of uncertainty: Reviewing the use and misuses of examination results.* A report commissioned by the Association of Teachers and Lecturers.

O'Neill, O. (2002). *A question of trust: The BBC Reith Lectures 2002*. United Kingdom: Cambridge University Press.

O'Neill, O. (2005) *Assessment, public accountability and trust.* Retrieved 25 May 2007 from http://www.cambridgeassessment.org.uk

Parker, S.L., & Parker, G.R. (1993) Why do we trust our congressman? *Journal of Politics*, n55, pp. 442-453.

Pinot de Moira, A. (2000) *Enquiries upon results - The board giveth .....* AQA Research Committee paper, RC74

QCA (2007). *GCSE, GCE, GNVQ and AEA: Code of Practice.* Qualifications and Curriculum Authority.

Simone, J. V. (2007) The rise and fall of trust in the medical profession. *Oncology Times*, v29 n8 p5-6.

Slovac, P. (1993) Perceived risk, trust, and democracy. *Risk Analysis*, v13 pp.675:682

Straten, G.F., Friele, R.D., & Groenewegen, P.P. (2002) Public trust in Dutch health care. *Social Science and Medicine*, v55 pp.227-234.

Taylor, R. (2007b) *The impact of e-marking on enquiries after results*. AQA Technical Report, RPA_07_RT_TR_050.

Thom, D.H., Ribisl, K.M., Steward, A.L., Luke, D.A., & The Stanford Trust Study Physicians. (1999) Further validation and reliability testing of the Trust in Physician Scale. *Medical Care*, v37 n5 pp.510-517.

Van der Schee, E., Braun, B., Calnan., M., Schnee, M., & Groenewegen, P. P. (2007) Public trust in health care: A comparison of Germany, the Netherlands, and England and Wales. *Health Policy*, v81 n1 pp. 56-67.

Warmington, P. & Murphy, R. (2004) Could do better? Media depictions of UK educational assessment results. *Journal of Education Policy*, v19 n3 pp.285-299.

Wiliam, D. (1996) Standards in examinations: a matter of trust? *The Curriculum Journal,* v7 n3 pp.293-306