

MEASURING TRUST IN THE EXAMINATION SYSTEM: SOME INSIGHTS FROM THE MEDICAL PROFESSION AND ELSEWHERE

Lucy Billington

“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).

“Trust is a vulnerable and valuable commodity, vaunted in the marketplace, acknowledged in every profession, yet perniciously difficult to quantify.”

(Thom, Hall, & Pawlson, 2004, p.124).

SUMMARY

If the examination system is to maintain its credibility within society, it is imperative that awarding bodies and their regulator make efforts to understand, measure, and engender public trust. Trust-based research, primarily from the medical profession, is reviewed to gain some insight into how best to develop a measure of trust in the examination system. A potential strategy for developing a measure of trust in the examination system is outlined, and the utility of such an instrument hinted at.

WHY MEASURE PUBLIC TRUST?

Trust is recognised as being central to the credibility of the examination system. Wiliam (1996), for example, 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).

In May 2008, the Office of the Qualifications and Examinations Regulator (Ofqual) was launched. From the outset, the new independent regulator has made it clear that it is committed to promoting public confidence in GCSEs and A levels. Over the next two years, Ofqual will be undertaking a ‘health-check’ on the reliability of tests, examinations and teacher assessments (Garner, 2008). Other Ofqual initiatives include additional monitoring of awarding bodies’ customer service arrangements and use of new technology in examinations (Ofqual, 2008);

highlighting the importance of public trust in final examination results *and* the processes and services that underpin them. The founders of Ofqual have acknowledged that reform to GCSEs and A levels, and the introduction of the Diploma qualification mean that now, more than ever, efforts must be made to monitor and engender the public's trust (Curtis, 2008).

Examinations, and the agencies responsible for administering and grading them, have been subject to a number of scandals e.g. the Scottish Qualifications Authority crises in 2000, the Edexcel crises in 2001, the A level results crises in England in 2002 (McCaig, n.d), and the failure of the delivery of this year's SATs results (Curtis, 2008). These scandals, coupled with the increased scrutiny by the British press of examination standards (Murphy, 2004), and the technicalities of grading examinations (Warmington & Murphy, 2004), seem likely to have affected the public's trust. Indeed, current events and media images have been identified as key in shaping trust in public institutions (Mechanic, 1996). 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 (Slovak, 1993).

In the past, assessment agencies have measured satisfaction with the services they provide (Chamberlain & Taylor, 2006), but have overlooked trust. Recent research within the medical profession has shown trust to be related to, but conceptually distinct from, satisfaction: *"satisfaction looks backward based on past experience, while trust looks forward, an expectation of future behaviour"* (Thom *et al.* 2004). One study, concerned with the patient-physician relationship, even found trust to be better than satisfaction at predicting which patients remained with their physician and adhered to medical advice (Thom, Ribsil, Steward, Luke & The Stanford Trust Study Physicians, 1999). Such findings imply that institutions, such as awarding bodies and their regulator, could use trust measures as evaluation tools, either along with, or instead of, satisfaction measures (Barr, Vergun, & Barley, 2000). Furthermore, trust measurements could be salient in informing policy positions e.g. the information that Ofqual could generate on the reliability of examination results to promote public confidence.

A further rationale for measuring trust is that high trust levels engender greater efficiency within institutions and reduce the need for costly contingency arrangements (Mechanic, 1996). Tschannen-Moran and Hoy (2000, p. 550), for example, note that *"as trust declines, the costs of doing business increase because people must engage in self-protective actions and continually make provisions for the possibility of opportunistic behaviour on the part of others"*. Distrust in the operations of an awarding body could lead to an increase in enquiries after results and even a loss of entries, as centres seek more 'trustworthy' alternatives e.g. the International Baccalaureate. Understanding and measuring public trust in the examination system would clearly be beneficial to awarding bodies and their regulator: it would help sustain the legitimacy on which the system is founded and prevent awarding bodies from devoting excessive amounts of time to grievance procedures which detract from their broader educational goals.

The purpose of this paper is to review *some* measures of trust used in other fields, such as the medical profession, with a view to generating ideas about how one could go about measuring trust in the examination system e.g. examiners, examination results, examination standards, awarding bodies and their regulator.

TRUST MEASURES

Early measures of trust

Some of the earliest measures of trust were experimental and conceived of trust as an observable behaviour. Deutsch (1958), for example, measured trust as a rational choice to act cooperatively in a two-person mixed-motive game. Mixed-motive games are commonly referred to as prisoner's dilemma (PD) games, following Luce and Raiffa (1957) who originally published a game of the same name. Such games are characterised by an opportunity for mutually beneficial cooperation, but also the temptation to act competitively and exploit another person's cooperation. In general, players are said to trust their partner if they make a cooperative choice, because they are making themselves vulnerable to possible exploitation by their opponent. Players are considered untrustworthy if they take advantage of a partner's cooperative choice to increase their gains at the expense of their partner.

Dawes, van de Kragt and Orbell (1990) estimated that before they began to decline in popularity in the early 1980s, mixed-motive games featured in over two-thousand studies. Many of these studies were used to test alternative strategies for breaking the cycle of suspicion and tension that had resulted in the arms race between the United States and the Soviet Union (Tschannen-Moran & Hoy, 2000). However, the validity of inferring trust and suspicion on the basis of observing participants' choices in mixed-motive games is questionable. Several researchers have argued that participants in PD games treat the situation as a competitive game rather than an opportunity to be trusting or trustworthy (Dawes et al., 1990; Fisher & Brown, 1988; Rotter, 1967). Furthermore, most trust relationships are ongoing, and take place in a social context in which the consequences of untrustworthy behaviour extend beyond the immediate situation (Tschannen-Moran & Hoy, 2000).

Generalised trust

A standard question purporting to measure generalised trust features on many large-scale surveys including the American General Social Survey, the European Values Survey, and the World Values Survey. Researchers typically define generalised trust as "*a default expectation of other people's goodwill*" (Miller & Mitamura, 2003). The question reads as follows:

Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?

Most people can be trusted

Can't be too careful

(Some questionnaires add a third category: "Don't know" or "Depends")

The question first appeared in a paper by Rosenberg (1956), published in the *American Sociological Review*. It formed part of a five-question index called a "faith in people scale", which included items about basic human nature. Overtime, it has become a stand-alone question, which has been used in a variety of studies to explore the relationship between trust and such diverse topics as religiosity (Schoenfeld, 1978), marital stability (Yoder & Nichols, 1980), and democracy (Fukuyama, 1995). Responses to this single question have also been

used by researchers to suggest that social trust is declining in the U.S (Putnam, 1993; Robinson & Jackson, 2001) and to rank-order countries by trust levels (Inglehart, 2000).

Whilst the question has been widely used in trust-based research, it should be noted that it is usually preferable to measure a concept, such as trust, using multiple-items rather than one. De Vaus (2002), for example, offers five reasons why researchers should use multiple items: 1) they are needed to tap into the complexity of most concepts; 2) multiple-item scales avoid some of the distortions and misclassifications which can occur by using a single-item to measure complex concepts; 3) multiple items increase the reliability of the measure (responses to a single-item could largely be a function of the wording of the question); 4) multiple items allow for greater precision, e.g. distinguishing between small variations in respondents' possession of trust; and 5) they can be used to summarise information obtained from a number of questions into one variable, which can be analysed with ease.

Indeed, Miller and Mitamura (2003) have queried the validity of the single-item used to measure generalised trust. They argued that the ambiguous wording of the survey item has led to misrepresentations of actual levels of trust, particularly in a cross-cultural context. Rather than asking respondents to choose between trust and distrust (as implied by almost all studies citing the question under investigation), respondents choose between trust and caution when providing an answer. The distinction between trust and caution becomes particularly evident when the question under investigation is divided into its two components. "Would you say most people can be trusted?" asks for an appraisal of other people's trustworthiness. The second half of the question asks people whether they believe "you can't be too careful". Unlike the first half, this part enquires about one's own behavioural preference rather than that of others. Thus, it taps into the respondent's willingness to be vulnerable and take risks.

The two conceptually distinct halves of the question make it difficult to interpret responses (Miller & Mitamura, 2003). A risk averse individual may believe that people in general are trustworthy, but still be inclined to be careful when dealing with others. Consequently, Miller and Mitamura suggest that the dual presence of trust and caution in the question has profound implications for the interpretation of the findings of past studies. Studies claiming that one society is higher in trust than another, for example, may be misled by the greater necessity for caution in one society than another. To test the validity of the commonly used survey question, Miller and Mitamura (2003) conducted a survey of students at UCLA (n=293) and Hokkaido University (n=169), Japan. Based on the assumption that Americans are more cautious than Japanese (due to higher crime rates and a more open social structure), Miller and Mitamura hypothesised that:

"Japanese will score higher on trust when measured as a dichotomous trust versus caution variable, but Americans will score higher when it is measured as a trust versus distrust variable".

(Miller & Mitamura, 2003, pp.64-65)

Scales concerning trust, interpersonal trust and interpersonal caution were employed in the study. Cronbach's alpha and principal components analysis revealed that trust and caution were, indeed, discrete factors. Moreover, caution levels accounted for the main differences between Japanese and American respondents. Miller and Mitamura's (2003) study demonstrates the importance of testing the validity of any pre-existing tool designed to measure trust, and of using multiple indicators of trust in order to guard against conceptual difficulties associated with a single-item measure.

Levels of trust

Trust has not only been measured as a general attitude or view of human nature, but also at various levels of social relationships. Several researchers (Mechanic, 1996; Calnan & Rowe, 2004, 2005), for example, have distinguished between interpersonal and social (or institutional) trust. Trust in persons is intimate in nature, emanating from earlier experiences with family and other caretakers. Social trust, in contrast, is more cognitive and abstract, and typically founded on inferences about shared interests and common norms and values (Kasperson, Golding, & Tuler, 1992). Mechanic (1996) makes an important point:

“This distinction between interpersonal and social trust is a simple way of characterising a more complex reality, but empirical research rarely allows us to generalize about more subtle variations”.
(Mechanic, 1996, p. 173-174)

Both types of trust are multi-dimensional and some dimensions may be more or less present than others. Writing of the medical profession, Mechanic (1996) gives the example that a patient may trust the competence of their physician but be less convinced of his or her personal commitment to their welfare. Similarly, social trust can be categorised into patients' sense about institutions, in general, or their perceptions about its component parts with which they have had experience (Mechanic, 1996). To offer a hypothetical example from the field of education, an individual may be suspicious of the “education system” in England, but trusting of their local school. Furthermore, trust in institutions may be cognitively differentiated according to personnel or performance (Mechanic, 1996) (e.g. an awarding body may be trusted to grade vocational qualifications, but not GCSEs or GCEs). In short, we may trust individuals and institutions in some respects but not in others.

Parker and Parker (1993) argue that the concepts of interpersonal and social trust are correlated and mutually supportive. Trust in an institution transfers to unknown personnel – one assumes that a respected and efficient organisation will appoint its professionals appropriately and will supervise and monitor them. Conversely, trust in personnel often generalises to their organisation and impacts on one's willingness to give them custom (Mechanic, 1996). Schlackman (1989) argued that organisations often understand these relations and frequently make efforts to assure the public of the quality of their employees while monitoring their employees to guarantee high standards of performance. The complex relationship between interpersonal and social trust makes it appropriate to review studies that have sought to measure trust relations at both the macro and micro level.

Studies measuring trust at the macro level

Scales to measure public trust in professions, institutions or systems tend to be less developed than those to measure trust in individuals (Hall, Dugan, Zheng & Mishra, 2001). Certainly, a recent measure used by MORI (Duffy, Downing & Skinner, 2003) to explore the concept of trust in public institutions is somewhat disappointing. Specifically, the survey examined three areas of local public services: police forces, hospitals and local authorities. Almost 2,000 respondents, aged over 15 years, were interviewed. The main question in the survey purporting to measure trust asked respondents to state, in general, how much they trusted each of the public agencies. Response categories included “not at all”, “not very much”, “fair amount”, and “great deal”. Reported levels of trust were later used in a regression analysis to identify the key drivers of trust in each service. Admittedly, the survey did ask more in depth questions regarding attitudes towards specific aspects of service delivery (e.g. efficiency, management,

perceived ability to learn from mistakes), the impact of high profile mistakes on trust, and perceptions of senior public sector managers, but the use of only a single-item to determine actual trust levels seems unsatisfactory.

Simmons and Betts (2006) documented the development of a quantitative measure of public trust in official statistics, stating that the report could be used as a reference for other organisations attempting to develop a similar quantitative measure of public trust in official figures. The module of questions was first asked on the July 2004 Office of National Statistics (ONS) Omnibus Survey, and subsequently underwent revisions following the findings of the survey, focus groups and interviewer feedback. The design of the quantitative measure was informed by a literature review of previous research and the findings of focus groups. Trust in official statistics was conceived as having two dimensions, each encompassing a number of aspects. The first dimension was the quality of statistical outputs, including the methodologies used, the accuracy of the figures, the competence of statisticians and the independence of the organisation producing the statistics. The second dimension concerned the delivery of statistical outputs, in other words, their dissemination, presentation and use.

The final questionnaire included sections on trust in general as well as trust in official figures (see Simmons and Betts, 2006 for the question set used by interviewers). The opening section, which Simmons and Betts (2006) termed 'contextual information', asked respondents about the sources of information they use to inform their opinions on current affairs, and their general interest in politics. These questions were included so that any associations between respondents' trust in official figures and these behaviours could be investigated. Questions about trust in people generally, in institutions, and in senior public officials were also included to give an indication of whether individual respondents were generally a trustful or mistrustful person. Responses to these questions were intended to allow for an exploration of differences in trust in official statistics between those found to be generally trusting and those who were not. Interestingly, general trust in people was measured using the question found to be invalid by Miller & Mitamura, (2003) (see the above section on 'Generalised trust'), highlighting the importance of fully researching the reliability and validity of a measure before adopting it.

Originally, the sections on trust in official figures had included the following question:

"In general, do you think that official figures give a true picture of what is happening in Britain these days?"
(Simmons & Betts, 2006, p.13)

The response categories were 'generally true picture', 'sometimes true', 'sometimes untrue' and 'generally untrue picture'. The question was intended to provide an indication of trust in official figures generally, in which it was left to respondents to decide which dimensions of trust (quality of statistics and delivery of statistics) they wanted to consider in their answer. However, following analysis of the June 2004 Omnibus data and focus groups, the question was dropped. The complexity of the public's views meant that a single measure of overall trust in official statistics was of limited use. Rather, it was thought that *"the different dimensions of trust were best looked at separately, which was possible using...questions...on trust in specific statistical series and attitudinal statements about particular aspects of trust"* (Simmons & Betts, 2006, p. 13).

Specifically, the questions pertaining to trust in particular statistics asked about trust in: hospital waiting lists, road casualties, burglaries, the Census, inflation and internet use. The order in which questions were asked was randomised to prevent any possible order effects, so a

respondents' opinion of one series of statistics could not influence their opinion of another. The questions comprised of attitude statements aimed to gauge people's overall confidence in statistics, without reference to specific statistics. The two key dimensions of confidence in statistics, quality and delivery, were explored within the statements. Attitude statements included: 'Official figures are generally accurate', 'Official figures are produced without political interference', 'The Government presents official figures honestly when talking about its policies', and 'Official figures published by the Office for National Statistics are more trustworthy than those published by other government bodies' (Simmons & Betts, 2006, p.20). Response categories ranged from strongly agree to strongly disagree.

The process described by Simmons and Betts (2006) to develop a quantitative measure of trust in official statistics highlights important methodological considerations for anyone wishing to develop a similar measure of trust in the examination system. The findings of focus groups and cognitive testing were instrumental in determining the way in which trust was conceptualised and the final question set presented.

"Cognitive interviewing methods explore the mental processes respondents use to answer survey questions. Specifically, these methods focus on four stages: how respondents understand and interpret the questions, how they recall information that applies to the question, the judgements they make as to what information to use when formulating their answers, and how they respond to questions."
(Simmons & Betts, 2006 p. 4)

The terminology used in several questions, for example, was altered to ensure the correct interpretation by respondents (e.g. 'official figures' was deemed more appropriate than 'official statistics', as was 'UK government' more than 'UK parliament'). It seems likely that terms associated with the examination system e.g. 'awarding body' and 'examination standards' would be equally difficult for respondents to grapple with and the meanings attributed to them should be explored.

Similarly, the final question set relied quite heavily on numerical scales used to measure levels of trust, which needed to be amended following cognitive testing. Simmons & Betts (2006) argue that numerical scales are preferable to scales with response categories such as, 'don't trust at all', 'trust a little', 'trust a lot', 'trust completely', because they are less subjective and provide a continuum. Initially, the scale of 1-10 was used, the rationale being that without a midpoint respondents would be unable to 'sit on the fence'. However, respondents frequently misinterpreted 5 to be the midpoint anyway. There was also evidence that respondents wanted to be able to answer 'zero', indicating no trust at all. The lowest extreme of 1 on the original scale was thought to indicate some trust, albeit a low amount. The scale used in the final questionnaire was 0-10.

Numerous studies have shown that response category formats are an important element in survey construction (e.g. Alwin, 1997; DuBois & Burns, 1975). Schwartz and Hippler (1987) suggested that respondents utilize the information in response categories, including phraseology, when deciding how to answer an item. Rating scales can be either bipolar (e.g. distrust -2 -1 0 +1 +2 trust) or unipolar in nature (e.g. Trust 1 2 3 4 5). The number of scale points used, numeric scale labels, and verbal stimuli have all been shown to affect the types of answers obtained (Peterson, 2000). Furthermore, variations in response categories have been found to interact with the mode of administration (Rockwood, Sangster & Dillman, 1997). The findings of Simmons and Betts (2006) highlight the importance of consulting the

literature on scale construction before developing a measure of trust in the examination system, and of thoroughly pre-testing items.

Another area of research that may be pertinent to public trust in the examination system is public trust in the medical profession. Efforts have not only been made to measure public trust in particular health care regimes (Calnan & Sanford, 2004), but also to draw comparisons in trust levels between different European countries (van der Schee, Braun, Calnan, Schnee & Groenewegen, 2007). Van der Schee *et al.* collected data on public trust in health care in Germany, The Netherlands, England and Wales, using a postal questionnaire. The questionnaire originated in the Netherlands, and was centred on possible dimensions of trust that had emerged from qualitative research (Straten, Friele & Groenewegen, 2002).

The specific items on trust in health care (as opposed to the general questions on overall confidence in today's health care system) were taken from a validated scale (Straten *et al.* 2002). Some of the questions in the public-trust-in-health-care scale intentionally used words such as 'always' and 'everything' e.g. *Patients always get the right medicine, Doctors can do everything*, to prevent clustering of answers in one category. 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 Appendix A.

The public-trust-in-health-care scale provides a framework for the development of an instrument to measure public trust in the examination system. Moreover, 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 to. Van der Schee *et al.* (2006), for example, found that differences in levels of public trust in health care systems were predominately due to cultural differences in the propensity to place trust, casting doubt over the utility of such measures in cross-cultural comparisons. Statistical analysis conducted exclusively on the data from England and Wales revealed that the key aspects of public assessment of 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 in England, it could be used to identify the dimensions on which levels of distrust are high and the key determinants of public trust. Resources could then be targeted to the appropriate dimensions so as to increase public trust in examinations.

Studies measuring trust at the micro level

Trust-based research in the medical profession does not only provide examples of trust being measured at the macro level (the level of the health care system), but also at the micro level (patients' trust in their physician). Thom and Campbell (1997), for example, conducted qualitative research exploring the patient-physician relationship. The aim of the research was to gain an understanding of how patients perceive trust of a physician, and how patients associate physicians' behaviours with their perceptions of trust. Four focus groups were conducted with a total of twenty-nine patients. Patients were recruited from three diverse practice sites in the San Francisco Bay Area to provide a broad range of experiences. The working definition of trust used in the focus groups was "*the patient's confidence that the physician will do what is best for the patient*" (Thom & Campbell, 1997, p.171). Within each group, participants were asked to

describe situations they had experienced that led them to trust a physician, and situations that had caused them to lose, or not to establish trust.

The focus groups were audio-recorded, transcribed, and coded using principles of grounded theory (Corbin & Strauss, 1990). The reported experiences of participants were coded into nine broad categories, each with several subcategories. Seven out of the nine categories related to the physician-patient interaction: 1) thoroughness in evaluation; 2) understanding the patient's individual experience; 3) caring; 4) providing appropriate and effective treatment; 5) communicating clearly and completely; 6) partnership building; and 7) honesty/respect for the patient. Two further categories were predisposing factors (e.g. age, sex, professional appearance), and structural/staffing factors (e.g. courtesy of office staff, on-call arrangements). The role of structural factors in the formation of a patient's trust in their physician again highlights the symbiotic relationship between interpersonal and institutional trust. Thom and Campbell (1997) concluded that the nine trust-related categories identified in their research point to ways in which physicians could be more effective in building and maintaining trust. Similar insights could be gained via qualitative research in the area of trust in the examination system. Moreover, the results of qualitative inquiry could be used to generate items for a quantitative tool to measure trust in the examination system (Goudge and Gilson, 2005).

Many other theorists within medicine have attempted to conceptualise patients' trust in interpersonal relationships (Anderson and Dedrick, 1990; Mechanic and Meyer 2000; Mechanic 1996; Lynn-McHale and Deatrck, 2000). Some focus on particular acts or obligations, while others stress the importance of personal attributes or character traits. Despite identifiable differences in conceptual schemes, Hall *et al.* (2001), argue that there is, nevertheless, extraordinary commonality among them. Specifically, Hall *et al.* (2001) posit a five-part configuration: 1) fidelity, 2) competence, 3) honesty, 4) confidentiality, and 5) global trust. *Fidelity* concerns the pursuit of a patient's best interest and not taking advantage of his or her vulnerability. It consists of caring, respect, advocacy and avoiding conflicts of interest. *Competence* refers to avoiding errors and producing the best results possible. The *honesty* dimension entails telling the truth and avoiding intentional falsehoods, such as blatant lies, half-truths, or deception by silence. *Confidentiality* entails the protection and appropriate use of sensitive or private information. The final dimension is *global* trust. Hall *et al.* (2001) argue that trust has a significant component that is irreducible – what one might call the "soul of trust". This *global* dimension is intended to capture the more holistic facet of trust. This five-part configuration may prove to be a reasonable template for dimensions of trust in the examination system. Certainly, the one-word descriptors are commonly used words within the area of educational assessment.

Thom *et al.* (1999) tested the validity and reliability of one measure of patient-physician trust, namely the Trust in Physician Scale developed by Anderson and Dedrick (1990). The Trust in Physician Scale is an 11 item, single score scale (see Appendix B), 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 0.90 and 0.85 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 *et al.*, 1999). However, since all data were cross-sectional, there was no information on the test-retest reliability of the scale or the predictive validity of the scale.

Using a more general, primary care population of male and female adult patients, Thom *et al's.* (1999) study was designed to: 1) examine item characteristics, internal consistency, and test-retest reliability, 2) investigate construct validity with respect to patient satisfaction, physician humaneness, patient-physician relationship characteristics, and general interpersonal trust, and 3) evaluate predictive validity at six months for satisfaction with care from the physician, continuation with the same physician, and self-reported adherence to prescribed medication. 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 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 was also found to increase with the length of the relationship, and was highest among patients who actively chose their physician, who preferred greater physician involvement, and who expected their physician to care for a larger proportion of their ailments. 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. 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 after results and appeals.

CONCLUDING REMARKS

Previous research that has aimed to conceptualise and measure trust is informative in terms of how best to develop a trust-measure for the examination system. The multidimensionality of trust makes it difficult to quantify. However, researchers from other fields have attempted to do so by first conducting qualitative research to gain an understanding of how individuals perceive their trust in a particular trust object e.g. official figures, health care systems, and physicians. It seems likely that the examination system will have its own unique dimensions of trust, such as marking accuracy, test validity, fairness and quality control. Trust in such dimensions is likely to vary (and may even vary by stakeholder group), with interpersonal factors (e.g. personal experience of dealing with awarding body staff) and structural factors (e.g. the introduction of new qualifications, such as the Diploma) interacting in the process of trust formation. It is suggested that qualitative research be conducted with key stakeholders (parents, students, teachers, employers, and admission officers for higher education) in the examination system to explore their perceptions of trust.

A quantitative measure of public trust in the examination system could be based upon the dimensions identified in focus groups. Measures of generalised trust should be incorporated into any instrument in order to enable comparisons to be made between people who are typically trustful and mistrustful. However, caution should be exercised when adopting measures used in other research. Cognitive testing of an initial question set could be used to ensure the correct interpretation of items by respondents. As with measures of trust employed in other fields, it would be imperative to assess the reliability and validity of a scale to measure trust in the examination system:

“A scale has validity if it properly represents the theoretical construct it is meant to measure. A scale has reliability if repeated measurements under the same circumstances tend to produce the same results”.
(DeCoster, 2005, p.1)

Constructing and validating a trust-in-the-examination-system scale would be a considerable undertaking. Once developed, however, it is almost certain that such a scale would be invaluable to awarding bodies and their regulator. It could be used to identify differences in trust levels between stakeholder groups, to target resources to enhance public trust, and as a tool to predict the behaviours of users of examinations.

Lucy Billington
October 2008

REFERENCES

- Alwin, D.F. (1997) Feeling thermometers versus 7-point scales. *Sociological Methods and Research*, n25 pp.318-40.
- 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*, n67 pp.1091-1100.
- Barr, D., Vergun, P., & Barley, S.R. (2000) Problems in using patient satisfaction data to assess the quality of care of primary care physicians. *Journal of Clinical Outcomes Management*, n7 pp.19-24.
- Calnan, M., & Rowe, R. (2004, November) *Trust in health care: An agenda for future research*. Paper presented at the Nuffield trust seminar.
- Calnan, M., & Rowe, R. (2005, December) *Trust relations in the 'new' NHS: Theoretical and methodological challenges*. Paper presented at the 'Taking Stock of Trust' E.S.R.C Conference, London School of Economics.
- Calnan, M. W., & Sanford, E. (2004) Public trust in health care: The system or the doctor? *Qual Saf Health Care*, v13 pp. 92-97.
- Chamberlain, S., & Taylor, R. A. (2006) *Customer satisfaction survey, 2006*. AQA Research Report, RPA_06_SC_RP_044.
- Corbin, J., & Strauss, A. (1990) Grounded theory research: procedures canons, and evaluative criteria. *Qualitative Sociology*, n13 pp.3-21.
- Curtis, P. (2008) Watchdog admits public trust hit by fiasco. *The Guardian*, 1 September.
- Dawes, R. M., van de Kragt, A. J. C., & Orbell, J. M. (1990) Cooperation for the benefit of us-Not me or my conscience. In J. Mansbridge (Ed.), *Beyond self-interest*. Chicago: University of Chicago Press.
- DeCosta, J. (2005) *Scale construction notes*. Retrieved 8 October 2008 from <http://www.stat-help.com/notes.html>.
- Deutsch, M. (1958) Trust and suspicion. *Journal of conflict resolution*, n2 pp.265-279.

- De Vaus, D. (2002) *Surveys in Social Research, 5th Edition*. Australia: Routledge.
- DuBois, B., & Burns, J.A. (1975) An analysis of the meaning of the question mark response category in attitude scales. *Educational and Psychology Measurements*, n35 pp.869-884.
- Duffy, D., Downing, P., & Skinner, G. (2003) *Trust in public institutions*. MORI Social Research Institute.
- Fisher, R., & Brown, S. (1988) *Getting together*. Boston: Houghton Mifflin.
- Fukuyama, F. (1995) *Trust: Social virtues and the creation of prosperity*. New York: Free Press.
- Garner, R. (2008) Investigation to test integrity of exam standards. *Independent*, 11 August.
- Goudge, J., & Gilson, L. (2005) How can trust be investigated? Drawing lessons from past experience. *Social Science & Medicine*, n 61 pp.1439-1451.
- 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.
- Inglehart, R. (2000) Culture and democracy. In L. E. Harrison & S. Huntington (Eds.), *Culture matters*. New York: Basic Books.
- Kasperson, R.E., Golding, D., & Tuler, S. (1992) Social distrust as a factor in siting hazardous facilities and community risks. *Journal of Social Issues*, n48 pp.161-187.
- Luce, R. D., & Raiffa, H. (1957) *Games and decisions: Introduction and critical survey*. New York: Wiley.
- Lynn McHale, D.J., and Deatricks, J. A. (2000) Trust between family and health care provider. *Journal of Family Nursing*, v6 n3 pp.210-230.
- McCaig, C. (n.d.) *The politics of exam crises*. Retrieved 4 November 2006 from <http://www.aqa.org.uk/support/iaea/papers/mccaig.pdf>.
- Mechanic, D. (1996) Changing medical organisation and the erosion of trust. *The Milbank Quarterly*, v74 n2 pp. 171-189.
- Mechanic, D., & Meyer, S. (2000) Concepts of trust among patients with serious illness. *Social Science and Medicine*, v51 pp.657-668.
- Miller, A. S. & Mitamura, T. (2003) Are surveys on trust trustworthy? *Social Psychology Quarterly*, v66 n1 pp.62-70.
- Murphy, R. (2004) *Grades of uncertainty: Reviewing the use and misuses of examination results*. A report commissioned by the Association of Teachers and Lecturers.
- Oqual. (May, 2008) *Getting it right: Promoting confidence in GCSE and GCE exams in England, summer 2008*. Ofqual report.
- Parker, S.L., & Parker, G.R. (1993) Why do we trust our congressman? *Journal of Politics*, n55, pp. 442-453.
- Peterson, R.A (2000) *Constructing effective questionnaires*. London: Sage Publications, Inc.

- Putnam, R. (1993) The prosperous community: Social capital and public affairs. *The American Prospect*, Spring pp.35-42.
- Robinson, R. V., & Jackson, E. F. (2001) Is trust in others declining in America? An age-period-cohort analysis. *Social Science Research*, v30 pp.117-45.
- Rockwood, T.H., Sangster, R.L., & Dillman, D.A. (1997) The effect of response categories on questionnaire answers. *Sociological Methods and Research*, n26 pp.118-140.
- Rosenberg, M. (1956) Misanthropy and political ideology. *American Sociological Review*, v21 pp.690-95.
- Rotter, J. B. (1967) A new scale for the measurement of interpersonal trust. *Journal of Personality*, v35 pp.651-665.
- Schlackman, N. (1989) Integrating Quality Assessment and Physician Incentive Payment. *Quality Review Bulletin*, v15, n8, pp. 23437.
- Schoenfeld, E. (1978) Image of man: The effect of religion on trust. *Review of Religious Research*, v20 pp.61-67.
- Schwarz, N., & Hippler, H.J. (1987) What response scales may tell your respondents: Informative functions of response alternatives. In H.J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology*. New York: Springer-Verlag.
- Simmons, E. & Betts, P. (2006) *Developing a quantitative measure of public confidence in official statistics*. Survey Methods Division, Office for National Statistics.
- Slovak, 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.
- Thom, D.H., & Campbell, B. (1997) Patient-physician trust: an exploratory study. *Journal of Family Practice*, v44 n2 pp.169-176.
- Thom, D.H., Hall, M.A., & Pawlson, G.P. (2004) Measuring patient's trust in physicians when assessing quality of care. *Health Affairs*, v23 n4 pp.124-132.
- 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.
- Tschannen-Moran, M. & Hoy, W. H. (2000) A multidisciplinary analysis of the nature, meaning and measurement of trust. *Review of Educational Research*, v70 n4 pp.547-593.
- 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.
- William, D. (1996) Standards in examinations: a matter of trust? *The Curriculum Journal*, v7 n3 pp.293-306
- Yoder, J. D., & Nichols, R. C. (1980) A life perspective comparison of married and divorced persons. *Journal of Marriage and the Family*, May pp.413-419.

APPENDIX A

Dimensions of the public-trust-in-health-care scale

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

APPENDIX B**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.