Virtuous Medical Practice: Research Report

Online Appendices

Working Paper: Designing the Jubilee Centre for Character and Virtues Moral Dilemma Tests

Working Paper:
Designing the Jubilee Centre for Character and Virtues Moral Dilemma Tests[1]

1. **Introduction**

In this working paper, we will give an overview of past attempts to understand morality in medicine using the moral dilemma approach, before turning our attention to how we used moral dilemmas in our study to understand virtue or moral character in medicine. Lastly, we will outline results from our study that for the first time attempted to design and analyse moral dilemmas in medicine from a virtue-perspective.

 2. **Measuring morality in medicine**

To anyone working in moral psychology, - education or - philosophy, Lawrence Kohlberg's (1981; 1984) work on moral development will be deeply familiar. Kohlberg posits that moral development from childhood into adulthood unfolds through a sequence of modes of moral thinking, from thinking of right action in terms of self-interest (pre-conventional thinking) to thinking of right action in terms of what is socially desirable (conventional thinking) to thinking of right action in terms of a set of moral principles that one can rationalise as universally valid (post-conventional thinking). The last form of moral thinking is clearly *deontological* or *principles*-based and, for Kohlberg, it constitutes the highest form of moral thinking. Kohlberg's Moral Judgement Interview (MJI) was designed to capture the extent to which individuals think post-conventionally and a paper-and-pencil version of the MJI, the Sociomoral Reasoning Measure (SRM) (Gibbs and Widaman, 1982) was later developed.

Kohlberg's research programme came to be very influential in the study of professional ethics. Especially Kohlberg's colleague James Rest's (1999) four component model of moral development played a crucial role in the field. According to Rest's model in moral psychology, moral action results from a combination of

- moral sensitivity
- moral motivation
- moral judgement
- moral character

The emphasis in this tradition is on studying *moral schemas* – typical, acquired and routine ways that moral subjects have of responding to moral problems. In order to study the moral schema that a person brings to a task, researchers in the four-component tradition devised a test of moral development, the Defining Issues Test (DIT). As originally conceived, the DIT was designed to measure only moral judgement schemas and not to measure the other three of Rest's four components. A second version of the DIT (the DIT-2) was later developed.

According to Kohlbergian scholars, it is conscious reasoning about what to do that drives moral action. As Kohlberg himself put it, '[h]e who knows the good chooses the good' (1981: 189). Working at the same time as Kohlberg, Blasi (1980), however, raised serious doubts about whether or not this is true. Blasi conducted a meta-analysis of studies on the link between moral cognition and moral action and found few correlations between Kohlbergian stages of moral reasoning and actual moral

---

[1] This working paper is adapted in part from Kotzee, B. and Ignatowicz, A. (manuscript) 'Measuring Virtue in Medicine', currently under consideration for publication.

behaviour. Partly in response to findings like these, the 'intuitionist' movement in moral psychology (associated with Haidt and others), stresses that it may not be moral reasoning, but, instead, moral *emotion* that drives moral action. According to the intuitionists, the relationship between moral reasoning and moral emotion should be reversed. It is not moral reasoning that plays the largest part in what the actor decides to do, but moral emotion. Haidt (2001) reports on studies showing that moral decisions are much more automatic and quick than previously assumed and that moral reasoning is not the well-spring of moral action, but are little more than *post-hoc* rationalisations of a position that an agent already holds. As Haidt holds, 'intuition comes first, strategic reasoning second' (2013: 286). Based on these insights Haidt and others have formulated what is called moral foundations theory (MFT) as an alternative to Kohlbergian theories of moral development and have developed a questionnaire, the Moral Foundations Questionnaire (MFQ) to assess respondents' moral orientation.

Kristjansson (2015: chapter 3) questions whether MFT should be seen as a virtue-approach to the empirical study of morality; in several respects, he holds, it is not. In a number of respects, however, the debate between rationalistic and intuitionist approaches to moral psychology clears the way for virtue-based approaches and sets the background to this study.

3.   Assessment tools for measuring ethics and professionalism in medicine

In order to understand how moral psychology has influenced the empirical study of medical ethics, we undertook a literature review to establish what are the most frequently used psychometric measures of professional ethics in medicine. We searched MEDLINE and PUBMED using combinations of the following keywords: 'assessment', 'measurement', 'ethics', 'professionalism' and 'virtue'[2]. Twenty four publications were retrieved from the literature and we used the reference lists of these to identify a further 47 studies. We also identified four previous reviews of assessment tools for measuring ethics and professionalism in medicine, with the most recent one from 2006 (Self and Baldwin, 1994; Bebeau, 2002; Baldwin and Self, 2005; Bebeau, 2006). Because these four articles review the empirical study of doctors' ethics exhaustively, we focussed in our own analysis on studies since 2006.

We categorised all studies according to their theoretical approach and found that – both before 2006 and since – most studies are cognitive in their orientation. We found that the most frequently used tests of medical ethics are the MJI, SRM, DIT or a closely related test. With only a few exceptions, empirical studies of medical ethics of all of these sorts have departed from a cognitive or rationalist perspective. By contrast, we could identify only one study from an intuitionist perspective (see discussion of Leffel *et al.* below) and could find only one empirical study from a virtue-perspective (Schulz *et al.,* 2013) even though ample theoretical discussion of medical ethics from a virtue-perspective exists (Gardiner, 2003).

3.1 How effective are existing ethics measures in medicine?

Self and Baldwin (1994), Baldwin and Self (2005) and Bebeau (2002, 2006) survey the literature on ethics assessments in medical education[3] and conclude that measures of general moral reasoning in the Kohlbergian tradition are the most influential in shaping thinking about how to assess doctors' ethics empirically. According to Bebeau (2002: 289), one important research question is whether professional curricula have a positive effect on moral development. Disappointingly, general professional curricula – or simply the experience of having taken a professional degree in medicine

---

[2] Criteria used to include a measure/assessment/test were: (1) pertains to medical education and/or some aspect of professionalism in medicine, and (2) empirical evidence of its use.

[3] …and professional education more broadly, in the case of the papers by Bebeau.

or a similar area – seem to contribute little to overall moral development. (Bebeau, 2002: 273 – 4). Bebeau (2002) discusses six studies using the DIT in medicine of which only one finds a moderate improvement in moral reasoning during the course of medical school; the other five studies found no improvement. More recently, Patenaude *et al*. found a statistically significant *decline* in moral development amongst 92 Canadian medical students as measured using the Moral Judgement Interview. (Patenaude, Niyonsenga, & Fafard, 2003). Helkama *et al.* (using the MJI) report similar results at a Finnish medical school (Helkama et al., 2003). Hren *et al.* (2011) even found that medical students regress from the post-conventional to the maintaining norms schema after entering the clinical part of the curriculum.

While experience of professional education in general does not seem to contribute to moral development, however, specific interventions may. Bebeau reports on seven studies investigating the effects of specific interventions – mostly dedicated ethics courses – on students' moral development. Of these, five reported significant pre/post test gains in moral development as a consequence of taking such courses. (Bebeau, 2002: 275) Baldwin and Self (2005) and Bebeau (2002; 2006) all hold that there is a positive relationship between moral reasoning in medicine and clinical competence and performance.[4]

Next to measures of general moral reasoning, some profession-specific measures of moral reasoning (containing dilemmas only relevant to one profession) exist. The best example is the 'Dental Ethical Reasoning and Judgement Test' (DERJT). In a 2010 survey of US Dental Schools, thirteen of the fifty-six dental schools which participated (i.e. 24.1%) said that they were using the DERJT in their admission procedures (Lantz, Bebeau et al. 2011).

Given the success of the DERJT in the field of dentistry, there have been attempts to design a similar measure for medicine. Caldicott, Faber-Langendoen, Bebeau, and Thoma (2008/2010) report on the development of the Medical Ethical Reasoning and Judgment Test (MERJT). The MERJT was piloted with 44 doctors and medical students in 2007-2008. All participants also took the DIT-2. Caldicott and colleagues found that the MERJT can distinguish between study participants and physician-ethicist experts in their ranking of best and worst action justifications and total best and worst choices and that the MERJT showed little redundancy with the DIT-2 (meaning that the MERJT measures unique information) (Pearson correlation = 0.32).[5]

Another intermediate concept measure for medicine is the Medical Intermediate Concept Measure of Ethical Reasoning (MD-ICM) (Pinijphon, 2009) Pinijphon studied the reliability and validity of the measure with 627 medical students at seven medical schools in Thailand. She found that the MD-ICM was able to distinguish medical students by experience, with more advanced students achieving better scores. Furthermore, she found a correlation between MD-ICM scores and the DIT-2 and found that direct instruction in medical ethics improves MD-ICM scores.

While most measures in the Kohlbergian and post-Kohlbergian traditions focus on moral reasoning, some tests also exist of moral sensitivity and moral reasoning. Bebeau, for instance, has developed two measures focussing on separate components of Rest's four component model: the Dental Ethical Sensitivity Test (DEST) was developed as a test of ethical sensitivity in dentistry and the Professional Role Inventory (PROI) was developed as a test of the professional's identification with their role. For Bebeau, this latter interest in identification with one's role performs the function of measuring one's moral *motivation*.

---

[4] Sheehan *et al.* (1980), Baldwin *et al.* (1996) and Adamson *et al.* (2000) reach similar conclusions.
[5] With thanks to Catherine Caldicott.

3.2 Why is a virtue approach to moral dilemmas needed?

While past work on the development of moral reasoning in the professions is impressive and some work has also been done on measuring moral sensitivity and moral motivation in the professions, measurements of reasoning, sensitivity and motivation *in isolation* do not capture what it is for a person to be virtuous *in the round*.

For Aristotle, a virtue is a trait of a person's character (*hexeis*); it is a developed, but (once developed) stable trait that influences the way a person acts from a moral point of view. As we have already seen (see section 4, above) the virtues that doctors identify as important are fairness, honesty, kindness, perseverance and teamwork. Aristotle held that each character trait of this sort consists of a different set of developed tendencies that a person has to do with the following things:

- to recognise or perceive moral situations correctly (to be sensitive to what is at stake in a situation)
- to respond emotionally to that situation in the right way (this may include being dispassionate in the right circumstances),
- to think well about what to do in the situation (either to know how to act or to reason appropriately about how to act)
- and to be motivated strongly enough to carry the right action through.[6]

All of these processes of sensitivity, emotion, reasoning and motivation need to be coordinated in action with a certain manner or style and virtuous action consists in all of these elements operating in harmony in a specific situation.

What we sought to do in the dilemma element of the study was to *build* on well-established work on moral reasoning and to design an *integrated* way to study the influence of virtue-considerations on how medical students and doctors respond to moral dilemmas in medicine.

Importantly, our study was *not* aimed at designing a validated psychometric *test of* or *measure for* virtue. Rather, the study used approaches from existing psychometric tests to *survey* how doctors justify, reason about or understand moral practice in their profession. In short, we wanted to understand which virtues influence doctors' moral decision-making and how these virtues interact

(1) with one another and
(2) with thinking in terms of rules and consequences.

Based on this survey work, psychometric measures that can be useful in medical selection, education and regulation may in future be designed; however, the design of such a measure was not one of the aims of this study.

4. The Jubilee Centre for Character and Virtues Moral Dilemma Tests

In designing the Jubilee Centre dilemmas, we modelled our approach on the 'Situational Judgment Test' approach in medical education and assessment. 'Situational Judgement Tests' (SJT's) are increasingly used in the selection and training of doctors in the UK. SJT's are tests of professionals' likely performance on representative tasks and contain hypothetical work-based scenarios in response to which the test-taker must judge possible courses of action. Advocates of SJT's hold that well-designed SJT's show good reliability and validity as a selection tool. They are capable of

---

[6] See Curren and Kotzee (2014), Fowers (2014) and Kristjansson (2015) for discussion of the psychology of virtue.

predicting job performance and fare better than other possible assessment tools (e.g. interviews, IQ tests or personality tests) in this regard. (Patterson *et al.*, 2009; Patterson and Ashworth, 2011)[7] In the UK, selection of medical graduates for foundation year training makes use of SJT's from 2013.

When constructing SJT's the following procedure is normally followed. First a thorough job-analysis is carried out in order to ensure that the tasks that feature on the SJT are representative. Next items are written – typically in panels. Lastly, a 'concordance' panel is convened. Concordance panels are typically expert clinicians (most often medical educators who are experienced SJT writers themselves). Panel members take the SJT's themselves as if they were respondents and their responses are analysed for agreement or 'concordance'. Items for which there exist high levels of consensus as to how they should be answered go forward for inclusion in the SJT, while those that do not are dropped. (Patterson *et al.*, 2009; Bergman *et al.,* 2006)

We constructed six representative ethical dilemmas that may arise in medicine. The design of this element of the study was conducted in two expert panel phases:

- Phase 1: A group of experts in medical education (N = 12) was convened to design six ethical situational judgement tests that would represent a range of ethical problems that a doctor could face in daily practice.
- Phase 2: A second group of experts – this time all medical educators who were recognised by their peers as experts (N = 15) – took the situational judgement tests as if they were participants. The experts were instructed to score the reasons that one could give for taking each course of action from 1 to 6, giving an indication as to the expert doctor's own moral reasoning when faced with one of the situations devised by the first panel.

The dilemmas were posed in such a way as to uncover what it was about a person's character or values that influenced them in their decision.

The degree of agreement/disagreement - or 'concordance' – between experts can be expressed precisely using Kendall's Coefficient of Concordance (or Kendall's W). Across all six dilemmas (and across two sets of ranked courses of action for each dilemma), concordance between experts varied from .450 to .869 (with 0 being no concordance and 1 being complete concordance).

Dilemma	Course of Action 1	Course of Action 2
1	.727	.491
2	.869	.500
3	.694	.812
4	.500	.535
5	.450	.506
6	.587	.463

Figure 3: concordance between expert rankers[8]

Friedman's Test was carried out to test for the significance of these results and indicated that the results are significant at the .005 level.

---

[7] Critics hold that SJT's show little incremental validity over general intelligence and personality. (Bertua *et al.*, 2005)

[8] Thanks to Ian Davison.

Bibliography

Baldwin, D. and Self, D. (2005) 'The Assessment of Moral Reasoning and Professionalism in Medical Education and Practice', in Stern, D. (ed.) *Measuring Medical Professionalism*, Oxford: Oxford University Press.

Bebeau, M. (2002) 'The Defining Issues Test and the Four Component Model: Contributions to Professional Education', *Journal of Moral Education,* vol. 31, no. 3, pp. 271-95.

Bebeau, M. (2006) 'Evidence-Based Character Development', in: Kenny, N. and Shelton, W. *Lost Virtue: Professional Character Development in Medical Education: Advances in Bioethics, vol. 10*, Oxford: JAI Press.

Bebeau, M., Born, D. and Ozar, D. 1993. 'The Development of a Professional Role Orientation Inventory' *Journal of the American College of Dentists* 60 (2), 27 – 33.

Bebeau, M., Rest, J. and Yamoor, C. 1985. 'Measuring Dental Students' Ethical Sensitivity' *Journal of Dental Education* 49 (4), 225 – 35.

Bebeau, M. and Thoma S. 1999. '"Intermediate" concepts and the connection to moral education' *Educational Psychology Review* 11, 343–360.

Bergman, M., Drasgow, F., Donovan, M., Henning, J. and Juraska, S (2006) 'Scoring Situational Judgment Tests: Once You Get the Data, Your Troubles Begin', *International Journal of Selection and Assessment,* vol. 14, no. 3, pp. 223-35.

Bertua, C., Anderson, N., and Salgado, J. (2005) 'The Predictive Validity of Cognitive Ability Tests: A UK Meta-analysis', *Journal of Occupational and Organisational Psychology*, vol. 78, no. 3, pp. 387-409.

Blasi, A. 1980. 'Bridging Moral Cognition and Moral Action: a review of the literature' *Psychological Bulletin* 88 (1), 1 – 45.

Borgstrom, E., Cohn, S. and Barclay, S. (2010) 'Medical Professionalism: Conflicting Values for Tomorrow's Doctors', *Journal of General Internal Medicine,* vol. 25, no. 12, pp.1330-1336.

Born, D., Bebeau, M. and Rozmenoski, S. 1995. 'Analysis of the Professional Role Orientation Inventory' *Journal of Dental Research* 74, 162 – 62.

Caldicott, C., Faber-Langendoen, K., Bebeau, M. and Thoma, S. (2010) 'Assessing Moral Reasoning in Medical Training and Practice: A Pilot Study', *Cambridge Consortium for Bioethics Education*, New York City, USA.

Calman, K. and Downey, R. (1987) 'Practical Problems in the Teaching of Ethics to Medical Students', *Journal of Medical Ethics*, vol. 13, pp. 153-156.

Campbell, A., Chin, J. and Voo, T-C. (2007) 'How Can We Know that Ethics Education Produces Ethical Doctors?', *Medical Teacher,* vol. 29, no. 5, pp. 431-436.

Cawley, M., Martin, J. and Johnson, J. 2000. 'A Virtues Approach to Personality' *Personality and Individual Differences* 28 (5), 997 – 1013.

Cowley, C. (2005) 'The Dangers of Medical Ethics', *Journal of Medical Ethics*, vol. 31, no. 12, pp. 739-742.

Coulehan, J. (2005) 'Today's Professionalism: Engaging the Mind But not the Heart', *Academic*

Medicine, vol. 80, no. 10, pp. 892-898.

Curren, R. and Kotzee, B. 2014. 'Can Virtue Be Measured?' *Theory and Research in Education* vol. 12, no. 3, pp. 266 – 82.

 Eckles, R., Meslin, E., Gaffney, M, and Helft, P. (2005) 'Medical Ethics Education: Where Are We? Where Should We Be Going? A review', *Academic Medicine*, vol. 80, no. 12, pp. 1143-1152.

Fowers, B. 2005. *Virtues and Psychology: pursuing excellence in ordinary practices*. Washington, DC: APA Press.

Fowers, B. (2014) 'Toward Programmatic Research on Virtue Assessment: Challenges and Prospects', *Theory and Research in Education*, vol. 12, no. 3, pp. 309-328.

Haidt, J. 2001. 'The Emotional Dog and its Rational Tail: a social intuitionist approach to moral judgement' *Psychological Review* 108 (4), 814 – 34.

Haidt, J. 2012. *The Righteous Mind: why good people are divided by politics and religion*. London: Penguin.

Hebert, P., Meslin, E., Dunn, E., Byrne, N. and Reid, S. 1990. 'Evaluating Ethical Sensitivity in Medical Students: using vignettes as instrument' *Journal of Medical Ethics* 16 (3) 141 – 5.

Hebert, P., Meslin, E. and Dunn, E. 1992. 'Measuring the Ethical Sensitivity of Medical Students: a study at the University of Toronto' *Journal of Medical Ethics* 18 (3), 142 – 7.

 Helkama, K., Uutela, A., Pohjanheimo, E., Salminen, S., Koponen, A., Rantanen, V. and Ntsi, L. 2003. 'Moral Reasoning and Values in Medical School: a longitudinal study in Finland' *Scandinavian Journal of Educational Research* 47 (4), 399 – 411.

Hren, D., Marusic, M and Marusic, A. 2011. 'Regression of Moral Reasoning during Medical Education: Combined Design Study to Evaluate the Effect of Clinical Study Years' *PLOS One* DOI: 10.1371/journal.pone.0017406

 Kohlberg, L. 1981. *The Philosophy of Moral Development.* San Francisco: Harper and Row.

 Kohlberg, L. 1984. *The Psychology of Moral Development.* San Francisco: Harper and Row.

 Kristjánsson, K. (2015) *Aristotelian Character Education,* London: Routledge.

Lantz, M., Bebeau, M. and Zarkowski, P. 2011. 'The Status of Teaching and Learning of Ethics in US Dental Schools' *Journal of Dental Education*, 75 (10) 1295-1309.

Lievens, F. and Patterson, F. 2011. 'The Validity and Incremental Validity of Knowledge Tests, Low-Fidelity Simulations and High-Fidelity Simulations for Predicting Job Performance in Advanced-Level High-Stakes Selection' *Journal of Applied Psychology* 96 (5), 927 – 40.

Lohfeld, L., Norman, G., Goldie, J., Schwartz, L., Eva, K., Cotton, P., Morrison, J., Kulmakan, K. and Wood, T. 2012. 'Testing the Reliability and Validity of a Scenario-Based Questionnaire to Assess the Ethical Sensitivity of Undergraduate Medical Students' *Medical Teacher* 34 (8), 635 – 42.

Patenaude, J., Niyonsenga, T. and Faffard, D. 2003. 'Changes in students' moral development during medical school: a cohort study' *CMAJ* 168 (7), 840 – 4.

Patterson, F., Carr, V., Zibarras, L., Burr, B., Berkin, L., Plint, S., Irish, B. and Gregory, S. 2009. 'New machine-marked tests for selection into core medical training: Evidence from two validation studies'. *Clinical Medicine*.vol. 9, no. 5, 1-4.

Patterson, F. and Ashworth, V. (2011) *Situational Judgement Tests: The Future of Medical selection?*, [Online], Available at: http://careers. bmj.com/careers/advice/view-article.html?id =20005183 [Accessed: 24 October 2014].

Pedersen, R. 2008. 'Empathy: a wolf in sheep's clothing?' *Medicine, Health Care and Philosophy* 11 (3), 325 – 35.

Pinijphon, P. (2009) *An ICM Approach to the Assessment of a Medical Ethics Intervention in Thailand,* Ph.D., University of Alabama.

Rest, J.R. (1979) *Development in Judging Moral Issues*, USA: University of Minnesota Press.

Rhodes, R., and D. S. Cohen (2003) 'Understanding, Being and Doing: Medical Ethics in Medical Education', *Cambridge Quarterly of Healthcare Ethics*, vol. 12, no. 1, pp. 39-53.

Schulz, K., Puscas, L., Tucci, D., Woodard, C., Witsell, D., Esclamado, R. and Lee, W. (2013) 'Surgical Training and Education in Promoting Professionalism: A Comparative Assessment of Virtue-based Leadership Development in Otolaryngology-head and Neck Surgery Residents', *Medical Education Online,* vol. 18, [Online], Available at: http://med-ed-online.net/index.php/meo/article/view/22440 [Accessed: 13 November 2014].

Self, D. and Baldwin, D. (1994) 'Moral Reasoning in Medicine', in Rest, J., Narvaez, D. (eds.) *Moral Development in the Professions*, Hillsdale (NJ): Lawrence Erlbaum.

Smith, S., Balint, J., Krause, K., Moore-West, M. and Viles, P. 1994. 'Performance-Based Assessment of Moral Reasoning and Ethical Judgement Among Medical Students' *Academic Medicine* 69 (5), 381 – 6.

Thoma, S., Bebeau, M. and Born, D. 1998. 'Further Analysis of the Professional Role Orientation Inventory' *Journal of Dental Research* 77, 116 – 20.

Tsai, T., Harasym, P.H., Coderre, S., Mclaughin, K. and Donnon, T. 2009. Assessing ethical problem solving by reasoning rather than decision making. *Medical Education* 43 (12), 1188-1197.

Wilkinson, T., Wade, W. and Knock, L. 2009. 'A Blueprint to Assess Professionalism' *Academic Medicine* 84 (5), 551 – 8.