



Measure of Goodness

Randall Curren

These are unpublished conference papers for the 'Can Virtue Be Measured?', held by the Jubilee Centre for Character and Values at Oriol College, Thursday 9th – Saturday 11th January 2014. These papers are works in progress and should not be cited without author's prior permission.

**School of Education
University of Birmingham
Edgbaston
Birmingham¹**

Measures of Goodness

Randall Curren

Judging virtues is an aspect of human affairs as automatic as reading faces and as minutely studied as a Jane Austin novel. How we go about it depends on our purposes and the constraints imposed by time, settings, resources, and the forms of contact, information, activity, and relationships involved. This talk will address some different purposes educators might have in aiming to assess their students' virtues, and the significance of those purposes for the methods chosen. It will consider the extent to which different forms of information justify attributions of virtue, suggest a triangulation approach, and consider the prospects for a virtue-focused test of moral response.

Can virtues be measured? Should they be measured? If so, in what ways and to what ends? On the face of it, the possession of virtues can be judged within limits, and taking the measure of others' virtues is so important to human social functioning that it is unavoidable and to some extent automatic and unconscious. It will be useful to note some aspects of informal, out-of-school judgments of persons' qualities, including moral virtues, before considering the prospects for conducting more formal assessments in schools.

Face Reading

We register the emotions that flash across others' faces in our own emotions, without ever having been taught to do so and often without being able to pinpoint what has caused us to feel what we feel. Subtle signs of aggression, untrustworthiness, or ill-will induce unease and mistrust, emotions we generally do well to heed, though our learned dispositions and judgments may get in the way. Expectations, implicit bias, and stress can also blind us to good qualities and signs of cooperation. We *can* read others' minds in their faces, and it is in our nature to do so, but we don't always get it right or draw the right inferences, even when lives are at stake.

Consider two very different cases:

1. Having been taught to help those in need, the kind hearted young woman opens her apartment door to the stranger whose demeanor gave her a bad feeling when she saw him moments before in the lobby. A dangerous serial predator, he forces his way in and brutally attacks her.¹
2. Primed by years of watching television series that depict their work as far more exciting and dangerous than it is, four police officers in plain clothes stopped their unmarked car when they saw Amadou Diallo standing quietly on the porch of his residence in the South Bronx on the night of February 3, 1999. Misreading the signs

¹ I am withholding details given in grand jury testimony and by friends of the victim to protect her privacy.

that he was terrified, thought he was being robbed, and was trying to offer his wallet as they advanced on him, they drew their semi-automatic weapons and fired 41 shots in what they imagined was self-defense.²

What we read, sometimes misread, and sometimes read but fail to act on in the heat of the moment can flash across a face too quickly to consciously parse in real time. Recorded and slowed down, the patterns of facial movements that display different emotions can be identified by observers trained in Facial Action Coding, and might conceivably be reliably identified by an *automated* coding system in the future. The Facial Action Coding System (FACS), developed by Paul Ekman and Wallace Friesen, is indeed already used by computer animators at Pixar and DreamWorks to create emotionally evocative animated characters, though I won't hazard a guess as to how surmountable the obstacles to a recognition system might be.³ Forty-three distinct human facial muscular movements have been classified, and about three thousand combinations of those distinct movements have been identified as displays of emotions. Many of these combinations of movements are difficult or impossible to fully suppress or make voluntarily, so reliably tracking their occurrence might have evidential value in discerning gradations of virtue, if characteristic patterns of emotional response to situations are one aspect of virtues and vices. If human emotional responses to actual situations and simulations are sufficiently similar, one could imagine simulation games enacted with facial monitoring and coding being one form of measure of a person's goodness.

Whether a technology of virtue assessment along these lines should be welcomed would depend upon how it is likely to be used and how adequate the underlying understanding of virtue is. A concern illustrated by the first case I have described is that even societies that are no longer *intentionally* patriarchal seem to be more focused in cultivating cooperative virtues and helping behaviors in girls than in boys, with the result that girls and women are more likely to be exploited and suffer victimization. If automated systems for assessing emotional dispositions became available, would they be gender-neutral in discerning not just compassion and good-will, but also prudence, fearing neither too little nor too much in the circumstances, and the

² This is recounted by Malcolm Gladwell in *Blink: The Power of Thinking without Thinking* (New York: Little Brown & Co, 2005), pp. 189 ff. I have departed from Gladwell's account by adding a probable but undocumented hypothesis about priming by fictional images of police work. The role of television depictions of police work in shaping police expectations and overreaction in routine situations is well documented in Didier Fassin's important anthropological study of the Paris police, in *Enforcing Order: An Ethnography of Urban Policing* (Cambridge: Polity Press, 2013).

³ I rely here again on Gladwell, *Blink*, 197 ff., who interviewed Ekman and his teacher, Silvan Tomkins. The primary works that virtue educators and psychometricians might profitably consult as points of departure are: Paul Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage* (New York: Norton, 1995); Paul Ekman and Wallace V. Friesen, *Facial Action Coding System, parts 1 and 2* (San Francisco: Human Interaction Laboratory, Dept. of Psychiatry, University of California, 1978).

confidence to assert one's own interests when it is appropriate to do so? If the answer to every part of this question is "yes," I can imagine contexts in which the information made available would have a salutary effect and others in which it would not.

Getting to Know Who is Good and Good at What

Readings of faces and body language are often conjoined with explanatory inferences from more or less extensive histories of verbal and non-verbal conduct, in more or less varying circumstances, witnessed by one to many observers who are more or less virtuous themselves and more or less intimately acquainted with the person being judged. In all of these respects, more is better, and taking the measure of moral virtues is not unlike taking the measure of other personal qualities. There are qualities we value and are more or less able to discern in our neighbors, friends, partners, children, and the workers, professionals, and leaders on whom we rely. It would be surprising if there were not some reliability, but also limitations and variability, in our ability to judge the presence of those qualities – limitations and variability beyond our native capacity to distinguish friend from foe and identify potential mates with qualities similar to our own. What is revealed in conduct and affect must be discerned, and may not be equally evident to all observers, any more than gradations of particular talents and dispositions to use those talents well would be equally evident to all observers. Experts in the domains of talent or excellence to be judged are presumably better judges of the relevant forms of goodness, but judgments about who is good in general and who is good at one thing or another are often entangled in the choice of one person over another in employment and other contexts. Knowing who will reliably perform well in expert tasks and who will reliably be honest may both be important, and the latter is not obviously harder than the former. In judging *ability* in a hiring decision, we would give some weight to test performance, if there is a relevant test, but in projecting likely *performance* we might rely more on the testimony of past supervisors and the degree of consistency evident in the academic record and other arenas of responsibility.

The observation about better and worse judges has significant implications for character education and assessment. On the one hand, it shrinks the perceived distance between measuring moral virtue and measuring other forms of goodness. On the other hand, it suggests we must tread carefully in conceptualizing the scope and methods of character education and their relationship to the expertise required of character educators and evaluators. Expertise on substantive moral doctrine is not what is needed, and claims of such expertise should be met with skepticism, but I'm inclined to think that qualities of ethical discernment and thoughtfulness *are* needed, along with skill in facilitating discussion and knowledge of moral development, learning, motivation, and the like.

The similarities between moral virtues and other forms of goodness are evident not only in accounts of moral virtue that emphasize the analogy with complex skills, but in what we know of the history of ideas of virtue in Greek antiquity. There are wider and narrower senses of the word "virtue" and its ancient Greek counterpart, "*arête*." In its wider sense, we speak of the virtues or

good features of all sorts of things: the qualities that suit them for some purpose or make them pleasing or admirable. One virtue of a good hammer is that the handle is easy to grip, so it doesn't fly from the hand when swung. Traits of persons can be virtues in this sense relative to specific activities or roles. Endurance is a virtue, or desirable trait, in a runner of marathons, courage a virtue of a soldier, or a man, if soldiering is something required of men as such. The idea of *human virtue* or the virtues of a human being as such seems to have originated in this way in Greek antiquity, as the manly virtues associated with the defense of vulnerable *polises* – strength, courage, cunning, and endurance – and later broadened to include traits less obviously essential to the possessor's success, but nevertheless essential to the internal functioning and stability of established societies – self-restraint, justice, and wisdom. Wisdom was understood to entail a kind of respect for reason and beings who reason, and thereby an ethic of mutual goodwill and norms of truthful reason giving. It was also understood, in one way or another, to be an essential component of being fully virtuous: true virtues, in the Aristotelian version of this idea, are guided by wisdom or good practical judgment.

On an Aristotelian understanding of moral virtues, they are complex dispositions or clusters of related dispositions. Habituation – or guided practice in pursuit of being good at certain things or in certain ways – is supposed to shape desires, emotions, perceptions, beliefs, conduct, and reason responsiveness as a causally related package. It is hard to see how it could succeed without supervision and coaching that calls the learner's attention to factors that make a difference to how one should act, provides a related vocabulary, and engages the learner in the forms of discernment, imagination, reasoning, and judgment on which good decisions are based. The point of practice is not simply for the learner to become reliably respectful of others and committed to good ends, but to develop the perceptiveness, imagination, and judgment needed to actually achieve good ends. If it is true that moral virtues are dispositional clusters formed in this way, and moral perceptiveness and judgment are among the trailing effects of habituation that also shapes moral motivation and commitment, then evidence of ethical perceptiveness and judgment would be evidence that moral motivation and dispositions to act well are also present in those who have been morally educated on Aristotelian principles. This suggests a relatively optimistic view of the prospects for assessing virtue in efficient ways in schools, but the variety of virtues and contexts in which they are expressed cuts in the opposite direction.

The extent to which virtues like courage and self-restraint can be acquired and measured in schools is surely limited by the limitations of context inherent to schools. How much supervised practice in self-restraint in the face of tempting pleasures or courage in the face of fearsome threats are we realistically going to offer students in schools, and how many corresponding contexts for the expression of those virtues are we prepared to test them in? When it comes to those virtues, we might plausibly test *understanding* of them and opportunistically observe them in action – or not in action! – when tempting and dangerous things that are not supposed to happen in schools do happen in schools. We won't be able to test *pivotal* cognitive aspects of those virtues, such as the accuracy of perceptions of danger in the heat of the moment (a pivotal

aspect of the Amadou Diallo case). We may be able to infer from students' behavior in school and their academic success that they are able to defer gratification or work hard even in the face of many tempting things.

The idea of administering *tests* is that students will be prompted to respond to stimuli of the school's choosing at predetermined times, and those stimuli and responses will usually be essentially verbal in nature and without real consequences. We can use practical tests in evaluating skill in laboratory procedures, musical and athletic performances, and other practical arts, but there are ethical and practical barriers to using such tests to assess virtues.

Why Try to Measure Virtues in Schools?

If by "measuring" virtue one has in mind the use of standardized instruments for making comparative estimates of the extent to which various students possess moral virtues, and make progress over time in acquiring those virtues, then the idea of measuring virtues warrants careful scrutiny. It is important to ask what the purpose of such measurement would be and how the results would be used.

I would like to start by dismissing the notion of extending the recent enthusiasm for high-stakes testing and accountability schemes into the realm of virtues. The idea that it is productive to reward, penalize, and motivate teachers and school leaders on the basis of their students' standardized test scores is remarkably ill-conceived, and it has proven to be counterproductive in practice. The notion seems to be that without such accountability schemes teachers are not sufficiently motivated to teach well, and that being more highly motivated will improve their teaching. Motivational psychologists disagree, and research on the effects of imposing such controlling structures on teachers is that it displaces the intrinsic motivation they bring to their work and induces anxiety that undermines their performance. They become more controlling and anxious in their interactions with students, and frame the value of schoolwork in more instrumental terms, with the result that students are less motivated and learn less.⁴ Attempts to "incentivize" performance can yield heightened but also dysfunctional motivation. This is only one of several detrimental aspects of testing and accountability schemes, but one that has special relevance for virtues. If one's interest is in cultivating virtues in children, then testing their acquisition of virtues as a basis for judging their teachers is one of the worst things one could do.

⁴ See L. Pelletier & E. Sharp (2009) "Administrative Pressures and Teachers' Interpersonal Behavior," *Theory and Research in Education* 7(2): 174-183; R. Ryan & N. Weinstein (2009) "Undermining Quality Teaching and Learning: A Self-determination Theory Perspective on High-stakes Testing," *Theory and Research in Education* 7(2): 224-233; M. Vansteenkiste, B. Soenens, J. Verstuyf and W. Lens (2009) "What Is the Usefulness of Your Schoolwork? The Differential Effects of Intrinsic and Extrinsic Goal Framing on Optimal Learning," *Theory and Research in Education* 7(2): 155-163.

Virtue involves an attachment to and pursuit of what is good because it is good. Attachment or *internalization* of values that yields healthy self-regulation, or fully integrated motivation, occurs when learners' needs for good relationships, competence, and autonomy are satisfied and they understand and accept the real importance [of something] for themselves" or have "identified with [its value] for themselves."⁵ The "autonomy supportive" contextual factors identified as favorable to fully integrated self-regulation include the offering of a *rationale* that is meaningful to the learner, respectful *acknowledgement* of the learner's "inclinations and right to choose," and a manner of offering the rationale and acknowledgement that *minimizes pressure and conveys choice*.⁶ The evidence suggests that high stakes testing of student virtues would directly undermine the social conditions in schools foundational to students' virtue acquisition and embrace of the inherent value in their learning generally.

What other purposes might one have for measuring virtues in schools? One might think that measuring virtues is essential to *evaluating programs* of character education. Or one might think that character education should be like any other form of education in having a *student evaluation* component, conceived as *formative, summative*, or both. Or one might think that measures of student goodness more systematic than those routinely used in schools could be usefully employed in decisions about school and classroom management: decisions about how to distribute students between different classes in the coming year, how to respond to disruptive behavior, and so on. I'll speak to matters of program evaluation and routine student evaluation in the concluding sections that follow.

Program Evaluation

No one should be field testing educational programs (or structural reforms), including programs in virtue education, without a basis in prior research and tested theory. The body of theory and research on motivation and contextual factors favorable to fully integrated internalization of values is one such basis for program design, as are the program evaluation studies through which possible components of programs have been found to be efficacious. There are reasons why it is nevertheless preferable that new programs should be field tested. I'll limit my remarks about this to two observations.

⁵ E. L. Deci & R. M. Ryan, "Motivation, Personality, and Development Within Embedded Social Contexts: An Overview of Self-Determination Theory," p. 89, in R. Ryan (Ed.) *The Oxford Handbook of Human Motivation*, pp. 85-107 (New York: Oxford University Press, 2012).

⁶ E. L. Deci, H. Eghrani, B. C. Patrick & D. R. Leone (1994) "Facilitating Internalization: The Self-determination Theory Perspective," *Journal of Personality* 62(1): 119-142, at p. 124.

First, there are reasons to prefer evaluating the program at multiple sites, if possible, to screen out confounding factors that may be in play at particular sites. The efficacy of a program is better tested through 10,000 subjects spread over multiple sites than 10,000 subjects at one site.

Second, the limitations of available methods for assessing virtue commend a combination of pre- and post-intervention measures, and the measures chosen need not be "student-level significant" measures designed to provide profiles of individual students' degree of virtue acquisition. The goal in program evaluation is to establish the efficacy of a program, not to compare students with one another (or with themselves over time), and the quality of information about programs may be enhanced through sampling strategies that do not attempt to learn the same things about every student. So, for instance, we might randomly select some students to discuss the ethical climate of their schools in focus groups, some before the intervention and others after, and code the discussion for frequency of salient normative terms. We might enlist selected classes on a similar pre- and post-intervention basis in writing essays on life plans or the traits valued in friends. Other written measures might be random samplings of student work in any aspects of the curriculum in which ethical learning interventions are introduced – a matter I address below. The progress of student ethical attunement and judgment might be tracked using a combination of such methods, and observations of student conduct and affect might be useful in establishing that the progress is not just cognitive. I am picturing here something comparable to the method for counting pro- and anti-social behaviors used to validate the PATHS curriculum developed some years ago in the US.

Evaluation of Student Ethical Learning

There are ways in which we already evaluate students' acquisition of virtues in schools, where the goals of educating them include not just skills and understanding, but commitment to certain goods – goods of inquiry, artistry, and the like – and consistency in pursuing and achieving those goods. "Effort" figures significantly in grades, and "effort" is a matter of striving toward the right things. This is not to say that the striving and commitment are moral striving and commitment to what is ethically valuable, but the task of evaluation in the two cases might be very similar. The question is where specifically moral virtues would lodge in the curriculum or extra-curriculum of schools in such a way as to provide a basis and vehicle for evaluating students' moral virtues. Apart from tests of what were once taken to be fixed, native abilities, what we test students on is what we have been teaching them.

So how would we teach moral virtue, such that what is practiced could be put to the test? I have made some cautionary observations about teaching and measuring virtues like courage and self-restraint in schools. Notice that those virtues don't look much like the complex skills to which Julia Annas has helpfully compared virtues.⁷ What Annas doesn't quite say is that the focus of guided habituation in her account is on the learner *deciding what to do* for herself. Situations in

⁷ *Intelligent Virtue* (Oxford: OUP, 2011).

life are complex in ways relevantly similar to the complexities faced by those with complex skills in the context of using those skills. The teacher must engage the student in recognizing and responding appropriately to the factors that are relevant. So the practice is substantially practice in recognizing and weighing the factors relevant to decisions and making those decisions, the rest being development of technique, as in violin performance or medical procedures. In order to test what students have practiced, then, we would – on an understanding of Aristotelian habituation that takes the role of good judgment seriously – test ethical discernment and judgment.

The analogy between general moral learning and learning the ethics of a sphere of professional practice is instructive. That is where I will begin, describing a form of Aristotelian habituation in ethical medical decision making. From there, I will briefly address three related forms of guided practice in ethical choice and the forms of student evaluation appropriate to them. My view is that any of these or similar forms of evaluation might play roles in evaluating programs in character education.

Ethical Coaching in a School of Medicine

If one were designing an Aristotelian approach to teaching ethics in medical schools, its centerpiece would be supervised practice in ethical decision-making.⁸ Students would participate in the monitoring of cases, and when they meet as a class they would discuss what they and others take to be ethically salient. The instructor would facilitate discussion, casting the students in the role of ethical consultants, and coach them as they think through the cases. The students would practice discernment, listening, and modes of consultation and decision-making conducive to good decisions and the development of dispositions of apt perception and wise counsel and choice. This would occur against a background of prior instruction in a code of professional ethics and the basis for that code in the fundamental goods at stake in medical practice. A key to success would be a classroom atmosphere that is non-judgmental. It is impossible to coach someone who does not reveal herself engaging in the activity to be coached. The goal is to cultivate professional integrity, and evaluation might best occur through longitudinal monitoring the quality of patient care, patient satisfaction, and frequency of malpractice lawsuits. In the context of the class, the most authentic measure of the kind of learning that occurs might take the form of essays in response to case scenarios, scored for their efficiency in identifying the ethically salient features and considerations. An alternative would be an oral “consultation” constructed and scored on the same principles.

⁸ See Margie Hodges Shaw, *Coaching as a Form of Instruction and Component of Medical Ethics Education* (PhD. Dissertation, Warner School, University of Rochester, 2011), available at: <https://urresearch.rochester.edu/viewContributorPage.action?personNameId=4321>.

The PATHS Curriculum

A starting point for character education is helping children become more attuned to the emotional dynamics of social interactions and more in the habit of thinking before they act – thinking specifically about the social and emotional dynamics relevant to what they do and the likely consequences of different choices. The PATHS curriculum was designed to do just this, and it used simple pictures of children in social transactions as a basis for teacher facilitated discussions of what the children in the pictures were feeling, might do, and should do. It was validated through an observational counting method, as noted above, but the learning stimulated by discussions of the pictures could presumably be tested by using novel pictures as prompts for free responses, scored for quality of noticing and understanding.

Critical Thinking Projects

In 1996, a colleague and I launched an undergraduate cluster in philosophy and teaching, that allowed students who had taken classes in critical thinking and philosophy of education to spend a semester as teaching interns in urban elementary schools. They directed a variety of critical thinking projects, usually focused on decisions students had to make in their lives, and involving developing reasoned essays and engaging in staged debates. The model we preferred was for students to be engaged in identifying ethically significant questions they wanted to address. One example of this was the question, “Should I join a gang?” It is plausible that students’ ability and commitment to thinking through personal decisions in light of relevant ethical considerations was strengthened by such learning. That’s what was practiced and coached, and their essays and debate performances (sometimes watched by the entire school community) were the apt objects of judgment that were used in evaluating their ethical learning. Supposing those were scored for quality of ethical discernment, cogency, and evidence of seriousness of purpose in cooperating in the opportunity to learn, that might be about the best evidence of progress in acquiring good character one could hope for in the short term in a school setting. The documentation of the college student intern’s efforts in all of this, and oral presentation of her work, might similarly be about the best short term evidence of her good character that educators could obtain without systematic observations that would be far beyond a school’s capacity to obtain.

High School Ethics Bowl

A final example, more formalized in its prompts and scoring system is the High School Ethics Bowl, an off-shoot of the Collegiate Ethics Bowl.⁹ Both are forms of staged competitions in which student teams respond to unexpected questions concerning cases scenarios they have previously had time to research and analyze. The questions that are most salient for our purposes are ones that pertain to the decisions of specific characters in the cases. The scoring is again

⁹ See <http://nhseb.unc.edu> for cases, rules, scoring guides and other details of the National High School Ethics Bowl.

focused on the quality of identification of ethically salient features of cases (completeness, emphasis, etc.) and cogency of the analysis and defense of the answer to the question. The Ethics Bowl competitions are now the basis for “experiential learning” on many college campuses and a growing number of high school campuses. It is arguably a model for cultivating seriousness and thoughtfulness about ethical matters, and testing the extent to which such qualities of character are acquired, though a measure of a person’s state of character as such it is not.

Conclusion

If the Aristotelian picture of habituation as producing a causally linked dispositional cluster of desire, affect, perception, and conduct survives scrutiny, we would have a grounded theoretical basis believing that measures of ethical perception and judgment such as these are more adequate as measures of virtue than we currently have reason to believe. Until then, we should focus on evaluating forms of student learning as best we can, and evaluate character education programs through mixed methods and in light of our best theoretical understanding and knowledge of the relationships between meeting children’s needs and enabling them to be good and live well.