



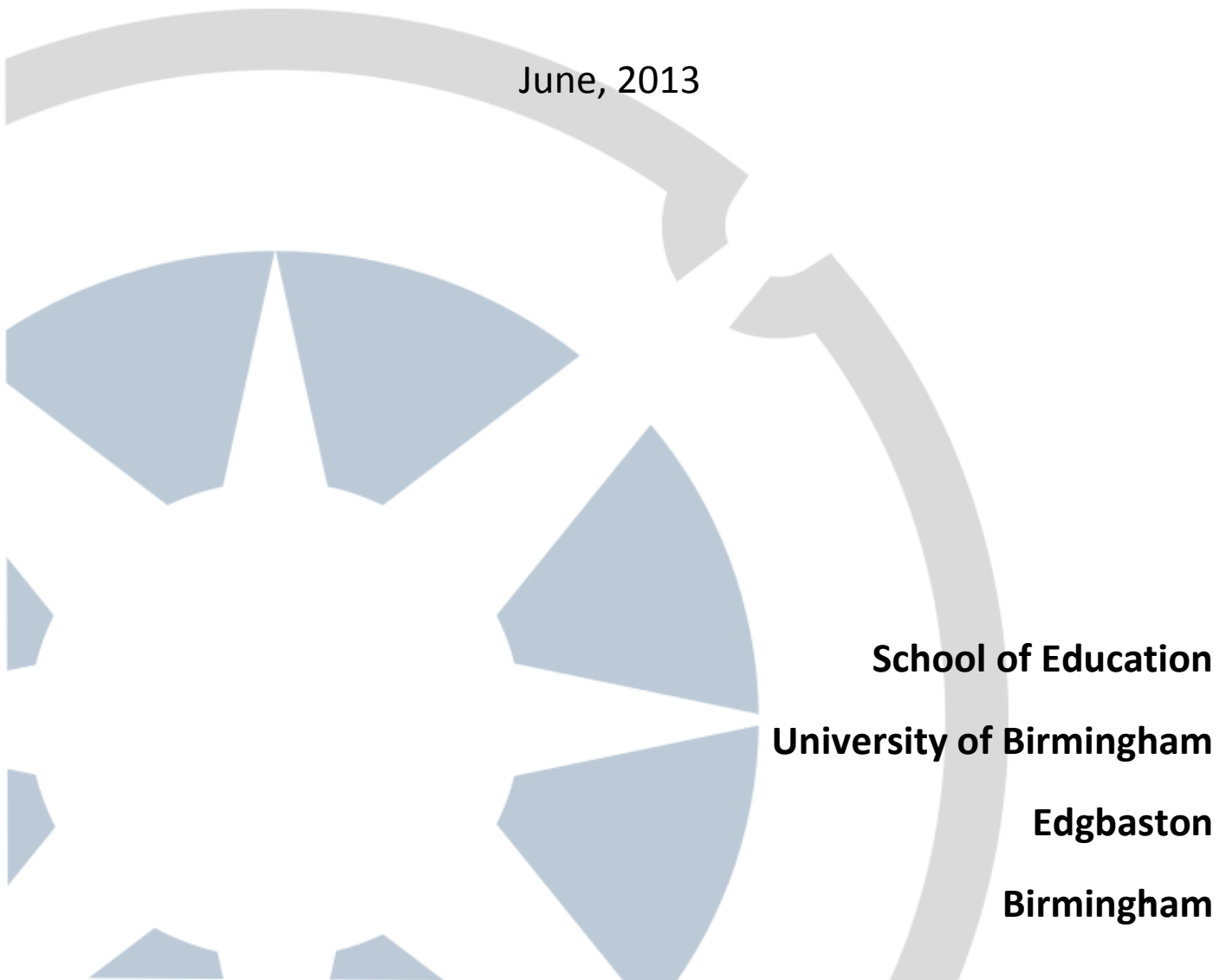
THE
JUBILEE CENTRE
FOR CHARACTER & VALUES

UNIVERSITY OF
BIRMINGHAM

Problems of Assessment in Education for Intellectual Virtue

Ben Kotzee

June, 2013



School of Education

University of Birmingham

Edgbaston

Birmingham



Problems of Assessment in Educating for Intellectual Virtue

Paper delivered at the Educating for Intellectual Virtues Conference, Los Angeles Marymount University, 20 – 22 June 2013

Ben Kotzee

1. Introduction

While attention to the teaching of intellectual virtue is gathering steam, little attention has been paid to the matter of its assessment. Even though it belongs to the slightly drearier regions of educational thought and practice, assessment is important to education. Without being able to assess to what extent students already possess certain forms of knowledge or skill, it will be hard to know where to start in our teaching; and without being able to assess for learning, we are not in a position to know whether our teaching has been effective. However, standard forms of testing – especially as we encounter them in the high stakes graded exams common in schools and universities – seem to be inimical to the teaching of intellectual virtue. If we are to teach students to be truly intellectually virtuous we must teach them to value knowledge for its own sake and not for the reward of a grade; virtuous practice must not only be exhibited on test day, but become an ingrained facet of intellectual character.

The paper proceeds as follows. In section 2, I begin by raising doubts about the possibility of assessing intellectual virtue. I ask whether assessing intellectual virtue is, firstly, possible and, secondly, advisable. I conclude in the affirmative. Next, I turn to the question of how to assess for intellectual virtue. In section 3, I consider one approach – the approach taken by researchers into epistemic development (or ‘personal epistemology’). In section 4, I consider another possible set of approaches – that of the critical thinking movement. In section 5, I ask what we may learn from approaches to measuring virtue in the moral sphere. In section 6 I summarise what we may conclude regarding the assessment of intellectual virtue from these approaches.

2. Doubts about assessment

I start by raising a number of doubts about the idea of assessing intellectual virtue. Criticisms of high stakes testing regimes within the educational systems of countries like the USA, UK and Canada are well-known. Amongst the complaints that are regularly made are that such tests force a narrowing of the curriculum, that they

advantage well-off groups over others, that they are regularly gamed, that they lead to teaching to the test and that they are unfairly used to reward and punish teachers. Ravitch (2010), for instance, sketches vividly how the standards movement in education in the United States (regarding which she was initially positive) has been 'hijacked' and turned into a movement concerning testing. Inasmuch as high stakes tests force students to game the test and to adopt an instrumental approach to education generally, the use to which forms of assessment such as exams and standardised tests are put in the classroom risk making students (and their teachers) not virtuous, but intellectually vicious.

Such complaints about how high stakes testing regimes distort education should concern any educator. As Elgin makes clear, however, concerns about high stakes tests often slide between at least four different issues: (1) whether the consequences that follow from high stakes tests are appropriate (whether the stakes are too high), (2) what high stakes tests should test for (the content of such tests), (3) the format of high stakes tests (whether, say, multiple choice tests, timed exams, essays, take-home projects, etc. is best) and (4) how much of what is educationally valuable can actually be tested for. (Elgin, 2004: 271 – 2) This paper concerns the second and third of these issues as it applies to the assessment of intellectual virtue, but, first, something must be said regarding the fourth issue – whether intellectual virtue is something that can be assessed at all or whether it is one of those things that we value, educationally speaking, but cannot test.¹ Elgin's view seems to be the latter. As she writes:

'It is hard to imagine a standardized test for sensitivity, originality, cooperativeness, or openness to new approaches. In itself, this is not a problem. But to the extent that educators become convinced that the results of high stakes tests are *the* measures of achievement, they may become blind to the value of objectives that cannot be measured by such tests.' (Elgin, 2004: 274)

Elgin articulates a worry that goes back at least as far as Aristotle. Aristotle held that, while it is certainly not impossible to know whether another (or oneself) is virtuous, it is very hard to make exact judgements regarding virtue – ethics is an 'inexact science'. (Irwin, 2000) While Aristotle is, in these passages, concerned with the inexactness of the study of moral virtue, there is little reason to suppose that the study of intellectual virtue would be any more of an exact science. Furthermore, when it comes to virtue, we are confronted with another principle that may pose a difficulty in assessing virtue - the principle of the unity of the virtues. If the principle were true, having one intellectual virtue requires having the lot. That would imply that assessing for intellectual virtue would have to be done on the level of looking for all of the virtues together. Lastly, no psychometric test of intellectual virtue exists. Instruments have been crafted to test for moral virtue (the best known being the Values in Action Inventory of Strengths or VIA-IS), but opinion divides regarding whether a psychometric approach truly succeeds in measuring full virtue as manifested in objective action (as opposed to the subject's assessment of whether they are virtuous) (see below for further discussion).

¹ The first issue – whether the stakes tied to assessment are too high – falls largely outside the scope of this paper.

Some researchers are sceptical that virtue can be studied empirically *because virtue has no stable effect on behaviour*. Drawing on results from social psychology, Harman (1999) and Doris (2002) suggest that moral character, *qua* enduring and general feature of a person, has so little effect on what people actually do, that one may as well say it does not exist. Rather, they hold that it is situation or context that determines action.^{2 3} Recently, this challenge has been extended to the field of virtue epistemology. (Alfano, 2012) Alfano appeals to empirical research on the link between cognitive performance and mood to suggest that what explains intellectually virtuous behaviour best is not a global trait like intellectual virtue, but situational influences:

‘It will turn out that when people behave in accordance with the intellectual virtues of curiosity, flexibility, and creativity, their conduct is often better explained in terms of situational influences like mood elevators than in terms of consistent global traits.’ (Alfano, 2012: 234 – 5)

Doubts about the project of assessing intellectual therefore come from two directions. Firstly, a view is possible that intellectual virtue may exist, but that it cannot easily be assessed for. As, for instance, White (1999) suggests, ‘rich’ understanding may not be amenable to paper-and-pencil testing at all, but may only be assessable on the basis of long personal acquaintance of the teacher with the student. Another possible view is articulated (although not defended) by Possin who points out that many university faculty hold that ‘...the *only* test that could measure one’s acquisition of cognitive abilities is *life itself...*’ (2008: 202) Secondly, it may simply be that intellectual virtue does not exist, or is so fickle that we may as well not trouble ourselves with detecting it. This is the situationist position of Alfano.

It is very important to notice, though, why we cannot simply give up on the project of assessing intellectual virtue.

Winch and Gingell hold that assessment is a necessary component of the educational process. They quote Anthony Flew to the effect that, if one is to teach anything, one must be concerned with whether one succeeds in teaching it, i.e. if one is managing to get across what one had in mind to get across. As a matter of necessity, anyone occupied with ‘teaching’ needs to give thought (while they are teaching) to whether their students are learning. This is assessment. (Winch and Gingell, 1996: 377) I am not so sure that finding out whether one’s teaching is successful is indeed a necessary element of teaching or, to put it differently, that one is not *teaching* if one is not also assessing at the same time. Especially when it comes to virtue, I can imagine a case like the following: the dying father uses his last

² Snow (2010: 99 – 116), however, outlines a number of problems with the situationist critique. Snow discusses six of the most important empirical findings that the situationists rely on to demonstrate that virtue does not exist or does not have a stable effect on behaviour (amongst them the Stanford Prison Experiment, the Milgram Obedience Experiment and Darley and Batson’s experiment on the helping behaviour of Princeton seminarians). For each of these, Snow identifies problems to do with the original experiments and/or problems to do with how these experiments are interpreted by the situationists. Also see section 5, below.

³ Another criticism of situationism can be found in Fowers (2005: 23 – 4)

months on earth to try and teach his children to be honest or fair or caring or whatever else he thinks is the greatest virtue in life. The father may never come to know whether he was successful, but this does not imply that he was not teaching his children honesty or fairness or care in all seriousness. As teachers we are in much the same position, especially with the higher intellectual virtues. We try and teach our students intellectual honesty, fairness, open-mindedness, rigour and so forth, but, typically, because we do not see their reasoning in action in their lives after university, we do not know whether our teaching has worked.

What I cannot make sense of is another possibility. Even if we may never know for a specific student whether they have acquired a specific intellectual virtue, we must be able to tell in principle and often enough. As Curren (2004) makes clear, in thinking about assessment, we are confronted with a variant of one of the big philosophical questions: the problem of knowledge of other minds. The problem is as follows. We assume that, because we teach them things, our students learn what we try to teach them (and learn it pretty much as we have taught). In order to be sure of this, we set them tests, such as writing an essay or working out problems. But how do we know that these assessments reflect accurately what students have learned on the basis of what they were taught? The answer is that we do not always know, but that anyone engaged in purposive teaching cannot be entirely sceptical about the possibility of assessment. As Elgin puts the point:

‘If we can know nothing about other minds, we cannot feasibly hope to bring it about that other minds (if they exist) change in worthwhile ways. Hence, if we are going to embark on education at all, we need to assume (a) that other people have minds and (b) that it is possible, somehow, to gain knowledge or justified beliefs about the contents of their minds.’
(Elgin, 2004: 275)

While we may not be in a position to have exact knowledge regarding how intellectually virtuous our students are after they have finished some course of study with us (as would be demanded if virtue were an ‘exact science’), the fact that teaching is a purposive activity rules out complete scepticism regarding whether one’s teaching has had some effect. The upshot is that, if teaching intellectual virtue is possible, then assessing it – in some form – must be possible too.

Besides being merely possible (in fact it is an assumption of teaching), coming to know whether our students become intellectually virtuous as a consequence of studying with us is also clearly educationally desirable. Firstly, it follows hard on the heels of the purposiveness of teaching that educational effectiveness depends on coming to know whether what one teaches has any effect and adjusting one’s teaching if it does not. Secondly, in order to demonstrate the worth of an intellectual virtues approach to teaching to policy makers (in order that the approach may be funded and promoted, or, at the very least, allowed), one will have to be in a position to demonstrate that this approach to teaching is at least as effective or superior to rival educational approaches.

A crucial task, then, is to be clear on what the project of assessing intellectual virtue is actually *for* and what demands tests of intellectual virtue should meet. Would such tests be designed:

1. to form part of a scientific research programme to illustrate that intellectual virtue exists and to prove that it has a stable effect on behaviour?
2. to illustrate the general success of a programme of teaching for intellectual virtue?
3. to illustrate how virtuous an individual student is or has become following some programme of teaching?

This will have an important effect on what tests for intellectual virtue eventually look like.

Importantly, the form of assessment that is demanded by the very possibility of teaching for intellectual virtue and by the twin demands of teacher effectiveness and public accountability is *not* that of the psychometric test of intellectual virtue (demanded by assessment strategy 1 and to an extent by strategy 3). Rather than taking psychometric testing as our model for the assessment of intellectual virtue, it may be much more appropriate to take an approach akin to *programme evaluation* (implied in strategy 2). In the evaluation of the success of teaching programmes, it is not always necessary to come to a detailed understanding of how each individual has responded to a programme – generalising across the group may be entirely appropriate.⁴ Norris *et al.*, for instance, hold that the educational assessment typically looks for objective change in measurable factors such as ‘behaviours’, ‘performance’ and ‘competency’. (2004: 284) Rather than striving for statistical generalisation about how an individual is likely to behave, Norris *et al.* hold that we should reconceive assessment as a form of inference to the best explanation where inferences are made about the contents of the mind of the student on the basis of what our assessments measure. This calls for forms of assessment that – while they do not support statistical generalisation to how someone is likely to behave – still supports us in inferring that our classes as a whole have had an effect.

Simply relying on teachers’ knowledge of their students might – as White suggests – sometimes do the trick. However, it doesn’t nearly always do the trick. Assessment study after assessment study shows how prone teachers are to rely on matters like first impressions in assessing students, how subject they are in their judgements to rely on markers to do with class, race or culture and how dependent their judgements are on the circumstances surrounding the assessment. While being cautious of falling back on a psychometric model, we can certainly do better than *that*. In the next four sections, I sketch some of the possibilities that present themselves.

3. Assessing epistemic development

The psychological study of the epistemic development of children and young people is heavily influenced by the pioneering work of William Perry (1970) into the epistemic development of university students. In recent years, this programme of study has come to be better known as the study of ‘personal epistemology’. Perry worked in the tradition of Lawrence Kohlberg and modelled his ‘Scheme of Intellectual and Ethical Development During the College Years’ on Kohlberg’s theory of the stages of moral development. According to Kohlberg, the moral development of children and young people follows a predictable course through six stages of

⁴ I thank Randall Curren for impressing this point on me.

increasing sophistication that can be simplified into three 'levels' of moral development:

1. Identification of what is right and wrong in terms of punishment or reward ('pre-conventional' moral thinking)
2. Thinking of moral actions as those that are socially accepted or that are prescribed by laws, rules or conventions ('conventional' moral thinking)
3. Adopting a principled ethical position that makes for universal application ('post-conventional' moral thinking).

Kohlberg was heavily influenced by Rawls and identified a Rawlsian orientation towards morality as the height of moral development.

Like Kohlberg thought that moral development takes the young person through a more-or-less predictable journey through different moral stages, so Perry thought that young's people's epistemic development progresses through nine stages. For ease of understanding, one may summarise these nine stages of the scheme into four broad levels of epistemic development:

1. dualism
2. multiplicity
3. relativism
4. commitment in relativism

(Feucht and Bendixen, 2010: 6)

In a nutshell, Perry postulated that college students arrive at university with the view that knowledge is absolute and that authorities are infallible sources of this knowledge (the view that Perry calls 'dualism') through the view that there are multiple sources of knowledge that may all make different but equally valid knowledge claims (what Perry calls 'multiplicity') through the view that what counts as knowledge is relative to some context ('relativism') to a sophisticated form of relativism in which the student is '...certain about the contextualised truth of a knowledge claim...' (Bendixen and Feucht, 2010: 6) ('commitment in relativism').⁵ Studies of students' personal epistemology typically make use of Schommer's 'Epistemological Questionnaire'. (1990)⁶ The questionnaire assesses the degree to which students agree or disagree with five hypothesised 'epistemological beliefs':

- a. Knowledge is simple rather than complex
- b. Knowledge is handed down by authority rather than derived from reason
- c. Knowledge is certain rather than tentative
- d. The ability to learn is innate rather than acquired
- e. Learning is quick or not at all.

The questionnaire probes the degree to which respondents hold these beliefs by asking them how strongly they agree with 63 items, each designed to represent one of the five main 'epistemological beliefs'.

⁵ A rival account in personal epistemology is that of King and Kitchener (1994). King and Kitchener sketches epistemic development in terms of a development of reflective thinking from pre-reflective to quasi-reflective to fully reflective styles of thinking.

⁶ Other instruments of note are the Beliefs about Learning Questionnaire and the Epistemic Beliefs Inventory.

Sample Items from the Schommer Epistemological Questionnaire (SEQ)

20. If you are ever going to be able to understand something, it will make sense to you the first time you hear it.

- Strongly agree Agree Neutral Disagree Strongly disagree

21. The only thing that is certain is uncertainty itself.

22. For success in school, it's best not to ask too many questions.

25. You can believe almost everything you read.

26. I often wonder how much my teachers really know.

(adapted from Boden, 2005: 298 – 9)

As a description of the attitudes that university students often hold towards what they learn at university – and the attitudes they take towards their professors, lecturers, tutors, etc. – Perry's observations (as operationalized by Schommer) contain more than a hint of truth. Certainly, students often do arrive at university believing that 'the tutor knows what is right and wrong' before coming to be troubled by such thoughts as 'different tutors expect different things of us' and even 'there is something that the tutors want us to believe, but I'd be darned if I knew what it is'.⁷ However, as an account of the nature of knowledge, the account falls short woefully.

Firstly, Perry's scheme conceives of epistemic development mainly as a rejection of realism in favour of increasingly sophisticated forms of relativism. However, none of the stages that he sketches contains enough detail to form a coherent epistemic position that a real person could hold. Thus, the most developed epistemic position – 'commitment in relativism' – contains elements that hint both at relativism and at a version of fallibilism that would be acceptable to any realist. Philosophically, Perry's stages are at best under-described and at worst confused. Secondly, that realist positions in epistemology are by their nature primitive and forms of relativism or constructivism more developed would seem – to most analytic epistemologists – to get things exactly the wrong way around. Far from being regarded a more sophisticated epistemological position, out-and-out relativism has had a torrid time in analytic philosophy since the 1960's and 70's (the time of Perry's work). (See, for instance, Siegel, 1987, Goldman, 1999 and Boghossian, 2006.) At the very least, one must acknowledge that there are less and more developed realist *and* relativist epistemic positions that one may adopt. Thirdly, Perry sketches reliance on authority to back up one's knowledge claims as epistemically primitive. As decades of work on the epistemology of testimony and expertise attest, however, we are all regularly in the position of having to rely on authority for our knowledge claims; rather than primitive, the epistemic moves involved in learning from others are complex and learning to judge authority or expertise is part of epistemic development. 'Personal epistemology' is clearly not satisfactory *epistemology*.

⁷ Apologies to Perry.

Next to the philosophical confusion, all of the psychological criticisms that can be made of Kohlberg's theory of moral development can also be made of Perry's theory of epistemic development.

Firstly, that epistemic development happens in the stage-like fashion that Perry sketches (or proceeds in stages at all) is unclear. Chandler and Proulx (2010) point out that some writers see the development of quite sophisticated epistemic beliefs as occurring during the preschool years, whereas others hold that proper development only takes place at university level. Either some in the area are '...confusedly calling radically different things by the same name, or someone has obviously gotten their facts badly wrong.' (Chandler and Proulx, 2010: 199) Furthermore, Wildenger, Hofer and Burr (2010: 238 - 9) point out that epistemic development need not be entirely one-directional – the subject may shift from absolutist to relativist views and back again during their epistemic development. (p. 238 – 9)⁸ Both conclusions call into question the assumption that there is a single trajectory according to which people's views regarding belief and knowledge inevitably develop.

Secondly, personal epistemology research focuses on children and young people's *professed beliefs regarding knowledge*. It does not probe their actual epistemic practices or how they genuinely conduct themselves in their thinking, but the extent to which they can articulate some theory about knowledge. In the absence of much training in philosophy, people's views regarding knowledge may be quite haphazard and even contradictory, but this does not make them relativists. Putting the point in a softer way, professed beliefs about knowledge does not necessarily capture how someone reasons. For one thing, they may be inaccurate in how they describe their actual epistemic practices. For another, these practices are not always influenced by professed belief about knowledge, but much more by factors that are unconscious or tacit.⁹

Thirdly (and relatedly), while personal epistemology research may capture something about the individual's professed beliefs regarding belief and knowledge, it does not capture what actual rules or procedures of reasoning someone follows in their academic study or in the course of their everyday thought or problem-solving activity. In order to study that, the personal epistemology tradition would have to study how children and young people approach arguments or problems or would have to study their knowledge and application of (formal and informal) logic. However, empirical investigation of personal epistemology proceeds solely on the level of asking children and young people their opinions on knowledge, belief and authority in the classroom.

Personal epistemology research may contain some useful hints regarding how one may investigate students' attitude to knowledge, their attitudes to their teachers and their attitudes to and confidence as independent learners. All of these are arguably very important to understand in gathering whether someone has acquired the virtues of being a good thinker. However, such self-assessment will hardly provide a full picture. Especially to studying the actual reasoning skills and styles that students bring to their thinking the methods of another field of research must be applied. This is the methods of the critical thinking movement.

⁸ Teachers of undergraduate philosophy would find this entirely unsurprising.

⁹ Rest *et al.* (1999: 6) make this point regarding Kohlberg's work, but it applies equally to personal epistemology research.

4. Assessing critical thinking skills

The study of thinking and how to improve it is most associated with the critical thinking movement. Building on Dewey's work on reflective thinking, authors such as Glaser, Ennis and Paul have identified basic critical thinking abilities (such as recognising arguments, analysing them, finding and criticising unstated assumptions, etc.) and proposed methods for measuring and improving students' critical thinking.

Possin (2008: 203 – 4) holds that the field today divides between the extremes of those who hold that critical thinking amounts to facility with (formal and informal) logic and those who hold that it only amounts to thinking in a reflective or self-critical fashion. Some (for instance McPeck, 1981, 1991) hold that critical thinking is highly discipline-specific: there is really no such thing as a general critical thinking ability, but only critical thinking in mathematics, critical thinking in history and so-on. According to yet another view of the subject, associated with the 'dialectical' approach of Van Eemeren and Grootendorst, critical thinking amounts to being skilled at persuasion, conflict resolution or debate. (Possin, 2008: 204) On this approach, critical thinking may be more akin to rhetoric than logic.

A consensus view is found in the Delphi Report (1990), according to which critical thinking is:

'... purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based.'

A number of tests of critical thinking exist.

Test	Format	Abilities tested
<i>Watson-Glaser Critical Thinking Appraisal Test</i>	Multiple choice	inference, assumptions, deduction, interpretation, evaluation
<i>Cornell Critical Thinking Test</i>	Multiple choice	deductive reasoning, fallacies, acceptability of premises, inductive reasoning, definition, implicit premise identification
<i>California Critical Thinking Test</i>	Multiple choice	analysis, evaluation, inferences, deductive reasoning, inductive reasoning
<i>Ennis Weir Critical Thinking Essay Test</i>	Essay	getting the point, seeing the reasons and assumptions, stating the point, offering good reasons, seeing other possibilities, responding appropriately to and/or avoiding common errors in

		thinking
<i>International Critical Thinking Test</i>	Essay	skills of identifying, explicating, and using the elements of thought, intellectual abilities, affective traits, intellectual standards ¹⁰

Today, the most commonly used tests of critical thinking ability are the Cornell and California tests. Multiple-choice tests seem to be preferred over essay tests due to issues to do with ease and reliability of scoring.

An important point to note about the standard tests of critical thinking is that they test for critical thinking skill or performance in a specific context – the ‘prompted situation’ of the classroom or exam hall. (Ennis and Norris, 1990; Norris, 2003) The great question is whether this critical thinking ability will also transfer to students’ lives outside the classroom. Put differently, while these tests can show ability at critical thinking, they do not necessarily show propensity to think critically. With this in mind, the critical thinking movement has – since the 1990’s – studied critical thinking behaviours or dispositions in addition to ability at critical thinking.

Studies (for instance Taube, 1997) found that one could distinguish between critical thinking ability and critical thinking disposition empirically and that most of the common multiple choice format critical thinking tests tend to capture the cognitive – or abilities – aspect of critical thinking. Disposition to critical thinking must clearly be studied using different methods. An influential early attempt at studying critical thinking disposition was by Facione and Facione (1992) who used a self-report inventory of their own design – the *California Critical Thinking Skills Dispositions Inventory* – to study respondents’ disposition to think critically. Through factor analysis of responses, Facione and Facione found that the disposition to critical thinking comprises the following separate dispositions: open-mindedness, inquisitiveness, systematicity, analyticity, truth-seeking, critical thinking self-confidence and maturity. (Facione, Sanchez and Facione, 1994)

More recently, Perkins *et al.* (2000) suggest a ‘dispositional view of intellectual traits’. Perkins *et al.* point out that most accounts of intellectual performance and how it is cultivated are ‘abilities-centric’. Such accounts pay attention to matters like how the brain enables certain intellectual performances, what the cognitive processes and meta-processes are that underpin a certain level of performance or what thinking strategies or skills effective thinkers adopt. However, Perkins *et al.* point out that real-world intellectual performance is often determined by something else besides pure ability. In real life – or ‘in the wild’ – our intellectual efforts have to be rationed and martialled to achieve our most important goals: matters like whether people are sensitive to real-life opportunities for thinking and – to be honest – can be bothered to think well often determines real-life thinking performance over and above pure ability.

¹⁰ A somewhat vague list. The authors of the ICTT prefer a broad definition of critical thinking. See <http://www.criticalthinking.org/pages/a-model-for-the-national-assessment-of-higher-order-thinking/591>

Ennis (1991) and Facione and Facione (1992), for instance, recognise this and have begun to outline what the dispositional ‘habits of mind’ are that drive effective thinking. They propose thinking dispositions such as open-mindedness, reasonableness, curiosity and metacognitive reflection as being of particular importance. (Perkins *et al.*, 2000: 272) Perkins *et al.* propose that all of these thinking dispositions have three components: the (i) ability, (ii) inclination and (iii) sensitivity to reason in a certain way. (2000: 273 – 4) In a series of studies, they find that the three elements are empirically separable. They suggest, furthermore, that shortcomings in thinking performance can often be attributed to low sensitivity (that is, not spotting opportunities for thinking or not spotting problems in someone’s thinking) rather than low inclination (or motivation) to think.

Ritchhart (also co-author on the Perkins *et al.* study) developed this work on intellectual dispositions into an account that comes very close to a virtue account. For Ritchhart (2002: 27 - 30) the main dispositions that make up ‘intellectual character’ are:

1. Open-mindedness
2. Curiosity
3. Metacognition
4. Seeking truth and understanding
5. Strategic thinking and
6. Scepticism

Ritchhart (2002: 37) also proposes a four-component model of an intellectual disposition (in place of Perkins *et al.*’s triadic model):

- Inclination
- Awareness
- Motivation
- Ability

Out of all the work canvassed so far, Perkins *et al.*’s and Ritchhart’s component models of intellectual dispositions come closest to describing what one may call ‘intellectual virtues’. Not only do they describe the dispositions that they identify with virtue-terms, but the component models also come close to a contemporary understanding of what a virtue is – an amalgam of sensitivity, emotion, reasoning ability and motivation. Perkins *et al.*’s empirical results are also encouraging, although it must be noted that their methods are intensive and would be hard to translate into a simple test.¹¹

Despite the promise it shows, one aspect of the critical thinking dispositions research should concern virtue epistemologists and that is that, in calling what they are interested in ‘dispositions’ authors such as Perkins *et al.* and Ritchhart risk misdescribing what they are interested in.

The philosophical literature on dispositions is extensive. Dispositions talk is best suited to describing the tendency for a physical material to behave in a certain way

¹¹ Likewise, the methods Norris (1989) proposes for a valid method of testing critical thinking that transfers out of the classroom (or ‘generalises’) – a ‘think aloud’ critical thinking interview – is intensive.

under certain conditions. Thus, a material is soluble iff it dissolves when placed in water (Carnap) is malleable iff it can be shaped, fragile iff it shatters given a certain impact, etc. Dispositions describe law-like regularities of the form:

An object is disposed to D in condition C iff it would D if it were the case that C.

The law in question is a natural law – it is a consequence of the physical make-up of the object that it behaves as it does in condition C.

Intellectual dispositions are different. Even people who are *models* of intellectual virtue are not *always* open-minded, curious, careful or the like. Whether they display these virtues depends on context for one thing and for another, the intellectually virtuous person may (perhaps in a weak moment) deviate from virtuous conduct while still being virtuous. Moreover, what makes it the case that a certain person is open-minded, curious, careful or the like is not a natural regularity but something about how the person is motivated to think. What makes it the case that that person is, say, curious, is that they are motivated by a love of truth and are therefore compelled – in a rational, not a physical sense – to investigate very many different things with a genuine interest in order to get their answer. Furthermore, exercises of intellectual dispositions are *normatively evaluable* in a way that physical dispositions are not. The point is that it is *rational or good* for the intellectually virtuous person to be curious or sceptical in the appropriate situation. By contrast but it is not rational or good for sugar to dissolve in water or for any other physical material to possess the physical dispositions that it does... the rational or normative evaluation simply does not apply to sugar.

For all of these reasons, thinking of intellectual character traits as mere dispositions obscures what they really are: developed propensities on the part of a person to bring their intellectual abilities to bear on certain problems in a certain way, while being motivated by the rational end of finding out the truth and being subject to normative evaluation. In a phrase, the term that Perkins *et al.* and Ritchhart should adopt to describe what they are interested in is ‘intellectual virtue’.¹²

One possible reason why no psychological study of the intellectual virtues has been conducted is that the concept of virtue has had a hard time in psychology. Much 20th century psychology regarded virtue – with its overtones of not only the description of people or their behaviour but also moral evaluation built in – as an unscientific notion. What can we learn from the psychology of virtue regarding how *intellectual* virtue can be studied?

5. Assessing moral virtue

Efforts to understand people in terms of their unique character or persona date back to antiquity. Plato held that the cardinal virtues that a person could exhibit are wisdom, courage, temperance and justice (or fairness). Aristotle had a more capacious conception of virtue; in the *Nicomachean Ethics*, for instance, he proposes courage, temperance, liberality, magnificence, magnanimity, proper

¹² Johnson and Blair (1991: 50) for instance, already hold that ‘...critical thinking denotes a moral/intellectual virtue...’

ambition/pride, patience/good temper, truthfulness, wittiness, friendliness, modesty and righteous indignation, and in the *Rhetoric* he adds more virtues of an emotional nature. Thinking about individual differences between people in terms that would have been recognisable to ancient thinkers was prevalent in psychology until the early 1900's, until the seminal work of Allport (1937). Allport regarded the concepts of 'character' or 'persona' as unscientific. Specifically, Allport regarded 'character' as a notion that was too value-laden. Allport sought a way to describe individual differences that was stripped of any value-laden terms and alighted on the notion of 'personality'. Starting with dictionary definitions of words used to describe personality, researchers after Allport (e.g. Cattell) refined the resulting long list of descriptions of peoples' personality (through the process of factor analysis) to five 'factors' (or facets or components) that tend to cohere internally. The research programme resulted in the 'Big Five' classification of a person's personality in terms of the broad categories of how

- open
- conscientious
- extroverted
- agreeable and
- non-neurotic

a person is.

For the purposes of moral or intellectual *education*, the usefulness of personality psychology is limited to describing people as they already are. If personality is entirely a non-moral matter, then there is no urgency – or, indeed, point – to attempting to shape pupils' personalities – students' personalities simply are as they are.

However, it has been argued (Cawley *et al.*, 2000) that personality has a moral dimension after all. It is, for instance, something we admire or approve of when someone is open, is conscientious or is agreeable... moreover, these are ways that we want children to be. This has led to a reevaluation of the notion of virtue psychology and attempts to study virtue from within personality psychology. Cawley *et al.* constructed a virtue scale containing 140 items testing for 140 distinct character traits or virtues. Having collected data about the extent to which people regard themselves as holding these character traits, Cawley *et al.* found that the 140 character traits could be classified under four main 'virtues':

- Empathy
- Order
- Resourcefulness
- Serenity

While Cawley *et al.* make a good case for the psychological study of virtue, the usefulness of their virtue classification scheme for moral education is limited. The four virtue terms that they identify are very general and 'order' is not a virtue term as we ordinarily understand it. Cawley *et al.*'s list of four virtues also does not mention virtues like honesty, courage or justice that we find important from a theoretical perspective.

A much more developed account is found in Peterson and Seligman. (2004) Peterson and Seligman distinguish six virtues and 24 'character strengths'. The character strengths are not virtues themselves, but are

'...the psychological ingredients – processes or mechanisms – that define the virtues. Said another way, they are distinguishable routes to displaying one or another of the virtues.' (Peterson and Seligman, 2004: 13)

Peterson and Seligman derive their virtues classification from a study of philosophical and religious texts in the Western, Indian and Chinese traditions. (2004: 33 – 52) According to Peterson and Seligman, the 'high six' virtues that can be identified in all of these traditions are: courage, justice, humanity, temperance, transcendence and wisdom. Full empirical support for the Peterson and Seligman classification is, however, lacking. Peterson and Seligman acknowledge themselves – based on their own factor analysis – that a classification into five rather than six virtues would be better supported. (2004) Other empirical studies also cast doubt on Peterson and Seligman's classification.¹³ Nettle *et al.* (2011), for instance, conclude that the virtues Peterson and Seligman identify do not have the hierarchical structure (of virtues composed of strengths) that Peterson and Seligman think.

The most widely used measure in the empirical study of virtue is the Values in Action Inventory of Strengths (VIA-IS) associated with the Peterson and Seligman research programme. The VIA-IS is a 240 item measure with 10 items designed to measure the extent to which a respondent possesses each of the 24 character strengths identified by Peterson and Seligman.

Sample items from the VIA-IS

Creativity When someone tells me how to do something, I automatically think of alternative ways to get the same thing done

- Very much like me
- Like me
- Neutral
- Unlike me
- Very much unlike me

Curiosity I am never bored

Open-mindedness I make decisions only when I have the facts

Love of learning I always go out of my way to attend educational events

Perspective People describe me as 'wise beyond my years'

(adapted from Peterson and Seligman, 2004: 629, table 28.1)

¹³ See Haslam *et al.*, 2004, Macdonald *et al.*, 2008; for an overview, see Nettle *et al.*, 2011

One problem with current tests of character (such as the Values In Action Inventory of Strengths or VIA-IS associated with the positive psychology movement) is that it is a self-report test. Respondents are asked to rate themselves on the extent to which they possess a certain character strength. Self-report measures like these are susceptible to both conscious deception on the part of the respondent and non-conscious deception; it is also in the nature of character that people are not always the best judges of their own character. This insight is not new and Aristotle, for instance, held that virtue must be evaluated from the outside by wise judges. (Kristjansson, forthcoming) Other methods for studying moral virtue shed more light, but are, naturally, more intensive. Such methods include observation, interviews, biographical or diary methods and peer-report methods (e.g. Hawkins *et al.* 2007).¹⁴

Importantly, most of these methods rely on a third-person perspective on the person being studied. In her critique of situationism about virtue, Snow (2010: 21 – 5) relies on observational studies by Shoda, Mischel and Wright (1994) to show that it is possible to identify personal traits that issue in consistent behavioural reactions as long as these take into account how the person being studied perceives the situation. Similarly, she appeals to work by Ford and Tisak (1983) to show evidence for the existence of ‘social intelligence’ and evidence that, next to academic intelligence, social intelligence has a separate effect on behavioural effectiveness. Again, Ford and Tisak could only establish this through using very many different methods in tandem, including self-reports, but also peer and teacher reports, an empathy test and interviews. (Snow, 2010: 66)

6. Assessing intellectual virtue

What are the implications for the assessment of the intellectual virtues? Let me draw out the main lessons.

In section 2, I argued that we have to be able, at least in principle, to assess intellectual virtue, but that it is natural to doubt whether we will always know in detail exactly how virtuous our individual students have become as a consequence. We could get a rough idea based on a sample of virtues from which we can infer virtue, but this is unlikely to have the level of detail needed for summative assessment of individuals. The upshot is that assessments of intellectual virtue should be designed to demonstrate that our teaching is effective, but not necessarily to grade individual students. Bearing in mind that it is often the race for grades that lead students to adopt an instrumental approach to their learning, this may be a fortunate result for our efforts to teach them virtue. However, vigilance will have to be maintained in order that whatever assessments we do design are not co-opted by administrators or faculty in attempts to grade intellectual virtue.

In section 3, I raised some serious doubts regarding whether the field of personal epistemology research will enable us to measure the degree to which a person is intellectually virtuous. Stage-models of epistemic development that equate high epistemic development with holding some set of beliefs about the nature of knowledge are unhelpful – especially if the highest form of epistemic development is

¹⁴ I thank Blaine Fowers for drawing this paper to my attention.

to believe in relativism. The point about epistemic development is not that the knower should come to have beliefs about knowledge; rather the point is that the knower conducts herself in certain ways, epistemically speaking. It is not beliefs about knowledge that are important, but actual conduct in thought. That said, elements of the personal epistemology research may be helpful in understanding students attitudes to learning and to their teachers, specifically.

In section 4, I used this insight to begin to explore the great educational tradition concern with conduct in thought, the critical thinking tradition. The critical thinking tradition is miles ahead of the intellectual virtue movement in this sense: The critical thinking tradition has a clear view on what critical thinking skills are. It contains clear instructions on how to teach for these skills in routines of improving critical thinking. It also has available a range of tests to test for them. By contrast, the intellectual virtue tradition only has available a list of intellectual virtues and the injunction that the teacher should show examples of these virtues in action and then encourage children to practice them. The learning model is one of showing paradigmatic examples and then practice. But what if the paradigm isn't comprehended or does not take hold? We cannot break down the task into easier sub-tasks as the critical thinking movement can. However, intellectual virtue does have one shining strong point: it provides a way to think about the problem of transfer. The classical problem with critical thinking is that we do not know that it always transfers out of the classroom and evidence is mixed regarding even whether students use the critical thinking skills that they have mastered in classes such as Logic 101 during the rest of their studies. Whatever test of intellectual virtue we adopt, it will have to have a clear focus on the issue of transfer of newly acquired intellectual abilities to situations outside the classroom.

In section 5, I briefly reviewed some empirical research on moral virtue in an effort to show both that one may study virtue empirically and how it may be done. While the most obvious method is to have the respondent report on what they perceive as their own virtues, space must be made for a third-person approach as well. This is, however, naturally more intensive and does not lend itself easily to a traditional paper-and-pencil format.

Throughout, I have attempted to make clear that what will ultimately determine the shape of an assessment strategy for intellectual virtue is the *purpose* to which the assessment will be put. While some form of scientific study of the intellectual virtues that can illustrate how intellectual virtue shapes individual cognitive conduct will no doubt be wonderfully exciting, the best model on which to build such a study – the scientific study of moral virtue – is in its infancy. Furthermore, we know the trouble that summative assessments and grades get us into. While the strategies that I have canvassed in sections 3, 4 and 5 are often taken from the scientific study of virtue and sometimes designed for individual assessment, the real challenge will be to adapt the form of assessment contemplated for the purpose we have in mind, which is to show that the intellectual virtues have a clear place in the classroom and that a teaching strategy designed around them works.

References

- Alfano, M. 2012. 'Expanding the Situationist Challenge to Responsibilist Virtue Epistemology'. *Philosophical Quarterly* 62 (247), 223 – 49.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Boden, C. J. 2005. *An Exploratory Study of the Relationship Between Epistemological Beliefs and Self-Directed Learning Readiness*. PhD-thesis, Kansas State University.
- Boghossian, P. 2006. *Fear of Knowledge*. Oxford: OUP.
- Cawley, M.J., Martin, J.E. and Johnson, J.A. 2000. 'A Virtues Approach to Personality'. *Personality and Individual Differences* 28, 997 – 1013.
- Chandler, M.J. and Proulx, T. 2010. 'Stalking Young Persons' Changing Beliefs About Belief'. In: Bendixen, L.D. and Feucht, F.C. *Personal Epistemology in the Classroom: theory, research and implications for practice*. Cambridge: CUP.
- Curren, R. 2004. 'Educational Measurement and Knowledge of Other Minds'. *Theory and Research in Education* 2 (3), 235 – 53.
- Davis, A. 1998. *The Limits of Educational Assessment*. Oxford: Blackwell.
- Doris, J.M. 2002. *Lack of Character: personality and moral behaviour*. Cambridge: CUP.
- Elgin, C.Z. 2004. 'High Stakes'. *Theory and Research in Education* 2 (3), 271 – 81.
- Ennis, R.H. 1991. 'Critical Thinking: a streamlined conception'. *Teaching Philosophy* 14, 5 – 25.
- Ennis, R.H. and Norris, S.P. 1990. 'Critical Thinking Evaluation: status, issues, needs'. In: Algina, J. and Legg, S.M. (Eds.) *Cognitive Assessment of Language and Math Outcomes*, Norwood, NJ: Ablex Press, pp. 1–42
- Facione, P.A. 1990. *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*. ('The Delphi Report') Millbrae, CA: The California Academic Press.
- Facione, P.A. and Facione, N.C. 1992. *The California Critical Thinking Dispositions Inventory (CCTDI): and the CCTDI test manual*. Millbrae, CA: The California Academic Press.
- Facione, P.A., Sanchez, C.A., Facione, N.C. and Gainen. 1995. 'The Disposition Towards Critical Thinking'. *Journal of General Education* 44, 1 – 25.
- Feucht, F.C. and Bendixen, L.D. 2010. 'Personal Epistemology in the Classroom: a welcome and guide for the reader'. In: Bendixen, L.D. and Feucht, F.C. *Personal Epistemology in the Classroom: theory, research and implications for practice*. Cambridge: CUP.
- Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality*, 75, 825-862.

- Ford, M. and Tisak, M.S. 1983. 'A Further Search for Social Intelligence'. *Journal of Educational Psychology* 75, 196 – 206.
- Goldman, A. 1999. *Knowledge in a Social World*. Oxford: OUP.
- Harman, G. 1999. 'Moral Philosophy Meets Social Psychology: virtue ethics and the fundamental attribution error'. *Proceedings of the Aristotelian Society* 53, 709 – 27.
- Haslam, N., Bain, P. and Neal, D. 2004. 'The Implicit Structure of Positive Characteristics'. *Personality and Social Psychology Bulletin* 30, 529 – 41.
- Hawkins, A. J., Fowers, B. J., Carroll, J. S. and Yang, C. 2007. 'Conceptualizing and Measuring Marital Virtues'. In: Hofferth, S. and Casper, L. (Eds.) *Handbook of Measurement Issues in Family Research*. Hillsdale, NJ: Erlbaum.
- Irwin, T.H. 2000. "Ethics as an Inexact Science: Aristotle's Ambitions for Moral Theory," in Hooker, B. and Little, M. *Moral Particularism*, Oxford: Clarendon Press.
- King, P.M. and Kitchener, K.S. 1994. *Developing Reflective Judgement: understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey Bass.
- Macdonald, C., Bore, M.R. and Munro, D. 2008. 'Values in Action Scale and the Big 5: an empirical indication of structure. *Journal of Research in Personality* 42, 787 – 99.
- McPeck, J. 1981. *Critical Thinking and Education*. New York: St. Martin's Press.
- McPeck, J. 1991. *Teaching Critical Thinking*. London: Routledge.
- Noftle, E. E., Schnitker, S. A., & Robins, R. W. (2011). 'Character and Personality: connections between positive psychology and personality psychology. In: Sheldon, K., Kashdan, T. and Steger, M. (Eds.) *Designing the Future of Positive Psychology: taking stock and moving forward*. New York: Oxford University Press, 207-227.
- Norris, S.P. 2003. 'The Meaning of Critical Thinking Test Performance: the effects of abilities and dispositions on scores'. In: Fasko, D. (Ed.) *Critical Thinking and Reasoning: current research, theory and practice*. Cresskill, NJ: Hampton Press.
- Perkins, D., Tishman, S., Ritchhart, R., Donis, K. and Andrade, A. 2000. 'Intelligence in the Wild: a dispositional view of intellectual traits'. *Educational Psychology Review* 12 (3), 269 – 93.
- Perry, W.G. 1970. *Forms of Intellectual and Ethical Development in the College Years: a scheme*. New York: Holt, Rinehart, and Winston.
- Peterson, C. and Seligman, M.E.P. 2004. *Character Strengths and Virtues*. Oxford: OUP.
- Possin, K. 2008. 'A Field-Guide to Critical Thinking Assessment'. *Teaching Philosophy* 31 (3): 201 – 28.

Ramberg, B. 2001. 'What Davidson Said to the Sceptic or: anti-representationalism, triangulation and the naturalization of the subjective'. In: Kotatko, P, Pagin, P. and Segal, G. *Interpreting Davidson*, Stanford: CSLI Publications.

Ravitch, D. 2010. *The Death and Life of the Great American School System*. New York: Basic Books.

Rest, J., Narvaez, D., Bebeau, M.J. and Thoma, S.J. 1999. *Postconventional Moral Thinking: a neo-Kohlbergian approach*. Mahwah, NJ: Lawrence Erlbaum.

Ritchhart, R. 2002. *Intellectual Character: what it is, why it matters and how to get it*. San Francisco: Jossey Bass.

Schoda, Y., Mischel, W. and Wright, J.C. 1994. 'Intraindividual Stability in the Organization and Patterning of Behaviour: incorporating psychological situations in the idiographic analysis of personality'. *Journal of Personality and Social Psychology* 67 (4), 674 – 87.

Schommer, M. 1990. 'The Effects of Beliefs about the Nature of Knowledge on Comprehension'. *Journal of Educational Psychology* 82, 498-504

Siegel, H. 1987. *Relativism Refuted: A Critique of Contemporary Epistemological Relativism*. Dordrecht: D. Reidel Publishing Company.

Snow, N. 2010. *Virtue as Social Intelligence: an empirically grounded theory*. London: Routledge.

Taube, K.T. 1997. 'Critical Thinking Ability and Disposition as Factors of Performance on a Written Critical Thinking Test'. *Journal of General Education* 46 (2), 129 – 64.

White, J. 1999. 'Thinking about Assessment'. *Journal of Philosophy of Education* 32 (2), 201 – 11.

Wildenger, L. K., Hofer, B.K. and Burr, J.E. 2010. 'Epistemological Development in Very Young Knowers.' In: Bendixen, L.D. and Feucht, F.C. *Personal Epistemology in the Classroom: theory, research and implications for practice*. Cambridge: CUP.

Winch, C. and Gingell, J. 1996. 'Educational Assessment: reply to Andrew Davis'. *Journal of Philosophy of Education* 30 (3)