

Pluralist economics curricula: do they work; and how would we know?

Andrew Mearman
Department of Accounting, Economics and Finance
UWE, Bristol
Email: Andrew.Mearman@uwe.ac.uk

Abstract:

This paper aims to illuminate the debate on pluralist economics curricula by examining ways in which such curricula are evaluated. The paper argues for pluralism as a general approach and as pedagogy. It argues that there is a plurality of pluralist curricula. It further argues that pluralist curricula have multiple goals, implying multiple criteria for their success. The paper then examines the potential for experimental methods to evaluate pluralist curricula. It is argued that for general and specific reasons, experimental methods are unlikely to be illuminating. Rather, the paper argues for a mixed approach, entailing different types of data, different metrics of success, and a variety of data collection and analysis methods. It is claimed finally that the existing (albeit scant) evidence on pluralist curricula already employs many of the principles of pluralist evaluation.

Keywords: pluralism, economic education, mixed methods, evaluation, experimental methods

JEL codes: A20, A22, B4, B5, C80, C9

Pluralist economics curricula: do they work; and how would we know?¹

Introduction

Pluralist Economics curricula are those that somehow combine multiple perspectives in Economics, usually in order to stimulate critical thinking and open-mindedness. Several authors have argued for pluralist curricula (see Moseley et al. 1991; Feiner and Roberts, 1995; Denis, 2009; Dow, 2009; Garnett, 2009; Freeman, 2009; Ferber, 1995; Feiner, 2002; Knoedler and Underwood, 2003; Underwood, 2004; Raveaud, 2009; Garnett and Mearman, 2011). It is claimed, *inter alia*, that pluralism provides students with a wider range of tools with which to understand a complex world, alerts them to the contested and uncertain nature of knowledge in that world, equips them better to make decisions and may subsequently improve their life chances (see O'Donnell, 2009). These specific pedagogical benefits of pluralism augment general claims about the benefits of pluralism as providing better Economics than that delivered by *any* monist programme (see Garnett et al. 2009). Pluralism, in short, has many connected benefits.

Interested critics have found such arguments convincing to varying degrees, but have recently asked for evidence to support the theoretical claims. However, while the theoretical arguments are coherent, sophisticated and well advanced,

¹ A version of this paper was a plenary address to the conference of the International Confederation of Associations for Pluralism in Economics (ICAPE), University of Massachusetts-Amherst, on 11 November, 2011. A subsequent version was delivered at UWE staff seminar on 15 March, 2012. I should like to thank participants for their comments. Thanks to Tony Flegg for comments. The usual disclaimer applies.

their empirical support is limited. Garnett and Mearman (2011) tried to address this lacuna, but evidence remains scant.

If evidence is to be sought, what should it be? One possible strategy is to use experimental or quasi-experimental methods, augmented by statistical analysis of quantitative data. This approach dominates in the literature on Economics education; but is flawed and has stimulated appeals for more flexibility in evaluation of educational innovations (see Davies and Guest, 2010). An alternative approach is based on pluralism, realistic evaluation and case-based enquiry (see Pawson and Tilley, 1997; Byrne and Ragin, 2009). These literatures suggest multiple meanings of what it means for educational interventions to 'work', how – and indeed if - this might be assessed conclusively.

This paper argues for a cautious approach to evaluation, which recognises the complexity inherent in evaluating the impact of a multi-dimensional intervention, with multiple possible types of effects, carried out in open environments. It argues for a pluralist approach to the evaluation of pluralist curricula; which avoids bold claims that an intervention has 'worked' or not. These arguments are rooted in pluralist epistemology and ontology.

The paper addresses a series of questions, around which the paper is organised: what is pluralism? Why might it be beneficial: both generally and in educational terms? How might we evaluate if pluralist curricula 'work'? What does it mean to 'work'? What are the pros and cons of experimental methods in trying to understand its efficacy? How else might we investigate the impact of pluralist curricula on students? Next, the basic arguments for pluralism will be laid out, and a

brief discussion of evidence for pluralist curricula is offered. Following that, experimentalist methods are critiqued, and finally a range of other options for evaluating pluralist curricula is discussed.

Pluralist curricula

Pluralism can be defined as an advocacy of plurality. However, plurality operates at many levels, including ontology, epistemology, theory, methodology, method and policy (Mearman, 2008). Moreover, the plural objects can relate in a variety of ways. They may interact, perhaps dialectically; or they may simply co-exist. The way this relation manifests is contingent on the beliefs of the person (or subject group) creating it and the wider institutional dynamics and external constraints on them. Thus, unsurprisingly, there is a plurality of pluralist curricula. All involve a commitment to open up students' minds; usually involve multiple perspectives on Economics; often involve discussing perspectives via debate; and, increasingly, synthesise sets of ideas.

Many of the arguments for pluralist curricula are rooted in claims about pluralism in general. In a complex world, and given cognitive and other epistemological limitations on theories, it would seem beneficial for students to learn multiple theories. As such, teaching pluralistically can yield better understanding of the world and of theories within it. Pluralist knowledge may be useful. Pluralism may involve competition between theories; and thereby better

explanations (perhaps) of recent prominent economic events such as the financial crisis and recession.

Indeed, it could be argued that problem-based approaches to teaching require pluralism because of the nature of reality. Nelson (2009) refers to this as the 'broader questions, bigger toolbox' approach to teaching: students learn by tackling real problems and deploy theory as they need to and see fit. In so doing, students will develop critical thinking, comparative thinking, and crucially skills of judgement. Further, diverse students require diverse methods, and this may include multiple theories.

Moreover, if students think pluralistically about non-academic matters, pluralism may appeal to them; Resnick and Wolff (2011) argue that pluralism accords with the inner turmoil and confusion of many students. As such, pluralism may help attract students to the subject. The student protests which generated the *Post Autistic Economics* movement, plus evidence from *Economics Network* surveys of students (see Mearman, 2007) suggest that there exists a student demand for pluralism.

Economists sympathetic to 'heterodox' economics accept the above arguments quite easily, perhaps because they appeal to their own politics, emotions or psychologies. Further, they support pluralism as an antidote to what they perceive as a less than tolerant environment within Economics. Finally, they may be encouraging their students to take a critical attitude in life and specifically towards existing institutions.

Furthermore, there also seems to be some sympathy for these theoretical arguments in more 'mainstream' circles. Mearman and Webber (2009) presented many of these arguments to the *Developments in Economics Education* conference and were well received. A special issue of the *International Review of Economics Education* was devoted to pluralism. These audiences – motivated by a concern for good teaching – seem persuaded of the *a priori* case for the benefits of pluralist curricula.

However, these same receptive people demanded to know what evidence there is that pluralism works. That demand partly reflects a broader call for evidence in Economics education (Davies and Guest, 2010). At the time that the question was posed, evidence was scant. Barone (1991) told convincing stories about the programme at Dickinson College; Mearman (2007) presented some evidence from student evaluations; O'Donnell (2009) presented other supporting evidence based on student responses. Subsequently, using international online survey data on student perceptions of Economics, and focus groups asking students about the effectiveness of Economics in creating understanding of real world issues, Mearman et al. (2012) found preliminary evidence of the efficacy of pluralism in achieving multiple aims. Their survey data suggest considerable heterogeneity of students. They also found that confusion, if perceived to serve no useful purpose, generated negative perceptions of Economics. Also, the word 'debate' did not appear frequently in survey responses as something students wanted more of. So this evidence is mixed in its support of pluralism.

As Mearman et al. (2011) showed, focus group data were more positive. This rich quotation sums up the case for a pluralist approach:

...I found it fascinating studying it ...we do have a bit of a difficulty [in that] we've been presented with all these different [views] ... But I think it's necessary at this stage to be presented with the alternative views, because we're not studying neoclassical economics, we're studying economics. So I think it is quite broad and I do actually enjoy the slightly conflicting way that some of it's being taught. But I do think that rather than saying this is... "Read this and it will tell you how that works." It's not doing that, it's saying "Read this and it may give you an insight, give you an understanding of some processes." ... But you have to bring a lot to it as well, it's not like studying maths where you can just learn the answers, you have to bring a lot to it yourself, I think, and it is a kind of... it is giving the tools to be able to do things with them rather than giving you the answers...

This quotation conveys that a pluralist approach creates engagement ('I found it fascinating studying it'); the student is happy with ambiguity ('But I think it's necessary at this stage to be presented with the alternative views, because we're not studying neoclassical economics, we're studying economics'); the student recognises the partial nature of knowledge ('Read this and it may give you an insight, give you an understanding of some processes'); the approach stimulates active learning ('You have to bring a lot to it yourself'); and the approach provides ways of thinking about Economics and the economy rather than imposing a view ('It is giving the tools to be able to do things with them rather than giving you the answers').

This evidence is suggestive, but exploratory and highly provisional. Advocates of pluralist education need to build the case more strongly that pluralist approaches can work. Subsequently, Garnett and Mearman (2011) organised a

symposium to mark 20 years since Barone's original article about his *Contending Perspectives* approach. They also appealed for greater use of evidence. Thus they asked their authors to reflect on concrete cases of teaching practice and to provide means of assessing whether those innovations had 'worked'. Specifically authors were asked 'what have our students learned?' Barone (2011), Resnick and Wolff (2011), McIntyre and van Horn (2011), Stilwell (2011), Lapidus (2011) and Amin and Haneef (2011) all presented further supporting evidence. Their evidence was of different types; and their meanings of pluralism were different. This crucial theme will be returned to below.

Thus far we have heard persuasive arguments for pluralism; that some form of evidence is demanded in support; and that so far evidence is rather underdeveloped. Furthermore, we have seen that existing evidence is multi-faceted, drawing on different techniques for data collection and analysis; and reflecting the complex nature of pluralism. Moreover, we have seen that educational goals influence considerations of what 'works'? The remainder of the paper expands on these themes. Later, one common route for evaluating curricula is discussed: experimental methods. Before that, we need to consider briefly what it might mean for curricula to 'work': we examine that question through the lens of educational goals.

What does it mean for curricula to 'work'?

A pluralist position might say that there are no universal standards by which efficacy can be judged. However, the list of benefits above would suggest ways of assessing pluralist curricula. There may be some benefits that can be regarded as held universally, e.g. critical thinking; however, even critical thinking can be rather selective and therefore not achieved fully. For example, some economists may only want students to think critically about real markets (if only they would behave...), governments (if only they would work...), or corporations (if only they were not corporations...). This leaves us in the position in which the efficacy of a programme can be assessed in terms of the instructor's aims. All instructors – or in a programme, each team of instructors – have their own aims, even if they do not know this.

Clarke and Mearman (2001, 2003, 2004) argue that Economics educators must consider their aims. They borrow the analytic separation of aims into liberal and instrumental/ist. Liberal aims are based around the opening of the student mind via critical thinking, independent analysis and comparative thinking, leading to the exercise of autonomous judgement. Instrumental aims tend to focus on more concrete outcomes, e.g. being able to understand x, relate theory y, solve specific problem z. They may go beyond this to aim at employability, control, etc. Liberal aims could be instrumentalist in a sense: by aiming to reproduce a liberal society. Liberal and instrumental aims can be complementary: say, learn a theory in order to evaluate critically it; or in conflict: e.g. indoctrination. Clarke and Mearman (2004) discuss whether an 'employability agenda' is consistent with liberal aims: they conclude that it may be.

Pluralist curricula often have plural aims. Arguably, of course, all curricula do: they try to achieve learning of specific material, learning outcomes defined in terms of critical and analytical thinking, and increasingly, the employability of students. This multiplicity of aims may simply be an artefact of institutional history and current needs. However, it is possible that an instructor intentionally constructs a programme with multiple aims. Furthermore, they may construct aims which synthesise liberal and instrumental aims. An example is the 'practical liberal arts' approach developed by Guarasci (and colleagues) (Guarasci, 2001). Using a Deweyan framework, Guarasci aims for liberal educational outcomes and instrumental outcomes through practical engagement with real situations.

If one has only either liberal or instrumental aims in mind, it may be possible to achieve them by teaching one set of theories in a particular way. For example, it is perfectly possible to teach only neo-classical Economics, as a vehicle for critical analysis (Clarke and Mearman, 2001). However, as Mearman et al. (2011) have argued, pluralist curricula may achieve these aims, and perhaps better than monist curricula, by forcing students to use judgement, and to weigh up alternatives; to offer a wider set of analytical capacities; to allow students to grasp a wider range of material. Crucially, they argue that pluralism imports external critique and insures against non-critical teachers. Additionally, by encouraging debate, pluralist courses teach crucial skills, which may increase productivity, and thereby employability. Thus, liberal and instrumental goals can be achieved simultaneously. These types of claims about pluralist curricula are made across the literature.

Thus: pluralist curricula could 'work' in numerous ways, some of which are consistent with each other, and others which may not be. From the above arguments, it is possible to claim that pluralist curricula are more likely to achieve synthesis between compatible aims. For example, one of the dangers of a monist approach is that indoctrination with an ideological function occurs, and therefore that the liberal aim of autonomous thought is lost. However, with a pluralist course, this is less likely because the student cannot merely learn a single view and thus must question one view in terms of another.

This does assume that pluralism operates at the necessary level: pluralism in theory may be more likely to lead to critical perspectives on theory and the student would develop an appreciation of the limitations of specific theories. Pluralism of methods used in Economics may develop within the student a critical appreciation of economic data, but would be less likely to achieve critical appreciation of theories. However, students might develop a critical appreciation of theory *per se*, particularly if they have been taught with multiple epistemologies or even ontologies. Different types of pluralism generate different types of effects.

Evaluating what works – experimental approaches

A common response to the above issues may be to evaluate educational interventions experimentally. Much of the educational literature on evaluation discusses experimental methods (see, for example, Opie, 2004; Morrison, 2009). Arguably, in Economics education, experimental methods dominate. This

dominance partly reflects the utility of experiments in evaluating teaching, and as a teaching tool. It also reflects a more general growth in experimental activity in Economics, as a means of generating data, simulating real-world scenarios, and testing theories. Furthermore, the appeal of experimentalism is itself rooted in its claims to scientificity. Moreover, as Pawson and Tilley (1997) note, experimental design is often favoured by proponents of evidence-based policy (EBP): they report that Martinson (1974) only included experimental studies in his systematic review of prison reform programmes. However, as the educational literature makes clear, experimental methods may not be appropriate or effective in evaluating interventions.

‘Pure’ experimental design in education and other social sciences is modelled on Campbell and Stanley’s (1963) OXO formula, in which an experimental group and a control group are tested (the Os) before and after an intervention (X) in a process which is only applied to an experimental group.² Ideally, the two groups are identical in terms of key characteristics: their members are ‘matched pairs’. In a pure experiment, the scientist selects members randomly. Furthermore, the scientist creates a controlled environment such that any difference between the two groups (post-intervention) is directly attributable to the intervention. As Morrison (2009) notes, such experiments can take many forms, including laboratory tests of various types.

² As Morrison (2009) notes, the size of the intervention may also be varied across the experimental group. So, a group of people may take different strength variants of a medicine, to check how much is needed to cure the particular ill. Is this applicable to pluralist curricula? As argued above, there are varieties of pluralism; however, it would be extremely difficult to design an experiment within an institution which allowed some students exposure to the very pluralist, others to the medium pluralist, and yet others to the monist. The ethical problems discussed below would be considerable.

In our case, one group (the experimental) would be taught pluralistically. The other group (the control) would be taught a conventional course of Economics. At the end of the course, both groups would be tested. The assessment tool chosen would depend on the goals of instructors and their predispositions to particular data types. So, for example, one option is simply to use a multiple-choice test to assess accuracy of knowledge of theoretical material. If pluralism has positive effects, one might predict that the average exam score of the subject group would be higher than the control group. This might be the case even if the exam tests understanding of mainstream material, since some arguments for pluralism claim that it might lead to better understanding of that. Whatever measure is used, at this point statistical tests are often employed to assess difference (Opie, 2004). It is not unusual to infer causality from these tests: the test scores are the dependent variable and the intervention (or absence of) is the independent variable (Opie). If random sampling has taken place, this causal inference could be extended to the population. Such claims are stronger when supported by meta-analysis, and by replication (Pawson and Tilley, 1997). Such evaluations are often used to support policy-based reform of public sector practices.

Good experimenters know that the above process involves several difficult steps. Not least is that, even under the above design, we cannot avoid what Holland (1986) identifies as the fundamental problem of causal inference: that one person cannot be in the control group and the experimental group at the same time. Even if the experiment could be designed so that the subject(s) were in the control group and *then* the experimental group (or vice versa), what Holland calls causal

transience may occur, i.e. it is possible that being in/out of the control group has an effect which is then felt when the subject is out/in the experimental group (see Morrison, 2009: 159).

Pure experimental design is, therefore, a rather difficult goal. For instance, it may not be possible to set up groups of matched pairs, or even to set up a control group. Even where it is possible to construct a control group, doing so is difficult. For instance, certain types of people will opt into the experimental group – e.g. they might inherently favour or be intrigued by pluralist economics – leading to sample-selection bias. To counteract this ‘volunteer effect’, the experimenter may split up the volunteers into an experimental and a control group (Pawson and Tilley, 1997). This ruse may work well in testing efficacy of medicines. However, it is unlikely to succeed in our case: students who had signed up for a pluralist course would likely recognise they weren’t receiving one, and the experiment would break down.

Additionally, where one group has access to a potentially advantageous intervention, yet another does not, serious ethical issues are raised (Sikes, 2004: 25). Experimenters are aware of these problems, without clear resolution. Campbell (1969) offers an attractive vision of an experimenting society in which there is a social contract whereby experimentees are compensated but also where the benefits of experimentation are shared, and thus the risks too. However, that view might be considered naïve.

Thus for various reasons – many practical – pure experimental design is impossible and alternative designs are tried. For example, two-group tests without random selection are possible; although this raises selection-bias issues again. One

main problem with matched pairs design is that the paired subjects have to be identical in terms of all other variables (or as many as possible). This, of course, is very difficult to achieve, not least logistically, as it would involve a huge data collection process prior to the experiment (Borg and Gall, 1996). In these cases, a non-equivalent pairs, quasi-experimental design may be chosen. As Opie (2004) notes, this is often the only available option in educational research. Lim (1998) used such a design to test the effectiveness of WinEcon. He did some pre-testing of aptitude, and of demographics, to use as control variables in explaining the post-intervention exam scores. Lim's study is an example of a field experiment. Field experiments seek to "achieve sufficient *control* to make the basic causal inference secure" (Pawson and Tilley, 1997: 6, emphasis in original). However, as Morrison (2009: 159) notes, field experiments suffer from many criticisms, such as that the participants are too different to make meaningful comparison possible.³ Partly as a response to that criticism, one-group experiments are done, in which the experimental group is examined pre- and post-test to see if the intervention has made any difference. These designs are clearly flawed but often are all that is feasible in many educational institutions. Consequently, most of the evaluations of pluralist curricula below are done on that basis. Indeed, given the nature of the research into pluralist curricula, the logistical – not to mention pedagogical, theoretical and political – barriers to conducting even one-group evaluations are considerable.

³ Even more *impure* are natural experiments in which no control over events is feasible, but where a policy intervention has occurred and one can try to study its effects. However, in comparison with the *pure* experiment their limitations are clear.

Critiques of experimental approaches

As discussed already, the 'pure' experiment is rendered extremely difficult, if not impossible to achieve, for a number of mainly practical reasons. Experimenters are aware of many of them; hence their design of quasi-experiments. They accept many of the weaknesses of experimental design, yet argue to retain the method, as being superior to the alternatives (Gordon, 1992). Furthermore, as noted, association of experiments with 'science' and with EBP may give them extra power to resist critiques – of which there are many.

Experiments may generate black box theories. For Pawson and Tilley (1997: xv) "...experimentalists have pursued too single-mindedly the question of whether a program works at the expense of knowing why it works". Thus, experimental results are merely patterns of data to be explained. Similarly, for Morrison (2009: 157) "...'what works' may fail to address causation at all; causation is about 'how something works', not only 'what happens'." Moreover, "...judging 'what works' is a matter of values and not only of performance" (Morrison: 172).

Furthermore, interventions are expected to work in all cases; and if not, not at all. Hence, Martinson (1974) reached the conclusion that 'nothing works' because he could not find universally applicable results from the experiments he examined. Yet experimental results often display variety: "the same program will thus work in quite different ways for different subjects and...the experimental method is simply not designed to appreciate such subtleties" (Pawson and Tilley, 1997: 31). A further

problem with Martinson's approach is that, if experimental methods are flawed in essence, then only comparing experimental results with others is in turn flawed: it would be advisable to compare them with results of different types (see Siakantaris, 2000). Relatedly, an experimental approach can lead to biased analysis – experimenters may only look for “confirming evidence, and... not seek, or find, rival, alternative explanations, nor ... seek falsification criteria” (Morrison, 2009: 146). Experimentation can close off possibilities of other hypotheses emerging (Morrison, 155).

Moreover, “...there is no approach that can ever guarantee universal learning success, however success is defined” (Hodkinson et al. 2007: 37). Moreover, short-term interventions may have no effect, or at least no immediate, obvious effect (Morrison, 2009: 141). Causal relationships may suffer temporal complication: smokers do not immediately get lung cancer (Morrison, 2009): ditto for educational interventions. Hence, pluralist courses, which may only run for a term, may not have discernable effects. Indeed, if one of the goals of pluralist curricula is to open minds, the effects of this are unlikely to be obvious in the short term at all, yet, as Opie (2004) points out, the longer the experiment, the more difficult it is to remove unwanted confounding effects of other factors. Indeed, not all smokers contract lung cancer. Not all students in pluralist courses will become intellectually emancipated.

Overall, Pawson and Tilley (1997: 218) conclude that experimental work is somewhat brute – despite claims of sophistication. They claim that evaluation too often proceeds with the experimental method, even if it is not the most appropriate for the question in hand. For them, experimental approaches tend to be driven by

the method, rather than the question at hand. Further, for Morrison (2009: 207) experimental results are driven by their underlying assumptions. Such comments resonate with the generalised critiques of economic methodology associated with Lawson (*passim*), Dow (*passim*) and others.

Underpinning those critiques are fundamental criticisms of the way economists – and in this case experimenters – view the world. The experimental method works on the premise that the scientist can control a situation within an environment to such an extent that they can trigger a cause, to lead to a discernible effect, rendering other causes negligible; and that they can do this a required number of times, leading to a consistent finding. Morrison (2009: 151) says this assumes too much control over the world. In such a world, “personality, creativity and so on are irrelevant” (Opie, 2004: 87). For others, experimentation is scientific or inappropriate naturalism – an over-confident positivism – in which everything can be evaluated (Scriven, 1980) and quantified (Harrison, 1998).

In contrast, many authors argue that the world is complex, exhibiting equifinality, emergence and may comprise causal webs or nets (Morrison, 2009: 206, 211). In such a world, causes may not be separable. Thus, the neat separation of cause and context presupposed by experiments may be infeasible. As Morrison (2009: 107) says, “[a]ny cause or intervention is embedded in a web of other causes, contexts, conditions, circumstances and effects, and these can exert a mediating and altering influence between the cause and its effect”.

It seems that educational environments are of this type, rendering interventions in those environments difficult. The complex educational ontology

includes “teaching, teachers, learners, learning situations, and wider historical, economic, social and political influences” (Postlethwaite, 2007: 161). For James et al. (2007: 11) “...teaching and learning cannot be decontextualized from broader social, economic and political forces, both current and historic, and that addressing this complexity directly is the most likely route to acquiring an understanding that will be useful to policy and practice”. In other words, life outside college cannot be separated from life in college, learning sites cannot be separated (Hodkinson et al. 2007), the individual and the social may not be separable (Hodkinson et al.) and, more prosaically, control and experimental groups may not be separable because of social media (Morrison, 2009: 156).

Crucially, “...striving to control the influence of extraneous factors by random assignation of participants to control and experimental groups...is ill-judged, as this prevents researchers from identifying those very conditions that might be contributing to the success or failure of a programme...i.e., precisely the sort of information that might be useful to policy makers” (Morrison, 2009: 154-5). Specifically, the experimental method designs out people from the process: yet these may be crucial in determining whether the programme works.

As Pawson and Tilley (1997: 33) argue, excluding agents ignores the cognitive leap taken by participants, which is precisely the reason why the intervention works. To be successful, programmes require volition; co-operation (or mindset) may be the crucial factor which determines success (Pawson and Tilley: 36); but by trying to avoid the volunteer effect, volition is designed out. Thus, in educational fora, the dispositions of teachers (James et al. 2007: 12) or students may

be assumed away. However, this is unwarranted, because learning involves a conscious effort, and this may occur because of pre-existing subconscious predispositions, which are themselves changed in the process of learning (Hodkinson et al. 2007: 35). Ultimately, learning is something people do, not something that is done to them (Hodkinson et al.: 34).

So, any effects of pluralism are complexly determined. They include understanding, knowledge, criticality, engagement and the development of other generic skills such as research, use of literature, data analysis, etc. These factors may be thought of as independent; but in reality they are not. As has been argued widely, learning may be said to presuppose some extent of engagement, which as in Zepke (2011) is itself determined by various factors, most notably the approach and abilities of the teacher, the curriculum, the nature of the learning environment, and other factors such as the backgrounds of the students, group dynamics and the like. If we accept that pluralist benefits may take some time to accrue (because they are concerned with deep cognitive and attitudinal tendencies), then the complexity increases, because factors such as the general background cognitive capacities of the student may be changing, as might be the nature of their external environment, their teachers and the subjects they take.

All told, these criticisms suggest that, in the case of evaluating pluralist curricula, experimental methods have serious flaws. There is at least a case for being more flexible in experimental design and combining experimental results with those of different types. At the minimum, experiments can be designed more flexibly, to allow for different types of data, data-collection procedures, and techniques of data

analysis. The intention here is not to abandon all tenets of experimentalism. Indeed, many of the studies cited below do retain a basic feature of experimental design: that there is a subject group which is observed for any effects of the intervention made.

However, if experimental design were abandoned, what else could be advocated? Perhaps predictably, the answer would be: to be pluralist. That follows because many of the arguments above suggest a complex ontology, in which single conceptions of the reality, single theories and single methods of evaluation are unlikely to yield conclusive answers. In such environments, mixed-methods, or one type of pluralist methodology may be desirable (Downward and Mearman, 2007). Specifically also, a variety of research designs may be preferable in assessing pluralist curricula. Significantly, perhaps, attempts to evaluate pluralist curricula, taken collectively, do appear pluralist in approach.

Pluralist approaches to evaluating pluralist curricula

In this section, an argument is continued for a pluralist approach to evaluating pluralist curricula. The arguments here are drawn from connected literatures on case-based research (Byrne and Ragin, 2009), realistic evaluation (Pawson and Tilley, 1997) and mixed-methods research (Downward and Mearman, 2007). These arguments stress that research should be based in real cases, utilise methods that take context seriously, and be nuanced in their claims made.

Case-based methods attach primacy to the case, which involves a process as well as an outcome (Mjøset, 2009). Cases are unique and thus special attention must be paid to the individual history and context of each one. This position may suggest an ideographic (as opposed to nomothetic (law-like)) approach to research, in which inference beyond the case is impossible. However, recent case-based methodology reflects an established tradition of contextualism, i.e. bridging the divide between aiming for knowledge of completely unique cases, and the goal of general theories in which context is removed (Harvey, 2009; Windelband, 1894; Mjøset; Byrne, 2009). This line of research also tries to break down other dichotomies: statistical/historical analysis (Mahoney and Terrie, 2009), individual/aggregate (Mjøset), and significantly, quantitative/qualitative (Byrne, 2009). In this approach, while generalisation is difficult because of causal and conceptual heterogeneity (Mahoney and Terrie), and should be limited, it is recognised that cases are cases of something (Goertz and Mahoney, 2009), and that as such it ought to be possible to build up knowledge by comparing similar cases, using historical case knowledge (Rihoux and Lobe, 2009; Burawoy, 1998). There is some support from the literature on grounded theory for this position (Mjøset). There is even some support for the notion of comparing dissimilar cases: indeed, some authors are interpretable as favouring a form of inference similar to what Keynes called negative analogy, i.e. finding similar conclusions in dissimilar contexts suggests that the finding has greater weight.

A significant strand of the above literature is arguments for mixed-methods research. These arguments are well established. Creswell and Plano Clark (2011)

offer a range of arguments for (and demonstrations of) mixed-methods research. Feminist Economics has made strong arguments for mixing methods. Downward and Mearman (2007) suggest ontological arguments for mixing methods and for pursuing 'triangulation': complex objects demand a combination of data types and analytical techniques, with the exact combination being driven by the nature of the question(s) being asked and the object(s) under study. Standard arguments for mixing methods are that quantitative methods can offer a valuable summary of the object under study but cannot achieve the depth of qualitative; or that all methods are inherently fallible and therefore that no single method of assessing achievement of learning outcomes could be comprehensively successful.

Also, as Pawson and Tilley (1997) note, pluralist approaches are well established in the evaluation literature. They discuss constructivist, pragmatist and realistic evaluation approaches. They argue for realistic evaluation, which employs no one standard 'formula'..." (Pawson and Tilley: xv), but like pragmatist evaluation, to some extent is characterised by a plurality of techniques and "on the craft skills of the researcher" (Pawson and Tilley: 15). However, Pawson and Tilley favour realistic evaluation because it is wedded to a well-developed ontological position, similar to the one discussed above, based on open systems of interacting causes, and the crucial role of context in generating outcomes. As such, "...realistic evaluation can utilise a range of research designs and so can be quantitative or qualitative, action- or outcome-oriented, contemporaneous or retroactive..." (Pawson and Tilley: 182).

The broad educational literature reflects many of these themes. We can recognise the notion that “...a case is an outcome preceded by a process that unfolds in time” (Mjøset, 2009: 47) in the myriad case studies in the literature, which present detailed, specific institutional analysis. James et al. (2007) is one example, which deploys many of the principles outlined above. Their methodology deploys a number of case studies, all of which recognise the specificity of each, yet are linked as cases of learning cultures. However, unlike experimental design, which can universalise, the mixed-methods designs discussed here aim for relatability, i.e. the ability to say something about other cases, where possible. The studies also therefore reflect the programme of realistic evaluation described by Pawson and Tilley (1997).

Mixed-methods approaches are well established in the educational literature: “Rarely is there only one way to go about things. To present research design as being a straightforward, technical matter of ‘horses for courses’, with researchers ‘objectively’ choosing the most appropriate, if not the only possible, methodology and procedures for a specific research project, would be misleading and even dishonest and immoral” (Sikes, 2004: 17-18). Further: “The more the researcher wishes to understand causal process, the more methods in combination are useful, each with their own time frames and timing of data collection” (Morrison, 2009: 169).

James et al. (2007) deploy a combination of qualitative and quantitative techniques. Their approach (see also Postlethwaite, 2007) largely reflects what Creswell and Plano Clark (2011) refer to as a quanQUAL approach: i.e. Quantitative

data are used extensively but qualitative research is emphasised. Postlethwaite (170) describes how a survey would generate quantitative results, prompting further qualitative investigation of the statistically significant ones. Having said that, Postlethwaite (174) stresses a balance between quantitative and qualitative, in that one places a check on the other: “We analysed the qualitative and quantitative data separately then compared insights, finding enough synergy to suggest that findings were not artefacts of one or other method”.

Mixed methods are manifest in the educational literature in various ways. Educational research utilises quantitative data and analysis, for instance through large-scale survey (using questionnaires) data, generating different data types, including bi- and tri-variate response data (Morrison, 2009) analysed using a range of statistical tests. However, the literature also deploys various qualitative data collection and analysis methods (Opie, 2004; Postlethwaite, 2007). For example, qualitative semi-structured interviews (Opie; Soh, 2001), focus groups (Hodkinson et al. 2007: 29), student learning journals (Soh), physical institutional artefacts (Hodkinson et al.), tutor professional journals (Postlethwaite), small group discussions (Opie) and other ethnographic methods (Morrison, 103) are all evident. Opie (77) cites the use of thought sampling, whereby during student group discussions, a bell rings and participants are invited to express their thoughts at that time. Hage and Meeker (1988) use ‘tracers’ to track participants’ development and thoughts on such things as “conformity, freedom, autonomy, staff-student relationships, resources, planning, assessment, teaching strategies, resistance and so on” (Morrison: 105), many of which are aligned with the liberal goals discussed

above. These, and other methods discussed above, fit the longitudinal approach necessary to assess the effects of educational curricula, given the time lags and the nature of the developments involved.

As well as recognising the need for various data types, educational research embraces the benefits of consulting different types of people in different places at different times, because “different students, different tutors, different college managers, employers, parents, and policy-makers, will have differing views about the outcomes that are desired...” (Hodkinson et al. 2007: 36). Postlethwaite (2007) describes how, in one large-scale project, a range of data collection and analysis techniques is employed at a number of different learning sites; for instance, covering different types of institution and types of course (see James et al. 2007; Hodkinson et al.: 25). Postlethwaite also stresses the utility of student perspectives data (171). Morrison (2009) demonstrates a range of reporting methods: another example of mixing.

The logic of mixing also applies to metric used, as already suggested. In the second section, it was noted that frequently in experimental designs, the efficacy of an intervention was tested via comparing exam results from the experimental and control groups. That raises the question of appropriate metrics for success. Given that, as discussed above, in pluralist curricula there tend to be multiple goals, a single metric is unlikely to be appropriate. Let us recall the liberal goals discussed above: analytical, comparative and critical thought. Critical thinking may be assessed through more detailed questions or scenarios, such as research essays or long written answers in exams. At this point the assessor’s judgement evaluates

whether and to what extent the student has demonstrated critical thinking, made effective criticisms, applied the theoretical tools appropriately, or reached an sensible conclusion.

It is possible to try to measure students' intellectual capacities using standardised tests. Intelligence Quotient (IQ) tests are perhaps the most famous example. IQ tests usually test numerical or verbal ability, but can also assess spatial, logical or visual abilities. Thus they may assess a wide range of the cognitive capacities in which we are interested here. However, IQ tests have been criticised as rewarding training, or rewarding those suited to the specific type of test. Kirby et al. (2002) use Bateman and Crant's shortened Personal Proactivity Scale to assess students' ability to use initiative, which could indicate autonomy⁴.

Similarly, there are standardised tests of critical thinking, such as the California Critical Thinking Skills test (CCTST). However, such tests may be subject to the same criticisms as IQ tests; and they also may be problematic to administer since they may take 50 minutes to run. If one ran them at the beginning and end of a semester-long course, they would take a large chunk of the time available for other activities. Unless they replace the main assessment tool, they may also cause assessment fatigue in students. Kirby et al.'s (2002) study shows these issues writ large: in order to construct reliable controls so as to measure proactivity reliably, they had to administer three additional tests, plus gather data on students' prior performance and experience. For many research studies, these additional requirements may be prohibitive.

⁴ This scale is a Likert type. Kirby et al. (2002) used a 7-point version of the scale in conjunction with control variables such as general mental ability which might affect the score on this scale.

Cottrell (2011) presents a number of exercises to develop critical thinking, which could be embedded into curricula, also as tests or measures of critical thinking ability. The tests she provides vary. Some of the tests are akin to those found on IQ tests: tests of sequence, logic, difference, opposition and similarity. However, these exercises are designed to be developmental and it would not be appropriate to use them as before and after tests of critical thinking. Other exercises used by Cottrell may be more fruitful, such as those which ask students to identify (often hidden) assumptions, and assess the validity of given syllogisms. More detailed exercises may involve reading passages (or even longer pieces) with a critical evaluation checklist. These are useful exercises but it is not clear how one might deploy them as before and after tests. One would expect that all students would perform better on such tests as they progress throughout a term; and that they would usually perform better if they did such a type of test a second time. It might be possible in a two-group experiment to test whether one group did relatively better. Again, there may be logistical barriers to this option: they take too long. There is nothing to stop any instructor from embedding such tests in their courses: they are not the preserve of pluralist teachers.

Some instructors eschew examinations as being unjustifiably weighted towards tests of memory, preparation, or training on how to react in those specific situations. Their other weakness in assessing developments in intellectual capacities is that they are static photographs of the student's mind: it is difficult to impute into them some notion of intellectual trajectory. This is clearly a problem in pluralist curricula, because, as argued above, their effects may be long in gestation. For this

and other reasons, some commentators and instructors (as discussed above) favour learning portfolios or diaries, which can record a student's thoughts – and crucially their thought processes – longitudinally through a module of study. To the extent that the instructor manages these portfolios, they can mimic continuous assessment coursework: for instance the portfolio could comprise simply a set of written answers to questions. This could include asking students to deal with similar questions of increasing difficulty. Students' development would be mapped by their marks and the feedback they received on their work. In the second case, if students lose marks for different types of negative aspect of their work, this suggests a process of change within them. In the case of the learning diary or portfolio, the instructor will still manage the construction of them to some extent; however, they also have the potential to be much more free and unpredictable in their form and content. In this case, the instructor has to engage in difficult qualitative analysis of the contents of the portfolio.

Existing work on pluralist curricula – revisited

Taken collectively, the literature evaluating pluralist Economics curricula appears to exemplify the above principles of pluralist evaluation and case-based research. All of the existing studies provide detailed institutional analysis of specific cases: pluralist curriculum studies tend to do this, partly through necessity. Though each one provides specific findings, it is possible for us to take the studies collectively to reach cautiously general findings, whilst recognising the nuance in each.

The studies generate a range of findings. Most of them, perhaps unsurprisingly, make claims about the efficacy of pluralist curricula. Students become more critical or questioning (Barone, 2011; Stilwell, 2011; McIntyre and vanHorn, 2011), more engaged (Stilwell; Resnick and Wolff, 2011; Barone, 2012) and more aware of multiple criteria for analysis (Barone, 2011) and of ethics and values (Barone, 2011; Stilwell; Resnick and Wolff; Lapidus, 2011). Students in pluralist curricula are also claimed to be happy with ambiguity (Stilwell; McIntyre and van Horn) and complexity (Stilwell). These are important claims as they address Earl's (2000) discussion of the potential to generate confusion in students in pluralist curricula. In addition to these claims of the achievement of liberal aims, specific instrumental benefits are claimed. For instance, that students' employability improves (Amin and Haneef, 2011; Stilwell; O'Donnell, 2009) as does their problem-solving ability (McIntyre and van Horn; Lapidus). They also gained specific insights from different schools of economic thought, such as about the importance of social structure in generating economic outcomes (Barone, 2011; Stilwell).

However, despite all these positive claims, these studies do not claim universal efficacy of pluralist curricula, and they acknowledge the barriers to achieving the desired aims. In that way, they echo Earl (2000). They all also recognise clearly the importance of institutional constraints and context. Thus, they reflect pragmatic and realistic evaluation and case-based method. For example, Lapidus (2011) is very thorough on the ways in which her pluralist curriculum did and did not work, and recognised its limitations. Barone (2012) acknowledged for example that in pluralist curricula, students do still tend to take on instructor biases.

This is perhaps unsurprising, but runs contrary to the claim that pluralism builds in criticism and removes the bias inherent in monist curricula.

Many of the studies are quasi-experimental, as they are able to compare a group on a pluralist course with a group doing a conventional course. However, inevitably some of the decisions on evaluation reflect necessary pragmatic responses to institutional constraints, or pedagogical assumptions, or perhaps even are accidental. For example, Barone (2011) acknowledges that his evaluation approach was driven by the lack of a formal assessment process.

Multiple goals are clear in these pluralist curricula, as stated explicitly by many (see McIntyre and van Horn, 2011; Lapidus, 2011; O'Donnell, 2009; Earl, 2000). Some studies make explicit reference to the liberal and instrumental goals discussed above (see Barone, 2011; Stilwell, 2011). Amongst the specific goals pursued are skills (Barone; O'Donnell; McIntyre and van Horn), often for employability (Amin and Haneef, 2011) (O'Donnell); learning criticality (Barone); engagement in community (Barone); cross-cultural competence (Barone); the recognition of contradictions (Lapidus); and exposure to original texts (McIntyre and van Horn). Furthermore, these goals are targeted in a range of course types, sometimes introductory (Stilwell; Lapidus), intermediate (Barone, 2011, 2012) or upper level (O'Donnell). Many of them were explicitly called 'contending perspectives'.

To explore these courses and their effects, a large range of data sources was used. Often these were elements of the (often mixed) assessment schema. Specific assessment tools deployed include student essays (Stilwell, 2011; Lapidus, 2011),

student evaluations (Lapidus); interpretive discussion (McIntyre and van Horn, 2011); thought papers; and factual checks (McIntyre and van Horn). In addition, special additional evaluation methods were used, for instance student questionnaires (Stilwell; O'Donnell, 2009; Barone, 2012); tracer studies of graduate employability (Abdul Rahman et al. 2005; Amin and Haneef, 2011); student evaluations (Lapidus) and other formal student feedback (Stilwell); and outside reviewer reports (Barone, 2011). Several studies invoke informal methods, such as anecdotal evidence (Barone, 2011), informal discussions with students (Barone, 2011; Lapidus; Earl, 2000), and informal contacts with alumni (Barone, 2011; McIntyre and van Horn). Many studies draw heavily on tutors' reflections (Barone, 2011; Resnick and Wolff, 2011; Stilwell, 2011; Lapidus), highlighting that the judgement of the instructor will be important in evaluating the efficacy of a course. Finally, several studies invoked indirect measures of success, for instance requests for supervisions for independent study projects (Barone, 2011); the longevity of programme and enrolments, local reputation, and the quality of staff-student relations (Stilwell, 2011).

Conclusions

This paper has investigated pluralist economic curricula. However, unlike most of the literature in that area, this paper – building on Garnett and Mearman (2011) – focuses on the process of evaluating pluralist curricula with evidence. Its main focus is on how to evaluate whether or not pluralist curricula do indeed 'work'. Pluralist

curricula are difficult to evaluate because they are often aimed at pluralist goals: and thus they are not amenable to easy assessment. Their evaluation requires considerable thought about the appropriate measures of success, which suggests that merely examining test scores may not capture much of the complexity of pluralist curricula. Further, the liberal educational goals which often drive pluralist curricula are difficult to assess within a single course; and thus the typical methods of evaluating teaching interventions may not be available. Moreover, reflecting much of the educational literature, it is argued that any teaching intervention exists in a highly complex environment, in which it is very difficult to discern clearly their effects on students' learning.

Thus, the paper supports considerable literatures in education, in general research methodology, and even in evaluation methodology, which argue against the dominant experimentalist approach. That approach, also favoured by those who favour EBP, has benefits that should not be abandoned yet is also severely limited in terms of providing evidence that reflects the complex educational contexts in which curricular changes occur. Instead, the paper supports a case-based approach rooted in a complex ontology and which recognises the limits of all methods. Thus it argues for a pluralist approach to evaluating pluralist curricula. Specifically, we argue for a wide variety of measures of success of curricula, a mix of data collection and data analysis methods, and a research design that embraces rather than designs out the heterogeneity and complexity of educational contexts. We have seen that existing empirical literature embraces the above principles; although in many cases these choices reflect strongly individual institutional constraints. However, Stilwell

(2011) develops explicitly a mixed-methods design, an exemplar for the approach advocated here.

References

- Abdul Rahman, S., S. Wok, H. Majid and N. Nor (2005). *Tracer study 2004: First degree graduates*, IIUM & Economic Planning Unit, Prime Minister's Department, Kuala Lumpur, Malaysia.
- R. Amin and M. Haneef (2011). 'The quest for better economics graduates: revising the pluralist approach in the case of the International Islamic University, Malaysia', *International Journal of Pluralism and Economics Education*, 2 (1): 96-113.
- Barone, C. (1991) 'Contending perspectives: curricular reform in economics', *Journal of Economic Education*, Vol. 22, No. 1, pp. 15-26.
- _____ (2011). 'Contending economic perspectives at a liberal arts college: a 25-year retrospective', *International Journal of Pluralism and Economics Education*, Vol. 2, No. 1: 19-38.
- _____ (2012). 'Student Attitudes Toward Economic Pluralism: Survey-Based Evidence', *International Journal of Pluralism and Economics Education*, forthcoming.
- Borg, W. and M. Gall (1996). *Educational Research; an introduction*, New York: Longman.
- Burawoy, M. (1998). 'The extended case method', *Sociological Theory*, 16 (1): 4-33.
- Byrne, D. and C. Ragin (2009). *The Sage handbook of case-based methods*, London: Sage.
- Byrne, D. (2009). 'Complex realist and configurational approaches to cases: a radical synthesis', in Byrne, D. and C. Ragin (eds.) *The Sage handbook of case-based methods*, London: Sage.
- Campbell, D. (1969). 'Reforms as experiments', *American Psychologist*, 24: 409-429.
- _____ and J. Stanley (1963). *Experimental and quasi-experimental evaluations in social research*, Chicago: Rand-McNally.
- Clarke, P. and A. Mearman (2001). 'Heterodoxy, Educational Aims and the Design of Economics Programmes', *Journal of Economic and Social Policy*, 5 (2).
- _____ (2003). 'Why Marxist Economics Should be Taught But Probably Won't Be', *Capital and Class*, Vol. 79 (February): 55-80.
- _____ (2004). 'Comment on C. Winch, "Economic Aims of Education"', *Journal of Philosophy of Education*, 28 (3).
- Cottrell, S. (2011). *Critical Thinking Skills; Developing Effective Analysis and Argument*, London: Palgrave Macmillan
- Creswell, J. and V. Plano Clark (2011). *Designing and conducting mixed methods research*, Thousand Oaks: Sage.

Davies, P. And Guest, R. (2010). 'What effects do we really have on students' understanding and attitudes? How do we know?', *International Review of Economics Education*, 9(1): 6-9.

Denis, A. (2009) 'Editorial: pluralism in economics education', *International Review of Economics Education*, 8 (2): 6-21.

Dow, S.C. (2009) 'History of thought and methodology in pluralist economics education', *International Review of Economics Education*, 8(2): 41-57.

Downward, P. and Mearman, A. (2007) 'Retroduction as mixed-methods triangulation in economic research: reorienting economics back into social science', *Cambridge Journal of Economics*, 31(2): 77-99.

Earl, P. (2000) 'Indeterminacy in the Economics Classroom', in Earl, P. and S. Frowen, (Eds.): *Economics as an Art of Thought: Essays in Memory of G.L.S. Shackle*, pp. 25-50, London: Routledge.

Feiner, S. (2002) 'Toward a post-autistic economics education', *Post-Autistic Economics Review*, Vol. 12, March.

Feiner, S. and Roberts, B. (1995) 'Using alternative paradigms to teach about race and gender: a critical thinking approach to introductory economics', *American Economic Review*, Vol. 85, No. 2, pp. 367-71.

Ferber, M.A. (1999) 'Guidelines for pre-college economics education: a critique', *Feminist Economics*, Vol. 5, No. 3, pp. 135-42.

Freeman, A. (2009) 'The economists of tomorrow: the case for a pluralist subject benchmark statement for economics', *International Review of Economics Education*, Vol. 8, No. 2, pp. 23-40.

Garnett, R. (2009) 'Rethinking the pluralist debate in economics education', *International Review of Economic Education*, Vol. 8, No. 2, pp. 58-71.

_____ and A. Mearman (2011). 'Contending Perspectives twenty years on: what have our students learned?' *International Journal of Pluralism and Economics Education*, 2 (1): 1-18.

Garnett, R., E. Olsen and M. Starr (2009). *Economic Pluralism*, pp. 262-77, London, Routledge.

Goertz, G. and J. Mahoney (2009). 'Scope in case study research', in Byrne, D. and C. Ragin (eds.) *The Sage handbook of case-based methods*, London: Sage.

Gordon, S. (1992). *History and Philosophy of Social Science*, London: Routledge.

Guarasci, R. (2001). 'Developing the Democratic Arts,' *About Campus*, January-February: 9-15.

Hage, J. and B. Meeker (1988). *Social Causality*, London: Unwin Hyman.

Harrison, S. (1998). 'The politics of evidence based medicine in the UK', *Policy and Politics*, 26 (1): 15-31.

Harvey, D. (2009). 'Complexity and case', in Byrne, D. and C. Ragin (eds.) *The Sage handbook of case-based methods*, London: Sage.

Hodkinson, P., G. Biesta and D. James (2007). 'Learning cultures and a cultural theory of learning', in D. James and G. Biesta (Eds.) *Improving learning cultures in further education*, London: Routledge.

Holland, P. (1986). 'Statistics and causal inference', *Journal of the American Statistical Association*, 81: 945-70.

James, D., G. Biesta, P. Hodkinson, K. Postlethwaite and D. Gleeson (2007) 'Improving learning cultures in Further Education?' in D. James and G. Biesta (Eds.) *Improving learning cultures in further education*, London: Routledge.

Knoedler, J. and D. Underwood (2003). 'Teaching the principles of economics: a proposal for a multi-paradigmatic approach', *Journal of Economic Issues*, Vol. 37, no. 3, pp. 697-725.

Lapidus, J. (2011). 'But which theory is right? Economic pluralism, developmental epistemology and uncertainty', *International Journal of Pluralism and Economics Education*, 2 (1): 82-95.

Kirby, E., S. Kirby, and M. Lewis (2002). 'A study of the effectiveness of training proactive thinking', *Journal of Applied Social Psychology*, 32 (7): 1538-49.

Lim, C. P. (1998) 'The effect of Computer-Based Learning (CBL) support package on the learning outcome of low-performance economics students', *Computers in Higher Education Economics Review*, 12(1): 19-26.

McIntyre, R. and R. van Horn (2011). 'Contending Perspectives in One Department', *International Journal of Pluralism and Economics Education*, 2 (1): 69-81.

Mahoney, J. and P. L. Terrie (2009). in Byrne, D. and C. Ragin (2009). *The Sage handbook of case-based methods*, London: Sage.

Martinson, R. (1974). 'What works? Questions and answers about prison reform', *Public Interest*, 35: 22-45.

Mearman, A. (2007) 'Teaching Heterodox Economic Concepts', in P. Davies (Ed.) *Handbook for Economics Lecturers*, Bristol: Economics Network of the Higher Education Academy. Available at <http://www.economicsnetwork.ac.uk/handbook/heterodox/>, accessed at 15:49, 23 December, 2010.

_____ (2008). Pluralism and Heterodoxy: Introduction to the Special Issue', *Journal of Philosophical Economics* 1 (2): 5-25.

_____ and D. Webber (2009). 'Assessment strategies for pluralism', DEE conference, Cardiff, September.

_____, A. Papa and D. Webber (2012). 'Why do students study Economics? A student-level empirical investigation' UWE, Bristol, mimeo

Mearman, A., G. Shoib, T. Wakeley, and D. Webber (2011) 'Does pluralism in economics education make better educated, happier students? A qualitative analysis', *International Review of Economic Education*.

- Mjøset, L. (2009). 'The Contextualist approach to social science methodology', in Byrne, D. and C. Ragin (Eds.) *The Sage handbook of case-based methods*, London: Sage.
- Morrison, K. (2009). *Causation in educational research*, London: Routledge.
- Moseley, F., C. Gunn, and C. Georges (1991) 'Emphasizing controversy in the economics curriculum', *Journal of Economic Education*, 22 (3): 235-40.
- Nelson, J. (2009). 'The Principles course', in J. Reardon (ed.) *Handbook of Pluralist Economics Education*, London: Routledge.
- O'Donnell, R. (2009). 'Economic pluralism and skill formation: adding value to students, economies, and societies', in Garnett, R., E. Olsen, and M. Starr (Eds.) *Economic Pluralism*, London, Routledge.
- Opie, C. (2004). 'Research approaches' in C. Opie (Ed.) *Doing educational research: a guide for first-time researchers*, London: Sage.
- Pawson, R. and N. Tilley (1997). *Realistic Evaluation*, London: Sage.
- Postlethwaite, K. (2007). 'Methodological appendix', in D. James and G. Biesta (Eds.) *Improving learning cultures in further education*, London: Routledge.
- Raveaud, G. (2009). 'A pluralist teaching of economics: why and how', in Garnett, R., Olsen, E., and Starr, M. (Eds.): *Economic Pluralism*, pp. 250-61, London, Routledge.
- Resnick, S. and R. Wolff (2011). 'Teaching economics differently by comparing contesting theories', *International Journal of Pluralism and Economics Education*, 2 (1): 57-68.
- Rihoux, B. and B. Lobe (2009). 'The case for Qualitative Comparative Analysis', in Byrne, D. and C. Ragin (eds.) *The Sage handbook of case-based methods*, London: Sage.
- Scriven, M. (1980). *The logic of evaluation*, Inverness, CA: Edgepress.
- Siakantaris, N. (2000). 'Experimental economics under the microscope', *Cambridge Journal of Economics*, 24: 267-281
- Sikes, P. (2004). 'Methodology, procedures and ethical concerns' in C. Opie (Ed.) *Doing educational research: a guide for first-time researchers*, London: Sage.
- Soh, L. (2001). *A classroom case study on ways to create a motivating learning environment*, unpublished MEd thesis, University of Sheffield.
- Stilwell, F. (2011). 'Teaching a pluralist course in economics: the University of Sydney experience', *International Journal of Pluralism and Economics Education*, 2 (1): 39-56.
- Underwood, D. (2004). 'Principles of macroeconomics: Toward a multiparadigmatic approach', *Journal of Economic Issues*, 38 (2): 571-81.
- Windelband, (1894 [1980]). 'Rectorial Address: Strasbourg', *History and Theory*, 19 (2): 169-185.

Zepke, N. (2011). 'Understanding teaching, motivation and external influences in student engagement: how can complexity thinking help?', *Research in Post-Compulsory Education*, 16 (1): 1-13.