

# **“Quality Criteria in Educational Research: Is beauty more important than popularity?”**

**Terence Karran (University of Lincoln)**

*Keynote Presentation: International Association for the Scientific Knowledge*

*Teaching and Learning Conference 2009:*

*“Excellence and Quality: Timeless Challenges in Education”*

*Oporto, December 7<sup>th</sup> -9<sup>th</sup> 2009*

## **Abstract.**

During the past decade, as national governments more particularly in the USA and the UK have scrutinised more closely the cost effectiveness and impact of research funding within higher education, they have become critical of the overall quality of educational research, in terms of its scientific rigour, its utility for practitioners and the manner in which it is assessed. This paper addresses the reasons why the quality of educational research was questioned and then examines the discussion framework for assessing quality which emerged. Following from this, the merits of possible internal and external criteria for the worth of educational research are considered. The links between these criterial sets and the function and purpose of qualitative and quantitative approaches to educational research are considered. Finally the question of whether educational research is an art or a science is addressed.

## **Introduction**

Over the last decade, criticisms have been voiced both in government and academia about the overall quality and level of excellence of research into education, and how such research is to be assessed. Educational research has been accused of being of low quality, more especially when compared with other disciplines like medicine, and hence of negligible value either to policy makers or practitioners in the teaching profession. Additionally educational researchers have been dismissed as being a partisan and narcissistic community, content with disseminating their arcane research findings to each other via obscure academic journals. For example, in 2000, David Blunkett, then the UK government minister responsible for education, had argued that “We need to be able to rely on ... social scientists to tell us what works and why and what types of policy initiatives are likely to be most effective” and that “issues for research are too supplier-driven rather than focussing on the key issues of concern to policy-makers, practitioners and the public at large” (Blunkett, 2000). Moreover, while giving evidence to the House of Commons Education and Skills Committee in 2003, Sir Howard Newby, the (then) Chief Executive of the Higher Education Funding Council for England and a former university Vice Chancellor remarked: “education in this country on the whole has a problem with the quality of the research, not with the amount of it.”

“It (i.e. education research) is not as good as it might be and I speak as a former Chairman of the Economic and Social Research Council.”(Newby, 2003) Similarly, in the USA, Shavelson and Towne’s assessment of scientific research in education undertaken for the National Research Council, reported on “the widespread perception that research in education has not produced the kind of cumulative knowledge garnered from other scientific endeavours. ... The prevailing view is that findings in education research studies are of low quality and are endlessly contested” (2002, 28). That similar concerns have also been expressed following reviews in (inter alia) Australia (McGaw et al. 1992) and France (Proust, 2001) would suggest that such criticisms are widespread.

Criticisms of the quality of educational research from within academia have been just as (if not more) trenchant. Michael Bassey, for example, was a member of the Educational Panel during the national Research Assessment Exercise of UK universities in 1992 and commented on his experience thus: “I am less certain that much of the research reported in the literature does extend theory, or illuminate policy, or improve practice in significant ways. I have a strong impression of individualism, of researchers working in isolation from each other, dabbling in an amateurish way at issues which are too big to be tackled by lone researchers. I consider that much educational research is in a dilettante tradition that looks like a game of trivial pursuits.”(1993, 6) Similarly, Kaestle’s examination of the role of US federal funding for educational research over 25 years caused him to pose the question “Why is the reputation of education research so awful?” and he found that educational researchers had a “reputation for irrelevance, politicization, and disarray” (1993, 30). In addition to the apparently poor quality of educational research, commentators also noted that educational research rarely, if ever, had any utility for policy-makers or teaching practitioners. Moreover in the UK, an analysis by Hillage et al (1998, xi) of research which was relevant to practitioners and policy makers found that “Where the research does address policy-relevant and practical issues it tends to:

- be small scale and fails to generate findings that are reliable and generalisable;
- be insufficiently based on existing knowledge and therefore capable of advancing understanding;
- be presented in a form or medium which is largely inaccessible to a non-academic audience; and
- lack interpretation for a policy-making or practitioner audience.”

Helpfully Pring (2000a, 496) provides a succinct summary of the shortcomings of educational research which are, he suggests, fourfold. First, “it is claimed that educational research does not answer the questions which civil servants and ministers want answers to—often very quickly”;

second, “research is said to be of little help to practitioners—to the teachers who, in their daily lives, for example, need well founded advice on how to teach this or that subject matter to these pupils”; third, “despite the amount of research (and the theses, journal articles and books are witness to there being a lot), it by and large remains fragmented and piecemeal”; fourth “much of the research, so we are told, is tendentious. Under the banner of scholarship, ideological wars are fought”. In summary, Oancea and Pring (2008, 16) argue that “educational research was found to be lacking in ... relevance; cumulativeness and coherence; and cost effectiveness.”

The academic research community countered these negative evaluations by arguing that such critiques are based on fundamental misconceptions of the nature of social science and the role and context of educational research, and ignore “the thoroughly practical character of teaching - the diverse and difficult-to-operationalise goals, the multiple variables and complex relationships involved – (which) may mean that research can rarely provide sound information about the relative effectiveness of different techniques” (Hammersley 1997, 154). Additionally, Bridges et al (2008: 8) have pointed to the difficulties of relating research to policy, more especially that : “The notion of research providing a basis for policy is especially problematic in so far as it suggests that the process begins with research which then points to the required policy. This is an empirically and logically unsound view of the nature of policy and its construction. Policy is an ongoing process: it is not a vacuum waiting to be filled.” Similarly, but more fundamentally, Carr (2003, 132) argued that that “it is mistaken to construe human conceptual learning, or knowledge-acquisition, as a quasi-naturalistic process (of behavioural modification or ‘cognitive development’) apt for investigation via some kind of empirical science: on the contrary, any meaningful (human) educational learning (rather than animal training) is a matter of normative initiation into socially constructed and/or constituted rules, principles and values that no statistically conceived processes could even begin to explain.” As a result, Edwards (2000, 299) chastised the critics of educational research for “exaggerating the prospects for a science of teaching.” The continuing debate which this criticism has engendered, although impressively impassioned (see, for example, MacLure, 2005), thereby has generated much heat, but consequently has tended to deflect attention away from a serious analysis of how such criticisms could, or should be addressed.

### **The Discussion Framework**

These criticisms, and the subsequent debate that resulted, led the U.K. Economic and Social Research Council to commission the Educational Studies Department at Oxford University in 2004 to study criterial judgements in education research on behalf of academics, practitioners and policy makers. The fourfold aims of the study were to:

- clarify the concepts employed in relation to applied and practice-based research;
- map the various models currently in use and then explore the philosophical underpinnings of the models;
- review the recent UK initiatives concerning applied research and practice based research;
- develop an understanding of quality to assist the development of quality criteria appropriate for different types of applied and practice-based research.

The paper by Furlong and Oancea which resulted (*Assessing Quality in Applied and Practice-based Educational Research: A Framework for Discussion*) attracted widespread interest in academia and government, both in the UK and beyond. In the UK it was widely circulated and discussed and, more significantly in terms of its impact on research policy, it was referred to explicitly in the criteria for the Education Unit of Assessment in the 2008 Research Assessment Exercise. The Discussion Framework identified “four dimensions of quality – (i) epistemic, (2) technological, (3) capacity building and value for people and (4) economic” (2005, 10), and in subsequent paper, the authors returned to Aristotle’s *Metaphysics*, and utilised his conceptions of “*episteme theoretike* (knowledge that is demonstrable through valid reasoning); *techne* (technical skill, or a trained ability for rational production); and *phronesis* (practical wisdom, or the capacity or predisposition to act truthfully and with reason in matters of deliberation, thus with a strong ethical component)” (2007, 124). Within each domain (and that of the economic), they further identify characteristic attributes, the results of this process have been summarised in Table 1 below.

Table 1 Domains of Quality in Educational Research

Episteme theoretike	Techne	Economic	Phronesis
Trustworthiness	Fitness to Purpose	Auditability	Plausibility
Advancement of Knowledge	Concern for enabling impact	Cost-effectiveness	Reflexivity, deliberation and criticism
Transparency/ Explicitness	Specificity and accessibility	Marketability and competitiveness	Engagement
Propriety	Saliency/timeliness	Feasibility	Receptiveness
Paradigm- dependent criteria	Flexibility and operationalisability	Added Value/ ‘brand’	Transformation and personal growth
Scientific Robustness	Social and Economic Robustness		

Source : Furlong and Oancea (2005) p.15, Oancea and Furlong (2007) p.133

The broader domains of quality in educational research adopted by Furlong and Oancea encompass all the elements of the widely accepted OECD (1994, 7) Frascati definition of research. Firstly, basic research (“theoretical work, undertaken to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any specific application or use in view,”)

applies to the Episteme category, and it is this domain which encompasses the traditional parameters of excellence in social science research, namely, methodological and scientific robustness to which academic researchers aspire. Secondly, the Frascati definition of applied research (“also original investigation undertaken to acquire new knowledge. It is, however, directed primarily towards a specific practical aim or objective”), applies to the Techne, Economic, Phronesis categories, and these domains embrace those elements of excellence which are sought after by practitioners and policy makers alike.

Furlong and Oancea’s contribution can be welcomed as an important foundation for the analysis of an inherently problematic area - as Hodkinson (2004, 11) points out “there are almost as many different lists of suggested criteria for judging research as there are writers about the issue”. Their schema includes criteria which could be used to assess ‘the worth of educational research’. However ‘worth’, used in this context, can have two separate but inter-related meanings. First, it can refer to excellence and high or outstanding intrinsic qualities – in this sense, it might be considered that research into education was ‘of worth’. Second, it can be used with reference to the relative value of a thing, in respect of the extrinsic utility to which it can be put, or of the estimation in which it is held. Hence in this sense the ‘worth’ of educational research may refer to its greater or lesser utility for those within the teaching profession – it may be considered ‘worthwhile’. These meanings are clearly linked, in that something which has intrinsic worth in the first sense of the word, may frequently, (but not necessarily) have extrinsic worth in the second sense.

### **Internal and External Criteria**

Responding to the Discussion Framework, Hammersley (2008, 6) argued that “In the case of academic research, the immediate audience is not practitioners but fellow researchers. The aim is to contribute to a body of knowledge, albeit one that relates to some issue of human concern. Thus, the likely validity of the findings is to be judged primarily by the research community.” Hence, for Hammersley, criteria for the assessment of the intrinsic value of educational research will be internal, and will be established by the research community from which the research arose, by a process of, inter alia, peer-review and subject discourse. Internal criteria refer to the robustness of the method adopted to address a research problem, in terms of adherence to established and accepted methods of excellence, as defined and utilised by the subject discourse community – any research which meets these criteria, *par excellence*, might be adjudged ‘beautiful’ (indeed, academics not infrequently may refer to an article, or a colleague’s work in this way). Thus, although such criteria have to be validated by inter-subjective judgements within a subject discourse community, the benchmarks against which they are judged are epistemological in nature. Such

criteria have much in common with the parameters utilised in 2008 by the UK Research Assessment Panel for Education (HEFCE, 2007, 13) for judging the worth of research texts.

By contrast, external criteria of the excellence of research have, as their reference point, the impact of research, adjudging research to be significant if it has made an impact (in terms of altering policies or practices) on one or more external agents. In this respect external criteria comprise the measures developed by Furlong and Oancea (2005) of capacity development and value for people, cost effectiveness, marketability and competitiveness which enable an estimation of social and economic robustness. In this sense, a research artefact which has a widespread impact can be considered to be 'popular'. These criteria are more akin to the new 'impact' element of the new UK Research Assessment Framework (REF - which replaces the RAE) which calls for "An assessment of demonstrable economic and social impacts that have been achieved through activity within the submitted unit that builds on excellent research" (HEFCE 2009, 20). Within the proposed REF, the relative weightings of the three constituents in the overall assessment of research are Outputs (i.e. Episteme) = 60%; Impact (i.e. Techne, Economic, Phronesis) = 25%; Environment = 15%.

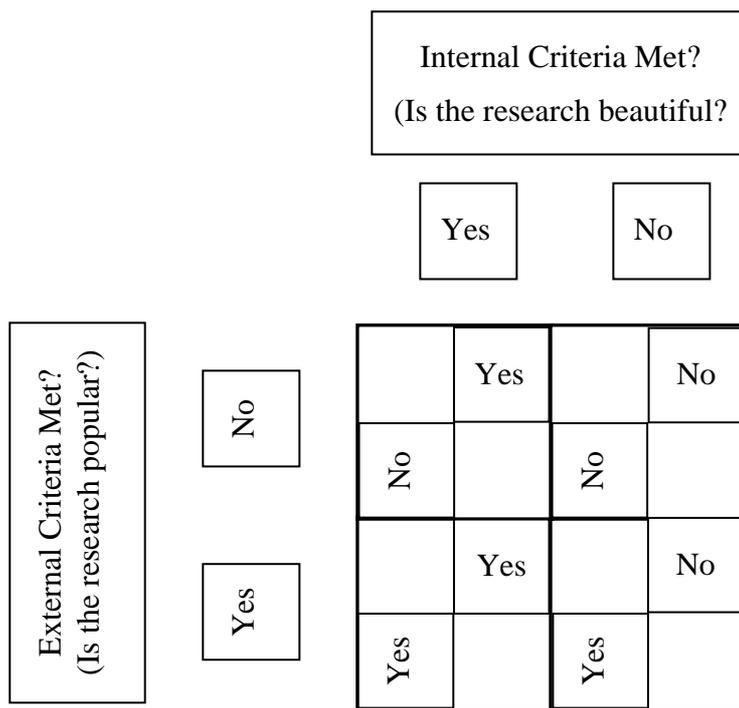
Given that the socio-economic impact of research may require a long term gestation, at any particular point in time it may be possible for a piece of educational research to have worth in the first sense using internal criteria (and hence be deemed 'beautiful') but not in the second sense, using external criteria (and hence not be 'popular'). Similarly, it may be possible for a piece of research to have high utility in informing policy or practice (and hence be considered 'popular'), but be deemed logically flawed or methodologically unsound (and hence lacking 'beauty') by academic researchers. A classic example of the latter is the work of the economist Milton Friedman, whose political philosophy, which called for a minimal role for government in favour of the private sector, heavily influenced the economic policies of the Reagan Presidency in the USA and the Thatcher government in the UK for more than a decade. When Friedman was awarded the Nobel prize for economics in 1976, this provoked a widespread protest by economists in academia who argued that his monetarists policies were incorrect; these criticisms were reiterated in 2007-9 when Keynesian economists blamed policies espousing Friedman's free-market philosophy for the credit crunch and the subsequent world-wide recession.

Hammersley's attempt to identify possible types of evaluative judgments, in the light of the Discussion Framework, threw up further methodological problems namely that "first, any comprehensive list is likely to be very long ... secondly, the criteria can serve as little more than reminders, because they cannot be transparent, ... because of necessary reliance on background

knowledge and expert judgement” (2005, 7). Clearly if a piece of work is to be judged by more than one criteria, then it is necessary to clarify the relationships (equal, superordinate, subordinate) between the criteria which requires agreement between the assessors, which is likely to be problematic. As Hammersley points out there are likely to be differences between assessments by the subject discourse communities of practitioner users and researchers. For example, in respect to the former “in relation to plausibility, ... what they take to be established knowledge will be different from that which is well established within the relevant research community. This is because they will have practical knowledge deriving from individual and collective experience” (ibid, 7). Hence determining the relative values of internal and external criteria is likely to be inherently problematic because different types of research have different purposes for different audiences, and hence giving a low value to a piece which is designed to have no practical benefit would be to unfairly discriminate against it. Moreover, as Spencer et al (2003, 4) point out, it is debatable “whether the concepts of quality used to assess qualitative research should be roughly the same as, or parallel to, or quite different from those used to assess quantitative research.” Additionally, Hammersley (2005, 5) points out that in the case of educational research “theories are not the only legitimate goal of inquiry. Instead we may aim at producing descriptions or explanations” but such research would likely be adjudged as of lower value than theoretical work, when internal criteria are adopted. This raises the question as to whether internal and external criteria may (or should) have a hierarchical relationship to each other, - the existence of a hierarchical relationship (and the basis for it) would clearly have significant implications for both the assessment of the excellence of educational research, and its future direction. In passing, it is worth noting that the proposed REF weightings, which suggest that the dominant criteria for judging the excellence of research should be internal, but that ‘impact’ external criteria are significantly important, resulted in a petition (see at: <http://petitions.number10.gov.uk/REFandimpact/>) being sent to the UK Prime Minister, Gordon Brown, urging him “to allocate funds for academic research solely on the basis of academic excellence and not on the basis of ‘impact’ or the judgements of ‘users’” in October 2009, which attracted over 2,500 signatures from the academic community in less than a month.

However, without necessarily determining a hierarchy between internal and external criteria, or making assumptions about agreements by the assessors, it is possible to examine the possible outcomes when a piece of research either meets or fails to meet both internal/intrinsic and external/extrinsic criteria, as below in Diagram 1.

Diagram 1 Possibilities for meeting internal and external criteria



Where both the internal and external criteria relating to a piece of research are met, then an assessment of the positive worth of the research would presumably not be questioned, and it could be deemed to be both ‘beautiful’ and ‘popular’. Similarly, when a piece of research fails to meet either internal or external criteria then, assumedly, it could justifiably be disregarded. However, problems arise in the other two cells of the table when one set of criteria are met, but the other set are not. From a strictly Positivist point of view, it could be argued that, in those instances where internal criteria are not met (e.g. the research lacks intrinsic rigour or is logically flawed) then, irrespective as to whether or not the research meets external criteria, it should be disregarded, which would imply a hierarchical order between the sets of criteria. However, such statements are problematic, not least because judgements about whether or not internal or external criteria are met are made by subject discourse communities, within which evidence (and, by extension, the worth of research) may be disputed. Moreover, as has been considered, all judgements about educational matters are inferential, and the relationship between evidence and judgement is often contentious. As Oancea (2007, 251) points out “the indicators used and the practices preferred in various evaluations of research are eclectic, reflecting the multitude of interests and demands that compete in setting the boundaries of accountability in research activities, rather than the nature of these activities themselves and of the particular forms of knowledge to which they contribute.” Rightly or otherwise, educational research has been criticised for being characterised by such disputes over the value of research. However, they are just as common in the assumedly more scientifically

rigorous disciplines. Lawrence (2003), for example, relates how Berridge and Irvine's 1984 paper on phosphoinositol and signalling, was originally turned down by the journal *Nature*, but subsequently published after an appeal and went on to become the second most quoted scientific article of the 1980s. (Lawrence's article provoked a chorus of complaints about the tyranny of the process of academic publication, see *Nature*, Vol 434: 479-80) Such cases demonstrate that the locus of a piece of research within the table (and hence its worth to both academic researchers and practitioners) may not be fixed but may vary over time.

### **Education as a Social Science**

The distinction between the merits of internal and external criteria in assessing research mirrors a broader divide within the educational research community between those who seek to emulate the pure sciences and pursue a scientific positivist approach (often via the use of quantitative methodologies), and those who (often espousing a qualitative approach) believe that such a pursuit is ill-judged and even counter-productive. The debates and disputes about the nature and purpose of educational research in the late 1990s coalesced around these two broad discourses described by Oancea (2005, 157f) thus: "one lamenting the misbehaviour of educational research from a managerial perspective (associated with a 'big science' model of knowledge production and an 'engineering' model of knowledge use), and the other attempting to defend it in the name of academic freedom and right to diversity, or to reinstate it through a humanistic model of knowledge transfer". The net result, Oancea and Pring (2008, 27) argue, has been a "dichotomy between quantitative and qualitative research designs - the former receiving the approbation of Government looking for the evidence for particular policies, the other generally embraced by practitioners but disdained by those who want general answers to generally conceived problems." Despite the fact that, as Harden and Thomas (2005, 265) point out "much research does not fit into neat categories of 'qualitative' and 'quantitative'", protagonists from both sides of the divide have been increasingly vocal in their support for one or the other framework. As Phillips (2005, 578) graphically relates "On the one hand, there are influential figures who countenance only rigorous scientific research; they use as their model of science the randomised controlled experiment or field trial, and they point to experimentation in medicine as the ideal model for educational research. The existence of this group of hardliners fills many other members of the research community with feelings of despair and utter hopelessness. On the other hand—at the other extreme pole of opinion—there are those who see the members of the first group as advocating 'their father's paradigm' ... that is hopelessly modernist, positivistic and imperialistic; ... This second position is so murky and fraught with danger that it is regarded by the advocates of scientific rigour as leading to the total extinction of the empirical research enterprise". Adopting such intractable stances

makes progress difficult, more especially when, as Pring (2000b, 258f) points out, “To understand an educational practice requires the careful analysis of the social situation—the underlying social rules, the interpretation of the participants, the values and aims embedded within the practice. Such ‘qualitative research’ is quite clearly necessary, and the absence of it leads to the gross generalisations and misleading science. On the other hand, such qualitative work, given what we know about human beings and about the social structures which constrain their activities, simply sets limits and gives greater refinement to the more general verifiable and (where possible) quantifiable claims which research should constantly be seeking.”

In attempting to find a way forward out of this impasse, Hargreaves (1996, 210) used medical research as a benchmark to which educational research should aspire, and pointed to the similarities between research into practice in medicine and education, more particularly that “both education and medicine are profoundly people-centred professions. Neither believes that helping people is a matter of simple technical application but rather a highly skilled process in which a sophisticated judgement matches a professional decision to the unique needs of each client.” In response Hammersley (1997, 154) declaimed “that there are some fundamental problems at the core of Hargreaves analysis. ... his reliance on the medical analogy is potentially misleading (as) [m]uch medical research does not involve the distinctive problems associated with studying social phenomena”. Similarly, Evans and Benfield (2001, 539) argue against steering “educational research in the direction of a ‘medical model.’ ... (as) ... [s]uch an approach will tend to reduce research questions to the pragmatics of technical efficiency and effectiveness”. Hammersley’s criticisms of Hargreaves’ medical analogy for educational research were trenchant, but perhaps misplaced and thereby generated a spirited rebuttal. Hargreaves (1997 409) retorted that “Hammersley’s error is to treat the research underlying medical practice as essentially homogeneous and as positivistic, working on physical phenomena in the interests of discovering universal laws and patterns of physical causation. Doubtless this applies to the root natural sciences and to some degree to the medical and clinical sciences. But with research into practice, the kind of research at the heart of evidence-based medicine, we are in the world of human beings making complex decisions. .... His depiction of the knowledge-base of medical practice is, in short, crude oversimplification.”

Hence Hammersley (2000, 225) and others have repeatedly argued that “It should be clear ... that there is a divide between pursuing scientific and practical research. They involve different goals and different immediate audiences; and the most effective approach to one generally involves serious costs from the point of view of the other. Thus, trying to do both kinds of research simultaneously

will often result in the requirements of neither being well satisfied.” However this is a finely drawn distinction which may be illusory in respect to education - as Neumann notes (2005, 185) “in professional disciplines the link between theory and practice is by definition inextricably close ... while the distinctions between pure and applied, theory and practice, appear neat and clear-cut for discussion purposes, in actuality they overlap and blur”. Moreover, it is doubtful that those undertaking research (in the pure or social sciences) view the apparent divide with such clarity. For example, the 2005 Nobel Prize for Medicine was awarded to Marshall and Warren for their discovery of the bacterium *Helicobacter pylori*. Their research was methodologically scientific, but their aim was very practical - how to prevent gastritis and peptic ulcer disease. Similarly in the social sciences, Alvin Gouldner, writing in the 1950s, suggested that theoretical social scientists also sought practical results for their research and believed that “Marx, Durkheim and Freud share the applied social scientist’s concern with bringing social science to bear on the problems and values of laymen with a view to remedying their disturbances” (1957, 94).

More helpfully, addressing what he refers to as “the “false dualism” of educational research” Pring (2000b, 257) argues that “social events and facts (and such are educational practices) can be explained in much the same way as physical events and facts can. The methods of the social sciences, with all their statistical sophistication, are brought to bear upon an understanding of education, and from the understandings gained those in charge of education, either at the policy or at the professional level, will know what interventions will make things work”. Hence although it cannot be denied that “it is not just that different people interpret physical reality different, but also that social reality is created by those interpretations. This quite clearly has a profound effect upon the nature of much of educational research that is concerned with learners acting within social traditions that shape their conceptions of reality” (Oanacea and Pring, 2008, 29), the important question is whether these differences are so great as to make the derivation of general principles impossible. It is clear that the commonality of interpretations enables social discourse to continue because if these different interpretations were very divergent, social life would tend to break down.

Although the purpose of using medicine as an exemplar for educational research was well intentioned, the differences between the two are such that criticisms suggesting that the medical model is inappropriate are valid. Medicine clearly has more in common with the natural and life sciences (especially biology and chemistry) than with the social sciences like education. Consequently, in examining the possibility of deriving general principles within educational research, the example of economics, rather than medicine, may be more helpful. As a social science, economics has many of the problems that currently beset educational research. However,

despite these limitations, economists have succeeded in using quantitative analyses to build micro and macro level theoretical models, covering an array of economic activities (e.g. theories of the firm, labour markets, international trade, technological change, price fluctuations, etc.) often with a high level of statistical sophistication and predictive ability. Unlike theories in the pure sciences, economic models are often conditional rather than causal statements, but this has not prevented the discipline from building a cumulative body of relevant knowledge which informs policy makers and practitioners (businesses, trade unions, banks, etc.) and which, it is argued, educational research has failed to do. John Maynard Keynes, the father of modern macroeconomics, although himself a theorist (he authored *The General Theory of Employment, Interest and Money*) nevertheless believed that “The theory of economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method rather than a doctrine, an apparatus of the mind which helps its possessor to draw correct conclusions” (1934, 6). The status of economics within the academic pantheon comparable to, but different from, the sciences of physics, chemistry, medicine, etc., was recognised by the inauguration of a Nobel Prize for the discipline in 1969.

### **Concluding Remarks**

Marie Curie was awarded the Nobel Prize for Physics in 1903 for discovering polonium and radium. Between 1898 and 1902, she and her husband managed to extract 1 gram of pure radium chloride from 8 tonnes of pitchblende. Such examples demonstrate that research that is experimental and quantitative often can be arduous and difficult. Undertaking quantitative educational research requires the collection and collation of data (often a problematic process), choosing and applying statistical tests for analysis and then interpreting the results – as Berliner (2002) rightly observes “Educational Research: The Hardest Science of All”. Donald MacIntyre, in his Presidential Address to the British Educational Research Association in 1996 acknowledged this by stating “One thing I know from several decades of experience is that I find it very difficult to do educational research well. It requires rigorous thinking, perceptiveness, imagination, self-awareness, social skills and self-discipline in such demanding combinations that I am usually disappointed with the quality of my own work. To judge from the many papers that I have to referee for research journals, other researchers also find it difficult to do well, and many seem to lack an understanding of the diverse basic disciplines required.”(1997, 129) Consequently, it may be easier (and, for some, more enjoyable) to demonstrate one’s academic élan and intellectual brio by declaiming another combative polemic demonstrating (inter alia) that the quantitative approach is inappropriate (or even harmful) to educational research. As Fuer et al. relate (2002, 6) “The history of educational research is not a simple tale of progress, and its story provides important insights for

its future. Educational research has a long history of struggling to become - or to ward off - science.”

Writing in 1902, John Dewey, the educationalist and philosopher observed, “There is another group of sciences which ...are more remote from a scientific status. ... the social and psychological disciplines. ... compared with mathematic physics we can employ the term ‘science’ only in a tentative and somewhat prophetic sense—the aspirations, the tendencies, the movement are scientific. But to the public at large the facts and relations with which these topics deal are still almost wholly in the region of opinion, prejudice, and accepted tradition”(1902, 4f.). Reviewing progress since then Phillips (2005, 582) states that “A charitable judgment is that ‘philosophy of educational research’ is roughly at the stage that much philosophy of science was at six decades or more ago when real examples of research, discussed with historical richness, were relatively rare,” suggesting that the current debate about the focus of educational research, and how it is to be assessed, has some way to run before a satisfactory conclusion may (or may not) emerge. This debate is an integral aspect of academic freedom and of the freedom of discourse essential to the advancement of knowledge, and therefore will not readily abate. Moreover the nature of, and need for, this debate points up the possibility that, in succumbing to the demand for research which adopts a particular methodology and (assumedly) provides tangible impacts, there is a danger that particular insights may be lost, moreover, such prescriptive actions are anathema to academic freedom. By its very nature, the location of new knowledge is unknown, although well-qualified and experienced academic staff are more likely than policy makers and politicians to know where it may be located. Trying to “manage” research in order to increase its impact is as impossible as trying to “manage” the weather - although reasonably accurate short term forecasts may be attempted. However, even if it is possible to measure previous rainfall (or existing research outputs) very accurately, it is still impossible to manage the weather (and university research) to make it more “productive” in terms of its impact.

Writing in 1849, somewhat unfairly, Thomas Carlyle, the essayist and historian described economics as “the dismal science” (1850, 531). Much has changed to the discipline since Carlyle’s day to render this description inappropriate, if not inaccurate. On the very last page of their undergraduate text book on Macroeconomics, Gregory Mankiw (Professor of Economics at Harvard) and his colleague Mark Taylor (2000, 609) have a concluding paragraph aimed at counselling their readers who are just commencing their studies of the subject. They state: “Economists and policy makers must deal with ambiguity. The current state of macroeconomics offers many insights, but it also leaves open many questions. The challenge for economists is to

find answers to those questions and to expand our knowledge. The challenge for policy makers is to use the knowledge we now have to improve economic performance. Both challenges are formidable, but neither is insuperable.” This measured advice is just as applicable to those starting to undertake educational research. There is a chance that, despite adopting a quantitative approach akin to that of the pure sciences, educational research may fail to achieve the academic status or predictive power of (say) economics. However, if it does not even try to adopt such an approach, then there is absolutely no chance that it can succeed. More significantly, perhaps, if we, as researchers and teaching practitioners, do not aspire to the highest level of scientific rigour and professionalism in our teaching and research, how can we expect those whom we teach and supervise to adopt such an approach to their own studies, and in their chosen professions?

Bassey, M., (1993) "Educational research in the universities and colleges of the United Kingdom: significant insights or trivial pursuits", paper presented at the Social Sciences and Humanities Research Council of Canada Conference on Evaluation, Social Science and Public Policy, 11 June 1993, Ottawa, Canada.

Berliner, David C., (2002) "Educational Research: The Hardest Science of All," *Educational Researcher*, 31(8): 18–20.

Berridge, M. J., Irvine, R.F., (1984), "Inositol trisphosphate: a novel second messenger in cellular signal transduction", *Nature*, Vol. 312: 315-321

Blunkett, D., (2000) "Influence or irrelevance: can social science improve government?" Speech made by David Blunkett, Secretary of State for Education and Employment, to a meeting convened by the Economic and Social Research Council on 2 February 2000, *Research Intelligence*, 12-21.

Bridges, D., Smeyers, P., Smith, R., (2008) "Educational Research and the Practical Judgement of Policy Makers" *Journal of Philosophy of Education*, 42(S1): 7-14.

Carlyle, Thomas, (1849 February) "Occasional Discourse on the Negro Question", *Fraser's Magazine for Town and Country*, Vol. XL., [reprinted in *The Commercial Review of the South and West* (later *De Bow's Review*), June, 1850, Vol. VIII (Old Series), Vol. II, No.4, p.527-538]

Carr, D. (2003) *Making Sense of Education* (London and New York, Routledge-Falmer).

Dewey, J., (1902) "Academic Freedom" *Educational Review*, 21(813): 1-14.'

Edwards, Tony, (2000) "'All the Evidence Shows ...': reasonable expectations of educational research", *Oxford Review of Education*, 26(3): 299-311.

Evans, J., Benefield, P., (2001) "Systematic Reviews of Educational Research: does the medical model fit?" *British Educational Research Journal*, 27(5): 527-541.

Feuer, M., Towne, L., Shavelson, R., (2002) "Scientific Culture and Educational Research", *Educational Researcher*, 31(8): 4-14

Furlong, J., Oancea, A., (2005) *Assessing Quality in Applied and Practice-based Educational Research: A Framework for Discussion*, Oxford University: Department of Educational Studies

Gouldner, A., (1957), "Theoretical requirements of the Applied Social Sciences", *American Sociological Review*, 22(10): 92-102.

Hammersley, M., (1997) "Educational Research And Teaching: A Response To David Hargreaves' TTA Lecture", *British Educational Research Journal*, 23(2): 141-161.

Hammersley, M., (2000) "Varieties of Social Research: A Typology", *International Journal of Social Research Methodology*, 3(3): 221 - 229

Hammersley, M., (2005) "Assessing Quality in Qualitative Research", paper for a TLRP seminar on Assessing Quality in Case Study and Qualitative Research, TLRP seminar series on *Quality in Educational Research*

Hammersley, M., (2008), "Troubling criteria: a critical commentary on Furlong and Oancea's framework for assessing educational research." *British Educational Research Journal*, 34(6): 747-762

Harden, A., Thomas, J., (2005) "Methodological Issues in Combining Diverse Study Types in Systematic Reviews", *International Journal of Social Research Methodology*, 8(3): 257-271

Hargreaves, D., (1996) "Teaching as a research based profession: possibilities and prospects", TTA Annual Lecture 1996, re-printed in (2001) *Leading Professional Development in Education*, (eds) Bob Moon, John Butcher, Elizabeth Bird, London: Routledge, pp. 200-210

Hargreaves, D., (1997) "In Defence of Research for Evidence-based Teaching: a rejoinder to Martyn Hammersley", *British Educational Research Journal*, 23(4): 405-419

HEFCE (Higher Education Funding Council for England), (2007) *Research Assessment Exercise Education Panel Assessment Criteria*, Swindon: HEFCE.

HEFCE (Higher Education Funding Council for England), (2009) *Research Excellence Framework: Second consultation on the assessment and funding of research*, Swindon: HEFCE.

Hillage, J., Pearson, R., Anderson, A., Tamkin, P., (1998) *Excellence in research in school*, London: Department for Education & Employment/Institute of Employment Studies.

Hodkinson, P., (2004), "Research as a form of work: expertise, community and methodological objectivity", *British Educational Research Journal*, 30(1): 9-26

Kaestle, C. F., (1993) "The awful reputation of educational research", *Educational Researcher*, 22(1), 26-31.

Keynes, J. M., (1934) Introduction to the *Cambridge Handbook of Economic Series*, Cambridge and New York: Cambridge University Press.

Lawrence, P., (2003), "The Politics of Publication", *Nature*, 422: 259-61

MacLure, M., (2005) "'Clarity bordering on stupidity': where's the quality in systematic review?", *Journal of Education Policy*, 20(4): 393 - 416

Mankiw, N. Gregory, Taylor, Mark P., (2006), *Macroeconomics*, New York: Worth Publishers.

McGaw, B., Boud, D., Poole, M., Warry, S., McKechnie, P., (1992) *Educational Research in Australia: Report of the Review Panel of Research in Education*, Canberra, Australian Government Publishing Service.

- McIntyre, Donald, (1997) “The Profession of Educational Research”, *British Educational Research Journal*, 23(2): 127-40.
- Neumann, R., (2005) “Doctoral Differences: Professional doctorates and PhDs compared”, *Journal of Higher Education Policy and Management*, 27(2): 173–188.
- Newby, Howard (2003) Verbal Evidence to the House of Commons Education and Skills Committee, 5th March, 2003
- Oancea, A., (2005) “Criticisms of educational research: key topics and levels of analysis”, *British Educational Research Journal*, 31(2): 157-183.
- Oancea, A., (2007) “From Procrustes to Proteus: trends and practices in the assessment of education research”, *International Journal of Research and Method in Education*, 30(3): 243-269.
- Oancea, A., Furlong, J., (2007) “Expressions of excellence and the assessment of applied and practice-based research”, *Research Papers in Education*, 22(2): 119-137
- Oancea, A., Pring, R., (2008), “The Importance of Being Thorough: On Systematic Accumulations of ‘What Works’ in Education Research”, *Journal of Philosophy of Education*, 42(S1): 15-39
- OECD, (1994), *Main Definitions and Conventions for the Measurement of Research and Experimental Development: A Summary of the Frascati Manual OCDE/GD(94)84*, Paris: OECD
- Phillips, D.C., (2005) “The Contested Nature of Empirical Educational Research (and Why Philosophy of Education Offers Little Help)”, *Journal of Philosophy of Education*, 39(4): 577-597.
- Pring, R., (2000a) “Editorial Conclusion: a philosophical perspective”, *Oxford Review of Education*, 26(3): 495 - 501
- Pring, R., (2000b) “The ‘False Dualism’ of Educational Research”, *Journal of Philosophy of Education*, 34(2): 247-60.
- Prost, A. (2001) *Pour un Programme Stratégique de Recherche en Education*, Paris, Ministère de l’Education Nationale.
- Shavelson, R. J., Towne, L., (Eds.), (2002) *Scientific Research in Education*, National Research Council Committee on Scientific Principles for Educational Research, Washington, DC: National Academy Press.
- Spencer, L., Ritchie, J., Lewis, J., Dillon, L., (2003) *Quality in Qualitative Evaluation: A framework for assessing research evidence*, Government Chief Social Researcher’s Office, London: Cabinet Office