What does ‘significance’ look like?
Assessing the assessment process in competitive grant schemes.

Lyn Yates
University of Melbourne
l.yates@unimelb.edu.au

Article has been published as:

Abstract:
This paper focuses on the writer’s experiences from 2002-2004 as the sole Education member of the Australian Research Council committee that assesses applications across the ‘social, behavioural and economic sciences’. It is drawn from a wider analysis of how judgments about research quality are produced across different spheres of education research activity, drawing particular attention to the characteristics of those who judge, their explicit and implicit criteria, and the textual markers of ‘quality’ that they work with. The article considers how the explicit categories for scoring; the characteristics of those appointed to the committee; the history of dominant research traditions, and a slippage between the categories ‘significance’, ‘national benefit’ and ‘national research priorities’ all influence the judgments and scores that eventuate from assessors and produce the final rankings.
Since the passing by the US Congress of the *No Child Left Behind* in 2000 and the establishment of the Institute of Education Sciences in 2002, most of the debate about research quality in the education research community has been concerned with explicit criteria, that is, with some new moves to mandate by technical definition what counts as quality – or indeed, adequacy (Feuer et al., 2002). The focus in conferences of national education research associations and in exchanges in education research journals has been on initiatives by a number of governments to define technically and specifically what methodological standards research needs to be meet to be designated ‘scientific’ and hence to count as ‘evidence’ when reviews are carried out, and to be eligible for funding. But whether or not criteria are defined technically, in competitive grant schemes, judgments about quality are never simply technical ones. Funding from such bodies is considered prestigious precisely because only some get selected even though almost all may be judged as technically sound. Competitive grant schemes such as those of the Australian ARC, or the New Zealand Marsden fund, or the British ESRC, or the Canadian SHREC, then are about judging comparative quality within a pool that do meet standards of technical eligibility. At some point judgments about relative significance and relative importance have to be made.

This article is a reflection on the processes by which such designations come to be made. It is based on a specific funding body (the ARC) at a specific time (2002-2004). At this time applications did not have to meet technical definitions of ‘science’ of the kind defined in the work of the US Institute of Education Sciences to be eligible for funding. As well, some of the specific features of the process I discuss (size and composition of the assessment panel, for example) are not identical with those in other countries, and are subject to change over time in the ARC itself. But the issues that I raise here, about who does the judging, and how implicit comparative judgments of significance and quality are made, will remain pertinent. Similar issues permeate too the workings of many internal research funding committees that operate within universities.

For the past three years I was the sole Education member on the twelve-member Panel of the Australian Research Council that deals with applications from the ‘Social, Behavioral and Economic Sciences’. Between them, the members of this panel assess all applications for ARC competitive grants both directly (by giving initial assessments and scores) and indirectly (by reviewing and making decisions about the corpus of assessments and scores from a range of other readers and the weighted score that results from these). This experience of assessing literally hundreds of applications; of engaging with ‘experts’ from other fields and seeing how they as well as other readers inside and outside Education assessed the grants I had assessed; of worrying all the time about what I was doing and about how the processes worked; heightened my interest in something I have been interested in for a long time – how in various contexts we come to decide that something is ‘good’. It led me to write a book (Yates, 2004) looking at that issue across different types of education contexts and considering the biographies and agendas and criteria that are brought to bear in them. This paper will focus specifically on my ARC experience and the issue of competitive grant schemes. But the comparative background that I mention here – that what counts as good or as significant for the purposes of a competitive grant application is not identical with what counts as good for commissioned research, or for practitioners, or for book publishers – is an important part of that story.

For ethical and legal reasons I will not be discussing any inside stories of what happened in panel deliberations or conversations I had with fellow panel members. But just working with
what is on the public record and with my own self-scrutiny of what was happening to me I think is sufficient to raise the issues I want to raise.

Some personal background:
Before being appointed to this committee, I had been successful in gaining some ARC grants, and also had experienced being unsuccessful with some applications. In one case the same longitudinal project was successful in its first application; unsuccessful when I applied for funding for the next stage three years later; and successful again when I resubmitted an application for the same stage one year later. My background is history, sociology, philosophy. I do qualitative research, sometimes on policies and documents, sometimes on small numbers of people. At times I draw on feminist and poststructural theories, and I am interested in the work being done by people who work with those theories. I have taught graduate research methodology courses for many years, and have written articles about the value of the type of research I do, and critical of some larger work, for example, the reliance on data-bases and the categories they work with. But the first year I was on the ARC committee, as I tried to get through the impossible workload of assessing some 140 applications (that routinely run to 40 or 50 pages) alongside my day job, I began to worry about how was I making decisions when I gave an application a score for significance. Some of the questions I began to think about were: ‘what makes something that is very small scale “significant”?’ ‘what makes something that is set up with a critical or deconstructive agenda “significant”?’ ‘how do I judge the issue of significance when I am reviewing an application totally outside my own disciplinary field?’ ‘how are other panel members and outside readers likely to judge these questions?’

In this article I want to say something about these issues by looking at two aspects of the process by which ‘significance’ gets judged in competitive grant schemes. First, who is doing the judging here – what sorts of people get selected to be assessors and ‘experts’ and how does that shape what counts? Secondly, what is being judged in terms of an application: what are the formal criteria and categories, and what textual representations allow assessors to decide the quality of the application?

Who are the judges and what is an ‘expert’?
As a member of the ARC panel, for the first time in my life, I was an officially designated ‘expert’. The panels were named, in the earlier years of my membership, the ‘Expert Advisory Committees’, and more recently, the ‘College of Experts’. The naming is not (or not only) chosen to flatter those who agree to do the work; it is chosen to signal in the public sphere some procedural propriety about how the government is going about the task of distributing money to researchers.

Who is an ‘expert’?
In the case of national competitive grant applications, to be constituted as an appropriate ‘expert’ an academic needs more than a doctorate and an institutional position: they need themselves to have demonstrated a profile within their field that constitutes them as a successful researcher and senior member of their field. That is, governments and semi-government bodies have to appoint people whose qualification to serve has a certain public form to it, and one that can be recognizable as appropriate by those outside the field in question. Some of the ways this might be demonstrated include having a higher than usual success rate in themselves attracting research grants; being frequently cited by others; holding senior roles in professional bodies of their disciplinary field; having senior university positions. In some schemes, these judges are selected solely from textual data-bases – such as
records of previous grant recipients or citation indices, but frequently some more direct processes of networking and nomination are also part of the decision. For example, universities or professional associations of researchers are asked to nominate people; or those currently on the committees use their own networks to solicit suggestions of appropriate successors; or there is an explicit process of triangulated checking about potential members. These days, higher level committees also often include one or more people to represent ‘user’ or ‘government value for money’ perspectives. Government-funded committees also often have to meet some agendas of representation in their overall composition – for example, in the Australian context, representation of different states, and, sometimes, gender.

Without considering individual personalities and idiosyncrasies and exceptions, what types of assessors and what types of experience do these processes tend to throw up? The process, as I have suggested, draws in people whose CV achievements make their appointment publicly defensible (a requirement for statutory bodies) – that is, they have achieved in certain conventional measures of this, in ways that are recognizable across fields. The way these assessors are selected tends to mean too that they are people who are involved in certain networks and have held certain positions (the kinds of things Bourdieu writes about in his analysis of French educational and social hierarchies, see Bourdieu 1998). Let us consider some of these characteristics and how they may play out in terms of what experiences members may bring to the act of judgment.

‘Successful’ themselves:
These are likely to be people more attuned to what are the signs and criteria of doing well, than to the problems of those who struggle; more attuned to the experience of the haves than the have-nots (and race, gender, ethnic, class differences may well be part of this – committees commonly under-represent women and have less sensitivity to the processes by which women become under-represented). For example, in terms of assessing track record of applicants, it is likely that a ‘good researcher’ will be judged to be one who has a coherent, steady, productive and building repertoire of achievement. Periods not in academic work, long periods without promotion, moves from tenured to untenured work, may all be judged with suspicion. I once read an outside assessor’s comment on an application from an early career researcher which commented unfavourably on the fact that the applicant had chosen to take a year off after their PhD to travel! The assessor said this might throw doubts on whether they would stick at a three year project if the grant were awarded.

Experience of holding senior position and doing related administrative work (being a Dean, chairing university committees, holding office in professional associations, etc.)
These are likely to be people attuned to working with administrative requirements and to judging very quickly any breaches of rules. In demographic terms they are likely to mirror the same demographic characteristics that have been documented in senior ranks of universities: for example, women and ethnic diversity under-represented; in smaller countries, academics of UK and USA origin over-represented – and all this will be reflected in their personal networks and associations and the experiences they bring to the judgment of applications. On the committee I served on, members serve 3 year terms, and there is an annual turnover of members, but it has never had more than 3 out of 12 women.
Senior positions, ongoing achievements in research and in career

These are likely to be busy people, doing many ‘important’ things other than judging research grants of others: they will have limited time, will be impatient with the lengthy reading involved which is often an extra to their main work, and is often not paid in more than a token way.

All these points are relevant to how applications will be read, and how they need to be written to be successful. To put this another way, the judgment of ‘quality’ of a researcher here may have some historical and gendered and discipline-biased biographical element (whether one’s own career was a linear one, for example). And the judgment of quality of the potential project will be influenced by the style of the writing of the application: how helpfully it conveys its key points to an inevitably busy reader who must skim many applications.

Specialists versus non-specialists: and the issue of ‘disciplinary mix’.

In the ARC system at the time of my membership, each application was assessed by two members of the expert panel, two outside assessors (‘readers’) who are in the disciplinary field of the application but may or may not be working with the methodologies and theories of the particular application (since their role is to read around 10 to 20 applications altogether, to produce ranking as well as scores); and up to four specialist assessors who are chosen for their knowledge of the application and may only read one or a few applications. ARC schemes, like many competitive grant schemes including smaller internal university schemes, deliberately set up a system whereby applications will be judged both by those with close expertise in the field of the grant, and some who are ‘outsiders’ to the specific area of the grant. Indeed the more money and publicly accountable funding involved in a competitive scheme, the more likely are judges and judgments to move beyond ones internal to the academic area of study, because in part the point of ‘competitive’ schemes is to make judgments about the relative merits of different fields of research, not just relative judgments within a given area. So the first point here is that for an application to be deemed ‘significant’ it must be able to make a case that convinces someone who knows the area well and also someone who may have no knowledge of the theories or research in the area at all – which is one of the delicate writing tasks the applicant must achieve. How does someone from another field judge that something is a ‘significant contribution to knowledge’? This is a question I will return to later.

The system is set up on a principle that there will be both ‘outsider’ and ‘insider’ judgements, including cross-disciplinary judgement, but there is often a historical element to what areas are grouped to be represented on the panel, how much representation different areas are given, and what is seen as ‘within a discipline’ as compared with across or in different disciplines. In the period I was a member, the committee assessing education grants was the ‘Social, Behavioural and Economic Sciences’ panel, and it was comprised of

- two psychologists (one ‘social’ and one ‘cognitive’)
- three economists (one representing ‘social policy’, one the industry or user representative, and one representing the discipline of economics)
- a management academic (who began as an economist)
- one sociologist
- one linguist
- one demographer
- one political scientist
• one health scientist
• one educationist.

For not very clear reasons, Anthropology, Legal Studies and Asian Studies at this time were located in a different committee, Humanities and Creative Arts; while Applied Linguistics was located in Social Sciences rather than Humanities, in which faculties it is often based.

Given that applications are always assessed by more than one committee member, and assessments overall are discussed and scrutinized by a committee as a whole, both the disciplinary mix of a committee, and the extent to which a particular type of work is represented can have some influence on whether applications are successful.

In the 1980s, Australia had a funding body specifically for Education research, the Education Research and Development Committee (ERDC). In Canada, within the body most equivalent to the ARC, the SSHREC, there are a number of panels specifically to cover different areas of education, as well as one for inter-disciplinary studies. An ARC-commissioned review of Education in 1990 had recommended that Education should be represented by at least two panel members, one with a quantitative and one with qualitative interests, but this was only finally achieved after the end of my term, in 2005, mainly as a response to the evident workload issue, that Education was receiving the highest number of applications per panel member. So, for education applications, at least one important assessor of whether they are doing something significant will be someone who comes from psychology or economics or linguistics or other disciplines represented on the panel. Sometimes both panel member assessors will be from outside education, since for conflict of interest and other reasons, not all applications in a discipline go to the member representing that discipline.

When new members are selected, preference is given to those whose disciplinary background is seen as having some breadth or cross-disciplinary experience; and to those who have been in positions where they have had to deal with judging across fields, for example on university committees. People will not be chosen if they are thought to be automatically biased against certain types of research. However, in terms of what they know intimately, in the time I was on the committee, there were never more than 3 out of 12 members whose background was not primarily quantitatively-based research; and even fewer who would be identifiable with any specifically radical research approaches. However this too has some historical dimensions. Traditions relating to socio-economic disadvantage and more radical perspectives on poverty are longstanding in the field and more of the potential research assessor community across disciplines would have some knowledge of these, than they would be highly attuned to feminist perspectives, or poststructural ones. Those willing to work as assessors are, on the whole, people who try to promote research, who try to be ‘fair’, who have a reasonably good knowledge of the range of work in their field and who do worry about encouraging new researchers and new forms of research. But their positioning in their own field is likely to be more towards the middle than at an extreme; they are unlikely to be seen by their colleagues as extremely radical; and their actual experiences are likely to have given them much more knowledge of mainstream methodological approaches than unconventional ones.

A further issue that arises is how assessors assess work in disciplines other than their own. In my experience (scrutinizing myself, and my experiences on a range of funding and promotion committees at different universities), having an economist or psychologist assess an education or nursing application, and an educationist assess an economics or psychology application is not simply a parallel activity. Journal titles of ‘straight’ disciplines (for
example, *British Journal of Psychology*) have a known meaning and prestige to people outside the field. Journal titles within a field (for example, *Journal of Curriculum Studies*) do not have a known meaning to outsiders – it is not apparent whether that is a ‘prestigious international publication’ or a low ranking one. Often the more specific the title, the more it is discounted.

If uses of citation indices and journal impact factors were to become mandatory, there would be a more uniform accountability in terms of how publication achievement was being ranked, but this too would be deceptive. In some fields and types of research, fewer publications (that is, larger, complex digested arguments) are the norm than in other fields. And citation measures are weighted to work that is of interest for a global market, and more particularly an American market, neither of which are unproblematic indicators of significant research for Australian education. For an education researcher then, decisions about mode and vehicle of publication, even decisions about article titles, may take different forms depending whether your prioritised audience is this grant application process or whether it is communicating to other researchers within your own field, or addressing a broader professional or public readership. To put this another way, ‘quality’ or ‘significance’, at least in its written form, is not identical for each of these three groups.

**What does ‘significance’ mean, and how is it represented textually?**

Audit culture is seeping into most aspects of university and research life, requiring tightly specified and procedural definitions and standards for everything from teaching quality, to ethics, to research standards. The move to mandate technical ‘scientifically based research’ principles is one particular form of this, and one that would be a significant modification of past practice, an indication that, unlike with scientists who are the putative benchmark, ‘experts’ in the social sciences cannot be trusted to assess methodological adequacy and must have these set as regulations to govern their practice. Governments are also becoming more involved in designating substantive research priorities. Commonly major competitive grant schemes have very detailed instructions to applicants and finely specified application procedures. Nevertheless, to this point in time at least (the tighter definitions of ‘scientific’ and ‘evidence-based’ may bring changes in the future), assessors are normally not instructed in any detail on how they are to judge quality of the researchers, and how they are to judge quality of the research project design itself.

Competitive grant schemes in most countries work only with fairly broad criteria or categories for scoring, and do not attempt to dictate closely how the assessor does this work. In part, these judgments are built into the design of the application form itself (for example, in terms of track record, what material the research team is asked to supply in terms of their achievements or track record) and into the procedures of assessment and ranking. For example, in one of the schemes I was working with (Discovery Grants), ‘quality’ was deemed to be made up 40% researcher quality (track record); 20% methodological rigour; 30% significance; and 10% contribution to the nation. This was how these categories were weighted in producing a final rank. But, within those categories, the scoring itself was a matter for judgment by each assessor, that is, the judgments are considered to be what the ‘experts’ doing this work ‘know’ and can bring to the task. Assessing quality in this sense requires a combination of technical and substantive knowledge and experience.

The category I want to focus on is ‘significance’. In the context of research, what does ‘significant’ mean? Does it mean important? or major? or crucial? or noteworthy? Different images are conjured up by these different synonyms. Does it mean ‘within the context of
research to date”? or ‘within a program of research endeavour?’ or ‘within the context of making a difference to the world, or improving people’s lives”? Again, the backdrop against which it is considered might substantially alter the comparative ranking of one grant against another. And does it mean ‘significant given designated political agendas and values for research’?

The system that I was part of was premised not on tightly defining how assessors do go about making their judgment of significance, but on having multiple insider and outsider reading and scoring which are then brought together. In principle, given that there is no perfect way to deal with the fact that we do not as academics have homogeneous judgements about what matters, this process does have some significant strengths in its checks and balances, and does deal reasonably well with some problems of value diversity and prejudice. But, in my experience, the system does bring with it a tendency to some commonsenses rather than others.

Scale matters:
Within the Education research community (at least until No Child Left Behind) it seems a long time since the legitimacy of case-study or action research had to be argued for. In everyday language ‘significant’ can mean ‘important’ or ‘striking’ or ‘breakthrough’ – it does not have to imply ‘bigness’. And yet, the first year I was on this committee, despite the fact that I can rehearse the problems of large-scale research at length, and have taught and written about small-scale and interpretive approaches for years, I did find myself reading applications for research that was only going to be done in a couple of school sites, or on a few people, and asking myself ‘is this significant?’ (that is, ‘is this important enough to get major national funding?’). I was also struck by finding that it was extremely common for outside assessors to raise questions about sample selection for applications involving small numbers, even though these questions seemed to be linked to a particular concept of ‘representativeness’, that seemed inappropriate to the methodology and design of the project. (To be fair, on the other side, just because a project was big, did not mean that it was necessarily seen as significant.)

A number of things are at work in producing this concern about scale and about sample selection. One is that not only are the majority of assessors in this particular scheme more experienced in quantitative and experimental traditions of research, but that this sense that size matters is also closer to some broad non-specialist commonsense about what is important. Although ‘significance’ can mean ‘important conceptual break-through’ as well as ‘large-scale’, the latter is quickly evident, whereas the former requires more difficult and careful reading and judgment. Another way of putting this is that a proposal that involves researchers across a number of countries conducting large-scale investigations of their education systems has a taken-for-granted appearance of being major, whereas a study involving only two schools, or a limited number of case-studies has to make a case as to why it is important. In terms of the ‘significance’ criterion, there is a more extensive writing task to be accomplished in the second case than in the first.

Moreover, regardless of whether SBR (scientifically-based research) principles are mandated, my own experience outside my specific research community suggest that the idea of ‘controlled comparison’ is still a strong lynchpin of what good research methodology looks like for most people. My perception, notwithstanding the long history of interpretive and critical traditions of research (let alone poststructural research), is that there is not a widely understood sense of how knowledge can be built from one or a few sites, or indeed, can be
built interpretively. There is a widespread perception that these days ‘good research’ involves doing a few case-studies as part of a larger project, to give a human narrative to the things that the big surveys show – but these are constructed as illustrations of the survey-derived picture, not as knowledge-building in their own right.

Elements of the ‘evidence-based’ movement have taken a stronghold because they are close to a widely held commonsense about what makes research trustworthy and important: that it sets up careful selection and controls and that it is of a scale that matters; that in principle the ‘results’ should not be dependent on the say-so of the researcher who did it. The textual task of establishing ‘significance’ and adequacy of methods is a simpler task when working with that established commonsense than when challenging it. And it is still more common for these issues about method to be posed to education researchers than for economists or psychologists to be asked to demonstrate that their findings will make a difference to teaching and learning in ordinary school contexts.

Competitive grant schemes are seen as funding schemes for serious research, research that makes ‘a contribution to knowledge’. In terms of assessment of grants a ‘contribution to knowledge’ to some extent is defined by its negative: it is not research that only benefits those involved in the research activity, but rather is research that will produce something that goes beyond the doing of it, will be something other than consultancy to fix up a particular problem now. But often implicit in the concern with ‘significance’ is a policy-related perspective on agendas: that big answers, international and national policies and research that speaks to these are more important than research that speaks to practitioners in particular situations, an issue that has been of very longstanding interest in the field of education research.

In this context, there is a further reason why justifying case-study or interpretive work as the content of a major grant application is more of a problem in Australia, and that is the small size of the funding pool even in what was previously called a ‘large’ grant scheme. In the UK and in Canada it is common for projects to be funded that involve numbers of different sites, engaged in parallel and complementary projects in which ‘case-study’ interpretive work is being done, each site (that is, each sub-component of the project) funded to a similar level to an overall Discovery grant in Australia. Where the average grant awarded is only $60,000 per year, as it was in my term on the committee, a researcher clearly has to make some choices between depth of involvement with the site or subjects and the numbers of sites or subjects involved. My concern was that because there is such competitive emphasis in the schemes, and such attention to value for money, this seems to encourage attention to coverage or scale, to doing more things, rather than to doing something that is limited with time-consuming attention. Given that in Australia large numbers of doctoral students in education do qualitative studies, and that in thesis examination there is more attention to the detailed quality of the work than to questions of overall significance to the field, this may lead to more of a problem in moving from the thesis to the grant than if they were working in fields like psychology or economics or history.

What kind of a discipline is education? Pinning down compared with opening up:
I was on a panel with a member of the Humanities panel once, answering questions at a university about what we thought made a difference in terms of applications that got up and those that didn’t get up. My number one point was ‘track record’. His number one point was ‘having a good idea’. In this article I am discussing the way judgements of quality are situated and embodied: who is involved and what combinations of people are involved
matter. Disciplines are historical artefacts that produce different ways of understanding what research questions look like, and what methods look like. I do think there are some different starting points in ‘humanities’ than in ‘social sciences’ for thinking about what good research ‘is’, and it is something we might think about more in terms of considering education as a field.

When applications are judged for their significance, assessors are dealing with projected work-to-be-done, rather than rating the quality of what the project actually came up with. In forecasting quality, they rely on two main things, track record of the researchers, and description of project design and methodology. Assessors do make judgments about whether a topic area matters, and whether this team will produce good results, but the focus on methods often attempts to pin down reliability. Have you adequately counted something in a controlled way? Have you pre-specified sufficiently the procedures you will use in your observation? Have you sufficiently pre-specified your analytic approach? The extent that assessors appear to want these details to be pre-given in some detail I find a bit worrying, not only because I find that hard to do (and I think there are disciplinary and paradigm differences in what is seen as appropriate), but also because it implies that the outcome or findings, for qualitative as well as experimental and survey work, are simply the fairly mechanical outcome of what is built into the design, rather than something that exceeds this. This also may be a peculiarity of the current Australian system, where all the assessment of rigour is essentially front-end, on the application prior to the research, rather than on later reports and accounts. The issue of finding new ways of seeing, of noticing and understanding unintended effects and processes is not well captured by an over-emphasis on front-end scrutiny of rigour of design.

Cross-disciplinary judgments:
Can research that has an explicitly critical or iconoclastic agenda be a ‘significant contribution to knowledge’? In recent decades, both within nations and cross-nationally, governments have attempted to take more control of research enterprise. There is more funding but also a more direct attempt to define research priorities, to ensure direct (and, in Australia at least, short-term) commercial or political benefit from the research that is done. The current schemes score ‘national benefit’ separately from ‘significance’, but I do not think I have read an application that does not make the case for both (and I would not advise anyone differently). I think that in the case of education ‘national benefit’ leaks into and overlaps what is scored in terms of ‘significance’. Short-term and applied significance, and relevance to political issues of the day, are more intrusively felt in education (that is, seen as mandatory) than in some other areas. This will be potentially be intensified with the attention to ‘impact’ in the proposed Australian Research Quality Framework (Yates 2005). Applicants may spell out what deeper knowledge or program of research they are trying to build, but my experience on many different cross-disciplinary committees is that everyone feels able to judge whether education research is ‘significant’ or not, regardless of whether they have any knowledge of what has already been done or not done.

Final comments
Often critiques of funding regimes are concerned about blatant directions and decisions in relation to political control of funding, and everyone swaps horror stories about supposed prejudiced individuals and assessors and results. In this article, I have been trying to reflect on emphases that arise regardless of individual attempts to be fair, as a result of processes and structures through which they are operated. Any system for distributing funds, or allocating success, builds ways of recognizing or preferring some representations and judging others as
lesser. There is no hands-off way of distributing competitive research funds and I have not only been complicit in the processes I describe, but would defend much of them given the constraints of funding and likely alternative scenarios for allocation. What I have been trying to focus on here is the way ‘significance’ works not as a matter of deliberate prejudice, but as an effect of the very processes that try to balance interests and select a range of insider and outsider ‘fair’ assessors. I have tried also to draw attention to an ‘educated commonsense’ at work, one that has methodological perspectives much more like the ‘evidence-based’ movement than those who object to this.

Given limited funding, there must be schemes where decisions about funding allocation are made. I have chosen in this article to focus on some aspects of the particular scheme I was involved in, one where forms of research that are heavily represented and approved in the education research community (for example action research) might be judged less favourably as a result of both explicit criteria (‘national significance’, ‘contribution to knowledge’) and the implicit experiences and orientations through which assessment judgements are made in it.

In the social world research questions and fields we work in are not tightly bounded; disciplines change and new ones form. Working with some form of insider and outsider assessment from a range of sources in my opinion is one reasonable way to try to judge quality. The experience of working on a panel of the ARC is one that gave me a great deal of sympathy for the constraints, interests and technical attempts to be fair with which the funding body was struggling (some might see this as ‘going native’ – yet another problem that may arise with a ‘college of experts’ structure). The role of the ARC and of similar bodies elsewhere does include advocacy for research interests to government and to public as well as assessment and regulation of researchers through the granting schemes. What I have tried to do here is show one specific set of experiences and reflections about judgments of significance and quality. I have tried to illustrate how assessments of quality are necessarily situated in history, process and embodied experiences. This does not make them meaningless or ‘merely relative’ or necessarily bad. But it does mean that we need to go on attending to who gets to do the assessments of (relative) quality, why, how and for what purpose.

Notes:
1. Parts of this article draw on that recent book (Yates, 2004), and the comparative argument that education research ‘quality’ is not a stable judgment across different arenas is elaborated in it.

2. Here are some guesses about reasons I may have been appointed to this committee, given my less than perfect record with my own previous grant applications, and that previous education appointees had been drawn from quantitative researchers in mathematics or science or assessment: (1) I was a past president of AARE; (2) I had managed not to offend many people in that association (or at least those involved in the cross-checking of names); (3) the ARC were embarrassed about the lack of women on the committee; (4) I had cross-disciplinary experiences; (5) possibly, I fitted their need to balance state/regional representation; (6) I had had previous ARC grants and (7) I had done previous assessment reports conscientiously and on time.

3. An absence of women on a committee is a potential embarrassment – though a number of the ARC panels at the time I was appointed did contain no women. There clearly was little
concern about attaining a reasonable gender balance, since the highest number of women on any of the panels was 3 out of 12.

4. Please note that this is not an empirical report on my fellow panel members (or necessarily of myself). In the period I am discussing, the committee chairs and executive directors spent some time cross-checking views in the professional community about individual assessors before appointing people, and my experience was that people selected often had wider than usual experiences and did attempt to be careful about prejudices, conflicts of interest and the like. Nevertheless I am trying to consider here what kinds of experiences may shape the broad group selected to be assessors, precisely because of what seem like ‘proper’ criteria that they be experienced, successful, knowledgeable.

5. This appointment relates to health applications that are social science-based in their orientation. Straight clinical or scientific studies in the health and medical area go through another research body, the NHMRC (National Health and Medical Research Council).

References: