

How do we recognise good research?

PSYCHOLOGY has evidently been vexed by qualitative research. Articles in *The Psychologist* and elsewhere have passionately argued whether it will save the discipline from the perils of positivism, or herald its banishment from the community of sciences into subjectivity and obfuscation (e.g. Gillett, 1995; Morgan, 1996). Of course, qualitative methods have long been employed by other disciplines, and are now being quietly embraced by medicine (Macnaughton, 1996; Mays & Pope, 1995).

Nevertheless, the vehemence of this debate in psychology is not surprising. After all, this is the science whose practitioners challenge not only each other's results but also their methods. However, to accept the debate as just another instance of the tensions inherent in psychology would be to miss an opportunity. Rather than becoming involved in the debate, this article will take a novel look at it and will argue that issues it has helped expose can clarify our thinking about how we evaluate research in general – quantitative *and* qualitative.

'Methodologism' in evaluating research

Disregarding the lofty epistemological positions that conventionally define the qualitative–quantitative debate, this article instead reflects a view that debate is better grounded in behaviour: psychologists' judgements of research quality. It emerges from my own need to assess the quality of research in the course of learning to write and review qualitative papers and to supervise and examine qualitative theses, against a background of quantitative research experience.

Criteria for judging quantitative research are familiar to psychologists, and can be reliably operated by examiners or referees anywhere. They can ask, for instance: Were groups allocated randomly? Was the sample big enough? Were appropriate statistics used? Notice that these judgements are about methods, or procedures. In effect, and without necessarily thinking in these



PETER SALMON discusses *anarchism, methodologism and the quantitative vs. qualitative debate.*

terms, the examiner or referee is applying an epistemology of 'methodologism'; research is good if it has been conducted according to certain methods.

There are growing signs of methodologism in judgements about qualitative research in psychology, too. Articles describing how such research should be done (Elliott *et al.*, 1999; Stiles, 1993; Turpin *et al.*, 1997) are sometimes used by writers and referees alike as checklists of methods to be ticked off; for instance, whether a tape-recorder was used, a field diary kept or an audit trail left. Just as methodologism sustains the quantitative tradition in psychology, it is beginning to shape its qualitative work.

The limits of methodologism Unfortunately, methodologism is a limited epistemology. It is a forlorn belief that quality can be guaranteed simply by following procedures. One problem is that many things that are called methods are not. For instance, aspiring qualitative researchers sometimes write that they 'used' triangulation. However, this term refers to an approach whereby researchers address their subject from different perspectives or with different data – or with different methods. So it is a category mistake to regard it as a method. Turning principles like this into qualitative 'methods' on a checklist has been criticised before (Barbour, 2001; Chapple & Rogers, 1998; Reicher, 2000). Its absurdity is clear when researchers write that they 'used' grounded theory or discourse analysis. A spade or a computer program or a statistical technique can be 'used' but an epistemological perspective, a way of thinking – which is what discourse analysis and grounded theory are – cannot.

It might be objected that 'using' a technique is just a manner of speech. But the words people use are the best guide we have to what they mean. Moreover, words shape, as well as express, what people understand, so new researchers who follow guidance to 'use' a technique might be less likely to seek to justify the quality of their work in a more robust way.

A serious limitation of methodologism is that it asserts rather than justifies the ascendancy of methods. When pressed to defend their allegiance to, for instance, controlled experimental design or interpretative analysis, researchers might cite the need for objectivity or for embedding meaning in the context. But these are themselves essentially methods and so the justification is circular. Feyerabend (1975, 1978) has pointed out this inherent circularity: justification for the value of scientific methods cannot logically emerge from the methods themselves. Natural scientists accepted the methods of natural science, not because they could prove their value, but because scientists and sponsors liked what they saw the methods deliver.

When pressed, methodologists might state an epistemological position to justify particular methods. Indeed, some qualitative researchers argue that investigators should be explicit about the epistemological basis of their work. There are two problems here. First, epistemologies describe the use to which methods can be put, not which methods can be used. For instance, qualitative methods can serve a researcher who believes that research *discovers* underlying reality as easily as they serve one who believes that researchers' interaction with research participants *constructs* reality.

Secondly, it is a rare researcher who thinks through an epistemological position before choosing a method. Such positions are more often post-hoc rationalisations of what has been done. In reality, researchers use methods for historical, ideological or practical reasons – they use the methods they have learned to use and that they *can* use. A more realistic and robust approach to evaluating research quality is needed, instead of the unholy alliance of methodologism and retrospective ‘epistemologism’.

An epistemology of quality
As well as identifying inherent limitations of methodologism, Feyerabend (1975, 1978) described an epistemology, grounded in scientists’ behaviour, that extends to evaluating quality. His ‘anarchist’ epistemology states not that rules are unnecessary but that all methodologies have limits, so the value of a methodology can be tested only by research that violates it. Scientific progress has resulted,

therefore, not from allegiance to methodological rules but from breaking them. Therefore the anarchist scientist might choose (or create) any of an infinite array of methods.

This leaves the question of how the choice is made in any instance. Feyerabend’s view is that the anarchist scientist is playful rather than precious with methodology. Methods are chosen simply because they interest the researcher or because the researcher values their products. For example, when I and colleagues recently set out to study paediatric patients’ pain, we soon noticed that our informal observations of nurse–patient interaction told us much more interesting things about pain than the questionnaires that we intended to use ever could. We could not describe our observations
using any methods
that we knew,

but by learning and adapting some ideas from qualitative methodology we were able to report what we thought that it was important to say (Byrne *et al.*, 2001). Guba and Lincoln (1982) argued that the fundamental axioms of quantitative and qualitative enquiry are arbitrary: whether to be quantitative or qualitative in any specific study should be decided by ‘fit’ with the phenomenon being studied. Guba and Lincoln did not, however, follow their own logic as far as did Feyerabend to infer that this decision is subjective. Feyerabend’s ideas provide an epistemological rationale for views that have emerged recently in psychology: that ‘good’ research is playful, that research that slavishly follows methodological rules stultifies the discipline, and that real scientific progress results from imagination, creativity and common sense, rather than merely deduction and induction (Rennie, 2000; Robinson, 2000).

Of course, we have to go further than this, because what is interesting, fun or valuable or seems to ‘fit’ for one researcher might leave another bored, uninspired or dissatisfied. Feyerabend’s anarchist view cannot tell us how to decide what is good research. Instead, it directs us to be honest and explicit about how we already decide this. It warns us that these judgements are inherently subjective and intuitive and reflect values that are broader than science. Therefore, a realistic debate about the ways in which psychological research should be evaluated will need to transcend parochial issues about, for example, qualitative vs. quantitative methods.

The test of this argument is whether applying it to the way we evaluate research in psychology can expose subjective values that, although normally neglected, do influence our judgements and so merit recognition and debate. The following section proposes some values of this kind that recent debate about qualitative vs. quantitative methods has helped to identify, which apply equally to both kinds of research. Readers might object that the values to be described are arbitrary or partial, and have been neither deduced from an epistemological position nor induced from a factor or thematic analysis. That is the point. They illustrate an epistemology that is based on subjectivity and behaviour: What do psychologists really look for when evaluating research?

The researcher should not try to mislead Since Rosenthal (1966) demonstrated investigator effects on

research outcomes, the reality that results of psychological research reflect the researcher as much as the researched has been inescapable. The scientist is a participant in, rather than observer of, the field of study (Potter, 1996). Some qualitative researchers celebrate this stance by arguing that researchers' experience, situation and motivation are integral to their research.

However, we still seek science rather than the biases of politics or rhetoric. Unfortunately, the boundary of science is not clearly defined, whether by the familiar practices that claim to ensure objectivity and reproducibility of quantitative research, or by the emphasis in qualitative research on disclosure of researchers' situation and perspective. Thus, reproducibility of quantitative findings is not assured by analytic techniques (Sohn, 1998) and is rarely tested in practice (Potter, 1996), and qualitative researchers' emphasis on reflexivity can be merely a rhetorical device for persuasive effect (Yardley, 1997). Therefore, the boundary between science and other kinds of persuasion remains ill defined (Potter, 1996), even though we apply it daily. The consumer of research is therefore left to make a moral as much as technical judgement that the researcher has not been 'too' biased.

Methods should be rigorous Like other disciplines, psychology values 'discipline' for its own sake. Researchers are expected to follow rules and conventions that make their task difficult and distinguish social science from journalism. 'Sloppy' research is deprecated.

Quantitative research is easily evaluated in this respect. There is a well-established methodological and statistical framework for recognising rigour, at the pinnacle of which is the randomised controlled trial (even though this design does not guarantee clinical generalisability, scientific advance or replicability; see Holmberg *et al.*, 1999; Sohn, 1998).

There are now many sources of similar rules and conventions that offer comparable rigour in qualitative research (Elliott *et al.*, 1999; Mays & Pope, 1995; Turpin *et al.*, 1997), and qualitative researchers sometimes cite strenuous 'methods' such as purposive sampling, grounded theory, multiple coding, triangulation and respondent validation. Regarded purely as methods, however, each of these can degrade rather than enrich research (Barbour, 2001).

Adhering to demanding conventions defines what it is to be 'a discipline'; it is inescapable that research is judged against such conventions. Sometimes, adopting particular conventions expresses an epistemological position. However, acceptance that we judge the level of 'discipline' for its own sake would allow more realistic debate about the importance of these conventions. It would also permit more general acknowledgement that research should be judged in its own terms (Reicher, 2000) and that authors should even choose how their work should be judged (Devers, 1999).

Analytic work should be done We expect to see that a researcher has not merely recorded and meticulously reported

'Often the problem is ignored, and the reader drowns in a sea of correlations or comparisons'

data, but that these data have been analysed. Qualitative researchers have well-developed approaches to ensuring that an analysis is developed until it 'works' – for example by achieving coherence and organisation or empowering the reader or participants (Stiles, 1993).

Paradoxically, it is in evaluating quantitative research that the role of analysis perhaps needs more debate. In the early days of quantification, when statistics were hard work and used preciously, data were reduced to a few statistics that were the analysis. Now, complex statistical procedures are available at a mouse-click and researchers can generate more statistics than they have data, which imposes new demands for further, post-statistical analysis. Often the problem is ignored, and the reader drowns in a sea of correlations or comparisons. Perhaps quantitative researchers need to debate taking a leaf out of the qualitative researcher's book, for example by seeking ways of organising results that indicate coherence or that focus selectively on findings that will 'empower' readers or participants. But such is the lure of the computer that some qualitative researchers also assume that citing a qualitative analysis computer program will reassure the reader that analysis has been done. In the absence of clear theoretical rationale and analytic thought, it will no more achieve this than SPSS will 'do' a quantitative analysis.

It should be possible to know when work is worthless Quantitative researchers are well used to stating aims as testable hypotheses. In reality, they are often not hypotheses, where they are written after the data have been analysed. Nevertheless, aims written like this are not purely rhetorical. They expose the coherence of the finished work to scrutiny, expose whether it has a message, and allow the reader to judge the work; aims that are trivial, unachievable or discordant with the design invite judgment that the work was futile.

New qualitative researchers are sometimes less helpful in inviting the reader to judge whether their studies have been successful or worthless. The aim frequently stated by students to 'gain a deeper understanding of' a phenomenon illustrates one potential problem. The word 'gain' defines the arbiter of success as the researcher (rather than reader), and 'deeper' is an ill-defined metaphor. By contrast, aims to 'describe', 'show how', 'develop some concepts to understand', 'develop a theory about' or even 'convey a more elaborate understanding about' a subject might more readily empower the reader to judge whether they were achieved. Students and examiners need debate about how we expect to be able to judge that qualitative research has failed, just as being open about the charade of hypotheses would allow more realistic debate about what we seek from quantitative research.

The work should matter to others, not just the researcher The argument in the previous section assumes that, by definition, research is not a self-indulgent activity for the researcher. It has to matter to others. Indeed, because research quality cannot come from adherence to methodological rules, it must instead be agreed by an audience – a scientific community (Rennie, 1998). This view leads to two further considerations for evaluating research.

The first is whether it is clear what the audience can take from it. Quantitative researchers rarely address this issue explicitly, often sheltering behind the 'generalisability' of their findings. However, there is a need for greater honesty about the products of their research. Even Cronbach (1975), whose alpha-coefficient is widely regarded as the stamp of permanence and generalisability on a questionnaire, warned that social phenomena are too context-bound to permit generalisation. Distinguishing

scientific from statistical inference (Holmberg *et al.*, 1999) is not the solution. Even if results are used to build and test theory rather than directly to generalise implications for a broader population, the theory's generalisability is at issue. One complication in evaluating the message of quantitative research is that it is almost invariably written up in a theory-testing way which, as noted above, is often pretence. Given the exploratory nature of much statistical analysis, and the habitual disregard of Type 1 errors, perhaps quantitative research should more openly be judged according to its ability to produce and develop ideas rather than test them.

Qualitative researchers are quick to disown generalisability, but are often less clear about what stands in its place. Accepting that work is 'exploratory' does not go far enough; it describes the motivation of the researcher, not what the receiver might take from it. Similarly, reference to 'transferability' in qualitative literature ducks the question unless it is clear what is being transferred. There is no shortage of possibilities, although they are not always made explicit. Concepts developed in a study can equip researchers or practitioners to think or act differently in future. Findings can help to identify new hypotheses. Findings can simply be incompatible with prevailing assumptions; what transfers is the need to rethink those assumptions. The practical orientation of medicine might explain why qualitative research has been accepted with so little controversy – clinicians can simply judge for themselves when a finding is useful (Corbin, 1998; Macnaughton, 1996). Indeed, it has been proposed that in medical research synthesis the message's importance should be weighted, not just the method's rigour (Edwards *et al.*, 1998).

A second consideration arises from acknowledging that research is for others. If this is true, work that never reaches its intended audience – because it is not published or presented – or work that once disseminated is never read or understood or has no impact on anything or anyone, could not be regarded as research.

It is hardly controversial that the views of a scientific audience define what counts as research. It is less readily accepted that this recognition is ultimately a social act; that is, what the audience accepts is what it *agrees* to accept. Acknowledging this might make explicit some interesting problems in evaluating research. For instance, if dissemination to an audience is

a component of the research process, can students' research be said to be complete and examinable before they have demonstrated their ability to communicate it effectively, for instance by writing it in publishable form for a specified audience?

'Qualitative researchers are quick to disown generalisability, but are often less clear about what stands in its place'

Who counts as a legitimate 'scientific' audience, and how big does it have to be?

Changing the terms of debate Psychologists' debate about qualitative research presents a big opportunity. At last we are looking at how to evaluate research. The lessons to be learned are not just – or even mainly – for qualitative researchers. Indeed, the qualitative–quantitative distinction becomes trivial in the context of the issues raised here. There is an

opportunity to become much more open about the subjective and intuitive criteria that we use to judge all research.

Many readers will disagree with arguments presented in this article, and few will regard the issues raised as exhaustive. That is to the good, because the aim was not to reveal yet another checklist of truths to prescribe what research should be. Instead the intention was to promote a debate that will expose the subjective and non-scientific values that guide our evaluation of scientific research. Explicit examination of the criteria by which research is assessed is central to the strength of a discipline (Devers, 1999). Psychology needs a debate that is grounded in the untidy and intuitive reality of what we do when we evaluate research, rather than in the more ordered and oversimplified logic of what we write to justify what we do.

■ *Professor Peter Salmon is in the Department of Clinical Psychology at the University of Liverpool. E-mail: psalmon@liv.ac.uk.*

References

- Barbour, R. (2001). Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *British Medical Journal*, 322, 1115–1117.
- Byrne, A., Morton, J. & Salmon, P. (2001). Defending against patients' suffering: A qualitative analysis of nurses' responses to children's postoperative pain. *Journal of Psychosomatic Research*, 50, 69–76.
- Chapple, A. & Rogers, A. (1998). Explicit guidelines for qualitative research: A step in the right direction, a defence of the 'soft' option, or a form of sociological imperialism? *Family Practice*, 15, 556–561.
- Corbin, J.M. (1998). Alternative interpretations: Valid or not? *Theory and Psychology*, 8, 121–128.
- Cronbach, L. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116–127.
- Devers, K.J. (1999). How will we know 'good' qualitative research when we see it? Beginning the dialogue in health services research. *Health Services Research*, 34, 1153–1188.
- Edwards, A.G., Russell, L.T. & Stott, N.C.H. (1998). Signal versus noise in the evidence base for medicine: An alternative to hierarchies of evidence? *Family Practice*, 15, 319–322.
- Elliott, R., Fischer, C.T. & Rennie D.L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British Journal of Clinical Psychology*, 38, 215–229.
- Feyerabend, P. (1975). *Against method*. London: Verso.
- Feyerabend, P. (1978). *Science in a free society*. London: Verso/NLB.
- Gillett, G. (1995). The philosophical foundations of qualitative psychology. *The Psychologist*, 8, 111–114.
- Guba, E.G. & Lincoln, Y.S. (1982). Epistemological and methodological bases of naturalistic enquiry. *Educational Communication and Technology Journal*, 30, 233–252.
- Holmberg, L., Baum, M. & Adami, H.O. (1999). On the scientific inference from clinical trials. *Journal of Evaluation in Clinical Practice*, 5, 157–162.
- Macnaughton, R.J. (1996). Numbers, scales, and qualitative research. *The Lancet*, 347, 1099–1100.
- Mays, N. & Pope, C. (1995). Rigour and qualitative research. *British Medical Journal*, 311, 109–112.
- Morgan, M. (1996). Qualitative research: A package deal? *The Psychologist*, 9, 31–32.
- Potter, J. (1996). *Representing reality: Discourse, rhetoric and social construction*. London: Sage.
- Reicher, S. (2000). Against methodolatry: Some comments on Elliott, Fischer, and Rennie. *British Journal of Clinical Psychology*, 39, 1–6.
- Rennie, D.L. (1998). Grounded theory methodology: The pressing need for a coherent logic of justification. *Theory and Psychology*, 8, 101–119.
- Rennie, D.L. (2000). Grounded theory methodology as methodological hermeneutics. *Theory and Psychology*, 10, 481–502.
- Robinson, D.N. (2000). Paradigms and 'the Myth of Framework': how science progresses. *Theory and Psychology*, 10, 39–47.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts.
- Sohn, D. (1998). Statistical significance and replicability: Why the former does not presage the latter. *Theory and Psychology*, 8, 291–311.
- Stiles, W.B. (1993). Quality control in qualitative research. *Clinical Psychology Review*, 13, 593–618.
- Turpin, G., Barley, V., Beall, N., Scaife, J., Slade, P., Smith, J.A. *et al.* (1997). Standards for research projects and theses involving qualitative methods: Suggested guidelines for trainees and courses. *Clinical Psychology Forum*, 108, 3–7.
- Yardley, L. (1997). *Material discourses of health and illness*. Routledge, London.