Finn Tschudi

Do Qualitative and Quantitative Methods Require Different Approaches to Validity?

In Steinar Evale (Editor) Issues of Validity in Qualitative Research. 1989. Lund, Sweden: Studentlitteratur
Do Qualitative and Quantitative Methods Require Different Approaches to Validity?

Finn Tschudi

Introduction

Whether research is carried out under (predominantly) qualitative or quantitative "tribal banners", interpretations and conclusions must be justified. This requires, in Popper's terminology, a process of "conjectures and refutations". In this process, plausible alternatives to favoured conclusions are winnowed out. Much can be (and has been!) said about this process. Since, however, its basic aspects will always be discretionary, any attempt at a complete formal account will be doomed to failure.

Stated at this abstract level, the process of validating claims is the same for all types of research. I will argue for this statement, hoping to show that this is not a trivial point, since there are a variety of important implications.

First I will outline the major features of the classical statement on validity in the quantitative, psychometric tradition — Cronbach and Meehl (1955) — and briefly attend to Campbell's internal validity. Prototypical examples of research in the quantitative validity tradition are compared with a prime example of qualitative research. I hope to demonstrate the basic similarities in the approach to validity.

The concern for validity should be seen in the light of assuring safeguards against the ubiquitousness of human error. Some relativists will, however, find any search for validity misplaced, and will rather pursue interesting "narratives". My emphasis on fallibilism leads to a warning against relativism, which may be seen as irresponsible hybris, and to a critical comment on postmodernism.

The next section briefly outlines how Campbell's (1977) fallibilism is tied to his important contributions to the standard literature on validity.
Several claims, however, have been put forward to justify a need for different strategies for validating claims in traditional social science research and what some researchers construe as an alternative “human science”. It is hoped that an examination of some of these claims may be a contribution to clearing away the gunsmoke in the quantitative/qualitative battlefield (cf Mehl, 1955, Reichardt and Cook, 1979, Tschudi, 1982, for other examples of conciliatory attempts). My protagonists, who have gained their accolades in the quantitative tradition, have recently shown an admirable openness to insights from qualitative researchers, especially in the field of hermeneutics. I outline the more humble role left for quantitative approaches.

In line with the theme of this book, however, it is more important to point out what the qualitative camp may stand to gain if they also cease fire and shake hands.

Salter (1989) has outlined several aspects of the researcher’s attitude and background that are important in a more general concern for validity. This should be supplemented, first by considering love, then by bringing in sociological determinants of viable research communities.

Classical statements on validity

Construct validity

Prior to Cronbach and Mehl’s signal contribution “Construct validity in psychological tests” in 1955, the dominant view of validity was predictive validity; i.e. that a test was valid to the extent that it correlated with a future criterion, or concurrent validity, to be assessed by correlation with a contemporary criterion. If we disregard the different time orientations of these two forms of validity, we see that they are identical, viz. validity equals correlation with a criterion. Period!

There had been “a great deal of dissatisfaction with conventional notions of validity” (p. 281), mainly because in many cases no criterion or, at most, highly fallible criteria existed. It was, for instance, recognized that teacher evaluations or school marks would not be valid criteria for an intelligence test, since these measures might be heavily influenced by the teacher’s general liking for a pupil.

The problem was especially acute for projective techniques, as for instance the Rorschach test. Such tests presumably tap levels of functioning which it would not be meaningful to “validate” by e.g. behavioural criteria.

“Construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not ‘operationally defined’” (p. 282), or “when an investigator believes that his instrument reflects a particular construct to which are attached certain meanings” (p. 290). Notice that this differs from a naive view that the test per se is “validated”. It is the use to which a test is put, how it is interpreted, which is validated.

“One does not validate a test but only a principle for making inferences” (p. 297).

“Construct validity calls for no new scientific approach” (p. 282). I will give a condensed Popperian account of how Cronbach and Mehl construe validation. This will then be illustrated by their “example of construct validation procedure” (p. 283). I will later argue that the general account is equally viable for qualitative research.

Scientific approach

The point of departure is an empirical observation which admits of several possible interpretations. Relevant empirical observations are drawn in to support a preferred interpretation and to refute alternative interpretations.

It should be emphasized that my present concern is to illustrate the general logic, while comments on specific procedures will be made in a later section. The general logic is best illustrated by giving a capsule description of how Cronbach and Mehl propose to investigate the construct validation of test X, a suggested measure of “anxiety”.

Illustrative example:

Observation: “Suppose measure X correlates .50 with Y, the amount of palmar sweating induced when we tell a student that he has a failed a Psychology I exam” (p. 283).

This observation is actually an aggregate of more elementary observations (individual measures of X and Y). At present this is not important, but I shall return to the problematic nature of this point of departure.

110
Preferred interpretation: Test X measures anxiety proneness. Supportive observations: Test X correlates .45 with fraternity brothers' ratings on "tenseness." Test X correlates .68 with Taylor Manifest Anxiety scale. Mean X score decreases among four diagnosed groups in this order: anxiety state, reactive depression, "normal," and psychopathic deviate. Alternative interpretation: X measures "academic aspiration." Observations refuting alternative: X has negligible correlations with social class, vocational aim and value orientation.

The example illustrates that construct validation implies construction of what Cronbach and Meehl call a nomological network – an interlocking system of (more or less justified) laws. The open-ended nature of construct validation is underscored by the discussion of the impact of negative evidence. "When observations will not fit into the network as it stands, the scientist has a certain freedom in selecting where to modify the network." (p. 290). The test interpretation may be erroneous, there may be an error in the theoretical network, or the experimental design may be inadequate for testing the hypothesis.

Construct validation does not lead to any yes/no judgement, but "the task is to state... the degree of validity the test is presumed to have." (p. 290). It should be added that, although factor analysis could conceivably provide a numerical estimate of "degree", Cronbach and Meehl take a dim view of this possibility (pp. 287–288). Assessing "degree" thus requires a judgement, and in this process the researcher should report "what evidence and reasoning lead him to believe [in the proposed interpretations]" (p. 297).

Internal validity

A large class of research projects has an understanding of causal relations as their aim. Imputing causal power to presumed "causes" is, albeit of major practical importance, fraught with uncertainty, or, in technical parlance, several threats to internal validity must be considered. I will briefly consider a classical example, discussed in great detail by Campbell (1969).

Observation: There was a large drop in fatalities from traffic accidents in Connecticut from 1955–1956.

Preferred interpretation: A crackdown on speed violations caused this drop. Supportive evidence: There was an increase in tickets for violations, probably also reduced speed. Alternative interpretation: History – other events – caused the drop, for example a change in weather conditions. Observations refuting alternative interpretation: Meteorological observations showed similar weather conditions; furthermore, neighboring states did not show a decrease.

There are a variety of other alternative interpretations, conveniently systematized by Campbell as a list of threats to internal validity; History, maturation, instability, testing, instrumentation, regression artifacts, selection, experimental mortality and selection – maturation interaction. Space does not permit an adequate treatment of these threats, but everyone concerned with causal interpretation should be thoroughly familiar with this list. I will comment below on the general value of such lists for qualitative research.

A comparative example from qualitative research

In a fascinating article, Runyan (1981) discusses both the specific problem of why Van Gogh cut off his ear and, more generally, the problem of justifying interpretations of life incidents.

Since Runyan leans heavily on Popper, it should be no surprise that my capsule rendering of "scientific approach" above equally applies to Runyan's analysis. To illustrate, I select just two of the dozen or more interpretations he discusses.

Interpretation 1: Van Gogh's self-mutilation was an unconscious strategy for holding on to his brother's attention. Just before he cut off his ear (Dec. 23), he learned that he could not – as he had expected – spend Christmas with his brother.

Supporting evidence: Van Gogh had two later mental breakdowns which also coincided with the threat of losing his brother's emotional and finan-
cial support (his brother’s marriage, and the birth of a child). Furthermore, there were other incidents in Van Gogh’s life with “a masochistic response under situations of rejection or loss of love” (p. 1074).

*Interpretation 2*: Van Gogh might have emulated Jack the Ripper, but “as a masochist... he would reverse Jack’s act by mutilating himself” (p. 1071).

*Lack of supporting evidence*: “This particular explanation depends upon the assumption that Van Gogh read these stories in the local paper, that he noticed the earcutting detail mentioned in 2 of the 15 stories, that it made a lasting impression on him, and that it influenced him the night he mutilated his own ear. This explanation depends upon a chain of assumptions, none of which has direct empirical support, which leaves this particular conjecture relatively unsubstantiated” (p. 1073).

More generally, Runyan discusses several possibilities when there is a plethora of conjectures:

1. One interpretation is uniquely true
2. Different explanations supplement each other, but some may be ruled out
3. All are true in some way
4. None is true
5. Truth is not relevant, it is hopelessly arbitrary to decide

Possibility 4 is not explicitly discussed by Runyan, but might well be included. It should serve as an injunction to be maximally bold in setting forth conjectures, and also underscore a necessary humility. Even though we might think we have arrived at the truth, all such statements should be regarded as provisional. A case in point from Runyan is that the disorders of George III recently have been convincingly “reinterpreted as symptoms of porphyria, a hereditary metabolic disturbance” (p. 1073).

Runyan points to 1 as an ideal: “critical testing... can lead to the elimination of many of the claims as implausible or highly unlikely. Ideally, this process will lead to a single well-supported explanation”, but he recognizes that “more than one explanation that is consistent with the available evidence may remain” (p. 1075). In the present case, he veers in the direction of accepting Interpretation 1 above.

For our purposes it is important to underline that he clearly distances himself from 3 above. “To assume that all possible interpretations ‘are ulti

mately members of one happy family is to abandon critical thinking altogether” (p. 1073).

Polkinghorne veers in the direction of “one happy family” with his notion of “syncretic” research, since this “proposes... syncretizing the results of the multiple inquiries into a unified and integral result” (quoted from Salner, 1989, p. 60). This seems to imply that he believes all the inquiries to lead to results containing some truth.

5 epitomizes a *skeptical attitude*.

Carried to their extreme, both 3 and 5 represent a *dissociation from science*. 3 by abandoning critical thinking, 5 by exclusively catering to this at the expense of “conjectures”. Gergen (1988) carries us into a borderline realm where “questions of truth and objectivity recede into obscurity”. This may further lead us to *postmodern thinking*, which, according to Kyale (1988), is characterized by “indifference”, and an attitude saying “stop making sense”.

To me Popper’s slogan “conjectures and refutations” epitomizes “precious norms”, to use Campbell’s (1984) phrase. These norms are not unique to, but characteristic of, science at its best, and will be further discussed below. To weaken or move away from such precious norms may carry great dangers. To clarify this statement, it is necessary to take a closer look at the background for Campbell’s signal contributions to the question of how to attain validity in empirical science.

**Ubiquitousness of error, fallibilism, and a critique of postmodernism**

Campbell’s philosophical position is spelled out in his (unfortunately not published) William James Lectures in 1977.

“i am a fallibilist and antifoundationalist. No part of our system of knowledge is immune to correction. There are no firm building blocks, neither as indubitable and therefore valid axioms, nor in any sets of posits unequivocal once made. Nor are there any unequivocal experiences or explicit operations on which to found a certainty of communication in lieu of a certainty of knowing.” (p. 18).
is highly unlikely that this represented a repressed wish. Following Hall, Eysenck argues that the purpose of a dream is not to conceal, but rather to reveal wishes. Dreams may be seen as metaphors, expressing what can not easily be put in words in a much richer language. An important methodological point is that Hall arrived at his conclusions by extensive studies of series of dreams. In this way he could infer recurring patterns and cross-check interpretations from a single dream.

It should further be emphasized that false interpretations may have grave social consequences. To underscore this point, some religious incest offenders have quoted the Bible to support having intercourse with their daughter(s). Perhaps the (implicit) reasoning has been something like: ‘Everything which is said in the Bible without being condemned must be a viable norm for my conduct. Lot slept with his daughters and this was not condemned. Therefore it is legitimate for me to sleep with my daughter.” Such reasoning neglects several facts. In Genesis, ch. 19 it is clear that it was the daughters who initiated intercourse by first serving their father wine so that he would not be aware of their intention (and the motive was not “sexual” but to conceive a child). Furthermore this interpretation neglects the elementary hermeneutical principle of considering other parts of the text for any interpretation of a given statement. In Leviticus, ch. 18, for example, there is a clear injunction against sex with anyone of close kin.

Such examples cannot be simply dismissed as due to “a sick mind”. Following “fallibilism”, none of us is beyond committing error. To take another example from Freud to illustrate the grave social consequences of false interpretations, Tennov (1975) has discussed in great detail negative therapeutic consequences for women who have had the misfortune of having their miseries interpreted in the light of Freud’s misogynous speculations.

Postmodernism, in its downplaying of truth as an ideal to strive for, may also pave the way for grave social consequences. This is illustrated in a philosophical novel by Jon Helleenes: “The Postmodern Asylum” (1986). He leads us into a surrealistic, Orwellian world, where untold terror is wrought. “Power is knowledge” Kvale (1988) says. Without any clear conception of “knowledge”, just power — in its most abusive forms — may remain. Apel (1988) sides with Habermas and also issues strong warnings against the French “wild men”, specifically the postmodernists Lyotard, Derrida and Foucault. According to Apel, this philosophy represents

This is summarized in Popper’s slogan: “We don’t know, we can only guess.”

Campbell goes far in emphasizing the need for creativity and boldness in making conjectures, and there he honors Peyerabend’s (1975) iconoclasm, as this can help in shaking us out of methodological complacency. Campbell (1960, 1974) has also strongly argued that at some level guesses imply “blind variation”, or variation not justified by present circumstances. (A critique that he could have gone further than he does is given elsewhere, Tschudi, 1986a). In order not to be lost in speculation, however, variations must later be subjected to “selection” — inevitably many of them are bound to be off target. So logically, a plethora of “conjectures” or “variations” also imply errors and must be balanced against “refutations” or “selection”.

Furthermore, the ubiquitousness of error is well documented in the literature. Campbell starts with Francis Bacon’s analysis of “idols” or “false images which plague the mind” (p. 77) or what we now call “cognitive biases”, and discusses a variety of evidence supporting the prevalence of such biases. He discusses for instance “oversimplification and the neglect of disconfirming evidence” (p. 81) and the related tendency to “overinterpret, to extract too much from happenstance coincidences” (p. 86). Unfortunately it is not possible here to review the extensive literature on cognitive biases. Miles and Huberman (1984) emphasize the importance of this literature for validating qualitative research. For reviews of the literature on cognitive biases see e.g. Tschudi (1975), Kahнемann, Slovic and Tversky (1982).

Here I point to some — more or less obvious — examples of erroneous qualitative interpretations. Eysenck (1985, p. 1970 ff.) discusses in detail Freud’s reconstruction of Leonardo’s psychosexual history where an early memory where “a vulture came down to me, and opened my mouth with its tal and struck me many times with its tail against my lips” plays a major part. On this basis Freud builds an elaborate edifice, clearly an example of overinterpretation, and there is the further embarrassing fact that “vulture” is a mistranslation of “kite, a small hawk-like bird”. The moral of this story is (to paraphrase Eysenck): Beware of “inverted pyramid” in interpretations!

Eysenck also has a devastating critique of classical Freudian dream interpretations. A woman’s dream of a horse being mounted may well have sexual connotations. But in a classical case, that of an engaged woman, it
Fallibilism and methodological pluralism

How can we arrive at viable knowledge from a fallibilistic point of view? Popper has a metaphor (here quoted from Campbell, 1977, p. 96–97) which provides an excellent rationale for methodological pluralism.

Building a theory can be compared to "building on piles driven into a bottomless bog. The trip-hammer that drives down a new replacement post is itself substanied by other posts collectively trusted for now, even though each post is as separately revisable in its turn. There is no true foundation, no permanent fixed fulcrum for leverage, but instead, collectively these individual fallible supports provide the only possible useful foundation from which prying out, replacing, and driving down can be done".

Any single method may now be compared to a single pile. By itself, a given method cannot give us any solid foundation, or in technical parlance: any method is infested with error. This requires any method to be cross-checked against other methods. Perhaps the most well-known contribution elaborated from this point of departure is the notion of "convergent and discriminant validation", (Campbell and Fiske, 1959). Convergent validation implies that different methods for the same construct should give relatively high intercorrelations, while discriminant validation implies that similar methods for different constructs should give relatively low intercorrelation.

Many proposed measures have shown a lack of discriminant validation. A classic case in point is the proposed construct of "social intelligence". Earlier research showed that generally various tests of social intelligence (which were supposed to tap the same trait) would have higher intercorrelations with a standard intelligence test than with each other. Obviously the tests of "social intelligence" did not measure anything different from intelligence, and thus lacked discriminant validity. This reasoning is elaborated in the "multitrait–multimethod matrix". Even if the whole formal apparatus is not brought in, the approach may still be useful as a conceptual tool.

A careful reading of Cronbach and Meehl shows that both convergent and discriminant validation are implied in their treatment of validity (see above), but the distinction is far more explicit in Campbell and Fiske (1959), and their contribution is recognized as an important amendment to Cronbach and Meehl.

Campbell and Fiske see their approach to convergent validation as an example of "triangulation", quoting Ayer and Feigl as precursors. This concept has recently gained much currency, also in the burgeoning literature exclusively devoted to qualitative methods (Miles and Huberman, 1984, Fielding and Fielding, 1986). The most extensive treatment is probably found in articles in Brewer and Collins (1981). Webb et al. (1966) provide a major source for inspiration on how to use triangulation.

A basic idea in triangulation is that the bias of any given method is best overcome by triangulating with a maximally different method. (A priori it is quite unlikely that two very different methods should share the same bias. If I read highly similar accounts of unwarranted police brutality in a marxist and a conservative newspaper, I am pretty confident that the police did show unwarranted brutality.) Let me give an example from my own research: I tried to cross-check conclusions about male psychology from my starting point – graffiti written on toilet walls (which may be seen as low status literature) – with high status literature (Tschudi, 1988).

If we converge that generally quantitative and qualitative methods are quite different, triangulation is a strong argument for combining the two approaches. An example is a recent thesis which I supervised (Norden, 1988). Norden used Ericksonian hypnosis to alleviate menstrual pains. The basic measure was a ten-point rating scale where 10 represented "the worst pain you could imagine". But are scale points comparable for all women? This question is of special significance, since many women experienced difficulty in assessing their pains quantitatively. Fortunately they also gave a qualitative description of their pains. Validity of both methods would be justified if we could show an unequivocal relation between the quantitative
and the qualitative descriptions. This was checked by having an independent observer "translate" each qualitative description into a quantitative rating. These observer ratings were then correlated with the women's own quantitative ratings which gave a correlation well beyond .90. We can thus conclude that in this case the qualitative and quantitative judgements were isomorphic.

Notice that the maxim of maximally different methods leads to a critique of, say, using one conventional test to buttress the conclusions from another test. Even if the tests seemingly had quite different content, they would have large "methods factors" in common, say, dependence upon verbal instructions, and a formal setting detached from practical life. Unfortunately unwarranted generalizations from psychometric test procedures have occurred, e.g. children misplaced in institutions for the feebleminded.

This "internalist" criticism of quantitative methods can, however, be directed with equal force against procedures exclusively relying on qualitative methods. By just relying on qualitative observations (e.g. from one person, or two persons similarly trained), capacities to function in the more formal settings typically implied by psychometric tests may be lost from view. While it is highly appropriate to criticize overreliance on quantitative tests for decisions, it is also the case that misplacements based on qualitative observations have occurred later to be corrected by quantitative assessment.

Discussion of critiques of validity in empirical science

In this section three challenges to my thesis of the similarity of approach to validity in quantitative and qualitative science are discussed. Finally I raise a puzzling question: Why is "logical positivism", long since a dead philosophy, still used as a whipping boy?

Ricoeur on verification versus argumentation

Ricoeur (1981, pp. 212–215) has argued that social sciences are different from empirical sciences in that "they are closer to a logic of probability than to a logic of empirical verification... showing that a conclusion is true" (p. 212). Argument is thus primary in social science, and not proof.

This point of view is picked up by both Kvale (1987) and Salner (1989, pp. 63–65), and accepted as a valid distinction by them. Kvale (1987, p. 48) for instance says: "It is then illogical to apply a logic of certainty developed for the observation of unequivocal behavioural facts to the interpretations of ambiguous meanings of texts." In fairness to Kvale, it should be added that the "illogic" is restricted to extrapolations from predictive and concurrent validity. But these forms of validity have already been subsumed under a more pervasive quest for construct validity (Messick, 1975). And it should be obvious that construct validity is an argumentative endeavour, cf. one of our earlier quotes from Cronbach and Meehl (p. 112) on evidence and reasoning leading to belief, and thus not leading to certainty or involving "proof". Campbell's fallibilism furthermore underscores the provisional nature of all our knowledge, and finally I note that Ricoeur also brings in Popper (p. 213) on the importance of falsification.

There is, however, a second aspect to Ricoeur's analysis in that argumentation is likened to judicial procedures, and "H. L. A. Hart shows in a very convincing way that judicial reasoning does not at all consist in applying general laws to particular cases, but each time in construing referring decisions." (p. 215, emphasis added). At first blush, this seems to go against Cronbach and Meehl's appeal to "laws" in a nomological network which, even though the formulations may be "crude, half-explicit (1955, p. 294) presumably have some general applicability. Should we forfeit attempts at generality and give juridical reasoning a privileged status as Ricoeur does, and which Salner (1989, p. 64) seems to accept?

In a discussion with Torstein Eckhoff, a Norwegian lawyer with special interests in sociology and the philosophy of law, he pointed out that lawyers really do not have any general competence in evaluating competing interpretations. There is some concern with what a "law maker really meant", but in practice a lawyer will be equally, if not more, concerned with the normative aspects of a law, how it ought to be interpreted. When
arguing about specific normative interpretations, there may often be
implicit recourse to presumed general "laws" about conduct, but such pre-
sumptions may, on analysis, turn out to be rather naive. Perhaps closer
analysis would not make Hart that convincing after all? Instead of passing
the buck to lawyers, we may with equal force argue that lawyers should
have more training in social science! However, what is obviously called
for is a mutual dialogue between lawyers and social scientists on these
issues.

Qualitative and quantitative, a one-way street?

Sometimes qualitative researchers like to pick at quantitative research.
This may reflect the fact that each approach "is at its best in its criticism of
the other, not in the invulnerability of its own claim to descriptive knowl-
edge" (Campbell, 1978, p. 204). For example, Wertz (1984) makes a
major point of the fact that "qualitative meaning is presupposed in estab-
lishing the reliability of measurement" (p. 187), and concludes that "quan-
titative research has often been so narrowly preoccupied with measure-
ment that the qualitative understandings it presuposes are naive, unques-
tioned and lacking in any formal rigour, rendering the quantitative research
based on them at best severely limited and at worst totally invalid."

This premise is even more strongly put forth by Campbell (1978), who
also explicitly gives an epistemological rationale for this claim, extending it
to all science: "The end product of a qualitative judgmental process...
underlies all quantitative knowing in the social sciences just as in the
physical (emphasis added). An unfortunate difference, however, is that
"crossvalidation of the quantitative by the qualitative is usually missing in
the social sciences", whereas "among laboratory scientists themselves [this
refers to physicists] this common-sense crossvalidation is in continual use"
(p. 192). In the preceding section Norden's research was discussed as an
example of such crossvalidation.

Campbell gives a variety of examples of the importance of including qual-
itative observations in evaluation research and concludes that if results from
the two approaches do not agree "I feel we should regard it possible that
the quantitative was the one in error," (p. 200). Parallel to Wertz's conclu-
sion, Campbell further states: "The belief that external evaluators are more
objective further prevents the sharing of qualitative experience. Undoubt-
edly the computer printouts often provide a pseudoscientific facade which
gullibly uses data for which the qualitative basis of the quantitative is in
fact invalid" (p. 201).

Why then bother with quantitative data at all? As Wertz puts it: "qualita-
tive research need not be supported or informed by quantitative knowl-
edge" (p. 202). Campbell answers:

"... because I believe that the quantitative, when based on firm and exam-
ined quantitative knowing, can validly go beyond the qualitative, and can
produce subtleties that the qualitative would have missed. Such results,
however, should again, for instance in participant group discussions, be
subjected to crossexamination, and if the results then appeared valid to the
qualitative observers, "the quantitative would have gone beyond the qualifi-
tive in a valid manner" (p. 204).

So, in Campbell's view, quantitative methods should be "sandwiched in"
between a qualitative foundation and a qualitative final examination. I
hope this conciliatory view on the relation between the two approaches
may serve as an invitation to collaboration and promote the status of qual-
itative research.

Context and validity

Polkinghorne (1986) claims that "traditional notions of validity ... imply a
system of concepts that is stable, context-free... yet human existence
points towards a conceptual system that is... context dependent..." (p.
129), and the general implication seems to be that the traditional notions
are thus inapplicable to "human science".

More specifically, he argues that a concept, for instance "political action",
may "vary according to personal, cultural and historical settings", and
"thus... the human sciences will not be able to develop a common set of
conceptions with which to study the human realm". The premise may cer-
tainly be granted, but the conclusion is a non sequitur. Polkinghorne does
not consider the possibility that the "set of conceptions" may fruitfully be
at a level of abstraction different from the concepts used by the persons
which are studied. A prime example in psychology is George Kelly's
(1955) "The psychology of personal constructs", which, precisely because
his professional constructs are very abstract, has found wide applicability,
not only in clinical psychology but also in cross-cultural studies and organ-
izational studies (see Brannister and Fransella, 1986 for examples). A related example is Tetlock’s (1985) studies of “cognitive complexity”, which have thrown light on a variety of important phenomena in politics. By focussing on this variable he is for instance able to show that American and Soviet foreign policy are interdependent in a way which defies any simplistic notions of e.g. American policy just being a response to “Soviet aggression.”

If, however, we take a closer look at Cronbach and Meehl’s example of construct validation, we see that Polkinghorne’s point of view implies a forceful criticism. The observation is a correlation, which is an average (an aggregate measure), and this is a highly “decontextualized” observation. In an important contribution to the literature on construct validity, however, Embretson (1982) introduces a distinction between “nomothetic span” and “construct representation”. The former refers to the network of relationships of a test with other measures. This is just a restatement of the old “nomological net”, but the important addition “construct representation” is “concerned with identifying the theoretical mechanisms that underlie task performance.” (p. 180). Here mathematical models are constructed for studying the components of performance when a person encounters an item. This requires task decomposition methods.

Embreton critically compares several mathematical procedures. It is interesting to note that a multidimensional extension of Rasch’s models by Embretson – which rest on more thorough conceptual and philosophical grounding than most psychometric work – was found most fruitful. Husserl’s “method of free variation”, frequently quoted in the qualitative literature, could be an important help in “task decomposition”. This seems to be an area where – in line with Campbell’s arguments outlined above – the quantitative and qualitative camps should join forces in order to get a valid picture of cognitive functioning.

Logical positivism: why whip a dead horse?

All the authors in the qualitative tradition quoted here raise more or less explicit battlecries against “positivism”; the most prevalent contention seems to be “unequivocal, quantitative facts” (Kvale, 1987, p. 47) or a belief in “real world facts”, “an epistemological assumption that is positivist in nature and antithetical to the assumptions of human science research”. (Salner, 1989, p. 47). “Positivism” has also – more or less sweepingely – been associated with traditional views of validity, and these views have then been found wanting, or downright irrelevant.

This ignores the fact that the chief architects of validity have never believed in positivism. The most prominent positivist was probably Carnap, but his pupil Quine definitely buried positivism in his famous essay “Two dogmas of empiricism” (1951/1953). I quote from the conclusion of this essay:

“The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges... A conflict with experience at the periphery occasions readjustments in the interior of the field... But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what to reevaluate in the light of any single contrary experience... but our natural tendency is to disturb the total system as little as possible... Even a statement very close to the periphery [for instance that there are brick house on Elm Street] can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws.” (pp. 42–43).

Clearly this is closer to a coherence view of validity than the more simple-minded correspondence view, though there may sometimes be pragmatic advantages to acting as if a correspondence view were appropriate. (See von Wright (1943) for a good discussion of the relation between these views.) In Quine’s view, no “fact”, not even the existence of physical objects can get any privileged epistemological status. “For my part I do, qua lay physicists, believe in physical objects and not in Homer’s gods... But in point of epistemological footing the physical objects and the gods differ only in degree and not in kind.” (p. 44).

As discussed by Campbell (1977), Quine – whom he quotes as the chief philosopher who has influenced him – strongly argued that logic and philosophy could not solve ontological and epistemological questions, but handed them over to empirical science. As Quine puts it: "Ontological questions... are on a par with questions of natural science." (p. 45).

Meehl (1986) is equally convinced that logical positivism has long since ceased to be viable: "The last remaining defender of anything like logical positivism was Gustav Bergmann, who ceased to do so by the late 1940s.

124

125
Some implications of emphasizing similarities of approach

I agree with Salner (1989, p. 66) that "human science research is not a set of procedures which can be learned", but "an applied epistemological position". This position should, however, be thoroughly informed by all the threats to validity (I have mentioned only a few) which Campbell and his coworkers have derived from the fallibilist position (see Cook and Campbell, 1979 for the most extensive presentation). Many of these threats can not be directly applied to qualitative research. They may, however, serve important heuristic functions in generating comparable lists uniquely relevant to a specific qualitative endeavour. A case in point is Becker's (1979) explicit recognition of Campbell's influence in outlining threats to be considered when asking "Do photographs tell the truth?"

Furthermore, there may be a more immediate heuristic value in using the well-known statistical concept "degrees of freedom" in an analogical sense in qualitative research, see Campbell (1979). The basic statistical idea is that we must have more observations than parameters in order to use the machinery of statistics. Analogically, a given theoretical interpretation ("parameter") should be checked by multiple observations. Keeping systematic track of the degrees of freedom, and giving special weight to confirming observations which appear highly unlikely from a common sense view (or an alternative theoretical interpretation), may be a useful strategy in qualitative research.

Individual and social considerations

Love and power

Salner (1989, pp. 65-69) draws attention to "qualities and abilities the human researcher needs to have in order to conduct studies which are likely to be regarded as contributions to our knowledge and understanding
of human life" (p. 65). I strongly endorse all her points, and I was especially pleased to see that knowledge of experiments and statistics was included, though I would have preferred a more explicit mention of familiarity with the traditional literature on validity as a prerequisite.

A more radical extension of her "qualities and abilities" is, however, called for. It is also necessary to consider how emotional qualities (as these are revealed in how the researcher approaches what is studied) flavour the inquiry and the validity of the knowledge arrived at. This problem was strongly brought to mind while I struggled to understand males in my previously mentioned graffiti research. I was struck with penetrating insights from both male and female writers on sexual arrangements in the standard literature on this topic, but many of them seemed bred of hate, rage or anxiety, and marred by onesidedness.

This raises the question of the importance of love for understanding and knowledge. "... love sees more than hatred. Hatred can be astute in perceiving every possibility for attack, but even though such astuteness may lead to penetrating insights... they always concern only part of the other and they are 'partial' to those parts which will serve as the points of attack. But hatred is unable to see its object in its totality" (Schachtel; quoted from Keller, 1985, p. 119).

In the essay "Dynamic objectivity, love, power, and knowledge", Keller starts by noting that Genesis may provide a deeper insight into knowledge than the traditional equation of knowledge and power. "Knowledge" is there used "simultaneously in the sexual and epistemological sense" (p. 115). Disregarding love "makes the equation between knowledge and power a sinister one, and at the same time allows objectivity to become contaminated with domination" (p. 116). Keller persuasively argues that the conventional opposition between love and power must be overcome. To express this in a formula:

Knowledge = love + power

The attitude of love and power is captured in such formulations as: "extend our body to include the object – so that we come to dwell in it", the subsequent reward being that "one feels silent and grateful because one was allowed to penetrate a layer of understanding which remained impene-trable to others" (p. 123).

It should be emphasized that the basis for Keller's insight was not the soft human sciences. She was trained as a mathematical biophysicist, her con-

victions being shaped by her work on "cellular slime molds" (p. 150-157), and a close study of Barbara McClintock's Nobel Prize work on DNA (p. 158-176). The main thrust of her argument is a plea for a loving approach in the physical sciences, and to the extent that we accept that, it should apply even more forcefully to the human sciences!

Sociological factors

Any consideration of validity will be incomplete without taking sociological factors into account. Campbell (1977, 1986) readily grants that a variety of other factors than the "natural world" enters the construction of scientific knowledge. He construes such factors as detrimental to validity (see below for examples) and raises the problem of: "under what social system of belief communication and target characteristics the role of the natural world would be maximized" (1986, p. 115). Some "precious norms", or beliefs which may be "validity enhancing factors" are:

The truth is to be found in the future, and not in ancient traditions

Conversely, "gerarchial authority", leaders who use their power to protect their own position from internal challenge from young rivals, and also "hostility and disgust towards outgroups" may all be detrimental factors.

Economic security for the researcher may be crucial, especially in evaluation research. If, for instance, continued support to the researcher depends upon positive results from a pilot project, this may be a serious threat to validity! (Campbell, 1984).

The study of validity-enhancing/detrimental factors is still in its infancy. There are many unresolved questions in this field, and comparative sociological studies of "successful" versus "unsuccessful" research communities are obviously called for (see e.g. Tschudi, 1966b). It is, however, hoped that a sensitivity to such factors may have beneficial consequences for research policy.

If a research community does not embody validity-enhancing factors, all teachings of methodology – from construct validity to Hussert's phenomenology – may be but empty rituals.
Concluding comments

My answer to whether qualitative and quantitative methods require different approaches to validity is a clear “no”. Arguments to the contrary have been critically examined and found wanting, mainly because they rest on a glib, superficial equation of quantitative methods (psychometrics) with positivism. This view ignores that for the last forty years “positivism” — with its simplistic views on “facts”, and its belief in unequivocal “operational definitions” — has not been a viable philosophical doctrine.

There are, however, many departments with simpleminded, “imperialistic” views on quantitative methods — leading qualitative researchers to recede into their own “human science” camps (as if predominantly quantitative researchers are “anti-human”?). I have put primary emphasis on Campbell’s views since he has been one of the chief architects of a perspective on validity which is generally espoused by quantitative researchers. A careful reading of his papers does, however, reveal a quite humble role for quantitative research in the total research enterprise. Hopefully this may pave the way for mutually beneficial cooperation instead of fruitless quarrels.

While I am quite enthusiastic about triangulation and multiple operations as examples of guidelines for such cooperation, exclusive emphasis on such concepts may lead us to miss deeper aspects of the research process.

Further inquiry into validity should pay primary attention to the relation between the researcher and the problems being studied. I have suggested the importance of “love”, or indwelling, which may be contrasted with, if not “hate”, then at least an attitude of distance, and lack of respect for the data. This distinction cuts across the dichotomy of quantitative vs. qualitative.

As a prominent example of an indwelling approach within the quantitative field I would mention the work of the Danish statistician Georg Rasch with whom I was once happy to spend some time (see Tschudi 1964). On the other hand a deluge of applications of standard statistical packages are carried out in an alienated way and quite often result in little, if any, contribution.

As an example of lack of indwelling within the qualitative field I have singled out Freud’s study of Leonardo da Vinci, where Freud clearly got lost in his own associations, at the expense of giving us further understanding of Leonardo.

Following Kelly (1955), whether we dub ourselves “quantitative”, “qualitative” or whatever, we are all personal scientists. A closer study of our relation to the problems we are faced with, and also to the social context in which the enquiries take place, is necessary to prevent us from getting lost in technicalities and quarrels.

References


