QUANTITATIVE AND QUALITATIVE:

A QUESTION OF COMPARABILITY

To appear in
International Encyclopedia of Political Science,
Bertrand Badie, Dirk Berg-Schlosser, Leonardo Morlino (eds),
Sage, 2011

John Gerring
Professor
Department of Political Science
Boston University
232 Bay State Road
Boston MA 02215
jgerring@bu.edu

Craig W. Thomas
Associate Professor
Daniel J. Evans School of Public Affairs
Box 353055
University of Washington
Seattle, WA 98195-3055
206-221-3669
thomasc@washington.edu
Perhaps no division in the social sciences is so persistent, so nettlesome, and so poorly understood as the division between quantitative and qualitative ways of knowing. The cleavage can be traced back to the first applications of statistics within the disciplines of economics, political science, and sociology, and became increasingly acute in the late-twentieth century as quantitative approaches gained in stature, grew in complexity, and pushed qualitative empirical analysis out of the limelight.

During this period, the division between qualitative and quantitative methods became associated – perhaps inappropriately – with the rival epistemological positions of positivism and interpretivism. Charles Smith (1989: 29) summarizes the now-familiar standoff:

On the one hand, there are those who argue that only through the application of quantitative measurements and methods can the social sciences ever hope to become ‘real’ sciences; on the other hand, there are those who claim that the subject matter of the social sciences is simply not amenable to quantification and all attempts to impose such measures and methods upon social behavior is just so much nonsense.

While there have been many attempts to shed light on this persistent division in the social sciences, work on this question is generated primarily by writers who occupy one of the two camps. These writers tend to be either strong partisans, or visceral opponents, of the “quantitative worldview.” A chronic dualism besets many of these debates, in large part because the distinction between quantitative and qualitative forms of descriptive and causal inference have been folded into the much larger and more “philosophical” debate between positivism and interpretivism.

Positivists (aka naturalists), usually identified with quantitative methods, present their perspective as hegemonic: there is, or ought to be, only one logic of inference. The conclusion of these scholars is either that there are no important distinctions between qualitative and quantitative work or, to the extent that such distinctions exist, they are to the detriment of qualitative scholarship. “When possible, quantify,” is the motto of this camp.

Defenders of qualitative work typically emphasize the limits of quantification and the insights that can be gained through an interpretive approach to social action. Rather than a unified logic, interpretivists suggest that there might be multiple logics at work in the social sciences. These multiple logics stem from epistemological or ontological commitments, which may themselves be culturally prescribed, political, or historical in origin.

A third position runs orthogonal to this debate. According to this view, the qualitative/quantitative division is inherently ambiguous, and hence a distraction from the real work of social science. In this spirit, a recent lexicon focused on qualitative inquiry notes, in the entry for ‘Qualitative,’ that “the adjective does not clearly signal a particular meaning” (Schwandt 1997: 129). Rather, it is used as a “blanket designation for all forms of [hermeneutic] inquiry including ethnography, case study research, naturalistic inquiry, ethnomethodology, life history methodology, narrative inquiry, and the like.” In sum, concludes the author of this dictionary of qualitative methods (with no apparent sense of irony), “qualitative research’ is simply not a very useful term for denoting a specific set of characteristics of inquiry” (Ibid. 130). The inherent fuzziness of the concepts qualitative and
quantitative thus prompts some scholars to argue that the so-called debate must be a red herring, i.e., a distraction from the real issues of social science methodology.

Thus, one finds three well-articulated positions: (a) naturalism (one logic of inquiry, based implicitly on the quantitative template), (b) pluralism (multiple logics of inquiry, with qualitative and quantitative work founded on different epistemologies), and (c) agnosticism (aka, let’s call the whole thing off).

We are sympathetic to the naturalist goal of a single, unified logic of inquiry; only in this fashion can the scientific objective of knowledge cumulation be achieved. However, we are dubious that the quantitative template provides an adequate framework for work across the social sciences. While the agnostic position is attractive, it seems to be mostly rhetorical. Scholars continue to divide themselves into camps – e.g., departments, journals, editorial boards – and these camps are often classifiable as predominantly qualitative or quantitative. These two concepts refuse to be banished, suggesting that there must be a there, there, somewhere.

As a way of cutting into this subject in a new fashion, we propose a new way of thinking about these issues: the principal factor separating qualitative observations from quantitative observations is the assumed comparability of the evidence at hand. Quantitative observations presume a population of things that can be readily measured, counted, and hence compared. Qualitative observations, by contrast, presume an empirical field where individual pieces of evidence are not directly comparable to one another, at least not to the extent that they can be precisely measured. In this sense, quantitative work is appropriately labeled nomothetic, and qualitative work idiographic. The key point is that the difference between these two kinds of observations rests on the presumed comparability of adjacent observations, not (at least not directly) on the size of the sample, the style of presentation (numbers or narrative), epistemology, ontology, or the formal structure of the method.

Our purpose in this entry, it should be stressed, is not to reinforce existing cleavages in the social sciences. Rather, we wish to draw attention to what we see as the most important underlying issue in these debates, an issue that has not received much recognition. We do not pretend that this is the only issue underlying the qualitative/quantitative divide, but we believe it to be the most fundamental. If analysis is based on comparison, the central methodological question is what we can reasonably compare, and how precise those comparisons can be. Here, there are legitimate differences of opinion, and they are not the sort that can be empirically proven.

**Comparability**

We understand comparable observations as members of the same population and therefore potential members of the same sample. They are examples of a similar phenomenon. They are apples and apples, rather than apples and oranges, to use the time-honored metaphor. Note that comparing apples and oranges is not prohibited; however, to do so we must adopt a higher-order concept – e.g., fruit – according to which apples and oranges are similar. This common-sense meaning of comparability is widely understood and agreed upon. But what does it mean for items to be comparable within the context of social science research? Surely, it is more than shared membership in an arbitrary linguistic category.

First, comparable observations must share a set of relevant descriptive attributes (dimensions). This is what makes them comparable. The observations need not demonstrate the same values for those attributes. Each observation in a sample may “score” differently on each attribute in either quantitative or qualitative terms – e.g., high/low, present/absent,
and so forth. But each observation must be *scale-able*, and the attribute must mean (roughly or precisely) the same thing across the contexts in which it is being compared. We label this *descriptive comparability*, seeing it is a fundamental feature of conceptual validity. The defining attributes of a concept must be valid across the designated observations. Otherwise, a concept is being “stretched.”

A second kind of comparability refers to the inter-relationship of two factors in a causal analysis, the cause (or vector of causal factors), X, and the outcome, Y. The specified X/Y relationship must hold across the chosen observations. We label this *causal comparability* (a looser version of *unit homogeneity*).

Thus, there are two kinds of comparability: descriptive and causal. The first is presumed in the second. If a sample of observations is assumed to be causally comparable, then it must also be descriptively comparable. In statistical research, the assumption of unit homogeneity makes this explicit, but it must also be true more generally, for causal comparability can exist only in the presence of descriptive comparability.

**Precision and Explicitness**

The underlying issues in the enduring qual/quant debate are tethered to the problem of comparability, as defined above. To “quantify” an observation is to formulate it in terms that can be *explicitly* and *precisely* compared across a large number of observations, i.e., where a concept can be expressed on a numerical scale, a metric, a variable (we use these three words interchangeably). Quantitative observations combine natural-language words (nouns, verbs, or adjectives) with numbers according to some pre-assigned metric. It is a question of measurement, which can be regarded as the assignment of numbers to phenomena in accordance with some uniformly applied rule.

Let us begin with a discussion of the concept of “precision.” Note that to simply re-code a dichotomous natural-language category as a series of binary numbers does not make it any more precise. Thus, 0/1 is no more precise than “pregnant/not pregnant.” However, numerical scales offer the possibility of greater precision when the number of categories surpasses the categories inherent in natural language, as well as in circumstances where these categories can be understood as positions on a continuous (interval) scale. To say that one room is “warmer” than another is comparative, but it is less precise than saying that one room is 70 degrees Fahrenheit and the other is 60 degrees Fahrenheit. Thus, in many situations the use of a quantitative idiom allows for more precise comparisons across units. In all situations, the use of a scale is *at least as* precise as natural language (in the sense that no precision is lost in the translation of words to numbers). Again, to use a quantitative idiom does not entail great precision; it entails the *possibility* of great precision (as well as a more explicit set of comparisons).

Precision should not be confused with *certainty*. Qualitative or quantitative statements may be uttered with more or less confidence. For example, one might say “I would guess that the room is 70 degrees” or “I would guess that the room is warm.” With quantitative statements, a mathematical indicator of uncertainty may accompany the point score. When we say that quantitative observations are more precise we are referring to the point estimate, not the degree of uncertainty (or dispersion).

As we said, quantitative statements are both more precise (at least potentially) and more *explicit*. This is because the very act of creating a numerical scale requires a set of explicit comparisons and an explicit comparison-set – a domain. Scales cannot be developed in highly specific contexts. Imagine developing a barometer of a single event, say, the
terrorist attacks that occurred on September 11, 2001. The idea is nonsensical because scales make sense only relative to classes of events. One can have a terrorism meter, but not a 9/11 meter (unless of course that event is being used as a metric for understanding other events, in which case it becomes a comparative metric). Weather can be measured precisely because, for one thing, there is lots of weather to measure, and temperature is thought to have the same general meaning in many different contexts. Granted, all scales are bounded; there is no universal scale (a scale applying to everything). Some things, like metaphysics, have no temperature; the concept of temperature (and whatever scale might be used to measure it) does not apply in this domain. The point is that relative to words, which are only loosely and implicitly comparative, scales are precisely and explicitly comparative, and their range is usually quite broad (otherwise, why bother to develop a systematic scale?).

Now, it is true that some natural-language adjectives, such as “warm,” are explicitly comparative. But most words are more ambiguous. This is apparent in the extra locutions that are necessary to render ordinary language comparative. One must clarify “warmer than,” “more chair-like than,” and so forth; whereas, to append such judgments to a numerical scale is redundant. (One does not say, “70 degrees F, warmer than 65 degrees F.”) Numerical scales are already comparative, and no matter what one does with a numerical observation it cannot lose its precise, explicitly comparative quality.

To be sure, if one labels an object with a noun – e.g., “chair” – one is implicitly (if not explicitly) comparing it with other objects: non-chairs. Language has this universal aspect; if we call something X, we imply that other things are not-X, or less-X or more-X. However, the comparisons are vague. It is unclear, for example, where a chair leaves off and a stool begins, for few words – and very few key words in social science – have crisp boundaries. More importantly, most words are multivalent; they have more than one attribute and consequently can mean more than one thing. Thus, to say that an object is “not-A” could mean a number of different things, depending on the attribute(s) that are intended by an author, or understood by the reader. Moreover, a word usually gains meaning by its context, and this context is undefined in settings other than that which the author is studying. Additionally, the other objects that are not-X are typically not defined, in which case the larger population of cases (the domain of the inference) remains implicit. Finally, words are contingent upon a particular natural language, and this imposes another sort of contextual boundary against comparison. (By contrast, the number “5” and the operator “=” mean the same thing everywhere – since the adoption of a uniform language of mathematics – and they also mean the same thing in all contexts that they might be employed.)

Frequently, natural-language comparisons are without any obvious comparative reference-point. The statement “Caesar crossed the Rubicon” is comparative in so many possible ways that it might be considered non-comparative: he did not cross it; he did not cross the Tiber; it was not Brutus who crossed the Rubicon; and so forth. If this is “comparative,” it is so in the most minimal sense. Yet, it is important for our purposes to recognize that comparison is a matter of degree. Qualitative observations can be more or less comparative, but quantitative observations are almost always more precisely and explicitly comparative.

One final clarification is in order. We have said that all quantitative statements about the world invoke a class of events; these form the basis for the metric. However, it does not follow that quantitative statements about the world are necessarily broader in scope than qualitative statements. Indeed, the very fuzziness of natural language issues a license to generalize – for one can avoid saying anything terribly specific – while the exactness of quantification may rein in the temptation to generalize. It follows that qualitative statements
can be either very restricted in scope (as in our previous example about the singular event of 9/11) or extremely broad. Saying something in words does not affect the scope of the inference. Saying something with a metric, however, presupposes a class of referents, which is to say it must make reference to more than one discrete event (and these reference-points must be fairly precise and explicit).

**Tradeoffs: Thick and Thin Description**

In principle, any qualitative observation can be converted into quantitative form, as attested by the plethora of methods and software designed to perform this function. e.g., NUD*IST software, Computer Assisted Qualitative Data Analysis Software (CAQDAS), various narrative-based methods, as well as more generic forms of content analysis. There is no such thing as a non-quantifiable observation because any single statement that can be made about one phenomenon could also be made about another phenomenon, thus providing the possibility of some sort of scale.

Yet, it is not clear that one would always want to make this transposition from words to words-with-numbers (variables). Indeed, there are usually costs associated with this conversion. The tradeoff may be understood in terms of precision and explicitness on the one hand, and depth (or richness) on the other. More concisely, the analyst has the option of describing thinly or describing thickly. Begin with the fact that words are usually multivalent; they generally carry a variety of attributes, some of which may not even be logically consistent. This is particularly true of key words in social science – e.g., democracy, justice, corporatism, political party, and so forth. When one of these words is converted into a measurable variable, which is to say into to a precise scale, a researcher is generally forced to drop one or more of its attributes. For not all of these attributes will be precisely applicable to the class of phenomena that the concept is now (quite explicitly) intended to cover. Recall that we are not necessarily describing an expansion in scope, for natural language can reach as far as mathematical variables. But in making the comparison precise and explicit, it is usually necessary to narrow the definition of the natural-language concept. To be sure, it could be that the intension of the natural-language concept is also quite a bit narrower than the full set of attributes normally (in ordinary language) associated with the term. An author is free to define a term as she sees fit; qualitative work is not wedded to ordinary language. The point is, in creating a variable one is forced to make explicit choices about which definitional attributes properly apply to a class of phenomena, and which do not. This is likely to prompt some narrowing of the semantic options. And this is why we consider the choice to quantify a concept as a move toward thin description. More explicit comparisons are being made, but they are narrowed down to one or several dimensions (the chosen attributes of the core concept).

Similarly, if one chooses to particularize, rather than to generalize, natural language is the obvious vehicle of choice. As we have pointed out, it is inappropriate to construct a scale when the class of instances under investigation is one or several. A scale presupposes a population. By contrast, a word can be used in a highly specific context; it does not presuppose an explicit comparison with other instances. This means that in describing the singularity of an event one is drawn towards the implements of natural language. The lack of perfect commensurability between words used in one context and the same words used in another context allows the researcher the facility to elucidate what is different – categorically (qualitatively), not marginally (quantitatively) – about that phenomenon. A very high score
on some scale can be (indeed, must be) indicated with a quantitative metric. But a very
different kind of score requires a word, perhaps a series of words.

In short, there are gains and losses in the transposition of words to numbers, and
vice-versa. What is interesting about this classic debate is that both may be described as
“reductionist.” Quantitative studies are often accused of reducing reality, and in the process
distorting that reality to fit the austere requirements of the quantitative format. Each piece of
reality must be sliced up into variables and these must be comparable across all observations.
Qualitative studies are accused of a different kind of reduction, in which a subject is shrunk
down to a highly particular context – the country, neighborhood, or event of special interest.

These contrasting notions of reduction/expansion are best understood as arguments
over comparability. Scholars inclined toward the tools provided by natural language are often
keen to explore a wide variety of different aspects in a particular setting. They wish to
explore multiple dimensions of one thing. (“Dimension” is employed as a synonym for
“variable” in this context.) Qualitative analysis is thus often focused inward, like a vast
funnel. Many comparisons are made, but they are all understood as features of the same
general topic, existing in one time and place. Natural language is adept for this purpose for it
is rich, textured, context-specific, and multivalent. It elucidates a wealth of details about a
person, event, or situation. This is why we find a natural affinity between qualitative tools
and ethnographic, historical, and – more broadly – interpretivist styles of research. By
contrast, scholars inclined toward a numerical understanding of the world are drawn toward
comparisons that are broad and thin. They intend to explore one particular dimension of
many things.

The interesting aspect of this familiar contrast is that both qualitative and
quantitative scholars perceive their work as conforming to the natural bend of the universe.
Qualitative scholars usually assume a case-centered approach. Different aspects of the same
cases are grouped; they go together. Quantitative scholars are drawn toward a dimensional
approach to comparability. A single aspect (dimension) of an entity is assumed to be
comparable across multiple cases. This is sometimes referred to as “variable-centered”
analysis. While for a qualitative scholar it would seem natural to explore everything about A,
for a quantitative scholar it would seem more natural to explore one thing about A, B, and
C. Underlying scholars’ choice of method are certain assumptions about cross-case
comparability. The tools we choose – words or numbers – are, in part, the expression of our
relative confidence in the ability to compare across entities in a given research context.

This is not to deny that many scholars use both words and numbers. The point is
that within a given context the likelihood of choosing one or the other strategy is influenced
by assumptions about case-comparability.

**CONSTRUING MEANING VERSUS ANALYZING BEHAVIOR**

The importance of comparison is illustrated in one of the most common defenses of
qualitative work in the interpretive mode. It is often said that qualitative analysis focuses on
human meanings while quantitative analysis focuses on the behavioral components of
human reality – i.e., actions, institutions, or events. From the interpretivist perspective, the
business of social science is one of construing meaning, not analyzing behavior. Yet, one
might reasonably inquire, why not study human meanings quantitatively – i.e., with scalar
measures – in addition to studying them qualitatively?

One rationale is that human meaning is constructed through language (which
establishes the categories by which we understand the world); therefore, it makes sense to
study these linguistic categories through other linguistic categories, rather than the (somewhat alien) categories of numbers. Herbert Blumer (1969: 132) concludes that the crucial limit to the successful application of variable analysis to human group life is set by the process of interpretation or definition that goes on in human groups. This process, which I believe to be the core of human action, gives a character to human group life that seems to be at variance with the logical premises of variable analysis.

The problem, though, is not that the analysis of linguistic phenomena must be carried out with linguistic tools; after all, the contemporary discipline of linguistics is heavily quantitative. Indeed, quantitative observations are necessarily rooted in language because all scales must be expressible in a linguistic category; every variable must have a name. What is it, then, that seems so problematic about collecting quantitative observations about semantic realities? Why is there no quantitative hermeneutics?

The problem, we suggest, is the core problem of the qual/quant dispute: the problem of representing human meaning across diverse contexts in an explicit and precise fashion. It is a problem of comparability, not of language per se, that is at issue. And what makes it so problematic is a basic feature of human life. The ways in which we make sense of our lived experience is extraordinarily diverse – through time, across cultures, and across individuals. It is difficult to reduce this complexity to standard categories, as quantitative knowledge requires.

Consider the question of human happiness (variously understood as welfare or quality of life), which has attracted increasing attention on the part of psychologists, sociologists, and economists. The following scale was developed by the United States Bureau of the Census to measure the quality of life: (a) terrible, (b) unhappy, (c) mostly dissatisfied, (d) mixed, (e) mostly satisfied, (f) mostly pleased, (g) delighted. The question we might ask ourselves is whether human happiness is accurately captured in these categories. This is to say, does a person who answers (b) unhappy have a lower quality of life than a person who responds with (d) mixed? There are potential problems of conceptual validity, and the issue is not simply a lack of sophistication on the part of researchers. More fundamentally, the problem is that human meanings – such as happiness – are resistant to uniform comparisons across diverse subjects. One feels much more comfortable with imprecise comparisons expressed in the looseness, and contextual specificity, of natural language. We might say, for example, that “Smith is happy,” implying that across some undefined population Smith’s level of happiness is, let us say, somewhere above the mean. However, we are probably reluctant to assign a precise score to Smith’s happiness because such a score would presume a precise comparison with all others in the population.

We do not wish to imply that all human meanings are un-quantifiable. Nor do we wish to imply that social scientists should abstain from quantifying difficult and ambiguous meanings such as human happiness. There may be much to learn from the quantification of abstract concepts that summarize a great deal of information about human experience. Our point is simply that the profitability of quantifying varies with the topic, and the core issue is the comparability of the phenomena across the chosen population. Certain topics are more recalcitrant, and these tend to be tied up with the generation of meaning (values, ideas, intentions), rather than observable behavior. It is noteworthy, for example, that some human intentions seem to be more quantifiable than others, and these tend to be those where we can assume a high level of comparability across individuals or across cultures. Thus, scholars routinely measure the concept of “economic voting” because the notion pursuing material interests seems comparable, and hence valid, across individuals and populations. By contrast,
the measurement of religious influences on voting behavior is much more difficult to quantify because it is more difficult to compare.

CONCLUSION

Approaches in the social sciences are often understood as qualitative or quantitative. Scholars differ in their opinions about the utility of this distinction. Some dismiss it as a red herring; others feel that there is good justification for the division. We feel that there is no plausible way of discarding the distinction but that it is greatly misunderstood. The key to this misunderstanding, in our view, is to be found in assumptions about comparability.

There is indeed a difference in basic-level assumptions between statements that are quantitative (i.e., understood through a numerical scale, a metric, or a variable) and those that are qualitative (expressed in natural language). Quantitative descriptive statements presuppose a class of comparable cases that can be compared in an explicit and precise manner. To measure phenomenon, X, is to impose a very specific metric on it, one that is explicitly comparative (since other phenomena in this same category are assumed to be score-able). Qualitative descriptive statements do not make any such presuppositions. There may or may not be an identifiable class of comparable cases that can be measured along some set of dimensions. Often, the assumption is quite the reverse – particularizing rather than generalizing. Thus, we argued that to quantify something is to compare in an explicit and precise manner. To qualify is to leave such comparisons open; one may or may not engage explicit comparisons with adjacent cases, and any comparisons made are unlikely to be very precise. While this might seem to indicate a distinct advantage for quantitative work, we also showed that there are costs to assuming a quantitative idiom. Not only must the cases be (actually) comparable, but there is usually some loss of information since words are usually multivalent and metrics are usually unidimensional (or at most combine several dimensions). It is not clear that we always gain in analytic leverage by moving from words to numbers. We do, however, make different sorts of comparisons.

The choice between math and natural language as tools of social science is, therefore, a highly consequential choice. Methodological tools help us to reconstruct the empirical world; they are not theory-neutral. In this respect, the division between math and language is akin to the influence that early anthropologists and linguists assigned to language. Different languages divide up the world into different packages; they encourage us to visualize things in different ways. So, arguably, do the different “languages” of mathematics and ordinary speech. Quantitative tools help us to compare, and hence to generalize; qualitative tools encourage us to differentiate.

It is quite another thing, however, to disentangle the causal priority of methodologies and ontologies. Do cultural anthropologists use qualitative tools because they envision a lumpy universe, or do they see a lumpy universe because they perceive it with qualitative tools? About all one can say with any degree of confidence is that there is a strong synergy between methods and ontologies. We presume that they are, at the very least, strongly reinforcing. This may help to account for the virulence and endurance of this central cleavage in the social sciences today.

In pointing to the problem of comparability as the key feature of this debate, we hope to contribute to a demystification of this longstanding distinction, perhaps even to a reconciliation of the two camps. It is not merely a matter of numbers versus words, or a debate about what can or cannot be quantified. More fundamentally, the venerable debate represents fundamental disagreements over how precise, explicit, and extensive social-
science comparisons ought to be. Those who resist numerical analysis, are dubious about the validity of comparisons. They see no need to enhance the precision or explicitness of comparisons because they do not seek to compare in the first place, or they seek a more restricted ambit of comparative reference-points. Those who embrace quantification are more comfortable with such comparisons. This debate has been engaged at many levels – across individuals, across levels of government, across cultures, across time-periods – and over many years. It shows no sign of dissipating.

Note: A version of this article with extensive citations to the literature is posted on Gerring’s homepage.