

VU Research Portal

Using Q methology in administrative ethics

de Graaf, G.; van Exel, J..

published in
Public Integrity
2009

document version
Early version, also known as pre-print

[Link to publication in VU Research Portal](#)

citation for published version (APA)

de Graaf, G., & van Exel, J. (2009). Using Q methology in administrative ethics. *Public Integrity*, 11(1), 63-78.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl



Using Q Methodology in Administrative Ethics

GJALT DE GRAAF AND JOB VAN EXEL

Abstract

Q methodology is seldom used by academics and practitioners in the field of administrative ethics, but it has important potential for empirical studies. Q offers a procedure and conceptual framework with which to study subjectivity in the social context. It has the advantage of bringing marginalized viewpoints to the fore but also has some drawbacks. The appendix provides a basic introduction to Q and shows how it can be used in research.

Administrative ethics uses a wide range of methodologies (Menzel 2005); many purely descriptive studies are either quantitative and use surveys, or qualitative and use case-studies. Q methodology, an approach for the study of subjectivity, is rarely used in academic research or organizational and policy practice related to administrative ethics.

Although not yet accepted in the field of administrative ethics, Q methodology is widely employed in other administrative sciences. As one example of its use, van Eeten's (2001) study of the Schiphol Airport expansion controversy in the Netherlands won best article of the year in the *Journal of Policy Analysis and Management*. He writes: "The 1999 mini-symposium in the *Journal of Policy Analysis and Management* shows that Q-methodology is finally being considered a candidate for the toolkit of mainstream policy analysis. The discussion seems to shift from fundamentalist debates over epistemology toward the question of the added value to the analysts' toolkit" (2001, 396). Q methodology has also been applied more widely than policy analysis in public administration. Prominent examples of researchers who have adopted it include Selden, Brewer, and Brudney (1999), who identified five conceptions of administrative roles and responsibilities; Gaines, van Tubergen, and Paiva (1984), who studied the perceptions of police officers about promotion as a source of motivation; and Cunningham and Olshfski (1986), who did a Q on opinions and perceptions of the administrator-legislator relationship.

Rarely, however, is Q methodology employed in the field of administrative ethics. Close to the field of administrative ethics, although not selected in Menzel's (2005) overview, is the study by Brewer, Selden, and Facer (2000), which used the technique on individual conceptions of public service motivation. Earlier, Hiruy (1987) wrote a Ph.D. thesis on the ethical orientation of public administrators. The respondents,

***Context is of the utmost importance
for much of the research in
administrative ethics.***

with the help of Q, reflected on ethical theories that emphasized personal, situational, social, and existential ethics.

Why aren't there more such studies in administrative ethics? And what exactly is the added value of Q for administrative ethics? The discussion in this article explores some of the weaker points of the methodologies that today dominate administrative ethics, a field also sometimes termed "public service ethics" (Lawton and Doig 2006) and "ethics and integrity of governance" (Menzel 2005). It then considers whether Q can help administrative ethics with respect to the weaknesses of current methodologies and also treats the possibilities of combining Q with discourse theory.

Qualitative and Quantitative Studies in Administrative Ethics

Context is of the utmost importance for much of the research in administrative ethics. Values get their meaning only within the context in which they are used (de Graaf 2003). What is meant by the value "freedom," for example, can only be determined by the context it is used in. That is why case studies are popular in the field. Often an explorative and inductive research strategy is chosen (Eisenhardt 1989), which is fitting when little is known about the phenomenon or when the phenomenon is so complex that the variables and their interrelationships are hard to define. Both of these considerations apply in much of the research in administrative ethics. Case studies offer the advantage of the rich detail of actual situations and their contextuality. Nonetheless, most are about only one case (e.g., Frederickson and Newman 2000). The question that remains is whether the conclusions can be generalized to a larger population.

Most of the published empirical research in the field consists of studies based on surveys. While surveys (especially with large, random samples) have the clear advantage of offering a basis for generalizing about populations, they also have well-known disadvantages. One is that the context of the researched phenomenon is lost. This is especially problematic in administrative ethics because the researched concepts are socially constructed and often heavily debated—concepts like "public ethos" or, even more problematic, "values." What are values? Where can we find them? Values are "essentially contested concepts": The proper use of these concepts and of concrete values (such as honesty) is never agreed upon (de Graaf 2003). The best that can be said is that values never come just by themselves; they are never unaccompanied. Values are always attached to a value manifestation. In other words, the meaning of these concepts depends entirely on the context in which they are used. The contextuality of values demands that quantitative methods for studying values introduce validity threats: It is hard to know, for instance, whether employees who *speak* about the same value *mean* the same value (cf. van der Wal et al. 2006).

Another example of the specific drawbacks of quantitative studies in the field is the research on whistle-blowing. Much of the correlation and regression research on whistle-blowing studies it outside its own unique context. In a sense, the reporting agent “disappears” along with the violation that is reported. This leads to the study of factors or variables that are certainly relevant to an understanding of whistle-blowing, but it draws attention away from the act of reporting and the reporter. Quantitative research does not tell the whole story about the decision process of the reporter; it necessarily ignores the characteristics and details of the context of each case. Quantitative work cannot account for contingency, which is so important for social research—especially in the contingent reporting cases—because of the complexity of the phenomenon. It also says little about the *reasons* for reporting. Most research shows, for example, that there is a correlation between group size and whistle-blowing (Miceli and Near 1988; Miceli, Dozier, and Near 1991), but causality cannot be derived from this finding because correlations do not provide causal links.

Before discussing whether Q can help with the identified weaker points of qualitative and quantitative studies, it must be clear what Q is and what its characteristics are.

What Is Q Methodology?

Q methodology provides a foundation for the systematic study of subjectivity—a person’s viewpoint, opinion, beliefs, attitudes, and the like (Brown 1993). It was introduced by William Stephenson in 1935 when he announced his inversion of the use of intercorrelations so that individuals were measuring themselves rather than being measured by a researcher (Smith 2000). Stephenson distinguished his method from R methodology (hence the name “Q methodology”), which provided (and provides) the basis for a science of objectivity in psychology (Brown 1986). “The letter R in R methodology is a generalization of Pearson’s product moment r , which has most often been used in the study of relationships among objective characteristics such as traits, attributes, abilities, and so forth” (Brown 1986, 57). In contrast to R methodology, Stephenson correlated people rather than test items.

In a Q study, people are typically presented with a sample of statements about some topic (e.g., the loyalty of administrators) known as the Q set. Respondents, or the P set, are asked to rank-order the statements from their individual points of view according to some preference, judgment, or feeling about them, mostly using a quasi-normal distribution. By this process of Q sorting, people give their subjective meaning to the statements, and in doing so reveal their subjective viewpoints (Smith 2001) or personal profiles (Brouwer 1999).

The individual rankings (or viewpoints) are subjected to factor analysis. If each individual had specific likes and dislikes, Stephenson (1935) argued, the profiles of the respondents would not correlate; if, however, significant clusters of correlations existed, they could be factorized and described as common viewpoints (or tastes, preferences, dominant accounts, typologies, etc.), and individuals could be measured with respect to them. Brouwer argued that one of the important advantages of Q is that questions pertaining to one and the same domain are not analyzed as separate items of information but rather in their mutual coherence for the respondent: “Subjective feelings and opinions are most fruitfully studied when respondents are encouraged

to order a good sample of items from one and the same domain of subjective interest (instead of just replying to single questions)” (1999, 35).

The results of a Q methodological study can be used to describe a population of viewpoints (Risdon et al. 2003). The factors resulting from Q analysis thus represent operant clusters of subjectivity; that is to say, they represent distinctions that are functional (as seen by the subject) rather than merely logical (as seen by the researcher) (Brown 2002, 1993). “Studies using surveys and questionnaires often use categories that the investigator imposes on the responses. Q, on the other hand, determines categories that are operant” (Smith 2001, 324).

Interested readers will find more information on the methodological background of Q in Stephenson (1953) and Brown (1980, 1986), a guide for Q technique in Brown (1980, 1986, 1993), and a recent discussion and review of applications in Smith (2001).¹ The Appendix provides a basic introduction to Q methodology, largely based on Brown (1980, 1993), that discusses the five steps in a Q methodological study: (1) definition of the concourse, (2) development of the Q sample, (3) selection of the P set, (4) Q sorting, and (5) analysis and interpretation.

Q and Administrative Ethics

Surveys generally claim to make objective measurements of some construct formulated about a population and assume that the differences are only quantitative (Smith 2001). Stephenson (1935, 18–19) presented Q methodology as an inversion of conventional factor analysis in the sense that Q correlates persons instead of tests. “Whereas previously a large number of people were given a small number of tests, now we give a small number of people a large number of test-items.” These test-items are always related to each other; therefore they are always within a context, and context is vital to administrative ethical research. Q can be very helpful in exploring tastes, preferences, sentiments, motives, and goals—the parts of personality that have great influence on behavior but often remain unexplored. Where single case studies have problems with generalizations, Q offers the opportunity to generalize clusters of viewpoints within a given population, clusters of subjectivity that are operant.

The field of administrative ethics is a broad area of study. Lawton and Doig distinguish six themes within it: (1) the public service ethos, (2) regulation of conduct, (3) trust, (4) individual behavior, (5) professionals, and (6) context. “These themes are enduring because individual, organizational and societal values continue to rub against each other” (2006, 21). Interestingly enough, empirical research questions in all six themes lend themselves to a Q methodological study. After all, as is now clear, Q methodology is well suited for studying empirical questions about the different views or standpoints that exist in any population. And these types of questions are asked in every one of the themes distinguished by Lawton and Doig (2006). Here are just a few examples of questions within these themes for which Q methodology is a good fit:

- Do civil servants have different understandings of “public ethos”?
- What form of trust do politicians find important?
- How do an organization’s members interpret the code of conduct?
- What different ethical intentions do public officials have?

- Where does the loyalty of public officials lie?
- What different ethical climates are perceived in a police organization, and does this perception depend on the officer's rank in the hierarchy?

As previous research using Q has shown (e.g., de Graaf 2005; Selden, Brewer, and Brudney 1999; van Eeten 2001), answering questions such as these with the help of Q methodology leads to empirically based theory (categories). And "theory based upon empirical evidence" is exactly what Lawton and Doig (2006, 28) call for in the field of administrative ethics. Q methodology gives the investigator the opportunity to examine and build theory without predeveloped categories. Indeed, it gives the investigator the opposite—an opportunity to construct categories based on information received only from the researched population.

Beyond looking for the mainstream viewpoints, it is the duty of academicians to also seek the *marginalized* viewpoints. Many surveys, however, conceal marginalized viewpoints (Dryzek 2005) because they "can be washed out in averages as across gender, party identification, and all manner of demographic categories, whose status is structural rather than functional" (Brown 2006, 374). Another interesting feature of Q methodology with advantages for administrative ethics is its "built-in features that, while not providing guarantees, certainly load the dice in favor of seeing things from the native's or any other point of view, marginalized or otherwise" (Brown 2006, 365). The elements of the P sample, once they have been Q sorted, reveal the inherent structure of a community's viewpoints and of the functional groups that contribute to it; marginalized viewpoints are revealed on the same level as mainstream viewpoints (Brown 2006, 374). This means that Q methodology can be used for empowerment (Brown 2006, 377), especially in the areas of race (Hunter and Davis 1992) and gender (Kitzinger 1999; Snelling 1999). For example, when the loyalty of administrators is studied with Q, a minority viewpoint may become apparent.

An examination of two concrete research questions in administrative ethics will show how Q methodology can contribute to the field by revealing what is not revealed by other methodologies. The first question concerns how public administrators reconcile competing values in their daily work. One could answer this question by studying a few administrators closely (interviews or observation) using a multiple-case study design. Based on such an approach, one could draw conclusions and build theory, but it would not be possible to derive generalizations, because there are only a few cases.

If the survey were conducted with 500 randomly chosen administrators, asking them about their values and how they deal with competing values, the results would be generalizable. One would then be able to derive how managers with different personal or contextual characteristics deal with competing values, but it would still not be possible to cluster the managers according to the way they reconcile competing values. This is exactly why Selden, Brewer, and Brudney (1999) did a Q methodological study to examine the role conceptions of public administrators and to ascertain how competing values are reconciled within each role. They found that the officials described five different roles, five different ways of reconciling competing values: steward of the public interest, adapted realist, businesslike utilitarian, resigned custodian, and practical idealist (Selden et al. 1999, 184).

Another example would be research on the differences in values between the public and private sectors. A qualitative approach could be chosen in which a limited

number of sector-switchers are interviewed and asked about their experiences, as in the study by de Graaf and van der Wal (2008). Although this study gave considerable attention to the context of the respondents, the conclusions are not easily generalizable because only sixty people were interviewed. Another approach would be to conduct a large survey of public and private sector managers and measure which values they deem most important in their decision-making (e.g., van der Wal, de Graaf, and Lasthuizen 2008). Differences between the sectors are then expressed in averages; for example, “contrary to the private sector, lawfulness, incorruptibility and impartiality are the most important public sector values.” In other words, an overall picture of the value differences emerges from the survey. It would also be possible, of course, to cluster the data of the survey, but in that case the *test-items* of the research would be clustered. The result would be clusters of statements that were valued similarly by the respondents. In a Q study, on the other hand, the factor analysis would be conducted over the *respondents*, and the results would be clusters of respondents who think similarly about the statements; in other words, different types of administrators could be distinguished based on what they consider the most important values in the decisions they make.

Q Methodology and Discourse Analysis

Thus far Q methodology has not been discussed in relation to epistemologies and ontologies, which usually indicates the assumption, explicit or implicit, of a positivist research tradition. However, Q can certainly be combined with (variants) of post-positivistic research, such as discourse theory. For example, the political scientist Dryzek (1990, 187) names Q methodology as a method to describe discourses. Examples of successful discourse analyses using Q methodology include work by van Eeten (1998), Dryzek and Berejikian (1993), and Thomas, McCoy, and McBride (1993). “Q study will generally prove a genuine representation of that discourse as it exists within a larger population of persons. . . . To put it another way, our units of analysis, when it comes to generalizations, are not individuals, but discourses” (Dryzek and Berejikian 1993, 52). It is worth stating once again that the researcher does not use predeveloped categories in examining the discourses. This gives investigators the opportunity to reconstruct the discourses in their own words, using only the words spoken by individuals in the discourse.

One methodological problem for discourse analysis is deciding which of the texts found within an organization are important and which are not—Putnam and Fairhurst (2001) call this the major question of sampling. Q methodology deals with it in an interesting and advantageous way: Instead of simply sampling all available texts, Q makes certain that those aspects most important in a discourse come to the fore.

The combination of some form of discourse theory and Q methodology is well suited to studies of the framing of ethical questions. There have been very few empirical studies on the framing of moral questions in administrative ethics (Lawton and Doig 2006; Menzel 2005), but the subject is important, because officials frame such questions on a daily basis. The fact *that* a moral question arises at a certain place and time is as interesting as *what* the question is, as is the fact that many moral questions are *not* asked. Every question asked receives an answer of some kind that has consequences.

One way to study the framing of moral questions is to study the discourses

of public officials. See, for example, Lawton's contribution to this symposium. Discourses necessarily contain both facts and values (e.g., Foucault 1980; Hajer 1995). Moral elements and factual statements are inextricably joined within a discourse (de Graaf 2006). How one looks at the world and how one perceives facts necessarily determine how one values. The "is" and "ought" influence each other in countless ways. In daily conversations, normative and factual statements are intertwined; people often do not realize how much their views of facts determine whether they see problems in the first place. But when discussions are studied more carefully, it can be seen that the "is" and "ought" are intertwined. Specific discourses raise different moral questions. Once managers of soccer clubs begin to talk about soccer as a "product," a relatively recent development in Europe, a new world opens up around the same old game with new opportunities, managerial problems, and moral issues (Hawkes 1998). Discourse analyses with Q methodology do not simply help us to understand *that* a certain moral question is asked; they also give the context wherein the moral problems lie and therefore their spectrum of possible solutions—what is or is not seen as a viable solution to a specific moral problem. As Hajer (1995, 54) observes: "Discourse analysis investigates the boundaries between . . . the moral and the efficient, or how a particular framing of the discussion makes certain elements appear fixed or appropriate while other elements appear problematic." A problem definition inevitably predisposes certain solutions, and vice versa. Schön and Rein argue that:

One way to study the framing of moral questions is to study the discourses of public officials.

When participants . . . name and frame the . . . situation in different ways, it is often difficult to discover what they are fighting about. Someone cannot simply say, for example, "Let us compare different perspectives for dealing with poverty," because each framing of the issue of poverty is likely to select and name different features of the problematic situation. We are no longer able to say that we are comparing different perspectives on "the same problem," because the problem itself has changed. (1994, 153)

Asking a moral question assumes that one knows what would constitute an answer to it. A discourse analysis with Q methodology can identify the rules and resources that set the boundaries of what can be said, thought, and done in a particular context or situation, what Foucault (1980) called "the conditions of possibility" of a discourse. "Thus, if we are to comprehend how decisions are made . . . it is by examining the conditions of possibility in relation to which these statements are formulated, that is, the often implicit institutionalized speech practices that guide what is and what is not likely to be said" (Mauws 2000, 235).

All in all, interesting opportunities in administrative ethics exist in which Q methodology can be combined with discourse theory. The method is not restricted to positivistic research traditions.

What Q Cannot Do

Like any other approach, Q has its drawbacks. Although more and more Q studies are administered via mail and the Internet, most Q sorting still takes place in an interview setting. And interviews, with their planning and traveling, take time.

Generating statements can also be time-consuming, especially when they are derived from interviews, as is often the case.

Q methodology also has the rather obvious disadvantage that it is designed to study subjectivities and thus is not suited to study matters of fact. This means that it cannot be used for many interesting research questions that ask for matters of fact to be measured.

Another problem often mentioned is that different investigators use different structures to compose their Q sets (see the Appendix), leading to differing Q sets from the same set of possible statements (the *concourse*). Many Q methodologists, however, do not regard this as a concern for two reasons. First, the structure chosen is only a logical construct used by the investigator; whatever the starting point, the aim is always to arrive at a Q set that is representative of the wide range of existing opinions about the topic. Second, irrespective of the structure and of what the researcher considers a balanced set of statements, it is the subject that eventually gives meaning to the statements by sorting them (Brown 1993). The limited number of comparative studies that have been carried out indicate that different sets of statements structured in different ways can nevertheless be expected to converge on the same conclusions (Thomas and Baas 1992).

Furthermore, because Q is a small-sample investigation of human subjectivity based on sorting items of unknown reliability, results from Q methodological studies have been criticized for their reliability (Thomas and Baas 1992).² The most important type of reliability for Q is replicability: Will the same condition of instruction lead to factors that are schematically reliable (i.e., represent similar viewpoints on the topic) across similarly structured but different Q samples when administered to different sets of persons? According to Brown (1980), an important notion behind Q methodology is that only a limited number of distinct viewpoints exists on any topic. Any well-structured Q sample containing the wide range of existing opinions on the topic will reveal these perspectives.³ Based on the findings of two pairs of tandem studies, Thomas and Baas (1992) concluded that skepticism about this type of reliability is unwarranted. The more common notion of statistical reliability regarding the ability to generalize sample results to the general population is of less concern here. As a small-sample methodology, Q is not well suited to cross-sectional or large-sample purposes: “administering Q-sorts is a clumsy way to count noses” (Brown 2002, 18). This holds true as well for any relations between factors and the characteristics of the people loading on these factors. From a purely statistical standpoint they are nonsignificant, although salient relations have been used as hypotheses for further research.

The procedure of a forced distribution has, according to some, the disadvantage of being an artificial procedure (Polit and Hungler 1999). Such a distribution is seen as unnatural, or as violating the independence assumption. It can be argued that this tends to exclude information concerning how people would ordinarily distribute their opinions, but studies comparing forced versus unforced distributions have shown that this has no bearing on the qualitative and statistical results (Block 1957; Brown 1971). “People can ‘tell a story’ only if they have the appropriate statements with which to tell it. Thus, the start of a Q study involves a careful and methodological review of the things people write and say about the topic in question” (Cross 2005, 212). In addition, there is a risk of bias at the interpretation stage. “To take the analysis beyond the most basic descriptive and counting exercise requires the

researchers' analytical skills in moving toward hypotheses or propositions about the data" (Cross 2005, 211).

Finally, an often-heard criticism concerns the fact that there is no clear guidance to the number of factors that should be chosen as the final solution of the analysis of the Q sorts, and therefore that the results of a Q study are sensitive to the researcher's personal opinions, expertise, and technical skill. Academic publications using Q methodology only infrequently present and compare alternative factor solutions, thereby making it impossible for outsiders to assess whether the presented solutions and research results are the most appropriate. This is valid and important criticism, albeit again not unique to Q methodology.

Summarizing, the main problems with Q methodology have to do with reliability and generalizability. It is important to remember that this technique measures the existence of subjectivities within a population, and not the exact distribution of clusters of opinion.

Conclusion

What is the potential for Q methodology in administrative ethical research and organizational and policy guidance? This question has been answered herein by way of an examination of the many possibilities for Q methodological studies in the field. Q clearly offers a procedure and conceptual framework well suited to studying subjectivity in the social context.

Although only touched upon in the discussion of marginalized viewpoints, the promise of Q methodology is greater than just describing subjectivity: It also offers new opportunities for prescription. After all, descriptions have consequences. By describing the different moral dimensions of their discourses, managers can, for instance, become more aware of the problems they and their colleagues see. In a sense, describing different viewpoints can be therapeutic. Viewpoint (discourse) descriptions can make parties more aware of their stances, factual and valuational. This might change how they talk and weigh their options. Awareness of different discourses leads to a different view of self-discourse. This kind of plurality is important⁴—as long as the other viewpoints make sense, of course, which might not be the case when the other viewpoints are theoretically constructed. And this is precisely one of Q's strongest points: It gives the investigator, whether an academician or a practitioner, the opportunity to examine and build theory without predeveloped categories.

NOTES

1. Various documents and manuscripts are available from the QArchive at the University of Wisconsin (www.uww.edu/personal/fac/cottlec/QArchive/qindex.htm) and from Peter Schmolck's QMethod page (www.rz.unibw-muenchen.de/~p41bsmk/qmethod). We also recommend the Web site of the International Society for the Scientific Study of Subjectivity (www.qmethod.org/).

2. Because there is no external criterion for a person's point of view, the issue of the validity of Q sorts does not apply (Brown 1980).

3. The test-retest reliability of Q sorts has been demonstrated to range upwards from 0.80 (Brown 1980).

4. Bauman (1991) noted this too when discussing the Milgram studies. The subjects

were inclined to do what a “scientist” in a white coat asked of them, even if that was against their moral feelings. But as soon as there were two “scientists” in white coats with *opposing* views and instructions, the subjects were no longer prepared to follow instructions to physically hurt other people. The first thing a totalitarian regime usually tries to do is silence opposing points of view: to not allow people moral choices or opposing moral points of view (discourses).

REFERENCES

- Bauman, Zygmunt. 1991. *Modernity and the Holocaust*. Ithaca: Cornell University Press.
- Brewer, Gene A., Sally Coleman Selden, and Rex L. Facer II. 2000. “Individual Conceptions of Public Service Motivation.” *Public Administration Review* 60, no. 3:254–264.
- Brouwer, Marten. 1999. “Q Is Accounting for Tastes.” *Journal of Advertising Research* 39, no. 2:35–39.
- Brown, S. R. 1978. “The Importance of Factors in Q Methodology: Statistical and Theoretical Considerations.” *Operant Subjectivity* 1, no. 4:117–124.
- Brown, Steven. 1980. *Political Subjectivity: Applications of Q-Methodology in Political Science*. New Haven: Yale University Press.
- . 1986. “Q Technique and Method: Principles and Procedures.” In *New Tools for Social Scientists*, edited by William D. Berry and Michael S. Lewis-Beck, 57–76. Beverly Hills, Calif.: Sage.
- . 1993. “A Primer on Q methodology.” *Operant Subjectivity* 16, no. 3/4:91–138.
- . 2002. “Structural and Functional Information.” *Policy Sciences* 35:285–304.
- . 2006. “A Match Made in Heaven: A Marginalized Methodology for Studying the Marginalized.” *Quality & Quantity* 40:361–382.
- Cross, R. M. 2005. “Exploring Attitudes: The Case for Q Methodology.” *Health Education Research* 20, no. 2:206–213.
- Cunningham, Robert, and Dorothy Olshfski. 1986. “Interpreting State Administrator-Legislator Relationships.” *Western Political Quarterly* 39, no. 1:104–117.
- de Graaf, Gjalt. 2003. *Tractable Morality. Customer Discourses of Bankers, Veterinarians and Charity Workers*. Rotterdam: Erim.
- . 2005. “Veterinarians’ Discourses on Animals and Clients.” *Journal of Agricultural and Environmental Ethics* 18:557–578.
- . 2006. “Discourse and Descriptive Business Ethics.” *Business Ethics: A European Review* 13, no. 3:247–258.
- de Graaf, Gjalt, and Zeger van der Wal. 2008. “On Value Differences Experienced by Sector Switchers.” *Administration & Society* 40, no. 1: 79–103.
- Dryzek, John S. 1990. *Discursive Democracy: Politics, Policy and Political Science*. Cambridge: Cambridge University Press.
- . 2005. “Handle with Care: The Deadly Hermeneutics of Deliberative Instrumentation.” *Acta Politica* 40:197–211.
- Dryzek, John S., and J. Berejikian. 1993. “Reconstructive Democratic Theory.” *American Political Science Review* 87, no. 1:48–88.
- Eisenhardt, Kathleen M. 1989. “Building Theories from Case Study Research.” *Academy of Management Review* 14, no. 4:532–550.
- Foucault, Michel. 1980. In *Power/Knowledge: Selected Interviews and Other Writings 1972–1977*, by Michel Foucault. Brighton: Harvester Press.
- Frederickson, H. George, and Meredith A. Newman. 2000. “The Patriotism of Exit and Voice: The Case of Gloria Flora.” *Public Integrity* 3:347–362.
- Gaines, Larry K., Norman Van Tubergen, and Michael A. Paiva. 1984. “Police Officers Perceptions of Promotions as a Source of Motivation.” *Journal of Criminal Justice* 12, no. 3:265–275.

- Hajer, Maarten A. 1995. *The Politics of Environmental Discourse*. Oxford: Clarendon Press.
- Hawkes, Toby. 1998. "Scoring an Own Goal? Ethical Issues in the UK Professional Soccer Business." *Business Ethics: A European Review* 1, no. 1:37–47.
- Hiruy, Minas. 1987. "Exploring the Perspectives of Ethics: The Case of Public Administrators in the United States." Kent State University, Kent.
- Hunter, Andrea G., and James E. Davis. 1992. "Constructing Gender: An Exploration of Afro-American Men's Conceptualization of Manhood." *Gender & Society* 6:464–479.
- Kitzinger, Celia. 1999. "Commentary: Researching Subjectivity and Diversity: Q-Methodology in Feminist Psychology." *Psychology of Women Quarterly* 23:267–276.
- Lawton, Alan, and Alan Doig. 2006. "Researching Ethics for Public Service Organizations: The View From Europe." *Public Integrity* 8, no. 1:11–33.
- Mauws, Michael K. 2000. "But Is It Art? Decision Making and Discursive Recourses in the Field of Cultural Production." *Journal of Applied Behavioral Science* 36, no. 2:229–244.
- Menzel, Donald C. 2005. "Research on Ethics and Integrity in Governance: A Review and Assessment." *Public Integrity* 7, no. 2:147–168.
- Miceli, Marcia P., and Janet P. Near. 1988. "Individual and Situational Correlates of Whistle-Blowing." *Personnel Psychology* 41:267–282.
- Miceli, Marcia P., Janelle B. Dozier, and Janet P. Near. 1991. "Blowing the Whistle on Data-Fudging: A Controlled Field Experiment." *Journal of Applied Social Psychology* 76:508–521.
- Polit, Denise F., and Bernadette P. Hungler. 1999. *Nursing Research Principles and Methods*. Philadelphia: Lippincott.
- Putnam, L., and G. Fairhurst. 2001. "Discourse Analysis in Organizations." In *The New Handbook of Organizational Communication: Advances in Theory, Research and Methods*, edited by J. F. and L. Putnam, pp. 78–136. Thousand Oaks, Calif.: Sage.
- Reber, Bryan H., Stanley E. Kaufman, and Fritz Cropp. 2000. "Assessing Q-Assessor: A Validation Study of Computer-Based Q Sorts Versus Paper Sorts." *Operant Subjectivity* 23, no. 4:192–209.
- Risdon, Andrea, Chris Eccleston, Geert Crombez, and Lance McCracken. 2003. "How Can We Learn to Live with Pain? A Q-Methodological Analysis of the Diverse Understandings of Acceptance of Chronic Pain." *Social Science & Medicine* 56:375–386.
- Schön, Donald A., and Martin Rein. 1994. *Frame Reflection: Toward the Resolution of Intractable Policy*. Boston: MIT Press.
- Selden, Sally Coleman, Gene A. Brewer, and Jeffrey L. Brudney. 1999. "Reconciling Competing Values in Public Administration: Understanding the Administrative Role Concept." *Administration & Society* 31, no. 2:171–204.
- Smith, Noel Wilson. 2000. *Current Systems in Psychology: History, Theory, Research, and Applications*. Belmont, Calif.: Wadsworth.
- . 2001. *Current Systems in Psychology: History, Theory, Research, and Applications*. Belmont, Calif.: Wadsworth.
- Snelling, Susan J. 1999. "Women's Perspectives on Feminism: A Q-Methodological Study." *Psychology of Women Quarterly* 23:247–266.
- Stephenson, William. 1935. "Correlating Persons Instead of Tests." *Character and Personality* 4:17–24.
- . 1953. *The Study of Behavior: Q-Technique and Its Methodology*. Chicago: University of Chicago Press.
- Thomas, Dan B., and Larry R. Baas. 1992. "The Issue of Generalization in Q Methodology: 'Reliable Schematics' Revisited." *Operant Subjectivity* 16, no. 1:18–36.
- Thomas, Dan, Craig McCoy, and Allan McBride. 1993. "Deconstructing the Political Spectacle: Sex, Race and Subjectivity in Public Response to the Clarence Thomas/

- Anita Hill 'Sexual Harassment' Hearings." *American Journal of Political Science* 37, no. 3:699–720.
- Van der Wal, Zeger, Gjalt de Graaf, and Karin Lasthuizen. 2008. "What's Valued Most? Similarities and Differences Between the Organizational Values of the Public and Private Sector." *Public Administration* 86, no. 2:465–482.
- Van der Wal, Zeger, Leo Huberts, Hans van den Heuvel, and Emile Kolthoff. 2006. "Central Values of Government and Business: Differences, Similarities and Conflicts." *Public Administration Quarterly* 30, no. 3:314–65.
- Van Eeten, Michel. 1998. *Dialogues of the Deaf: Defining New Agendas for Environmental Deadlocks*. Delft: Eburon.
- . 2001. "Recasting Intractable Policy Issues: The Wider Implications of the Netherlands Civil Aviation Controversy." *Journal of Policy Analysis and Management* 20, no. 3:391–414.
- Van Tubergen, Norman N., and Robert A. Olins. 1979. "Mail vs. Personal Interview Administration for Q Sorts: A Comparative Study." *Operant Subjectivity* 2, no. 2:51–59.

Appendix: How Does Q Methodology Work?

Performing a Q methodological study involves the following steps: (1) definition of the concourse, (2) development of the Q sample, (3) selection of the P set, (4) Q sorting, and (5) analysis and interpretation. Discussion of each step follows.

Definition of the Concourse

In Q, the concourse refers to "the flow of communicability surrounding any topic" in "the very stuff of life, from the playful banter of lovers or chums to the heady discussions of philosophers and scientists to the private thoughts found in dreams and diaries" (Brown 1993, 95). Concourse is a technical concept (not to be confused with discourse) for the collection of all the possible statements the respondents can make about the subject at hand. The concourse is thus supposed to contain all the relevant aspects of all the discourses. It is for the researcher to draw a representative sample from the concourse at hand. The concourse may consist of self-referent statements (i.e., opinions rather than facts), objects, pictures, and so on. A verbal concourse may be obtained in a number of ways: interviews, participant observation, popular literature (media reports, newspapers, magazines, novels), and scientific literature (articles, essays, books). The gathered material represents existing opinions and arguments, i.e., what laypeople, politicians, representative organizations, professionals, and scientists have to say about the topic; this is the raw material for a Q. Although any source may be used (and many have been), "the level of the discourse dictates the sophistication of the concourse" (Brown 1993, 95).

Development of the Q Set

A subset of statements is drawn from the concourse and subsequently presented to the participants. This Q set (or Q sample) often consists of forty to fifty statements, but it can contain fewer than forty or more than fifty (e.g., van Eeten 1998). According to Brown (1980, 186), the selection of statements for inclusion in the Q set is of crucial importance, but remains "more an art than a science": The researcher uses a structure for selection of a representative miniature of the concourse. Such a

structure may *emerge* from further examination of the statements in the concourse or may be *imposed* on the concourse based on some theory. Whatever structure is used, it forces the investigator to select statements widely different from one another in order to make the Q set broadly representative (Brown 1980). Finally, the statements are edited where necessary, randomly assigned a number, and printed on separate cards, the Q deck, for Q sorting.

Selection of the P Set

As discussed, a Q study requires only a limited number of respondents: “All that is required are enough subjects to establish the existence of a factor for purposes of comparing one factor with another. . . . P sets, as in the case of Q samples, provide breadth and comprehensiveness so as to maximize confidence that the major factors at issue have been manifested using a particular set of persons and a particular set of Q statements” (Brown 1980, 192–194). The P set is usually smaller than the Q set (Brouwer 1999). The aim is to have four or five persons defining each anticipated viewpoint, which often number two to four, rarely more than six. The P set is not random. It is a structured sample of respondents who are theoretically relevant to the problem under consideration—for instance, persons who are expected to have a clear and distinct viewpoint regarding the problem and, in that quality, may define a factor (Brown 1980). Eventually, the number of persons associated with a factor is of less importance than who they are; in the total population the prevalence may be much higher (Brown 1978).

Q Sorting

The general procedure is as follows (Brown 1993): The Q set is given to the respondent in the form of a pack of randomly numbered cards, each card containing one of the statements from the Q set. The respondent is instructed to rank the statements according to some rule—the *condition of instruction*, typically the person’s point of view regarding the issue—and is provided with a score sheet and a suggested distribution for the Q sorting task. The score sheet is a continuum ranging from *most* to *most* (“most disagree” on one end and “most agree” on the other, for instance),⁵ which form the endpoints of a distribution that usually takes the form of a quasi-normal distribution. The kurtosis (i.e., the state of flatness or peakedness of the curve) of this distribution depends on the controversiality of the topic.

In case the involvement, interest, or knowledge of the respondents is expected to be low, or relatively few statements are expected to be salient, the distribution should be steeper to leave more room for ambiguity, indecisiveness, or error in the middle of the distribution. In case respondents are expected to have strong or well-articulated opinions on the topic at issue, the distribution should be flatter to provide more room for strong (dis)agreement with statements. Usually, respondents are requested to adhere to the distribution provided.⁶ The range of the distribution depends on the number of statements and its kurtosis. According to Brown (1980), most of today’s Q sets contain forty to fifty statements and employ a relatively flattened distribution with a range of –5 to +5.

Respondents are asked to read through all of the statements carefully to gain an

impression of the type and range of opinions at issue. They are instructed to begin with a rough sorting while reading by dividing the statements into three piles: generally agree (or like, find important, etc.), generally disagree, and neutral, doubtful, or undecided. The number of statements in each pile is recorded to check for agreement-disagreement balance in the Q set. Next, participants are asked to rank-order the statements according to the condition of instruction and to place them in the provided score sheet. It is generally recommended to follow the Q sort with an interview in which the sorters are invited to elaborate on their points of view, especially on the most salient statements, those placed at the extreme ends of the continuum. This information is helpful for the interpretation of factors later on.

Many feel that because the sorting procedure is complex and unfamiliar to the lay public, it requires administration in an interview setting. Van Tubergen and Olins (1979), however, argue that Q studies may just as well be conducted by mail. They found results from Q sort self-administration to be highly congruent with those from face-to-face settings. Reber, Kaufman, and Cropp (2000) performed two validation studies comparing computer- and interview-based Q sorts and concluded that there is no apparent difference in the reliability or validity of these two methods of administration. Interviews, however, usually enable the researcher to understand the results better, and this often leads to a more penetrating interpretation. The present authors would mail a Q sort only if the relevant sample had a wide geographical distribution or administrative costs (including time) posed a hardship.

Analysis and Interpretation

Brown (1980, 1993) provides a comprehensive overview of the analysis of the Q sorts. Because many software packages are now available to perform the analysis, the overview here will be limited.⁷

The analysis of the sorts is a purely technical, objective procedure—and is therefore sometimes referred to as the scientific base of Q. First, the correlation matrix of all Q sorts is calculated. This represents the level of (dis)agreement between the individual sorts, that is, the degree of (dis)similarity in points of view between the individual sorters. The correlation matrix is then subjected to factor analysis to identify the number of natural groupings of Q sorts by virtue of similarity or dissimilarity—that is, determining how many basically different Q sorts are in evidence (Brown 1993, 1980). People with similar views on the topic will share the same factor. A factor loading is determined for each sort, expressing the extent to which each Q sort is associated with each factor. The number of factors in the final set depends on the variability in the elicited Q sorts.⁸ Taking along more than the number of factors anticipated in the next step of the analysis (factor rotation) to preserve as much of the variance as possible is recommended: “Experience has indicated that ‘the magic number 7’ is generally suitable” (Brown 1980, 223).

This original set of factors is then rotated to arrive at a final set of factors. Rotation may be either objective, according to some statistical principle (like varimax), or theoretical (or judgmental), driven by theoretical concerns, some prior knowledge or preconceived idea of the investigator; or an idea that came up during the study (e.g., from a salient Q sort or during a follow-up interview).⁹ By rotating the factors, the investigator muddles about the sphere of opinions and examines it from different angles. A judgmental rotation looks for confirmation of an idea or a theory,

a theoretical rotation for an acceptable vantage point by statistical criteria (though the investigator has to judge the acceptability of this solution). Rotation does not affect the consistency in sentiment throughout individual Q sorts or the relationships between Q sorts, it only shifts the perspective from which they are observed. Each resulting final factor represents a group of individual points of view that are highly correlated with each other and uncorrelated with others.¹⁰

The final step before the factors are described and interpreted is the calculation of factor scores and difference scores. A statement's factor score is the normalized weighted average statement score (Z-score) of the respondents that defined that factor.¹¹ Based on their Z-scores, statements can be attributed to the original quasi-normal distribution, resulting in a composite (or idealized) Q sort for each factor. The composite Q sort of a factor represents how a hypothetical respondent with a 100 percent loading on that factor would have ordered all the statements of the Q set. When the factors are computed, one looks back at the Q sorts to see how high their loadings are on the different factors. If a respondent's factor loading exceeds a certain limit (usually $p < 0.01$), it is called a defining variate (or variable).¹² The difference score is the magnitude of difference between a statement's score on any two factors that is required for it to be statistically significant.¹³ When a statement's score on two factors exceeds this difference score, it is called a distinguishing (or distinctive) statement.¹⁴ A statement that is not distinguishing between any of the identified factors is called a consensus statement.

Factor scores on a factor's composite Q sort and difference scores point out the salient statements that deserve special attention in describing and interpreting that factor. Usually, the statements ranked at the extreme ends of the composite sort of a factor, called characterizing statements, are used to produce a first description of the composite point of view represented by that factor. The distinguishing and consensus statements can be used to highlight the differences and similarities between factors. Finally, the explanations Q sorters gave during the follow-up interview can be helpful in the interpretation of the factors, in ex-post verification of the interpretation, and as illustration material.

This introduction to Q methodology is largely based on Brown (1993, 1980).

NOTES TO APPENDIX

1 Sometimes a continuum range from *least* to *most* on the same judgment item is used. For theoretical reasons, however, "most" to "most" (with absence of feeling in the middle) should be used wherever possible (Brown 1980). Alternative items that enable Q sorters to express their point of view next to "(dis)agree," for instance, are "important," "relevant," "desirable," and "attractive." The range of the continuum must match the conditions of instruction provided to Q sorters.

2. This forced distribution is practical but not necessary; it hardly has any effect on factors emerging from the data (Brown 1980).

3. For instance, PCQ by Stricklin (www.pcqsoft.com) and PQMethod by Schmolck and Atkinson (freeware: www.rz.unibw-munchen.de/~p41bsmk/qmethod; WebQ is also here).

4. The number of factors in the final set can be anticipated by (1) the number of original factors with at least two significant loadings, or more stringent, factors of which the cross-product of its two highest loadings (ignoring sign), exceeds twice the standard error; (2) the number of original factors with an eigenvalue (i.e., the sum of the squares of the factor loadings) in excess of 1.00.

5. Objective rotation is based on the structure of the data and therefore referred to as an objective or rational procedure. Theoretical (or judgmental) rotation gives more room to the aims and subjectivity of the investigator, who is nevertheless constrained by the structures that emerge from the data.

6. Secondary statistics include (1) factor eigenvalue, (2) percentage of total variance of the correlation matrix, and (3) communality, the sum of squared factor loadings per respondent, representing the part of a person's response that is associated with the factors (s)he has in common with the other respondents.

7. The weight w is based on the respondent's factor loading f , and is calculated as $w = f/(1 - f^2)$. The weighted average statement score is then normalized (with mean of 0.00 and standard deviation of 1.00) to remove the effect of differences in numbers of defining respondents per factor, and making statements' factor scores comparable across factors. Statements with a Z-score larger than 1 (or smaller than -1) are referred to as *characterizing* for that factor.

8. The limit for statistical significance of a factor loading is calculated as the multiplier for the desired level of statistical significance divided by the square root of the number of statements in the Q set (multipliers: 3.29 for $p < 0.001$; 2.58 for $p < 0.01$; 1.96 for $p < 0.05$).

9. The difference score is based on the standard error of the factor scores (SE) and a multiplier for the required level of statistical significance. See Brown (1980) for full detail.

10. Although a statement may be distinctive between two factors, usually a statement will be printed out as distinguishing only if it distinguishes one factor from *all* the other factors.

ABOUT THE AUTHORS

Gjalt de Graaf, Ph.D., is an assistant professor of public administration and organization in the Department of Public Administration of the VU University Amsterdam. His work has been published in *Public Administration Review*, *Journal of Business Ethics*, *Perspectives on European Politics and Society*, *Social Sciences & Medicine*, *Public Administration*, *Public Integrity*, *Business Ethics: A European Review*, *Administration & Society*, and *Health Policy*.

Job van Exel is a lecturer in the Department of Health Policy and Management, Erasmus Medical Center, Rotterdam. He received his M.Sc. in economics from Erasmus University, Rotterdam, in 1995. His work has been published in *Journal of Health Economics*, *Health Economics*, *Journal of Socio-Economics*, *Social Science & Medicine*, *Health Policy*, and *Gerontologist*.