

Integrating Qualitative and Quantitative Data: What Does the User Need?

Anthony P.M. Coxon

Key words:
qualitative-
quantitative
integration, mixed-
mode archiving,
cognition, diaries

Abstract: The convergence between quantitative and qualitative approaches is fragile. Nowhere is this more evident than in the attempt to lodge mixed-mode data. Assumptions about the "basic" form of the data dictate what is considered relevant. Quantitative archiving assumes no more than qualitative appended material, and qualitative archiving sits lightly on the structured nature of the quantitative data. Problems consequently arise at the level of data-collection and of retrieval and analysis. By reference to two mixed quantitative-qualitative projects, this contention is illustrated (Occupational cognition and Sexual diaries) where decisions about representation of the data dramatically affect the possibilities of retrieval and analysis in context.

Table of Contents

- [1. Qualitative and Quantitative: Contradictions, Contraries, or Choices—The False Dilemma?](#)
- [2. Research example \(1\): POOC \(Project on Occupational Cognition\), Edinburgh, 1972-75, Occupational Hierarchies \(COXON & JONES 1978, 1979a,b\)](#)
- [3. Research example \(2\): Project SIGMA \(Socio-sexual Investigations of Gay Men and Aids\) 1982-95: Sexual Diaries](#)
- [4. Some Conclusions](#)

[References](#)

[Author](#)

[Citation](#)

This paper is designed to be both contentious and pragmatic. It is *contentious* in the sense that it expresses possibly tendentious views of the dangers of an over-narrow qualitative focus, which can lead us to forget that much data is both quantitative and qualitative, and that the key issue is therefore one of integration. It is *pragmatic* in that the points that I shall make arise from my experience as:

- Creator, Lodger and Documenter of qualitative and quantitative data,
- Teacher and Researcher using qualitative and quantitative data, and
- Ex-Director of both a primarily quantitatively-oriented Institute (the Institute of Social and Economic Research) and ex-Acting Director of the Qualitative Archive (Qualidata) at the University of Essex. [1]

1. Qualitative and Quantitative: Contradictions, Contraries, or Choices—The False Dilemma?

The qualitative and quantitative contrast has had a history long before its social science use, but since its latter-day proclamation by Paul LAZARFELD and in LERNER's (1961) *Quality and Quantity*, it has now become accepted as an unquestioned taken-for-granted of social science research. I want to argue that

the distinction is not only internally incoherent and misleading, but needs eradication and replacing by more sensitive and useful distinctions. [2]

To what does the distinction refer? In usage, it can refer to several levels of discourse, and can be used differently within and across levels—hence its incoherence. Some of these levels in which the qualitative and quantitative contrast appear include:

- Paradigms or methodological approaches: in effect, Variable-centered *versus* Meaning-centered approaches.
- Epistemological positions: Positivism in its common-usage sense (or rather, Empiricism) *versus* Interpretativism, and this in turn paralleled by nomological and causal methods *versus* rule-following and reasons-based explanation.
- Data conceptualization: formal measurement approaches (such as algebraic representation and variable/attribute levels of measurement *versus* natural-language and speech as data. (It is ironic that within the former "quantitative" approach, "qualitative" is used to mean non-metric or non-continuous measurement.)
- Data collection: here, the contrast is usually between systematic questionnaire-format *versus* discursive data-elicitation, but can often be as trite as "closed-ended" *versus* "open-ended" question format.
- Data analysis models and approaches: following the form of data-conceptualization, appropriate analysis usually follows the General Linear Model and its statistical variants *versus* thematic, semantic and content analytical procedures. [3]

These distinctions are fairly commonplace, but there are two further aspects which are probably more potent and influential in research practice:

1. *Affiliative identification*: where one side of the contrast is taken as a label to identify those who follow what is considered the appropriate or authentic research procedures and to unchurch those outside it. As in other churches, the greatest negativity is reserved for those who will not accept the division, rather than for the opposition!¹ While dialectical development is probably necessary to legitimize new approaches, in the long run it is the development of CAIDAS, Computer-Assisted Integrative (or "qualitative/quantitative blind") Data Analysis, which is needed for a social science research environment. Indeed, quantitative data analysis can be considered as not significantly different in kind to interpretative procedures in qualitative analysis (a point well made by KRITZER 1996).
2. *Software hegemony*: programs developed and used for data analysis certainly follow user-demand in many ways, but they also embody implicit preconceptions of what is "appropriate" data and reinforce the above distinctions by effectively dictating what is and what is not taken into

1 The statement "Arid, statistical, formalistic positivism" can be spat out with the same disdain as "warm, pink, and fluffy qualitative approaches"!

consideration.² More dangerously they are in danger of implementing KAPLAN's (1964) "Law of the Instrument": give a small boy a hammer and he will find that everything he encounters will need pounding at the expense of more appropriate activities. This applies as much to the research practitioner as to the child. [4]

It could even be argued that the search should be on for a new paradigm to supplant the qualitative and quantitative distinction in social science, and if so I would advance the claims of Artificial Intelligence, which has successfully integrated both—as have other "cognitive" disciplines—in the search for adequate representation of more subtle and complex data-structures which today's developments demand: belief-systems, semantic networks, fuzzy-logic categorization. [5]

But that may be to look ahead too far. In this paper I want to concentrate rather on the current user's frustrations³ and argue for not separating software and data storage. In this I have no desire to deny the reality or importance of the quantitative *tradition* as a focus, a paradigm, nor to deny the importance of (primarily) *qualitative* methods. But I do wish to stake a claim for integrative middle ground as well. Integrated data also have their needs, and it is true that these are more liable to be met, and are probably more appropriately met within the qualitative tradition than elsewhere. [6]

2. Research example (1): POOC (Project on Occupational Cognition), Edinburgh, 1972-75, Occupational Hierarchies (COXON & JONES 1978, 1979a,b)

POOC was a project designed to investigate the conceptions and images of the occupational world and subjective aspects of occupational structures. One part of a range of cognitive "tasks" performed with occupational titles and descriptions asked of respondents in Edinburgh was concerned with the "Method of Hierarchies". The example here concentrates on data generated by the method. All data collection was done in an interview situation and tape-recorded. In brief, the subject was given 16 occupational titles, and asked to construct an inclusive hierarchy of levels of occupations. First he⁴ was instructed to pick out the two most similar, and say why they were so. Then he was asked either to pick out another pair, or add ("chain") another pair to the existing pair. This process of pairing, chaining and joining existing clusters continued until all occupations were in one cluster, thus creating an implied hierarchical clustering scheme (Johnson 1967) consisting of a set of inclusive clustering of occupations in increasing generality. At every stage he was encouraged to verbalize, and these reasons (grounds, bases)

2 It is interesting that the Harvard Package DATATEXT (ARMOR 1969), which included both survey *and* content analysis procedures, was followed and supplanted by the Chicago package SPSS, which triumphantly championed the survey component only.

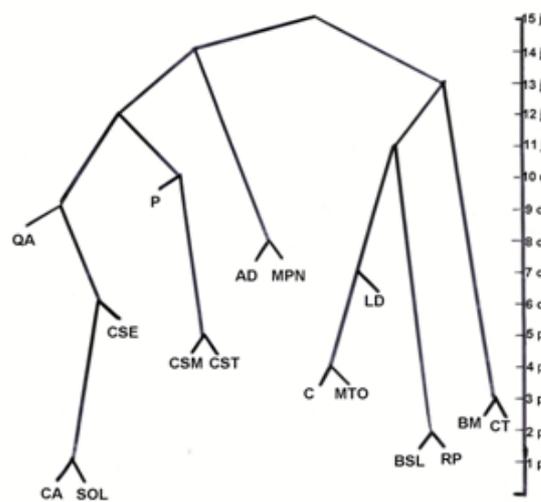
3 Perhaps the best advice to today's user is to keep discourse linear, and save it in ASCII format, because any more enriched data format will lead to problems. This more than anything else shows the "user-hostility" (rather than the oft-claimed "user-friendliness") of most software and incompatibility between packages.

4 All subjects were men, so the male pronoun is used throughout.

were an integral constitutive part of the data. In the analysis it was essential to know at what levels occupations were joined, what predicates were used to describe them, and conversely also to know for a given predicate, the level and the sub-structure to which the reasons applied. In data coding for analysis (and archiving!) the following steps were necessary for each subject's data:

1. the hierarchy was encoded as a number-bracketed sequence;⁵
2. this was checked for consistency and output to file as a matrix in which the entries $x(i,j)$ were the ultrametric distance between occupations (i) and (j) (the "quantitative" information);
3. the verbal material, with pointers to the subject and the level, was separately transcribed and stored as a text file (the "qualitative" information); and
4. for analysis,
5. "quantitative" programs such as SPSS and MDS(X) were used to aggregate and analyze the hierarchies, and
6. "qualitative" content analysis programs such as Inquirer (STONE et al. 1966) were used to analyze the verbal material. [7]

The hierarchy and the verbal material relevant to level 15 (the final join) are given in Figure 1 (and see COXON 1983).



[Final Join at Level 15 of the two main branches. What follows is the transcribed tape-recording relevant to this point]

{Respondent}
... I think that I would put the Psychiatric Nurse [MPN] and the Ambulance Driver [AD] in this category as well, um, I realise there's a contradiction here - I've got people with professional qualifications and perhaps academic skills, um, and an ambulance driver, but the reason I mentioned (i) Policeman [P] *vis a vis* the Commercial Traveller [CT] and Barman [BM]; it seems to me these people have responsibilities in relation to the community which they're working in, that perhaps the Commercial Traveller and Barman doesn't have.

Figure 1: A subject's hierarchy and verbalization at level 15 [8]

Programs had to be specially written to effect stages 2, 3 (although nowadays "qualitative" software has the ability to encode the pointer references) and, most

5 Since this was the era of the punched card, the representation in the Figure as a number-bracketed sequence was actually a stack of cards forming occupational clusters prefaced and followed by the appropriate level-numbered card.

crucially, programs had to be constructed to allow information from 4(A) and 4(B) to be related and retrieved in context—no mean task. [9]

To show how important "qualitative retrieval in quantitative context" is, consider the research problem of examining whether the Themes established as most prevalent in the occupational narratives generated in doing the Hierarchies task—Money, Training, Caring and Responsibility—are generalizing or particularizing themes. We thus need to know not simply how often a Theme occurs in these data but, more relevantly, where *in the hierarchical structures* they occur, and relate their occurrence to the overall consensus hierarchy. The answer is presented in Figure 2, where Theme occurrences are located in the hierarchical position in which they occur.

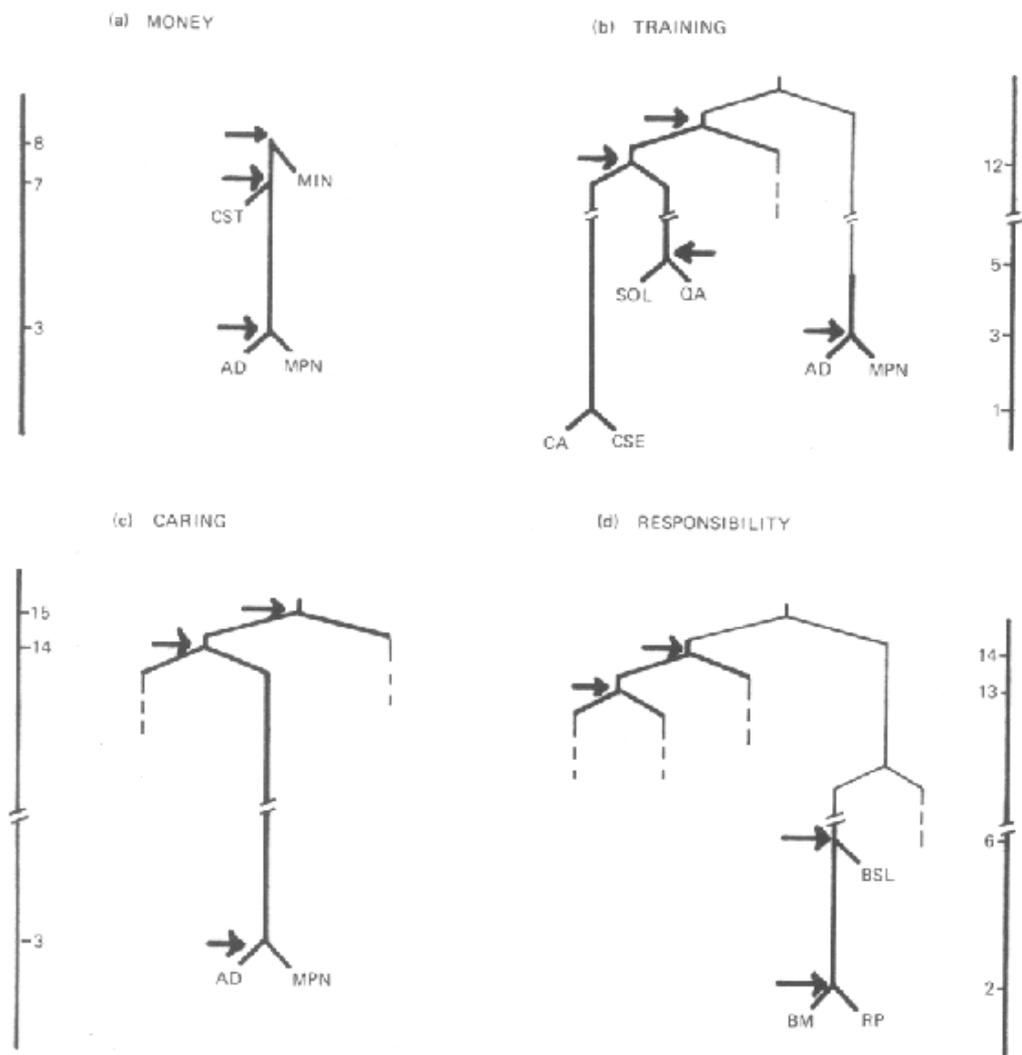


Figure 2: Local structure of thematic applicability⁶ [10]

6 COXON and JONES (1979) *Class and Hierarchy*, Ch4. Themes usage differs by context both *level* (of generality) and *subsumption* (instances of occupations to which it applies in the hierarchy). Arrows denote points at which a theme is mentioned or implied.

When archived, the Archive could not accept the original data and would accept only 4(A) data, having no facilities for storage of 4(B) or for lodging the bespoke programs.⁷ Consequently it is impossible for Archive users to access the "qualitative" data at all, and even if they could (in a Qualitative Data Archive?), they would be unable to relate it to the hierarchical content in which it occurred and thus replicate or extend the findings. [11]

3. Research example (2): Project SIGMA (Socio-sexual Investigations of Gay Men and Aids) 1982-95: Sexual Diaries

Project SIGMA (DAVIES et al. 1993; COXON 1996) consisted of a longitudinal study primarily designed to monitor gay and bisexual men's sexual behavior and lifestyle in England and Wales in the early days of the AIDS pandemic. In order to complement self-reports of sexual activity, the method of diaries was developed and adapted to give more detailed (and more accurate) information. These natural-language structured sexual diary data⁸ form the basis of this second example. It is similar to the first in that it involves structured data, but dissimilar to it in that the issues are different and diaries are more commonly regarded as qualitative data. [12]

The Project had developed a common schema for representing and analyzing sexual activity (COXON et al. 1992) and diarists were alerted to the necessary components when filling out their daily diary; see Figure 3/(1)/(FORM).

7 Some progress is being made in giving access to coded data. A project of QUALIDATA, Essex (Edwardians Online; technical paper <http://www.qualidata.essex.ac.uk/edwardians/about/online.asp>) is using XML format appropriate for interchange that will enable sophisticated online searching and information retrieval from encoded texts (any structural or content features of data, such as interview text), and which is potentially applicable to other qualitative datasets. It could usefully define and ensure a common archival standard/preservation format.

8 See full details at <http://www.sigmadiaries.com/>.

1: DIARY RECORD (Free text):

- (FORM) # The Time, The Place, The Partners (from partner list)
Then, describe the session in your own words
Remember to mention exactly what happened to the "come" (ejaculate) and always mention the use of condoms.
List any accompaniments you use (poppers, lubricants, drugs, sex toys...)
Mention how much you drank each day if it is associated with sex.

___Day We'd been drinking in a gay bar in Amsterdam, where we met. He was 30s, from Utrecht, into leather. We went back to my hotel room just after 1a.m. and after sharing a joint, went to bed ...
First we kissed, then I sucked him and he then sucked me and after that we moved into a '69' position and he came. Then I moved to fuck him and he put a condom on me with KY over it. I entered him (him sitting on me) and he was wanking himself, and we started using poppers and then I came in him and then he came over my chest. After pulling out, we kissed for a while and then went to sleep.

2. ENCODED ACTIVITY:

MDK AS PS MS,NM (AF,CN/I HW,NI)/p MDK.

n.b. This string is equally "textual data" and as much a sentence as a conventional linguistic sentence and can be analysed in the same way (e.g. KWIC, Co-occurrence, Theme ...)

3. ENCODED AND data-base structured:

=====

NO: CF887 |TYPE: II|STATUS: N |DAY: Fri |DATE: 26/08/88

TIME: 13.00 |PLACE: My hotel room, Amsterdam

PERSON: C3 [30s, male, Utrecht, into leather]

=====

ACT: MDK AS PS MS,NM (AF,CN/I HW,NI)/p MDK

POPPERS: Y |CONDOMS: Y |LUBS: Y (ky)

OTHER: Met in gay bar; smoking marijuana

DRUGS: Y

OTHER CODES:

=====

Figure 3: Three variants of the same textual (diary session) data [13]

A panel interview schedule contained questions which were constructed on the same principles as the diaries, so that subjects' accounts of their sexual behavior would be comparable between both methods. Thus "quantitative" (interview-based closed-ended questions) and "qualitative" (natural-language diary entries) data were designed to be complementary and integrated, at least for the purposes of comparative validity (COXON 1999). Because the structure is so specific, conventional software could not represent, let alone analyze, the diary data. Therefore the natural language data (Figure 3/(1)) had to be selectively encoded (Figure 3/(2)) and then entered in a flat database (Figure 3/(3)). Once again, special-purpose software SDA (Sexual Diary Analysis, see website cited in endnote vi) had to be constructed to analyze the data, though many of the analysis operations are common enough, such as identifying "word" stems, prefixes and suffixes and counting their occurrence. Even the more complicated analyses, such as looking at the contextual variation in risk activity, have clear parallels in qualitative data analysis. In part this is because the coding scheme was (deliberately) akin to language in its structure, with parallels for sentence

(sexual session), component words (sexual acts) and inflections (insertive/receptive modality; ejaculation). But despite our best endeavor, such software could not be persuaded to do more than the most basic operations, such as KWIC. More serious was the fact that there was no practical (low-cost!) way that the diary events could be related to the corresponding interview data. [14]

The current state of affairs is that the original data existing in anonymized microfiche (but not machine-readable) form, have been lodged via Qualidata at Wellcome Contemporary Medical Archives Centre, London. The interview data are lodged at the UK Data Archive and are due to be accessible via NESSTAR and the coded diary data languish un-lodged, but will in time be lodged in the UK Archive. Some non-trivial record linkage will make integrative analysis possible in the future.⁹ [15]

4. Some Conclusions

This paper is predicated in part on the hazards of taking the qualitative and quantitative distinction too seriously. Although the contrast certainly reflects a real enough methodological divide among social science practitioners, and software has been constructed to implement one side to the exclusion of the other, it can be a dangerous divide which militates against integrated styles of research and actually prevents integrated data analysis, pious platitudes about the importance of integrated research notwithstanding. [16]

Some problems and trials in implementing the integrated approach have been outlined in the paper. No-one, myself least of all, would claim that the examples reflect an extensive or even typical situation, but the argument is *a fortiori*—if a well-equipped sociologist had and has such difficulties in carrying out similar research (and archives have difficulty in lodging the data), how much more so the ordinary practitioner! Indeed, as I have said above, those proposing such an approach should think twice and weigh the cost (in every sense) before embarking. Whether data archives should take a pro-active role in this, I shall leave to others to argue; I am simply maintaining that hegemony of software producers will ensure a safe qualitative and quantitative divide continues to exist, and a reactive role will simply reinforce it. I have remained impressed by the finding emanating from a survey of computing-competent social scientists in the early 1970s (conducted by the UK SSRC), which found that the most common single activity engaged in was taking a data-set and writing programs to modify it, often repeatedly, to meet the requirements of different software packages. I have my suspicions that the situation has not changed much in the intervening years! [17]

⁹ We are grateful to the U.K. Medical Research Council grant G0001216 for making this possible.

At this point, two related issues remain:

- How does one best archive data that have already been collected? And the more serious and relevant challenge ...
- How can the teaching and practice of data collection and analysis (and archiving) be re-structured so that it *assumes* integration, or at the very least does not treat it as an eccentric practice of those who are old enough, or wise enough, to know better. [18]

These issues, of past and future significance, need to inform the promulgation of integrative, *and* qualitative and quantitative, research and its archived records. [19]

References

- Armor, David & Cousch, Howard (1969). *DATATEXT Manual*. Cambridge, Ma: Harvard University.
- Coxon, Anthony P.M. (1983). Subjects' Accounts and Occupational Predication. In G.-Nigel Gilbert & Peter M. Abell (Eds.), *Subjects' Accounts* (pp.15-23). London: Gower.
- Coxon, Anthony P.M. & Jones, Charles L. (1978). *The Images of Occupational Prestige*. London: Macmillan.
- Coxon, Anthony P.M. & Jones, Charles L. (1979a). *Class and Hierarchy*. London: Macmillan
- Coxon, Anthony P.M. & Jones, Charles L. (1979b). *Measurement and Meaning*. London: Macmillan.
- Coxon, Anthony P.M.; Davies, Peter M.; Hunt, Andrew J. & Weatherburn, Peter (1992). The Structure of Sexual Behaviour. *Journal of Sex Research*, 29(1), 61-83.
- Coxon, Anthony P.M. (1996). *Between the Sheets: Sexual Diaries and Gay Men's Sex in the Era of Aids*. London: Cassell.
- Coxon, Anthony P.M. (1999). Discrepancies Between Self-report (Diary) and Recall (Questionnaire) Measures of the Same Sexual Behaviour. *Aids Care*, 11(2), 221-234.
- Davies, Peter M.; Hickson, Ford C.I.; Weatherburn, Peter; Hunt, Andrew J. with Broderick, Paul J.; Coxon, Anthony P.M.; McManus, Thomas J. & Stephens, Michael J. (1993). *Sex, Gay Men and Aids*. London: Falmer.
- Johnson, Stephen C. (1967). Hierarchical Clustering Schemes. *Psychometrika*, 32, 241-254.
- Kaplan, Alan (1964). *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler.
- Kritzer, Henry M. (1996). The Data Puzzle: The Nature of Interpretation in Quantitative Research. *American Journal of Political Science*, 40, 1-32.
- Lerner, Daniel (Ed.) (1961) *Quality and Quantity*. Glencoe, Ill.: Free Press.
- Stone, Philip J.; Dunphy, Dexter C.; Smith, Marshall S. & Ogilvie, Daniel M. (1966). *The General Inquirer*. Cambridge, Ma: MIT Press.

Author

Tony Macmillan COXON

Present position: Honorary Professorial Fellow,
Graduate School of Social and Political Studies,
University of Edinburgh.

Major research areas: multidimensional scaling;
health studies (gay men and Aids).

Contact:

Tony Macmillan Coxon

Tigh Cargaman
Port Ellen

Isle of Islay
Argyll PA42 7BX
Scotland, UK

E-mail: apm.coxon@ed.ac.uk

URL: <http://www.tonycoxon.com/>

FQS 6(2), Art. 40, Anthony P.M. Coxon:
Integrating Qualitative and Quantitative Data: What Does the User Need?

Citation

Coxon, Anthony P.M. (2005). Integrating Qualitative and Quantitative Data: What Does the User Need? [19 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2), Art. 40, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0502402>.