Qualitative Sociology, Vol. 19, No. 2, 1996

Coder Training: Theoretical Training or Practical Socialization?

Tony Hak and Ton Bernts

Usually the effectiveness of coder training as a means to improve the inter-coder-reliability of the coding of responses to open-ended questions is considered a result of (successfully) communicating the coding scheme to coders. However, the actual practice of coder training has never been studied empirically. In this article we present fragments of a transcript of a coder training that suggest that inter-coder-reliability is improved not only by communicating the coding instructions to coders (theoretical training) *but also by* socializing *coders into practical rules which are not part of the coding instructions and are not warranted by them. Hence, it cannot be excluded that the improvement of the inter-coder-reliability by means of coder training is a* training artifact: *an artificial outcome affected through the training process. It follows that, in each particular case, the researcher must make plausible that the actual coding process has yielded valid data.*

1. INTRODUCTION

Coder training is a standard procedure in sociology, in particular when a method of content analysis is used. Although methodological studies and textbooks on methods provide instructions regarding coder training, there is no *empirical* evidence of what coder training consists of in practice. This lack is the corollary of another gap in our empirical knowledge of sociological practice, namely coding itself. Little is known of coding problems encountered and solved in everyday sociological practice. The literature describes coder training primarily as a procedure for improving the study's

Direct correspondence to Tony Hak, Sociology, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

inter-coder-reliability by means of theoretical instruction and practical advice. It is doubtful, however, whether these instructions provide the basis for the actual practice of coder training. This doubt arises from the scarce empirical literature on the coding process (Cicourel 1964; Garfinkel 1967; Katz and Sharrock 1976; Rehbein and Mazeland 1991) which suggests that coding decisions are very complex interpretive activities that can be learned only to a limited degree through training. Although it seems reasonable to suppose that some knowledge about the coding process must be available to the trainer before a training can be provided, in fact this knowledge does not exist. Coding is, in the words of Rehbein and Mazeland (1991, p. 166), "a very specific form of linguistic action of which almost nothing is known. Often it will be practiced intuitively. It is, thus, a complex blind spot in the analysis of scientific communication". It is the substantive focus of this paper to describe coding and coder training as 'a specific form of linguistic action'.

Describing some fundamental features of coding practices leads directly to a criticism of how coding and coder training are dealt with in textbooks on methods and in sociological practice. Lack of systematic knowledge of coding problems and how these are solved relates intrinsically to a lack of knowledge of how the quality of the coded data can be assessed and improved. The effectiveness of coder training as a means to improve the inter-coder-reliability is usually considered a result of (successfully) communicating the coding scheme to coders. In this paper we present fragments of a transcript of a coder training that demonstrate that inter-coder-reliability improves not only (or even not primarily) through communicating the coding instructions to coders *(theoretical training)* but also by *socializing* coders into practical rules which are not part of the coding instructions and hence are not warranted by them. This finding suggests that, in practice, coding decisions are less well controlled than generally assumed in research reports. This is the methodological focus of this paper.

2. TWO LEVELS OF INTERPRETATION

The substantive focus of this paper derives from our reading of the scarce empirical literature on coding practices (such as Cicourel 1964; Garfinkel 1967; Katz and Sharrock 1976; Rehbein and Mazeland 1991) which suggests that coding is a very complex interpretive activity that can be learned only to a limited degree through training.

While observing coders coding clinic records, Garfinkel (1967, p. 20) saw them

> assuming knowledge of the very organized ways of the clinic that their procedures were intended to produce descriptions of. More interestingly, such presupposed knowledge seemed necessary and was most deliberately consulted whenever, for whatever reasons, the coders needed to be satisfied that they had coded 'what really happened'. This was so regardless of whether or not they had encountered 'ambiguous' folder contents.

Thus, in order to be able to do their work, coders assume an order 'out there' of which the data is an appearance but that is not available to them, neither in the data to be coded nor in the coding instructions. In another context, not of coding clinic records but of attributing sense to utterances in a conversation, Garfinkel (1967, p. 40) observes that

> sense cannot be decided by an auditor unless he knows or assumes something about the biography and the purposes of the speaker, the circumstances of the utterance, the previous course of the conversation, or the particular relationship of actual or potential interaction that exists between user and auditor. The expressions do not have a sense that remains identical through the changing occasions of their use.

The coder must 'see the system' in the data in order to be able to maintain the relevance of the coding instructions and the correspondence between these data and the codes that must be assigned. Garfinkel (1967, p. 20) concluded from his observations of coders' practices that "agreement in coding results was being produced by a contrasting procedure with unknown characteristics". By studying the procedures that made up these practices, Garfinkel (1967, pp. 20-1) soon found "the essential relevance to the coders, in their work of interrogating folder contents for answers to their questions, of such considerations as 'et cetera', 'unless', 'let it pass', and 'factum valet' (i.e., an action that is otherwise prohibited by a rule is counted correct once it is done)". Garfinkel calls these considerations *ad hoc* considerations and the coders' use of them *ad hocing*. He reports (1967, p. 21) that attempts to suppress *ad hocing* while retaining an unequivocal sense to the instructions produced bewilderment on the part of the coders.

Katz and Sharrock (1976) have presented data taken from a transcript of a conversation between a coder and a researcher in which the coder explained to the researcher which answers he experienced as difficult to code and how he has resolved these difficulties. Their conclusions confirm Garfinkel's observations (Katz and Sharrock 1976, pp. 269-70)

> The coder's presumption that he has a mastery of the language enables him to locate problematical instances of talk and to resolve them by assuming that they are, in the end, further instances of talk in that language over which the coder has a mastery. [.....] The assumed mastery of the language is embedded in an assumed competence in the community and knowledge of the social world that generates the talk.

It must be recalled that Cicourel has devoted much of his work, in the 1960's in particular, not only to confirming and detailing this kind of ob-

servation but also to determining the consequences for the practice of sociology. He attempted to improve that practice by requiring the sociologist to explicate the "normal and taken-for-granted cultural information" (Cicourel 1978, p. 255) that is used in analysis. It is important to note that ethnomethodologists usually do not agree with Cicourel. Katz and Sharrock, for instance, state explicitly that they

> are not at all interested in taking a 'corrective' attitude toward coding here. We are neither concerned to criticise coding as a way of doing sociological research nor to evaluate the things those we are studying do as examples of 'good' or 'bad' coding practices according to any ideal standards governing that activity. Our concern is entirely in understanding how those whose activities we are examining carry things off as, for them, satisfactory ways of going about the work of coding (Katz and Sharrock 1976, p. 246).

We fail to see, however, why the descriptions of Katz and Sharrock, or our results for that matter, should not be used as arguments for advocating that a far more careful attention be given to the manner in which data are coded, and that the researcher be "much more explicit and self-conscious than is customary in making available to his audience the context and grounds for his interpretations" (Wilson 1970, p. 706).

The problem for coding and for coder training, as documented in the literature reviewed here, is summarized by Mishler (1986, p. 4) as follows:

> Because meaning is contextually grounded, inherently and irremediably, coding depends on the competence of coders as ordinary language users. Their task is to determine the 'meaning' of an isolated response to an isolated question, that is, to code a response that has been stripped of its natural social context.

Having noted that there is considerable individual variation in frames of reference, values, and levels of understanding, Mishler states that

> the actual work of coding cannot be done reliably until coders build up a set of shared assumptions, specific to the study, that allow them to implement the code in a mutually consistent way. The development of such a coders' subculture is the most significant by-product of training and periodic reliability checks of coders. Often these assumptions are ad hoc, reflect coders' everyday understandings and competences as language users, and tend to remain tacit in the research process (Mishler 1986, p. 4).

It is this coders' subculture, which usually remains tacit in the research process, that we want to discover and describe in the case study presented here.

3. THE QUALITY OF CODED DATA

The quality of assigned codes—for answers to open-ended questions in our case; see below — depends on two values: validity and (inter-coder)

reliability. The *validity* of the coding process is the degree to which the theoretically relevant features of the answers are represented in the codes. The quality of the relation between the answers and the codes, however, cannot be discovered in a way that is independent of coding itself. It is precisely the aim of coding to establish whether, and how, a theoretically relevant feature is observable in the answer. This means that the validity of the coding process and of the resulting codes cannot be determined in an independent way. Validity is a matter of *argumentation.*[1]

In contrast to validity, (inter-coder)*reliability* can be assessed easily, without any reference to the substance of the coding process. It is possible to assess unequivocally whether a coding of an answer is identical to another coding of the same answer, whether done at another time by the same coder or at the same time by another coder. Inter-coder-reliability can, thus, be established and discussed without any reference to actual features of the coding process. Reliability is a matter of *computation.*

Usually the validity of the coding process is considered to be dependent on the quality of the coding *instructions,* i.e. on the quality of the researcher's argumentation, whereas inter-coder-reliability is considered as being dependent on coders' *application* of these instructions. According to the literature, inter-coder-reliability can be improved by such means as coding by pairs of coders, the development of more detailed instructions, selection of professional coders (e.g. the researcher's colleagues or graduate students) and, last but not least, coder training. It is remarkable, however, that this literature gives hardly any instruction about the way a coder training must be organized in order to have the required effect.[2] When the inter-coder-reliability has been improved through a coder training, this result is usually interpreted as an effect of the improved quality of the study's concepts and their operationalization, *and* of the coders' improved understanding of the meaning of the concepts. This improvement is considered as the result of *theoretical* training resulting in a more consistent interpretation of the coding instructions. This is, to give only one example, the presupposition of the following remark in a textbook: "Any discrepancies will indicate an imperfect communication of your coding scheme to your coders" (Babbie 1989, p. 360).

Two comments can be made here. First, the literature does not recognize sources of coding discrepancies other than coders' inability to interpret the instructions adequately. It, thus, presupposes that it is possible, even necessary, that the researcher can make the coding instructions sufficiently transparent for allowing an unequivocal coding of each answer. This presupposition, however, neglects the fact that coders not only must interpret the coding instructions but the answers as well. Second, although this is less obvious, it is assumed that the absence of discrepancies in codes

(i.e., when inter-coder-reliability is high) guarantees that the assigned codes are warranted by the coding instructions. The fact that two coders assign the same code is assumed to be a sufficient guarantee that the researcher himself, coding the same answer according to his instructions, would have chosen that same code. Having the same coding result is, thus, considered as evidence of theoretical soundness. In this way inter-coder-reliability functions implicitly as a measure of validity: when the coding *instructions are valid* (that is, when the instructions are theoretically warranted) and when the coding process produces *reliable codes,* it is assumed that these reliable codes are the result of a *valid application* of the instructions.

It is, however, quite possible that a high degree of inter-coder-reliability is the result of a training artifact in the sense Muskens (1980, pp. 124-6) gives to the term

> Artifacts are outcomes of those practices, inherent to a variety of distinct levels of a research design and its execution, which lead to distortions in the data and thus to unmonitored conclusions about reality: the research design and its execution themselves cause certain measurement results and relationships, not the reality studied (cf. Rosenthal and Rosnow 1969). The term training artifact covers those results of content analysis where findings about the texts analyzed are caused by inadequate selection and training of coders. [.....] One can imagine that at a certain point, [coders] no longer judge on the grounds of a *'nature' judging skill,* but do so as *'brainwashed' automatons* reacting to stimuli and accurately assigning codes in the way they expect the researcher would like them to. [.....] Whether or not in a concrete example of research one can speak of a training artifact cannot be confirmed. The training artifact, i.e., its influence on the research data, is a *qualitative* question. One will, [.....], have to make *plausible* the idea that the application of content analysis is not distorted by any training artifact. Only then it is useful to calculate coefficients of agreement and to draw conclusions on the basis of their values concerning the reliability of the research. Because, however high the reliability expressed in coefficients of agreement is, a *training artifact* yields *invalid data.*

The possibility, hypothesized by Muskens, that the ultimate agreement between coders effected by a particular process of selection and training is an *artifact* and yields *invalid data,* forms our methodological focus for the following empirical case study of what coder training in practice consists. Our question is whether in this case inter-coder-reliability is improved by theoretical instruction only (which is the received wisdom of textbooks though it has not been studied empirically), and whether this is done in a way that is theoretically sound. Our assumption is that a description of what really happens in a coder training will show that other effective means to improve inter-coder-reliability are used which, however, cannot be warranted in terms of the theory that underlies the coding instructions. In other words, we expect to find that coder training is an occasion in which coders are socialized into practical rules which are not part of the coding instruc-

tions and hence are not warranted by them. Our research question thus is:

Does coder training consist only of training in the correct application of the coding instructions (theoretical training) or also of coders' socialization into practical rules which are not warranted by the coding instructions?

4. METHOD OF THIS STUDY

In the course of a study of opinions on equity in health care (Bernts 1991) the authors assisted in the coding of answers to an open-ended question. Bernts' study consisted of a survey of a representative sample of Dutch households. Respondents were asked to answer a series of (closed-ended) questions on preferences regarding sanctions in health care (e.g., differential treatment of smokers and non-smokers) and attitudes to health, health risks and health care. The hypothesis was that attitudes would explain preferences. One of the attitudes assessed was the *definition of health,* a variable that was conceived as an indication of the *demand* side of health care. It was expected that definitions of health as a means to personal creativity *(expressive* definitions) would be correlated with a low degree of risk-solidarity in health care (e.g., by differentiating health insurance premiums according to lifestyle), whereas *instrumental* definitions of health (as a means to being able to do one's work) would be connected with high risk-solidarity.

The respondents' definitions of health were measured by means of an open-ended question derived from a French study (d'Houtaud 1978). The (Dutch) question was a translation of its English version: "What is, according to you, the best definition of health?" (d'Houtaud and Field 1984, p. 34). It was expected that the answers could be coded unequivocally as either belonging to the class of expressive definitions of health or to the class of instrumental definitions. However, an initial attempt by Bernts to categorize answers as either 'expressive' or 'instrumental' failed, mainly because many answers appeared to consist of multiple parts whose sense could differ in terms of expressivity and instrumentality. Therefore, he developed a multidimensional coding scheme (see Appendix) in which definitions of health are considered as answers to three questions:

— how does one perceive health? (CRITERION)[3]
— how does one get or maintain health? (MEANS)
— what is the value or the aim of health? (VALUE)

A respondent's statement may entail an answer to all three questions, but often it answers only one or two of them. Within the broad category of CRITERION, Bernts distinguishes three aspects: SIGN (is the definition positive, negative, or both?), POINT OF VIEW (is it subjective or objective?), DOMAIN (is it physical, or more than that?).

It is important to note that these categories and their classes have a direct relation to the theory on 'instrumental' and 'expressive' definitions of health, which is explicitly discussed by the researcher in a note to the instructions (see at the bottom of the Appendix): "A 'positive' SIGN, a 'mental' DOMAIN, a 'subjective' POINT OF VIEW, a 'harmonious' MEANS and a 'growth' VALUE are taken as indicative of an *expressive* definition of health. A 'negative' SIGN, a 'physical' DOMAIN, an 'objective' POINT OF VIEW, a 'preventive' MEANS and a 'work' VALUE are taken as indicative of an *instrumental* definition of health."

The training procedure was as follows. The researcher first gave the coders the instructions with no formal introduction. The researcher and two coders then coded about 40 (out of 800) answers in order to test their ability to use the coding instructions. It appeared that many answers were assigned different codes. This presented sufficient reason for the researcher to organize a meeting in which these differences were discussed. In that meeting, the coding instructions were explained and additional coding rules were formulated. A second coding of the same answers, however, resulted again in many differences between coders. In a second meeting even more detailed instructions were formulated. This resulted eventually in a sufficiently high inter-coder-reliability (more than 95% identical codes between pairs of coders, over all 800 answers). The process described, consisting of two rounds of coding each followed by a meeting in which coding decisions and coding instructions were discussed, has all the features that, according to the literature, are characteristic for coder training. Therefore, an analysis of these two meetings (which both were tape recorded and transcribed for analysis) can illuminate the practice of coder training.

We are interested, first, in how coding discrepancies are discussed and how agreement is sought and achieved. Second, we are interested in whether achieved agreement (on the correct coding of a specific answer) affects how other answers will be coded, either because the agreement entails a clarification or a specification of the coding instructions (theoretical training) or because it implies the application of rather informal rules (practical socialization). Our starting-point is that the purpose of this coder training was to improve the inter-coder-reliability and that this was achieved. Our effort then is to determine how this was achieved and

whether the way it was achieved had effects pertinent to the theory that underlies the coding instructions.

5. STRATEGIES IN CODER TRAINING

Our assumption that only matters of inter-coder-reliability (and not matters of theoretical soundness) would be discussed in the coder training, was confirmed in the transcript. When the three coders had coded an answer in the same way, the code itself was never a matter of discussion. Extract 1 gives some examples.[4]

Extract 1

Researcher:	uh 2502 .. is a .. 3
Coder 1:	yes
Coder 2:	that's what I have too
Researcher:	2511 is a 4
Coder 1:	yes
Researcher:	2522 is a 2 of MEANS
Coder 1/2:	yes
Researcher:	3041 is a 3
Coder 2:	yes
Coder 1:	yes

This extract shows that the quality of a code is not a matter of discussion when coders agree on what the correct code is. Coders only check whether they agree; they do not check, at least not explicitly, whether the agreed code is warranted by the coding instructions. In our transcript we do not find any exception to this rule that there is discussion between coders only in those cases in which at least one coder had coded differently. This documents our assumption that an apparently reliable code is considered to be theoretically correct.

5.1. The Interpretation of Answers

There is a necessary preliminary stage in the coding process which is not covered by the coding instructions, a stage in which an answer's 'sense' or a respondent's 'intention' is to be assessed. A coder's reading of an answer's sense cannot be controlled, because it is dependent on the coder's (non-professional) lifeworld. Interpretive problems in this stage cannot be solved by consulting the coding instructions, nor for that matter by the theory underlying the instructions. When the coders encounter a discrepancy in their coding of an answer, the first question attended to is how the difference can be explained. This implies that the coders must give accounts for the codes they have given. These accounts

will be discussed, and that discussion will eventually lead to an agreement between coders on the correct reading of the answer's sense. There is, however, no coder *training* in a strict sense. In these cases, there is neither a specification of the instructions nor a formulation of rules for future cases. The reason is that there is no other useful device available than members' everyday methods of making sense, which hardly can be influenced by instruction.

An example is presented in Extract 2. The answer discussed here is "Not going to the doctor too often, determining yourself what is good and what isn't", and the problem is that the coders do not agree on what this answer *means*. The researcher and coder 2 interpret the answer as a description of a MEANS to improve and maintain health. Coder 1, in contrast, refers to everyday talk in which 'going to the doctor' is seen as a CRITERION of 'having complaints'. In the discussion, the researcher accounts for his code by referring to everyday talk as well: "For instance there is the saying 'if you go to the doctor too often they talk you into all sorts of disorders' isn't it?" In his interpretation, 'not going to the doctor too often' is not a document of the absence of complaints (a CRITERION) but rather a MEANS to maintain health. Thus, both refer to everyday talk in order to account for their interpretation of the answer:

Extract 2

Researcher:	this is a 1 of MEANS . harmony with the environment
Coder 1:	I have something quite different
Researcher:	yes what is your code?
Coder 1:	I had a ... uh .. CRITERION because I consider 'not going to the doctor too often' as absence of complaints and 'determining yourself what is good and what isn't' I consider as uuuh vitality so that is VALUE 1
Researcher:	well I think that it is dubious .. I think that it is both a CRITERION and a MEANS because for instance there is the saying 'if you go to the doctor too often they talk you into all sorts of disorders' isn't it
Coder 1:	yes but .. 'determining yourself what is good and what isn't' .. you read it as not allowing the doctor to determine what happens
Researcher:	precisely
Coder 1:	yes yes .. oh yes then then I do not consider it as a CRITERION any more no .. no then I consider it a MEANS indeed yes ... yes I agree completely on that .. then it becomes clearly something different from 'not going to the doctor too often'
Researcher:	yes so it is a MEANS .. yes?
Coder 1:	yes

Eventually the researcher's interpretation is accepted by coder 1. The reason is not that there is a better fit between the researcher's quote ("They talk you into all sorts of disorders") and the answer 'not going to the doctor too often' than between the coder's quote and the answer. The superiority of the researcher's proposal is that it allows for assigning only one code to the answer (coder 1: "Yes but .. 'determining yourself what is good and

what isn't' .. you read it as not allowing the doctor to determine what happens").

This extract, thus, shows how an initial difference of interpretation has been resolved. The means used to establish agreement consisted of (a) paraphrasing the answer in such a way that both parts of the answer could be read as versions of one underlying pattern ("Doctors talk you into all sorts of disorders") and (b) claiming that the underlying pattern can be attributed to respondents ("There is a saying .."). It must be emphasized that analytically this disagreement and its solution have no relation whatever to the coding instructions. Although the problem can only arise because of the requirement to make a distinction between two kinds of definitions of health, MEANS and CRITERION, the problem is not one of the application of these categories. The problem is rather one of determining the sense of an answer, which is independent of the coding process proper. This explains why the presented discussion neither resulted in a specification of the coding instructions nor in a formulation of rules for future cases. The problem is, thus, considered as restricted to that particular answer only, and hence as unique.

This confirms Garfinkel's (1967, pp. 19-20) observation that coders must *ad hoc* in order to determine an answer's sense. *Ad hocing* consists of finding an *(ad hoc)* rule that in one way or another can account for the decision made. Because that rule cannot be found in the coding instructions, it is the coder who must provide it. In the example above the coding decision is accounted for by invoking the *ad hoc* rule that an answer is coherent and consistent and that, therefore, one code for the answer is preferable to two codes, one for each of the two parts of the answer. This is the only reason why the paraphrase "If you go to the doctor too often they talk you into all sorts of disorders" is preferred. It is clear, however, that the application of this economy rule is neither part of the coding instructions nor can be grounded in the respondent's intentions.

It goes without saying that the problem that an answer's sense is indeterminate is partly due to the fact that answers (in this study) have been isolated from the context in which they were given, i.e. the setting in which the respondent has answered the questionnaire. That the coders are conscious of the absence of the natural context, is apparent from the fact that coders sometimes explicitly mention that answers could have been interpreted more easily if more information had been available about the respondent.

An example is the discussion in Extract 3 about the interpretation of the answer "Feeling well and being able to move easily". It is discussed whether 'being able to move easily' is only a CRITERION ('being able to move well' as a document of good health) or a VALUE ('being able to

move well' as an aim for which one needs good health) as well. In this discussion it is mentioned that knowing the respondent's age would have been of help in determining the sense of the answer. Similarly to the previous example, the problem is solved by choosing the alternative by which the complete answer can be covered by one code, although the researcher considers this a low quality solution. He would have preferred a solution in which the respondent's age had been decisive.

Extract 3

Researcher:	this is a CRITERION ... 'feeling well and being able to move easily'
Coder 1:	yes ... I have put also a VALUE 1 ... look ... similar to 'hard work' isn't it ... and so here 'to move' ... we have decided that hard work that if .. if the category VALUE can be used too ...
Researcher:	shall I ... shall I just look what is the respondent's age ((laughing)) if he is 75. then it probably must be a CRITERION
Coder 2:	((laughing loudly))
Coder 1:	I consider it a CRITERION .. but I think .. the point is . is it a requirement to use the category VALUE
Researcher:	yes I see, but if it is a CRITERION, 'being able to move easily,' then it becomes clear ... so it is meant explicitly as uh that one is able to rise from a chair easily and the like
Coder 2:	((still laughing))
Coder 1:	o yeah yeah you mean yeah
Researcher:	'being able to move easily' ... it may mean ... I am still able to walk to the cupboard without too much pain
Coder 1:	yes
Researcher:	or does the respondent mean moving easily ... in social relations ... so it is necessary to know whether the respondent is 25 ... such a person would probably mean another thing than a person of 65
Coder 1:	I do agree that strictly speaking it is not uh ... VALUE ... it is an addition to the CRITERION
Researcher:	I too, I think that it is meant that way
Coder 1:	yes
Researcher:	let's make a joke, let's look what is his age ...
Coder 1:	yes ... I agree .. but don't do that ... then you introduce circular reasoning . then you will find later on that a CRITERION is mentioned more often by the elderly . yes that is only because you have coded it that way

In the examples given the researcher does not attempt to instruct the coders about how to determine the sense of answers. The obvious reason is that there is no other useful device available than members' everyday methods of making sense, which hardly can be influenced by the researcher in a coder training. In these examples, therefore, there is no coder *training* in a strict sense. The only thing achieved, besides an agreement on the coding of these answers, is that it is demonstrated that an economy rule can be used in order to improve inter-coder-reliability. Decisions in these kinds of cases can neither be influenced by instructions nor by training. Because different readings of the same answer are equally defensible, there is no way out of *ad hocing*.

5.2. Theoretical Training

The coding process consists of two different stages. After the stage in which an answer's sense is determined (discussed in section 5.1. above), this sense has to be connected to a coding category. Whereas the everyday interpretive means used in the first stage cannot be influenced (let alone be improved) by instructions, the procedures used in this second stage are open to training. In the extracts presented below it is shown how the discussion functions as 'training'. The theory underlying the coding instructions is explained and the coding instructions themselves are specified. Because these specifications are the joint product of the coders, it is likely that they will be applied in a more or less consistent way by different coders. This results in an improvement of inter-coder-reliability. The specifications are correct to the degree that the researcher relates the specifications to his theory. Thus, the application of coding instructions to the 'sense' of answers can be 'trained' and this can be done in a way that does not damage, but instead can enhance, the quality of the coding process. Take the following example.

We have seen in Extract 2 how the researcher and coder 1 agreed on interpreting the answer "Not going to the doctor too much" as the expression of a MEANS to maintain health. But, which of the codes within the category MEANS has to be chosen now? Extract 4, which is the continuation of Extract 2, shows how the coders find a solution to this problem. The coding instructions provide for three kinds of MEANS, 'harmony with the environment', 'prevention/hygiene' and 'other means'. Coder 2 denies that the advice to avoid doctors fits in the category 'harmony with the environment'. He proposes the category 'other means' instead. Coder 1, in contrast, attempts to support the researcher's interpretation by finding a way of connecting the answer with the category 'harmony': "With harmony is meant that the environment is adapted to the health aim as well". But he admits immediately that "there is no environment here". The researcher takes this as an occasion for explaining what the coding instructions mean: "Adapting the environment is a kind of autonomy [.....] it is sort of meant that way". He gives an example of autonomy (having less work) which he apparently considers to be an example of adapting the environment and hence of 'harmony with the environment'. Coder 2, however, denies the applicability of this reasoning because he reads the answer "Not going to the doctor too often" as merely 'criticizing' doctors. In his view 'harmony with the environment' presupposes rather 'positive' aspects such as 'having contact with people' and 'a stroll in the country'. He admits that both 'self-determination' and 'harmony' have a positive value, but this does not imply that self-determination can be considered a form of harmony.

Extract 4

Coder 2:	but what kind of MEANS? in that case I would choose 'other means' uh ...
Coder 1:	yes but you can also uh for instance interpret it in Illich's terms as safeguarding autonomy and then it may be something more of ... with harmony is meant that the environment is adapted to the health aim as well... yes that's here ... but strictly speaking there is no environment here, is there?
Coder 2:	no I choose 'other means' ... it is merely 'other means' ... it is a kind of criticizing the medical class yes
Researcher:	well adapting the environment is a kind of autonomy, isn't it? you are right it is sort of meant that way
Coder 1:	yeah yeah uh
Coder 2:	yes but I do not see harmony with the environment
Researcher:	and that one says as well for instance I would like to have less work or ... I understand ... that you...
Coder 2:	but harmony with the environment can only be concluded from rather positive .. things such as uh having contact with people or things like that .. maintaining health .. by uuuh a stroll in the country or things like that? ... rather than by criticizing doctors?
Coder 1:	yes but self-determination is of course a very positive value
Coder 2:	yes but does this necessarily mean harmony or even harmony with the environment ... self-determination?
Researcher:	well it uh it uh it uh the definition is difficult indeed
Coder 2:	look he is determined .. it is a determined person that that is clear

This example shows that, even when the 'sense' of an answer has been determined (see 5.1.), there is still the problem of how this sense can be related to the abstract categories of the coding instructions. The solution of this problem depends on the successful interplay of two procedures, a bottom-up procedure in which the answer can be shown to be a member of a larger class of statements and a top-down procedure in which the coding categories must be specified in terms of more concrete subcategories. In Extract 4, coder 2 uses the bottom-up procedure. This results in more general categories such as 'criticizing doctors' and 'determined person'. In contrast, coder 1 attempts to find specifications of the coding category in order to reach a point at which the answer can be read as one of its possible specifications. Both fail because they cannot find a way to connect the bottom-up generalizations ('criticizing', 'self-determination' and 'determined person') to the top-down specifications (such as 'adaptation of the environment').

After an extensive discussion (which is not presented here) about the correct interpretation of the code 'harmony with the environment', coder 1 formulates an interpretation that eventually allows them to find a solution:

Extract 5

Coder 1:	prevention and hygiene is functional and this is everything with a rather intrinsic value
Researcher:	yes
Coder 2:	I like that word intrinsic value

Researcher:	okay
Coder 1:	it may be stated in a negative form as well .. not doing exaggerated exercises
Researcher:	yes you see also that that ... bipolar structure ... the fact is that I want to have it in a sense because uh ...
Coder 2:	if you take this formulation I will agree ... intrinsic value and uh autonomy and that kind of things

This discussion is eventually concluded with the agreement that 'harmony with the environment' must be read as 'intrinsic value' (as opposed to 'functional' or 'instrumental'). Both the agreement on the coding of this answer and the general agreement on the correct reading of the category 'harmony with the environment' have been achieved by the introduction of the one word 'intrinsic'. Its achievement is having bridged the distance between the words 'harmony' and 'determined person' in a way that appears to be acceptable to coder 2. This solution is even more than acceptable to the researcher because it accentuates an underlying structure of the coding instructions: "That bipolar structure ... the fact is that I want to have it in a sense". Although it is not formulated explicitly, the researcher's satisfaction certainly derives from the bipolar structure of his research question in which only two types of definitions of health figure, 'expressive' and 'instrumental'.

There is an interesting difference between how the agreement between the coders was achieved in this discussion about the appropriateness of a coding category and how it was achieved in the discussion about the answer's sense (in 5.1.). We have seen that the problem of determining the answer's sense was treated as a unique problem of which the solution had no bearing on other cases. The discussion about the interpretation of the coding category 'harmony with the environment', in contrast, has resulted in a specification that can be used in other cases. The chance that this category will be applied by coders in a more or less consistent way will be increased, which results in an improved inter-coder-reliability. Moreover, because the researcher had an active part in the discussion, in which he explicitly referred to the underlying theory (particularly in parts of the transcript not presented here), the fit between this interpretation and the researcher's theory and intentions is safeguarded.[5]

5.3. Practical Socialization

In this section we discuss other rules of applying coding instructions which are aimed at improving inter-coder-reliability. We refer to rules that are developed in the course of the coder training, but are not grounded in the coding instructions. By applying these kinds of rules, agreement on coding decisions can be achieved without grounding them in a specific in-

terpretation of the theory. Because these rules are not a part of the coding instructions proper, their use is usually not accounted for in the research report. Typically, these rules specify how an answer can be coded by taking only its form (rather than its 'sense') into account. These kinds of rules short-circuit the problems discussed in the previous sections (5.1. and 5.2.). If the answer's form is the only relevant criterion for the code it will receive, it is not necessary any more to determine the answer's sense nor to connect this sense to a coding category interpretatively.

It is clear that this manner of non-interpretative coding improves inter-coder-reliability. Because, however, coders are not machines, they often cannot resist reinterpreting these rules. Thus, in applying these rules coders will appear to be *ad hocing* again. In Extract 6 the application of the rule that the presence of the phrase 'be(ing) able to' is a sufficient condition for coding the answer as a VALUE is discussed. This discussion started with a disagreement on whether this rule should be applied to the answer "Feeling fine in one's body and being able to do normal exercise". In Extract 6, coder 1 extends this discussion to the coding of the answer "Feeling well and being able to move easily" (discussed also in Extract 3 above).

Extract 6

Researcher:	I think that .. if the word 'being able to' occurs in the answer we can almost always choose VALUE if it is possible .. uh .. to read the answer in that way
Coder 1:	thus in that case we don't take CRITERION
Researcher:	yes ... yes only 'feeling fine in one's body' is the CRITERION ..
Coder 1:	then the same would apply to 'feeling well and being able to move easily'
Researcher:	uh yes 'being able to move easily' .. it was ambiguous .. yes .. does it mean moving socially or really .. I mean uh .. physical movement .. being able to move knees and legs and the like
Coder 1:	yes I think .. I agree that one can read it as moving socially but it is also very difficult to consider it not a CRITERION
Researcher:	yes I would say moving easily is .. being able to move is an experience as well

This extract documents two cases in which the coders hesitate to apply the rule that 'being able to' must be coded as a VALUE because it conflicts with another (more informal) rule, namely the economy rule that two codes for one answer must be avoided. Applying the economy rule would result in not coding the answer as a VALUE. Instead the code CRITERION would be chosen in both instances. In order to account for this kind of decisions an additional rule was formulated in the subsequent discussion (not presented here): "The phrase 'being able to' will be coded as a VALUE, unless it can be argued convincingly that it is not a VALUE, e.g. when 'being able to move easily' must be seen as an expansion of 'feeling well" '. Although its formulation ('argued convincingly') suggests that the quality of the code prevails, in practice this rule boils down to merely giving

priority to one formalism (the economy rule) rather than another (the standard code for 'being able to').

Extract 7 documents a similar discussion about the correct coding of "feeling well". It is clear that it must be coded as CRITERION but the question is what kind of CRITERION, 'subjective' or 'objective' or both? The justification of the rule that well-being must be coded as 'objective' is questioned because it is seen as counter-intuitive. The result of the discussion is that the rule is maintained and even has received a validation: "We must assume that 'feeling' is not at stake here".[6]

Extract 7

Coder 1:	you say it is an experience too and hence subjective .. but in the instructions you cite well-being as an example of objective
Coder 2:	is well-being uh uh .. I have coded it as subjective all the time
Researcher:	yes .. mental and physical well-being .. I think it is the formulation of a kind of norm ... is it a mere subjective experience and hence a subjective criterion or is it a .. a definition that is applied socially .. someone is healthy if he feels well [zich welbevindt] .. it is not very clear how well-being can be determined .. but yes uh
Coder 1:	yes but I cannot see a difference with uh being able to move .. then
Coder 2:	I made a mistake .. I have coded well-being as subjective all the time .. it was not clear for me
Researcher:	wait a moment .. 'a state of mental and physical well-being' that's a typical .. is it subjective? .. I would choose objective uh .. an objective criterion .. because this criterion applies to others as well whereas .. 'I like the feeling' .. that's yes .. yeah ..
Coder 2:	you consider mental and physical well-being as objective?
Researcher:	yes it is 'being' isn't it .. being well
Coder 2:	yes okay
Researcher:	yes or do you think that it ...
Coder 1:	no I think in this case uh uh .. literally it says 'feeling' [bevinden] but that's only .. we must assume that 'feeling' is not at stake here
Researcher:	yes uh it's a decision .. we could make three categories but then ... we have only few cases .. at a certain point we must .. let's attempt to make a kind of dichotomy .. I have chosen to consider well-being .. to locate it on the more objective uh side

In a similar way the coders discussed the rule "Apply the category 'body' if, and only if, the 'body' or a synonym is mentioned explicitly". According to one of the coders it is 'clear' that many answers refer to the body even if the body is not mentioned explicitly. According to the researcher, however, "[I]t is interesting to look at explicit statements, because otherwise you can read the body in any answer. By looking at whether it is mentioned explicitly we can differentiate the answers". This is another case in which an *a posteriori* validation is provided for a practical rule which is introduced only in order to improve inter-coder-reliability by avoiding the interpretation of answers. The application of this rule implies that a new research question is introduced: "It is interesting to look at who mentions the body explicitly". It cannot be denied that this might be an interesting research

question. The problem here is that this was not a research question at the beginning of the coder training. The original research question referred to 'expressive' and 'instrumental' definitions of health. The researcher fails to clarify how the coding of the *explicit* mentioning of the body relates to the concept of health definition that he wanted to measure with this open-ended question in the first place.

Characteristic of practical rules as discussed in this section is that they are formulated by the researcher and that coders 1 and 2 only discuss problems which arise in their application, in particular when these rules are seen intuitively as invalid. The researcher then 'saves' the rules by accounting for their theoretical justifiability. The result is that coders 'understand' why they can apply the rules, even if these might be seen as invalid. This allows a rigid and uniform application which explains why they are effective means to improve inter-coder-reliability.

Summarizing this section, we have seen that practical rules are formulated in order to solve the problem that coders can neither be instructed nor trained in reliably determining the sense of answers. The solution consists of instructing coders in such a way that they be able to code mechanically according to the form of the answers. The practical rules in question are explained and justified to coders by the researcher (by claiming, for instance, that a certain categorization is 'interesting') but not to the readers of the research report who are led to *believe* that the yielded data are the results of an unproblematic, direct, and therefore valid, application of the coding instructions. In other words, we have discovered a training artifact.[7]

6. CONCLUSIONS

Our question was: *Does coder training consist only of training in the correct application of the coding instructions (theoretical training) or also of coders' socialization into practical rules which are not warranted by the coding instructions?*

We can conclude that in answering this question we must make a distinction between two types of coding problems, a distinction which is notably absent in the literature. There are problems in interpreting an answer (determining its 'sense' or the respondent's intention) on the one hand, and problems of interpreting the coding instructions on the other hand. These two types are not only different analytically, but are treated quite differently in coder training as well. The typical solution of the second type of problem (regarding the interpretation of the coding instructions) is theoretical training, i.e. the specification of the instructions and the socialization of coders into its theoretical logic. The first type of problem, however, can-

not be solved in this way. The reading of respondents' intentions in answers is an everyday competence that has no direct relation to professional views and instructions. Differences in interpretations of answers can only be superseded by active dialogic search for a common ground in specific cases, not by general instruction. Such differences can only be prevented by rules that permit coders to neglect the answers' sense, i.e., by rules that refer to the form (e.g. the presence of certain expressions) only. Because, however, those informal rules cannot be grounded in the coding instructions, they have a negative effect on the codings' validity.

It is an interesting phenomenon that coders easily accept practical rules that are counterintuitive in terms of the everyday 'sense' that answers make to them. Despite this problematic nature of practical rules, the researcher has no trouble in convincing the coders that they must be used. The researcher explains rather than defends these rules. This could be considered as just an effect of the unequal relationship between an instructor (the researcher) and his trainees (the coders). In the extracts discussed above, however, we see no signs of such a relationship of subordination. We see rather that coders comply as equals, as colleagues.[8] Therefore, we consider the coders' compliance with rules proposed by the researcher not as their accepting the researcher's version as a 'privileged' one but rather as their identifying with the researcher's position. As sociologists, the coders 'know' that matters of reliability are overriding both in research practice and publishing and, hence, understand the researcher's logic.

Our finding that a complete coding depends on two types of interpretation, each of which presents specific problems and hence requires specific solutions, is congruent with the results of previous studies of processes of professional interpretation. Hak (1992), e.g., concluded that psychiatric diagnostic procedures consist of two interpretive steps, a first step in which problematic behaviour is interpreted in an everyday way (i.e., labelling it as deviating from an everyday norm) and a second step in which those everyday labels are systematized into professional diagnoses. Procedures such as standardization of diagnostic criteria and professional training can improve the second step of the psychiatric diagnosing process, but its first step cannot be improved because it is dependent on everyday (lay) opinions.

Apparently, this is a fundamental and general feature of professional classificatory work. For the particular case of sociological coding, this distinction is—as far as we know—only made by Rehbein and Mazeland (1991, p. 167). They refer to two kinds of data: "interpretations of utterances which are grounded in coder's everyday knowledge and the subsequent categorization in accordance with the coding system".

The answer to our question, thus, is that both strategies (theoretical training and practical socialization) inevitably must be applied in coder training in order to achieve an improvement of inter-coder-reliability. It cannot be excluded that a high degree of inter-coder-reliability is achieved by informal rules that are not accounted for in the research report. Our findings, therefore, give further ground to Muskens' (1980, p. 126) suggestion that satisfying results of a coder training in terms of inter-coder-reliability can be due to a training artifact. However, because we have studied only one case of coder training (consisting of two sessions), we cannot generalize this finding to all instances of coder training.

What are the implications of our conclusion? The most important implication is that researchers must document each decision rule that is developed in the course of a coder training (or, for that matter, in any form of instructing coders) and in the course of the coding process. This means that coded data can never be considered as 'given' but instead must be seen and treated as 'produced'. The implication for editors of professional journals is that they must not accept papers for publication in which information on coding consists merely of a discussion of the coding instructions and a presentation of measures of reliability. In research reports it must be shown also how specific coding problems have been solved and what kind of practical rules have been applied.

Our findings allow us also to formulate the relation between quantitative and qualitative research in another way than is usually done by sociologists who are mainly involved in quantitative research. They tend to see qualitative methodology as an approach that may be combined with quantitative methodology (e.g. in an exploratory phase of the research, or as another method of measurement in triangulation, or in the development of an instrument) but that, at any rate, is quite distinct and separate from the quantitative approach. Our findings suggest that the quality of coded data in a quantitative approach can only be known and safeguarded (and communicated to readers) through a qualitative assessment of the coding practices (or, for that matter, of the interview practices and of the way respondents, or interviewers, select the pre-coded options in a questionnaire) that have produced these specific data. It appears, thus, that quantitative research is intrinsically dependent on qualitative decisions in the coding process. This implies that good quantitative research must scrutinize and explore its application of qualitative methods. It must, in fact, acknowledge its dependency on these methods (cf. Rehbein and Mazeland 1991, pp. 215-7). At the same time this provides us a definition of (good) qualitative research. Whereas in quantitative research only the validity of the researcher's guidelines (e.g., the validity of the coding instructions in relation to their underlying theory) is discussed, not their application in prac-

tice, in qualitative research the quality and the theoretical status of each datum is assessed carefully.

ENDNOTES

1. This applies also to cases in which the construct validity of a measurement is established by means of a statistical correlation with other variables, because this procedure depends on the argumentative plausibility of the expected correlation.
2. An exception is a half page (!) in Sonquist and Dunkelberg (1977, p. 88). These authors recommend, among other means, that "in a group, the coders and the research staff should go over each [interview], discussing procedures for dealing with different data structures and the concepts underlying the procedures and decision rules".
3. Throughout this article, both in the transcripts and in the body of our text, we will use capitals when we refer to the categories of the coding instructions: CRITERION, SIGN, POINT OF VIEW, DOMAIN, MEANS, and VALUE (see the Appendix).
4. The extracts presented are translations of edited versions of the (Dutch) transcripts. Because we are not interested in an analysis of the conversation itself, but only use the fragments for a discussion of some aspects of coder training, many details of the transcripts (such as hesitations, repetitions and interruptions) are not relevant for the purpose at hand.
   The coding instructions to which the coders refer in their talk are presented in the Appendix to this paper.
5. We are tempted to suggest that the researcher's efforts to control the fit between his theory and the specification of the category 'harmony with the environment' (in terms of 'intrinsic value') have resulted in a valid interpretation of the coding instructions. As stated above, however, validity cannot be measured, and is a matter of argumentation. Our claim in this case is that the specification of the instructions that has emerged in the coder training can convincingly be explained by the researcher to his audience.
6. In order to understand this discussion it is necessary to know that in Dutch the expression 'welbevinden', a rather formal word which literally means 'feeling well', has almost the same sense as the English 'well-being'. The expression's literal meaning opens the possibility to question the degree to which 'well-being' [welbevinden] is a 'feeling', and hence 'subjective', as well.
7. Note that the researcher *could* have reported that and why the rule "Apply the category 'body' if, and only if, the 'body' or a synonym is mentioned explicitly" was developed. This, however, would have imposed upon him the onus of explaining how this rule fits his theoretical distinction of expressive and instrumental attitudes to health (which would have been a rather difficult task).
8. This is particularly clear when coders anticipate readers' criticisms, such as, e.g., at the end of Extract 3, where coder 1 explains why the researcher must not look up a respondent's age because this would imply 'circular reasoning'.

REFERENCES

Babbie, E. 1989. *The Practice of Social Research*, Fifth Edition. Belmont, CA: Wadsworth Publishing Company.

Bernts, T. 1991. *Leven zonder zorg*. Lisse: Swets & Zeitlinger.

Cicourel, A. 1964. *Method and Measurement in Sociology*. New York: The Free Press.

Cicourel, A. 1978. "Interpretation and summarization: issues in the child's acquisition of social structure." Pp. 251-81 in *The development of social understanding,* edited by J. Glick and K. A. Clarke-Stewart. New York: Gardner Press, Inc.
Garfinkel, H. 1967. *Studies in Ethnomethodology.* Englewood Cliffs, NJ: Prentice Hall.
Hak, T. 1992. "Psychiatric records as transformations of other texts." Pp. 138-55 in *Text in Context,* edited by G. Watson and R. M. Seiler. Newbury Park, CA: SAGE Publications.
d'Houtaud, A. 1978. "L'image de la santé dans une population lorraine: approche psychosociale des représentations de la santé." *Rev. Epidém. et Santé Publ.* 26:299-320.
d'Houtaud, A. and M. Field. 1984. "The image of health: variations in perception by social class in a French population." *Sociology of Health and Illness* 6:30-60.
Katz, B.A. and W.W. Sharrock. 1976. "Eine Darstellung des Kodierens." Pp. 244-71 in *Etnomethodologie. Beiträge zu einer Soziologie des Alltagshandelns,* edited by E. Weingarten, F. Sack and J. Schenkein. Frankfurt am Main: Suhrkamp.
Mishler, E. G. 1986. *Research interviewing.* Cambridge, MA: Harvard University Press.
Muskens, G. 1980. *Frames of meaning—are they measurable?.* Ph.D. thesis, Katholieke Universiteit Nijmegen.
Rehbein, J. and H. Mazeland. 1991. "Kodierentscheidungen." Pp. 166-221 in *Verbale Interaktion,* edited by D. Flader. Stuttgart: Metzlersche Verlagsbuchhandlung.
Rosenthal, R. S. and R. L. Rosnow (eds.) 1969. *Artifact in behavioral research.* New York/London: Academic Press.
Sonquist, J. and W. Dunkelberg. 1977. *Survey and Opinion Research.* Englewood Cliffs, NJ: Prentice Hall.
Wilson, T. P. 1970. "Conceptions of interaction and forms of sociological explanation." *American Sociological Review* 35:697-710.

APPENDIX

Coding Instructions (Translated from Dutch)

Definitions of health can be considered answers to three questions:
—how does one perceive health? (criterion)
—how does one get or maintain health? (means)
—what is the value or the aim of health? (value)
A respondent can answer all three questions in one statement, but often he answers only one of them.

CRITERION: three aspects can be distinguished:

SIGN	is the definition positive, negative (not ill; no physical complaints) or both?
POINT OF VIEW	is it subjective (experience) or objective (being, having, well-being)?
DOMAIN	is it physical, or more than that?

codes:

SIGN	1. positive 2. negative 3. both
POINT OF VIEW	1. subjective 2. objective 3. both

DOMAIN
1. physical
2. physical and mental
3. undetermined

MEANS: 'harmony' means that one adjusts the environment to the health aims (adaptation of the environment), whereas 'prevention' rather represents adjusting oneself to the requirements of health (assimilation to the environment).

codes:
1. harmony with the environment
2. prevention/hygiene
3. other means

VALUE: 'vitality' refers to health as subservient to individual aims (among others), whereas 'work' refers to the aim of adequately adapting oneself to societal requirements.

codes:
1. vitality/growth
2. work/functioning
3. other value

Comment: A 'positive' SIGN, a 'mental' DOMAIN, a 'subjective' POINT OF VIEW, a 'harmonious' MEANS and a 'growth' VALUE are taken as indicative of an *expressive* definition of health. A 'negative' SIGN, a 'physical' DOMAIN, an 'objective' POINT OF VIEW, a 'preventive' MEANS and a 'work' VALUE are taken as indicative of an *instrumental* definition of health.