



Leah McClimans

The problem of purpose in quality of life research

Originally produced as part of the programme of the AHRC Contingency and Dissent in Science. Copyright © 2006 Leah McClimans, CPNSS.

You may cite this version as:

McClimans, Leah (2006). *The problem of purpose in quality of life research* [online]. London: LSE Research Online.

Available at: <http://eprints.lse.ac.uk/archive/00000794>

Available online: June 2006

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

<http://eprints.lse.ac.uk>

Contact LSE Research Online at: Library.Researchonline@lse.ac.uk

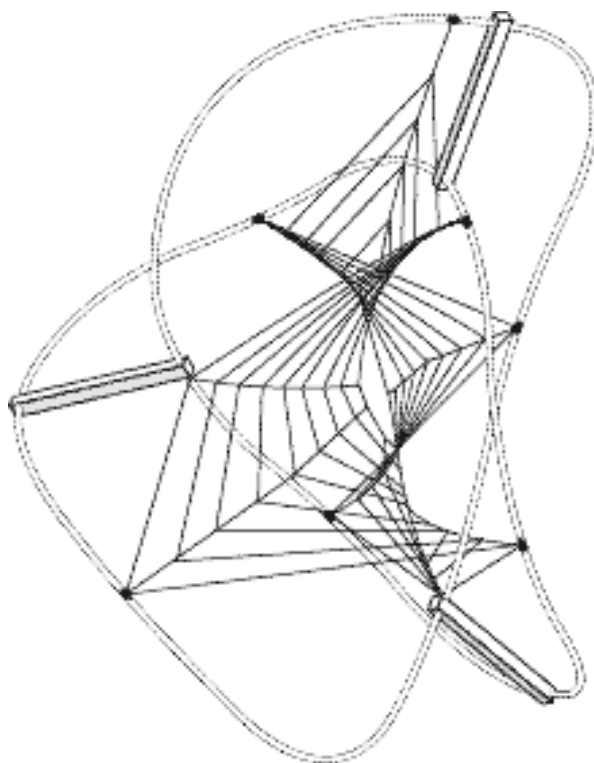
Centre for Philosophy of Natural and Social Science

Contingency and Dissent in Science

Technical Report 02/06

The Problem of Purpose in Quality of Life Research

Leah McClimans



Editor: Damien Fennell

The support of The Arts and Humanities Research Council (AHRC) is gratefully acknowledged.
The work was part of the programme of the AHRC Contingency and Dissent in Science.

Published by the Contingency And Dissent in Science Project
Centre for Philosophy of Natural and Social Science
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

Copyright © Leah McClimans 2006

ISSN 1750-7952 (Print)
ISSN 1750-7960 (Online)

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of the publisher, nor be issued to the public or circulated in any form of binding or cover other than that in which it is published.

The Problem of Purpose in Quality of Life Research

Leah McClimans

London School of Economics

I

In this paper I investigate one aspect of the validity of a certain type of subjective assessment of health and illness or as they are sometimes called Health-Related Quality of Life (HRQoL) measures. These assessments are typically a series of questions directed to individual patients in the form of a questionnaire and their aim is to discover the subjective experience of a cohort of people at different times, for instance before and after a surgical intervention. These measurements can be, and often are used in conjunction with clinical assessments to gain information on health outcomes for policy implementation or revision. For example, to determine which hospitals or surgeons are doing well and which ones are doing poorly or to determine which interventions are most effective.

Utility vs. Non-utility measures

A useful distinction among HRQoL measures can be made between utility measures and non-utility measures. Utility measures are always index measures, that is they reduce a patient's quality of life to a single number which represents that health state's utility; usually 1 indicates perfect health and 0 indicates a state equivalent to death. The utility values are derived from prior preference scoring from the general population, using the usual techniques.¹ The EuroQol (EQ-5D) is an example of a utility index measure of HRQoL.

Non-utility measures can be profile measures, which measure multiple dimensions of a construct and give us multiple numbers to represent a single health state, or index measures (which essentially measure just one dimension). Non-utility measures are arbitrarily weighted according to researcher expertise and layperson input and the scores represent the severity of whatever

¹ For example, using standard gambles or time-trade offs.

dimension they are meant to measure. Unlike utility measures which reduce all constructs to a single metric of utility, non-utility measures each develop their own metric of whatever dimension is being measured.

Utility index measures have typically preoccupied moral philosophers because of the role they play in the scarce allocation of health resources. Non-utility measures are less directly involved in resource allocation and more directly involved in giving us a picture of patient experience and thus fleshing out our understanding of health and disease. With one exception philosophers have almost entirely ignored non-utility measures. I take up the issue of these measures because I think they appropriately direct our attention first to the question ‘What is quality of life?’ as opposed to the question ‘How much utility is this health state worth?’

Non-utility measures, however, are not without their own set of difficulties for although they direct our attention to the question ‘What is quality of life?’ whether they in fact give a good answer to it is another story. For the rest of this paper I want to discuss one of the problems we encounter when attempting to answer this question, namely the problem of ascertaining the purpose of various questions in the questionnaires.

Silent Bias

For a HRQoL or subjective assessment measure to be valid it must accurately measure its target construct. Sometimes these are specific to a particular disease and sometimes they relate to general health. One way validity is undermined is if a respondent answers untruthfully or if she misconstrues the meaning of the question and answers according to her misunderstanding. These problems generally fall under the heading ‘respondent bias’ and there are various ways to handle it. In certain socially sensitive cases researchers’ may hide the purpose of a question so that respondents will answer truthfully. To combat problems of misunderstanding researchers often aim to make their questions simple and straightforward.

In practice it’s almost impossible to get rid of all the respondent bias and the most researchers hope to do is minimize it where possible and use auxiliary hypotheses to explain poor validity when it occurs. The problem with this is that respondent bias doesn’t always show up in

classical validity tests. In fact the only time it does show up is when respondents answer contrary to already established hypotheses as to how they should answer. But respondent bias also occurs when respondents answer *in line* with established hypotheses, but for *different reasons* and sometimes in response to alternatively interpreted questions. In these instances respondent bias is invisible from the perspective of classical test theory, but nonetheless threatens the validity of the measure. In the last 15 years a smattering of papers have investigated the extent to which this kind of ‘silent’ bias affects some of our most well-respected and well-validated questionnaires and with surprising results.

If we are stuck with prejudice or bias, then we must learn how to live with it. On the other hand, if we can get rid of it, then our energies ought to go towards removing it. Currently most of the time and energy in this area goes towards its removal. In what follows I look at two papers on respondent bias and argue that the invalidity which occurs here is primarily due to problems of understanding and interpretation. I then hypothesize about ways we might deal with this problem and in doing so I argue that the problems of interpretation go deeper than respondent bias for they also affect researchers’ understanding of their research question, and thus their application of the measure. Finally I suggest a dialogic solution to both of these problems.

II

The Nottingham Health Profile and the Problem of Purpose

In 1993 Donovan *et al.* published a paper entitled, ‘Assessing the need for health status measures’ which examined the propriety of using subjective health status measures to determine the health needs of local populations. Donovan and his team focused on the validity of the Nottingham Health Profile (NHP), but they were keen to emphasize that the issues raised were are not exclusive to it, “but apply to the genre”.² By looking at individual responses to statements in the questionnaires they wished to discover if individuals’ scores truly represented their views concerning their own health, for only if it did would its particular use to determine local health needs make sense.³

² Donovan, J. L. *et al.* (1993), p. 159.

³ Ibid.

In their study respondents were asked to fill in part of the General Household Survey, the NHP and two measures of health and health care developed by the Rand Corporation. They were then asked to participate in an audio-taped, semi-structured interview about their health care and any issues raised by the standardized questionnaires.⁴ Donovan *et al.* focused on the NHP because at the time this was the leading generic, standardized subjective health status measure on the market. As Donovan *et al.* remark it replaced the Sickness Impact Profile (SIP) as the leading generic measure because it was perceived to be ‘better’ in terms of validity and more sensitive in terms of responsiveness to change.⁵ The NHP is scaled with categorical judgments in the form of ‘yes/no’ responses to short statements and it measures three dimensions: perceived physical, social and emotional health.

During the interview section of the study Donovan *et al.* found three main areas of confounding. I will review them one by one. In the first case, they suggest that the yes/no response alternatives were too limiting for the respondents’ answers. For instance, the following are some of the questions on the NHP:⁶

Things are getting me down:	yes/no
I have pain at night:	yes/no
I have unbearable pain:	yes/no
I take tablets to help me sleep:	yes/no

To the first question, one respondent answered, “I won’t let them if I can. Can I put sometimes?” Here the categorical scaling of the question is obviously too limiting, reflecting perhaps a need to implement some kind of continuous judgment scale that would allow for a range of response options. But changing the response scale would not solve all the problems. In the other three questions respondents had rather different sorts of difficulties, ones that reflect more than a problem with the categorical answers. For instance, to the question about pain at night one respondent replies that she doesn’t so much have pain, but discomfort. Here the

⁴ Ibid.

⁵ Ibid, 161.

⁶ Although the ‘questions’ on the NHP, and many other standardized questionnaires, are not explicitly in the form of an interrogative it is common practice in this literature to refer to them as ‘questions’.

question seems to be whether discomfort counts as pain in this case. Similarly, with respect to the question about unbearable pain, a respondent says that the pain is only unbearable when she has a backache. We might say that because this is a generic questionnaire measuring general health, it isn't clear to her if an occasional backache is sufficient to count as unbearable pain in general. Finally, to the question about taking tablets to help with sleep, one respondent answers, "I take tablets at night for the cramp and they help me sleep. What do I put there?" In this case it doesn't seem clear to the respondent what the question is asking her: is the question a causal one—does she take tablets *because* they help her sleep—or is it interested in *whether* she takes them, for whatever reason? Here we might say that the respondent does not understand the question's purpose or meaning—*why* is she asked this question; what do the researchers want to know?

Donovan *et al.* say about the above examples that, "The statements [in the NHP] made people think about aspects of their health, but then constrained their responses...The forcing of responses into predefined categories negated people's desires to negotiate the meanings of health and illness." And earlier they comment that although the yes/no dichotomy of the NHP leads to straightforward analysis, it "does not allow people to express what they really feel". I agree with Donovan *et al.* that categorical scaling constrains responses, and I also agree that it does not allow for respondents to question what, for instance, pain means in certain contexts. Moreover, to the extent that respondents are left to answer these questionnaires on their own, and then researchers analyze them in terms of *their* meanings and *their* understanding of the questions, then, yes, they do not allow patients to express what they really feel. But I disagree with the conclusion that Donovan *et al.* seem to imply, namely that the problem is with the categorical scaling *per se*. The problem is not solely with the yes/no answers: the problem is rather that the respondents are not given enough of a context in which to make sense of what various terms, or whole questions, in the NHP mean, and moreover, they aren't given the opportunity to question that context and therefore the particular meaning of these terms. Once it is clear what it means to have unbearable pain, both in terms of the researcher's interest and the respondent's experience, then the yes/no format is not problematic. The problem is that the respondents are not allowed to ask questions, but only to answer them.

The second aspect of confounding is described as contradictory or arbitrary filling out of the questionnaire. Donovan *et al.* provide two examples. In the first case, to the question, ‘I find it hard to bend’, one respondent says, “I do find it hard to bend, but I’m not ticking yes there.” This is interpreted by Donovan *et al.* as contradictory, a result of the pressure to give socially acceptable responses.⁷ But while this may be the case, it is not necessarily so. To see why not consider Donovan *et al.*’s second example of arbitrary or contradictory confounding. In their second example Donovan *et al.* cite the following question and its response: ‘Worry is keeping me awake at night’ “Well yes, but it’s only stupid things. I lie awake thinking. I’ll put no because I’m just being stupid.”⁸ Unlike the bending example, this respondent actually explains why she has decided to mark ‘no’. Yes she worries, but she only worries about stupid things because she’s “being stupid”.

Although Donovan *et al.* don’t see it this way, this example does not appear to be a contradictory answer. It seems instead that the respondent takes her sort of worrying to be different from the sort in which the NHP is interested: her worrying is petty and “stupid”, while perhaps the worrying in which the questionnaire is interested involves existential doubt and acute torment. To be sure, she could be wrong about what the NHP is after, but that is not the point. The point is that we could paraphrase this respondent’s answer in terms of their first example in this category: “I do lie awake worrying at night, but I’m not ticking yes.” and nonetheless understand it as making sense given a particular understanding of what the question meant: her answer need not be contradictory. As I have said before, if we don’t know why the respondent can’t bend or is lying awake and how they understand the question, it’s difficult to know how to understand their answer.

Finally, Donovan *et al.* cite cases in which they construe people as misunderstanding the purpose or meaning of the statement, and thus responding according to problems unassociated with their health. For instance, to the statement, ‘I find it hard to reach for things’, one respondent

⁷ Donovan, J. L. *et al.* (1993), p. 159.

⁸ Ibid.

answered, “I do find it hard to reach for things, yes, because I am short.” This question however, isn’t asked with respect to the respondent’s height, but rather with respect to her health.

Although Donovan *et al.* are right—this misconstrual is indeed a problem with understanding the purpose of the question—I have argued that most of their examples share the same problem. In the first set of examples, we saw that the question about taking tablets to sleep was confusing because it was unclear whether its purpose was to discover *why* the respondent took tablets or *whether* she took them. In this case the respondent answered with a request for clarification. In second set of examples I argued that although respondents recognized that in one sense they did experience the phenomena in question, their understanding of the meaning or purpose of the question in the NHP perhaps precluded them from answering positively. In this current example, the respondent has answered the question, as we might say, literally—she hasn’t taken into account the difference between what the NHP might mean by the question and what the grocer might mean—and in this case we say that she’s misunderstood the question’s purpose.

Despite my interpretive disagreements with Donovan *et al.* as to the nature of the problem in each of the above examples, they do make several important points. When respondents understand questions differently from the researchers who analyze them, then the conclusions of the study can suggest morbidity that does not actually exist—as in the respondent who answered positively to the statement about whether she found it hard to reach things. Conversely, the study can also underreport morbidity. This can happen for the same reasons—respondents take the question to mean something different from what was intended. For example, a respondent may take ‘unbearable pain’ to mean ‘unbearable pain all the time’ and therefore tick no despite having it occasionally. Although I agree with Donovan *et al.* that the validity of the NHP is threatened in just these ways, I disagree over why it is threatened. Unlike Donovan *et al.* I suggested that each of their examples represented a problem with the respondents' ability to understand the purpose of the questions.

Towards the end of this article Donovan *et al.* observe how subjective health assessment instruments become “fashionable” and then as a result the consensus holds that they are more reliable and valid than their predecessors. They note that like the NHP which had once replaced

the Sickness Impact Profile (SIP) at the time of their article the Short Form-36 (SF-36) was starting to replace the NHP.⁹ Moreover, they intimate that the assumption of the SF-36's supposed increase in reliability and validity over the NHP is unwarranted. But Donovan *et al.* wrote this article in 1993 and we might think that in the last twelve years subjective assessment has improved, and not merely because the latest measure appears to be “fashionable”.

SF-36 and the Problem of Purpose

With this in mind I now turn to a study published in 2002 by Sara Mallinson entitled, ‘Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire’.¹⁰ Like Donovan *et al.*, Mallinson is worried about problems with the validity, which psychometric properties cannot detect—specifically she is worried that people do not hear questions in the manner they were intended, and thus their answers are misinterpreted by researchers analyzing the questionnaire. Mallinson conducted her research on individuals 65 years old or older from the North West of England who had recently been referred to community physiotherapy or occupational rehabilitation. She conducted two audio-taped interviews, one before and one six months after their therapy had begun. All interviewees were asked questions that covered socio-demographic characteristics, self-reported health problems, expectations of treatment, the SF-36 and two global health evaluation questions.

The SF-36 is a generic subjective assessment measure with 36 questions scored in eight dimensions. It is now the most widely used measure of its kind in the UK and it has been translated into 45 languages for use around the world. Moreover, there is an extensive literature that suggests it is one of the most reliable and valid short-form questionnaires available.¹¹ Because of its extensive testing and wide use Mallinson focuses on it to illustrate how pervasive these problems of invalidity can be.

For this paper Mallinson focuses on the responses from two of the eight dimensions: the physical functioning scale and general health perceptions. In the former respondents are asked how their health limits them in a variety of everyday activities. They are asked to respond to each question

⁹ Ibid, p. 161.

¹⁰ Mallinson, S. (2002), p. 13.

¹¹ Ibid.

with one of three answers: ‘Yes, limited a lot’, ‘Yes, limited a little’, ‘No, not limited at all’. In the latter dimension respondents are first asked to assess their general health as ‘excellent’, ‘very good’, ‘good’, ‘fair’ or ‘poor’. They are then given four statements about their general health and asked to rate them on a Likert scale as ‘definitely true’, ‘mostly true’, ‘not sure’, ‘mostly false’, or ‘definitely false’.¹²

As Mallinson makes clear in the beginning of this paper it is the problem of understanding meaning—how researchers hope the question is understood and how the respondent in fact understands it—that drives her research.¹³ Thus, for the most part, I’m going to focus on those of her findings that make this problem most clear. In the physical functioning dimension Mallinson found two problematic areas where this was particularly at issue. In the first instance respondents are asked three questions about how limited they are with respect to walking various distances: more than a mile, half a mile and 100 yards. Mallinson found that respondents often asked for clarification because they were unfamiliar with the distances involved. Instead of conversing in technical distances she found that they needed familiar landmarks in order to make a judgment. Once they understood that a mile was “from here to Tesco and back” they could easily answer the question.¹⁴ But without the help of an interviewer—the normal situation in self-completed questionnaires—these same individuals would have to guess these distances. Guessing is notoriously unreliable and puts the validity of the questionnaire into question.¹⁵

In the physical functioning dimension, Mallinson also found what she calls “vague questions”. Here when respondents were asked if they could walk half a mile, the problem was not how far half a mile was, but rather whether that half mile was up-hill or on the flat, whether it was fast or slowly walked. Consider one respondent’s answer:

I can walk down to the garden centre but there’s no way I could get back because it’s up-hill, and as soon as I, I can’t walk up that hill so it depends which, if you’re talking about on the flat, slowly, not talking or carrying anything...I can

¹² Ibid, p. 14.

¹³ Ibid, p. 12.

¹⁴ Ibid p. 15. I think it is important that this problem is not construed as one specific and limited to the old and infirm. Technical measurements—distance, volume, length...etc.—are constantly a problem in my everyday life. I have very little sense of any measure that I don’t explicitly use in that way on a regular basis, for instance I have only a very vague idea of how long 100 yards is.

¹⁵ Mallinson, S. (2002), p.16.

walk around the shopping precinct and round the supermarket because you're going slowly and you're stopping and looking at things and you're not talking to anybody.¹⁶

We have seen the problem of “vague questions” before in Donovan *et al.*'s paper. Recall the questions about pain in which respondents, although familiar with pain, were unsure what counted as pain in that context. In the SF-36 Mallinson found this problem in questions not only about walking, but also questions about lifting and bathing.¹⁷

Turning now to the general health perceptions dimension, the first question respondents are asked is to rate their general health. Here Mallinson found that respondents understood the meaning of health in a myriad of ways. Although it is well-known that concepts like health and illness are difficult to define and that people define them differently, the validity of this question requires that respondents understand it in the same way as one another and in the same way as the researcher. Mallinson describes the difficulties with this question as “comparative” problems because she found that the main differences in conceptualizing health came from what respondents compared their health to.¹⁸ So with regard to the question, ‘In general would you say your health is: excellent, very good, good, fair, poor’, some respondents understood this as saying ‘would you say your health *as opposed to others in the population* is...’ others understood it as ‘would you say your health *now as opposed to other times in your life* is...’ and still others understood it as ‘would you say your health *as opposed to others your age* is...’¹⁹ Without an explicit contrast, it is unclear what health means in this question.²⁰ Requiring respondents to decide for themselves what it means is tantamount to asking at least three different questions without any idea which one a particular respondent answers.

Finally Mallinson looks at problems involving what she calls “inclusions and exclusions” in conceptualizing health. These examples resemble those found in the final two sets of examples given in the previous study, in which respondent interpretations of questions can lead to

¹⁶ Ibid.

¹⁷ Ibid.

¹⁸ Ibid, p. 18.

¹⁹ Ibid, pp.18-19.

²⁰ See Bas van Fraassen's work on explanation and why-questions for a more extensive discussion on the importance of contrast-class for understanding questions. Van Fraassen, B. (1980), pp. 126-46.

surprising, and perhaps confusing answers that only make sense given the context of their understanding. In Mallinson's example, to the question 'In general how would you say your health is?' one respondent answered, "My health is good. It's the spinal atrophy that's the problem". As Mallinson notes, this man separated his specific chronic health problem from his sense of being a 'healthy' person, which is what he thought the surveyor was interested in.²¹ Another way to put this is that he interpreted the purpose of the question differently than perhaps the researcher intended. A similar problem occurred with the question that asked respondents whether they thought they got ill more easily than other people because individuals differed in how they interpreted the term 'ill'. For instance, many thought it referred to everyday problems and excluded chronic health problems as these two responses illustrate: "I'm not bodily ill, its mobility more than anything isn't it?" and "I don't think so—apart from my diabetes and it doesn't bother me that. I've got used to it."²² These answers are not the result of social desirability or denial so much as logical responses to different interpretations of the question. Mallinson sums up the problem well:

...they include and exclude problems in different ways and may unexpectedly weigh all kinds of information before answering. In effect, people are responding from different premises to each other and from the surveyor. This inevitably affects respondents' intentions/meanings in selecting a particular response option and makes it difficult for the surveyor to interpret their answers.²³

Mallinson's observations of respondents' answers to the SF-36 resembles, in many ways, my interpretation of Donovan *et al.*'s observations of the NHP. In both cases we see respondents' difficulties understanding what particular terms mean because, as Mallinson puts it, they are vague or unfamiliar. In these cases respondents ask questions about the questions and without an interlocutor they are left to guess. We also see in both articles cases where respondents don't understand the purpose or meaning of a question in the same way as the researchers do. In such cases respondents may ask for clarification or, as in the examples above, they may not appear confused, but provide answers that at first glance don't make sense or seem unlikely. In these latter situations Mallinson suggests, and I agree, that we need to understand the context of the

²¹ Mallinson, S. (2002), p. 19.

²² Ibid.

²³ Ibid.

respondents' answers before we dismiss them. At the end of her paper Mallinson concludes that the issue of understanding meaning is "absolutely central to understanding subjective views" and that without further research into how respondents' understand survey questions it will be difficult to establish the validity of subjective assessment.²⁴

It would seem that this same concern is reiterated in a recent article by Tennant *et al.*, who worry that when we use the SF-36, especially to make comparisons across various health or disease states, we run the risk that "patients who have had different experiences interpret it differently"²⁵. For instance a 'yes' response to a question about tiredness can represent a very different response for a healthy person than it does for someone with rheumatoid arthritis.²⁶ And as Mallinson has shown it can represent a different response even for people of the *same* relative age and health state. As Tennant *et al.* illustrate these interpretive differences may explain why data collected with the SF-36 sometimes provide surprising results: for instance, individuals with psoriasis have worse scores than people with arthritis, cancer and myocardial infarction.²⁷ Then again, they may *not* explain these differences because we cannot assume that people with psoriasis *don't* have a poorer quality of life compared to arthritis or cancer patients just because they have *clinically* less severe problems. What does seem clear is that the results obtained from these questionnaires, at least the NHP and the SF-36, often do not tell us much, and perhaps they won't until we get a better understanding of how respondents interpret questions.²⁸ In the next section I want to look at one theoretical attempt within the subjective assessment literature that tries to further this understanding.

²⁴ Ibid, p. 20.

²⁵ Tennant, A. (2004), S24.

²⁶ Ibid.

²⁷ Ibid.

²⁸ Although it is difficult to know how often the kind of misunderstanding illustrated in these two studies occurs generally, the frequency with which it occurred here indicates that it may happen often enough to affect the overall results of other questionnaires as well. Certainly the authors of these studies believe the problem to be widespread and significant.

III

Principles of Discourse

In J. Tanur's book *Questions about Questions*, Herbert Clark and Michael Schober try to explain respondent bias in a chapter entitled 'Asking Questions and Influencing Answers'.²⁹ They begin with the claim that we will be able to resolve the puzzles that respondent bias presents to questionnaire design only when we understand why it occurs and they go on to explain respondent bias in terms of certain peculiarities of language use with respect to discourse. The main premise of their discussion is that language use is not about words and what they mean, but about people and what they mean.³⁰ Clark and Schober later write, "What counts [in language use] is not the meanings of words per se, but what speakers mean by using them."³¹ Here they seem to make a distinction between the meaning of words and the meanings words have when speakers use them, emphasizing the importance of the latter for language use. What Clark and Schober seem to be appealing to is a distinction I have tacitly made throughout this chapter: that to understand questions we need more than familiar words: in addition we need to understand the context of those questions. Their task is to outline five ways we typically come to understand our interlocutors in normal discourse and how this affects the answers in standardized questionnaires.

They begin this discussion with the principle of "common ground".³² Here they seem to make the point that words become meaningful against a shared context; change the context and you often change what we understand a word to mean. In referring to a similar idea Larry Wright provides the following example. Take the sentence, 'The cat is on the mat.' Typically we conjure ideas of an animal on a rug. The context that provides this interpretation is perhaps a domestic setting. But we can imagine a different context, say that of a construction site, where the same sentence refers to a piece of machinery parked on a blasting mat.³³ It is the awareness of the "common ground" that we share with others that allows us to use ambiguous language

²⁹ Clark, H. H. and Schober, M. F. (1992), p. 43.

³⁰ Ibid, p. 15.

³¹ Ibid, p. 16.

³² Ibid, pp. 17-18.

³³ Wright, L. (2001), p. 5.

unconsciously and with clear intent. Perhaps this is most obvious in the frequent use of indexicals like ‘here’ and ‘now’.

“Common ground”, Clark and Schober write, “is essential in interpreting everything people say.”³⁴ In the examples from earlier we saw some of the consequences of an absence of “common ground”. For instance, respondents often misunderstood the intent of a question because they assumed a different context than the researchers did. Perhaps Donovan *et al.* provide the clearest case of such a misunderstanding in the case of the question, ‘I find it hard to reach things’. Here the respondent answers yes because she is short. She misinterprets the question as referring to an everyday context as opposed to a medical one.

Two further principles associated with “common ground” are the “accumulation of common ground” and “grounding”.³⁵ As we saw above Clark and Schober believe that some kind of common ground is necessary to every conversation, but as a conversation progresses more common ground is accumulated and this happens through what Clark and Schober see as the process of grounding. Grounding occurs in the myriad of ways we have of assuring one another that we have understood what has been said. For instance, a speaker is encouraged that she has been understood when an interlocutor, for instance, nods her head, displays positive facial expressions, but most importantly if she paraphrases what was said or shows her understanding in what she says next.³⁶ In other words, the most important thing for grounding is that the interlocutors make sense to one another as the conversation progresses. As this happens the formation of new common ground—an *intersubjective* ground—accumulates and the interlocutors now use this as well in deciding what to say next and in understanding what was meant by the other.

Because in standardized questionnaires there aren’t any mechanisms for people to check their understanding, they end up making assumptions about the meaning of questions. Clark and

³⁴ Clark, H. H. and Schober, M. F. (1992), p. 18. This is the case not only where “common ground” leads us to unconsciously expect that our meaning is understood, but also in cases where we recognize that a “common ground” is missing. This recognition often explains what may seem to others who share our “common ground” as unnecessary or lengthy description.

³⁵ Ibid, pp. 19-20, 24-5.

³⁶ Ibid, p. 25.

Schober write that respondents assume that the researcher chose her wording so that the meaning would be obvious: thus respondents impute whatever meaning seems right to them.³⁷ This is obviously the problem with the woman's interpretation of the question, 'I find it hard to reach things'. In everyday dialogue her response would have been quickly met with a correction, "No, I meant do you find it hard to reach things after your by-pass surgery." In addition, we can see how the process of grounding might illuminate areas of ambiguity or vagueness in places that the *researcher* did not anticipate, and moreover perhaps cannot immediately make more precise. For instance, as we saw in the previous section Mallinson found that the SF-36's question on general health gave rise to at least three different interpretations depending on the contrast class respondents used to understand it: 'would you say your health *as opposed to others in the population* is...'; 'would you say your health *now as opposed to other times in your life* is...'; 'would you say your health *as opposed to others your age* is...' Researchers may not have recognized 'health' as needing a contrast to make it clear, and upon that recognition they may not be sure which one to use—they may not be sure which question will best illuminate their guiding construct.

In their final two principles Clark and Schober attempt to make more precise what exactly is involved in their understanding of "common ground". The first of these is "perspective".³⁸ Perspective affects the kinds of answers that make sense given the description under which we express something. For instance, if I refer to my friend as crazy in relation to her zealous cleaning, then I imply that her cleaning is somehow not normal and not admirable. If I then ask, "Why is Jennifer so intense?" Clark and Schober contend that you are unlikely to contradict me and say, "She isn't intense". Compare this to a situation in which I ask the same question but the conversational implicature is admiration and respect. In this instance, according to Clark and Schober our interlocutor is unlikely to answer, "because she's an obsessive compulsive", but more probably something like "because she has takes a lot of pride in her environment". As interlocutors, they argue, we acquiesce to perspective because it is polite and because to do otherwise is to challenge your interlocutor's judgment.³⁹ Therefore, it is more likely that you will answer the question with respect to my description—that is you will take that description for

³⁷ Ibid, p. 23.

³⁸ Ibid, p. 20.

³⁹ Ibid.

granted—and understand your options accordingly. Moreover, in granting my perspective and answering my questions with sensible answers, it becomes part of our common ground that she is intense in a way that is not admirable.

Issues of perspective can also have the effect of creating normative overtones. We can see this explicitly in another example from Mallinson's paper. In this example, respondents were asked how their health affected lifting or carrying groceries. The perspective here assumes that they do in fact carry groceries. But in at least one case a respondent's husband always got the groceries for her, not because she was ill, but because that was the way they did things.⁴⁰ The perspective from which this question is asked assumes that carrying groceries is something everyone does from time to time: the response options follow up on this assumption because they only allow one to rate the degree to which one is limited—it does not allow you to explain that the question does not apply.

In this example we see a slightly different aspect of “perspective”, one that Clark and Schober, I think, underestimate. They tend to emphasize how perspective affects the direction and the particular understanding of a conversation, but perspective also opens up the opportunity to demur. When little is at stake, social pressure to be polite may mean that we tend to accept the descriptions others provide, but in many cases, including the one above, we challenge the descriptions others offer and assert our own perspective. Standardized questionnaires don't allow this sort of negotiation. Respondents often must either leave the question blank or accept the perspective of the question.

Finally, Clark and Schober discuss “common purpose”.⁴¹ “Common purpose” like “perspective” is another one of Clark and Schober's attempts to articulate what is involved in their notion of “common ground”. Common purposes are important because they shape the direction of the conversation as well as what people mean by what they say. Clark and Schober write that the evidence suggests that “common purpose” is essential in figuring out a speaker's meaning.⁴² To make clearer what exactly is involved in this idea, recall Wright's example: ‘The cat is on the

⁴⁰ Mallinson, S. (2002), p. 17.

⁴¹ Clark, H. H. and Schober, M. F. (1992), p. 22.

⁴² Ibid, p. 22.

mat.’ Here the context tells us whether the sentence refers to an animal or a machine, but it should also give us insight into *why* someone is saying this. For example, if you’re in the kitchen and your partner walks in and suddenly says ‘the cat is on the mat’ you may know that he’s referring to Cooper, but you may not understand the significance of what he’s saying. Was he looking for the cat? Is the mat off-limits and he’s expressing frustration? Is this code for, ‘now’s a good time to give Cooper his meds’?

Locating the Purpose of the Questionnaire

Wright puts the importance of purpose a bit differently than Clark and Schober when he writes that, “...the main thing a context supplies to help us understand each other is *motivation*.”⁴³ Moreover, we understand the motivation or purpose of a conversation, when we understand the *question* to which the discussion is the *answer*.⁴⁴ Thus we understand the significance of ‘the cat is on the mat’ when we understand it as the answer to the question, ‘Where is the cat?’ or ‘What is Cooper doing?’ or ‘When should we give him his meds? Depending on which question was asked, whether implicitly or not, your reaction will be different. If Cooper was lost and now he’s been found, you might express relief; if he’s lying on your Persian rug, you might run off with the squirt bottle; or if he’s half asleep and now is a good time for the meds, you might take them from the kitchen counter and walk to the mat.

But Wright’s suggestion that we understand motivation when we understand the question to which a sentence or discussion is the answer, is misleading if we then think that the question which supplies the motivation (Where is the cat? What is Cooper doing? Should we give him his meds?) can stand alone; that we can understand it without contextual support. We need look no further than one of the examples from Donovan *et al.*’s study to see that this is not the case. Recall the example in which the woman was unsure whether the question about taking tablets to sleep was interested in *why* she took the tablets or simply *whether* she took them. In this case the yes/no answers are not made completely clear in light of the question. For this question to be clear I suggested that the respondent needed to understand the purpose or motivation of the question itself. If we apply Wright’s question/answer structure to this problem, then the

⁴³ Wright, L. (2001), p. 142.

⁴⁴ Ibid. Although here Wright is specifically referring to how we understand the motivation of an argument, the idea applies generally. For a discussion of this more general application see Gadamer, H. (2003), pp. 370-1.

respondent's only hope is to understand the question about taking tablets to sleep as the *answer* to a further question.

This request to understand individual questions in the questionnaire as answers to a larger question follows the design of profile measures quite nicely since we begin the construction of these questionnaires with a broad question, say for instance, 'What is the quality of life of those suffering from depression?' And then through factor analysis or some other method we break this construct down into a few dimensions—perhaps social and emotional functioning—each one of the dimensions is then composed of questions, like the one about taking tablets to sleep for instance, that further breakdown the respective dimension.

I am suggesting that it is the overall research question that gives the appropriate context with respect to which individual questions make sense and this seems right given the construction of the measure. If for example the sleeping tablet question had been embedded in a conversation with a psychiatrist about depression, the respondent would probably have recognized its unstated purpose, which is to ascertain if she is taking the pills *because* they help her sleep. She therefore would have answered no, since she only takes them for a cramp.

But notice what is involved even in this conversation if the respondent is to recognize the purpose and give the correct answer. In the case above she first must understand the connection implied between taking pills to sleep and what it is to be depressed, namely that taking pills to sleep may indicate depression. But this requires a particular understanding of depression, in this case it seems to intimate a condition that people would rather not experience. Whether or not this is how we ought to understand depression is not the issue here, what is at issue is how similar the implicit understanding of both psychiatrist and respondent must be for the question and answer to be mutually understood and the appropriate information conveyed.

To be sure, recognizing the purpose of a question is something we do often and do with ease. But with standardized health assessments there are at least two difficulties with this: first it is widely known and agreed that constructs such as health and well-being—constructs that questionnaires commonly purport to measure—are contentious and difficult, if not impossible to

define.⁴⁵ Pushing the problem of understanding motivation to these concepts doesn't make our problem any easier. Second, researchers sometimes try to hide or neutralize the purpose of their questions, as we saw earlier, in order to deal with the confounding of social norms/desirability. Moreover, researchers often want to use the data from one questionnaire for purposes different than that for which it was originally designed. For example, in the National Audit of Nasal Polyposis and Rhinosinusitis the study was originally designed to distinguish between the levels of performance among different hospitals, but once the data was collected it was also used for prognostic modeling.⁴⁶ Thus researchers sometimes have good reasons to discourage us from making the connection between a particular question and its relation to the overall construct. But if Clark and Schober are right that understanding the purpose of a question is essential to understanding its meaning, and if understanding meaning is essential to giving the correct kind of answers, then it is no surprise that respondents have trouble answering questions according to researchers' understanding of them. I will have more to say about the significance of this problem a little later.

Clark and Schober end their discussion of respondent bias with the following remark, "To understand surveys and the data they produce, we must see survey interviews as a type of discourse...Only then will we resolve many of the puzzles of survey design."⁴⁷ Still, they don't tell us how understanding them as a type of discourse will lead us to these resolutions. They are clear that the principles people use in ordinary language do not get left behind when filling out questionnaires, and they illustrate how an understanding of these principles can lead us to explain and predict the kinds of respondent bias that typically plague researchers. But they don't tell us how this knowledge should help us get rid of or adjust for it. They certainly don't seem to advocate *abandoning* standardized questionnaires, so we are left to speculate on how their insights may help.

One solution that some researchers support with respect to problems of response shift—the phenomena whereby a respondent's internal scale on which they make judgments about their health alters, as their health changes—is to construct an instrument that measures response shift.

⁴⁵ Mallinson, S. (2002), p. 18.

⁴⁶ Personal correspondence with Dr. John Browne, 28 November 2005.

⁴⁷ Ibid, p. 43.

One would then use the measure of response shift in addition to the health assessment measure to get a valid and reliable reading of the measurement's target construct.⁴⁸ One can imagine a similar response with respect to the insights of Clark and Schober: we might construct measures of confounding due to principles of language use and then use these in conjunction with the health assessment in order to adjust for respondent bias. Although I don't know of any attempts of this sort, I want to look at a subjective assessment of cataract patients and argue that even if we *were* able to do this, we would still not prevent issues of invalidity with respect to problems of understanding meaning.

IV

The Visual Function Index and the Problem of Purpose

In the June 2005 issue of the *Canadian Journal of Ophthalmology* Lorne Bellan attempts to answer a question similar to Donovan *et al.*'s in a paper entitled, 'Why are patients with no visual symptoms on cataract waiting lists?' In the case Bellan studies, 30% of patients on waiting lists for cataract surgery had a score on the Visual Function Index (VF-14) of 91 or more out of 100, where 100 indicates no visual complaints. Moreover, there have been multiple studies that indicate that measures of functional impairment are the best indicators of degree of impairment and potential gain from surgery.⁴⁹ Thus 30% of patients are on waiting lists to have cataract surgery even though the subjective assessment questionnaire used to monitor and prioritize this surgery indicates that they don't need it.⁵⁰ Bellan's paper tried to understand this phenomenon better.

The VF-14 is a series of questions asking about the degree of difficulty performing common visual tasks like watching TV or driving a car. Respondents answer the questions on a continuous adjectival scale with four grades ranging from no difficulty to unable to perform. In Winnipeg, Canada all people scheduled to undergo cataract surgery must fill-in a VF-14 before surgery. Between January and May 2002 all patients filling-in the questionnaire and reporting no functional impairment were asked to join this study. 149 individuals agreed and they were

⁴⁸ Sprangers, M. A. G. and Schwartz, C. E. (1999), pp. 1507-15.

⁴⁹ Bellan, L. (2005), p. 434.

⁵⁰ Ibid.

then asked three questions, 1) Are there any other problems with your vision that you are experiencing that I haven't asked you about? 2) Please tell me the reason, as you understand it, why you have been scheduled to have cataract surgery? and 3) What activities do you think will be easier for you after your surgery? Of the 149 patients, 108 were having surgery because of symptoms not specified on the questionnaire, 28 were doing it purely based on the doctor's advice and 13 were asymptomatic.⁵¹

In January 2003 the same patients were contacted again to assess their satisfaction with their surgery. They were asked four questions, 1) How satisfied they were with their vision in the eye that had undergone surgery? 2) Had they found that the vision had been more impaired than they thought before surgery? 3) Did they feel that their vision had improved after cataract surgery? and 4) Would they be willing to repeat this type of surgery again?⁵²

Out of the 149 patients 105 had completed their surgery at this time of these patients 85% were very or extremely satisfied with the surgery, 75% felt that their vision had markedly improved and 94% were willing to repeat the procedure. Only 9% were unsatisfied or said their vision had not improved or would not repeat the surgery.⁵³ The upshot of this study then, is that most of the patients who had no functional impairment according to the VF-14, in fact did have some sort of impairment, and that impairment was great enough that surgical correction created significant benefit.

Thus the VF-14 — a subjective health assessment — underreports morbidity. In fact Bellan concludes that “the VF-14 cannot reliably and accurately identify all patients who are likely to benefit from cataract surgery.”⁵⁴ But he goes on to say that this conclusion is consistent with the findings of the original study in which the VF-14 was developed, since no one ever said it should be the only measure used to determine the need for cataract surgery.⁵⁵ Thus the VF-14's inability to identify all patients likely to benefit from cataract surgery is not evidence of its invalidity as a measure rather it gives us invalid results in this case because it was applied to a situation for

⁵¹ Ibid, pp. 434-5.

⁵² Ibid, p. 435.

⁵³ Ibid.

⁵⁴ Ibid, p. 437.

⁵⁵ Ibid.

which it was not made to handle. The problem is not with the measure itself, but with the decision to use it in this study. How should we characterize that decision?

Bellan implies that the sole use of the VF-14 in the Winnipeg study as a measure to discriminate between those who would and would not benefit from cataract surgery, was inappropriate. Thus the reason why individuals with no visual symptoms are on cataract waiting lists is because in fact they do have visual symptoms, the VF-14 is merely insensitive to those symptoms. Moreover, anyone familiar with the original design of the VF-14 should have known in advance of the Winnipeg study that the VF-14 would not accurately discriminate; as he says these results are consistent with those in the original study in which the VF-14 was developed. This interpretation of the VF-14 in relation to the Winnipeg study suggests that its use was an error in judgment, which could have been avoided if the researchers designing it had done their homework.

But is this the only way to understand what happened in the Winnipeg study? Might we salvage the reputation of the study designers and learn something important about the decision procedure involved in deciding when to use what measure in a particular study? In the introduction of his paper Bellan says that measures of functional impairment are the best indicators of visual impairment and potential gain from surgery. The VF-14 is a measure of functional impairment. Secondly, he leads us to conclude that it's because the VF-14 is a functional measurement and those measurements are the best indicators of impairment and gain from surgery that Winnipeg uses it.

Could it be the case that 'functional impairment' is a vague concept, like health and quality of life, which accurately applies, not only to the VF-14 but to other measures as well? And could it also be the case that although all such measures are measures of functional impairment they nonetheless measure the construct differently (some may measure functional impairment with respect to different health/disease states, some with respect to different kinds of interventions)? We might say that each measure has different strengths and weaknesses and thus in each case 'functional impairment' takes on a different meaning relative to the measure's potential. But if a measure's construct is inherently vague, then it seems that we can only understand a measure's

potential insofar as we use it. Thus it stands to reason that the more we use a measure like the VF-14 the more we will understand its nature as a measure of functional impairment.

In deciding to use the VF-14 in the Winnipeg study a charitable interpretation would suggest that the best evidence at the time of the design suggested that as a measure of functional impairment the VF-14 was the best indicator of impairment and gain from cataract surgery. In the course of the study and through Bellan's research we came to see some of the limitations of the VF-14 as a measure of functional impairment with respect to this population. Although these precise limitations are consistent with the developmental study of the VF-14 they were not prescribed by it. On this interpretation the study designers had no way of knowing in advance that the VF-14 would overlook 30% of the people with visual symptoms. They *may* have had reason to be cautious since they were using it in a slightly different context than it had been used before, but it seems as though they *were* cautious in the sense that the 30% of people who did not show any impairment still received surgery.

I suggest that what we ought to take from the study on the VF-14 is not that researchers made an obvious mistake in applying the VF-14 to the Winnipeg situation, but that they learned more about what the VF-14 does; that they made our understanding of it more determinate. This claim implies that prior to this study our understanding of the VF-14 was vaguer and more indeterminate. The VF-14 is a measure of functional impairment and it was designed for that purpose. But this alone doesn't tell us everything we want to know about it, in fact it doesn't even tell us all the circumstances in which we are licensed to use it.

Earlier I suggested that the questions in the NHP could be made more determinate—and thus we could decrease respondent bias and the consequent invalidity of the measure—if respondents understood their purpose. And then I went on to suggest that the research question or construct was where we should look for that purpose. I still think that this is right, but now we can see the limitations in this suggestion and a major difficulty at the heart of HRQoL or subjective health assessment, namely that the research question or construct is itself vague, even to the researchers who design and implement it.

In itself, the vague nature of the research question or construct is not surprising since much of the subjective assessment literature widely acknowledges the contentious and difficult nature of these constructs. But what is surprising is the sometimes implicit notion that the questions in the questionnaire become clearer as we break them down into smaller chunks, thus even if we aren't entirely sure what functional impairment means the questions in the various dimensions *are* clear. But as I've tried to show, the individual questions in the questionnaires are only as clear as our mutual understanding of them, and this understanding is aided in large part by our recognition of their purpose. When the purpose of these questions is as contentious as the measurement of health, illness and functional impairment then we can't expect the questions themselves to be completely clear—and by clear I mean we can't expect everyone to understand them in the same way.

V

Conclusion

I argued in section II that the source of invalidity in the NHP and SF-36 stems from problems of understanding meaning and interpretation. After looking at the work of Clark and Schober on the principles of language use, I suggested that one possible solution to these problems was to measure deviations of interpretation and to use this measure to adjust the score on the original health assessment. This procedure would then make the measurement valid. But as we saw in the last section with the study of the VF-14 even supposedly valid measures may not provide sound results if they are applied to the wrong task. Thus, even if we could measure discrepancies of interpretation, we would only push the problem back one step. The original problem remains that if we want to draw substantive conclusions from the results of our questionnaires we have to deal with the difficulties interpretation presents, namely understanding the meaning of the questions these questionnaires pose.

To be sure, it seems reasonable that the more we use the VF-14, as in the case above, the clearer researchers will be about what it *can* measure, and consequently what the question to which the VF-14 is the answer means. But using the VF-14 means using it on people, it means asking

people to answer questions, the purpose or context of which is to a certain extent unclear to all involved. Some will view this as an ethics violation: using people as guinea pigs to determine what a construct means, while in the meantime using the measure, with our inadequate understanding of it, in situations which will affect the decisions made about people who are ill. In fact I think this quandary is the one that S. M. Hunt is trying to articulate in her scathing editorial in *Quality of Life Research* where she criticizes the justification of HRQoL measures in terms of operationalization. She writes:

Sometimes the use of a measure will be justified by stating that quality of life is being operationalized in a particular way. However, operationalization is not, normally a part of *applied* research, rather it is part of the process of investigation into a concept, *in order* to develop measurement techniques where appropriate.⁵⁶

Operationalization here refers to the development of a concrete and fruitful research question from a broad and abstract area of interest.⁵⁷ Hunt's criticism, like my suggestion above, is that in quality of life research operationalization is part of the applied work—that researchers further develop the research question as they use the measure. For Hunt the development of the research question should be done before the measure is implemented, not during. With respect to HRQoL questionnaires generally she writes:

...it is virtually unprecedented that a term which lacks clear definition, upon which there is no consensus in the research or the clinical community, which is measured by a motley array of questionnaires of doubtful appropriateness and dubious validity, is actually being applied in situations which will affect the decisions made about people who are ill. This is not only unwise it is clearly unethical.⁵⁸

Hunt's criticism is best understood in the first instance, as an *argument* that this is indeed what occurs, and secondly as a criticism that it shouldn't, since researchers certainly don't see this as how they develop their research questions.⁵⁹ Although Hunt doesn't provide such an argument, I

⁵⁶ Hunt, S. M. (1997), p. 206.

⁵⁷ Green J. and Browne J. (2005), p. 25.

⁵⁸ Hunt, S. M. (1997).

⁵⁹ Ibid. pp. 23-31

have tried to do so in my analysis of the VF-14. But unlike Hunt, the conclusion I want to draw from this argument, is not that researchers ought to *avoid* the operationalization of their research question, but rather that they ought to *embrace* it. Let me explain.

In the analysis of the VF-14 we saw that researchers' understanding of their research question is, to a certain extent, indeterminate. But as we saw in the discussion of the importance of purpose with respect to understanding meaning, an accurate understanding of the research question establishes the context in which the questionnaire's questions get their meaning. If we turn to the analysis of the NHP and the SF-36 we see that the issues of respondent bias fall into three general categories 1) vague or indeterminate words like 'pain' and 'mile' 2) lack of explicit contrasts as in 'health as opposed to what' and 3) an overwhelming number of questions whose problems were due to troubles in conceiving its purpose or motivation. All of these problems are due to difficulties in understanding the questions. To be sure, we can't always blame this on the indeterminate quality of the research question, for as I said earlier, it is often the case that researchers intentionally neutralize the purpose of their research agenda in order to get respondents to answer truthfully or so that the data can usefully serve multiple purposes. But after all we have said, we may argue that this procedure creates more bias than it eliminates. In any case, I think we can solve the problems with respondent bias and make the research question clearer by allowing respondents to question researchers, and researchers to question respondents.

In suggesting this solution I don't have idealistic notions of ideally informed respondents and incredibly open-minded researchers. What I do have in mind are problems with understanding which lead to invalid measures and dubious results. Because researchers aren't entirely clear about the meaning of their research question, they can't be entirely sure what the questions in their questionnaire mean. As we have seen respondents are also unsure what these questions mean. I think they can help each other. In asking questions that result from their basic confusion, respondents can open researchers' eyes to how their questionnaire functions; in getting more precise about the meaning of the individual questions in the questionnaire, they will in turn get a more precise understanding of the research question as a whole and vice versa.

Perhaps we can solve some of our invalidity problems — specifically those that occur due to

complications brought about through interpretation — if we see health assessments, not as answers to questions which have but one meaning, but as facilitators of a dialogue which help us become clearer about what the questions themselves mean. In conceiving of questionnaires in this way, the invalidity found in the NHP and SF-36 should diminish because individuals can make clear what they mean by their answer. Sometimes this clarification will result in a correction of the question's purpose and at other times it may result in the question's elimination, expansion or redirection. The problems found in the study of the VF-14 should also diminish because researchers will have continuous, as opposed to, at best, sporadic, feedback from respondents that will make the research question more precise. By this I mean it will make the conditions in which it is a fruitful measure more precise.

Finally at the end of section III I wondered what lesson we should draw from Clark and Schober's conclusion that respondent bias stems from attempts to apply principles of language use to standardized questionnaires. As should be clear by now, I think we should not compensate for this by trying to make language use standard, but rather use questionnaires as a proxy for language use.

References:

- Bellan, L. (2005), 'Why are patients with no visual symptoms on cataract waiting lists?', *Canadian Journal of Ophthalmology*, 40: 434.
- Clark, H. H. and Schober, M. F. (1992) 'Asking questions and influencing answers', in J. M. Tanur (ed.) *Questions about questions: Inquiries into the cognitive bases of surveys*, New York: Russell Sage Foundation.
- Donovan, J. L. *et al.* (1993), 'Assessing the need for health status measures', *Journal of Epidemiology and Community Health*, 47: 159.
- Gadamer, H. (2003), [1975] *Truth and Method*, 2nd edn, trans. J. Weinsheimer and D. G. Marshall, New York: Continuum.
- Green J. and Browne J. (2005), 'Framing a Research Question' in J. Green and J. Browne (eds) *Principles of Social Research*, New York: Open University Press.
- Hunt, S. M. (1997), 'The Problem of Quality of Life' *Quality of Life Research*, 6: 205-212.
- Mallinson, S. (2002), 'Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire', *Social Science and Medicine*, 54: 11-21.
- Sprangers, M. A. G. and Schwartz, C. E. (1999), 'Integrating response shift into health-related quality of life research: a theoretical model', *Social Science Medicine*, 48:1507-15.
- Tennant, A. (2004) 'Application of Rasch Analysis in the Development and Application of Quality of Life Instruments', *Value in Health*, 7 suppl. 1: S24.
- Van Fraassen, B. (1980), *The Scientific Image*, Oxford: Oxford University Press.
- Wright, L. (2001), *Critical Thinking an Introduction to Analytical Reading and Reasoning*, Oxford: Oxford University Press.