
Abstract
Although texts recommend the generation ‘rich data’ from interviews, no empirical evidence base exists for achieving this. This study aimed to operationalise ‘richness’ and to assess which components of the interview (eg. topic, interviewee, question) were predictive. 400 interview questions and their corresponding responses were selected from 10 qualitative studies in the area of health identified from university colleagues and the UK Data Archive database. Analysis used the text analysis program LIWC and additional rating scales. Richness was operationalised along five dimensions. ‘Length of response’ was predicted by a personal, less specific or positive topic, not being a layperson, later questions, open or double questions; ‘personal richness’ was predicted by being a healthy participant and questions about the past and future; ‘analytical responses’ were predicted by a personal or less specific topic, not being a lay person, later questions, questions relating to insight and causation; ‘action responses’ were predicted by a less specific topic, not being a layperson, being healthy, later and open questions. The model for ‘descriptive richness’ was not significant. Overall, open questions, located later on and framed in the present or past tense tended to be most predictive of richness. This could inform improvements in interview technique.

Abstract word count: 189 words

Document word count: 6,178 words

Key words: interviews, rich data, qualitative methods
The role of topic, interviewee and question in predicting ‘rich’ interview data in the field of health research

Introduction

Over the past few decades there has been an increased interest in the use of qualitative methods across the social sciences, particularly psychology and sociology (Richardson, 1996; Miles and Huberman, 1994; Silverman, 1997; Seidman, 1998; Seale, 1999; Smith, 2003; Coyle, 2007; Seale, Silverman, Gubrium and Gobo, 2007) with greater opportunities for qualitative-based PhDs and the incorporation of qualitative methods into undergraduate and postgraduate courses. This interest has coincided with the rise of health related research and qualitative methods have been used across a number of diverse aspects of health to explore patients’ experiences of an array of physical illnesses, the reasons behind their health related behaviours, their use or non use of health services and in turn to explore health professionals’ own practices.

Qualitative analysis, in contrast to the numerical coding of quantitative data, generally involves collecting data in its natural verbal or written form, gaining a high level of insight into how individuals make sense of events and experiences (Coyle, 2007; Smith, 2003; Stangor, 2004). It is also acknowledged that the researcher plays a vital role in bringing their own interpretation to the research process, interacting with that of the participant (Burman, 1994; Coyle, 2007; King, 1996; Parker, 1994; Rapley, 2001). Qualitative methods facilitate a deep understanding of participants’ inner mental processes and reveal phenomena that may not have emerged from quantitative study (Denzin, 2000). They have been criticised, however, for lack of objectivity and traditional scientific control, and the possibility for researchers to interpret meaning that
There are many different forms of qualitative data which elicit different analysis techniques and are founded upon different epistemological positions ranging from constructionist through to positivist (Silverman, 1997; Ashworth, 2003; Smith, 2003; Seale et al, 2007). In discourse analysis, for example, the focus is on the way in which individuals specifically construct their speech in interactions with an emphasis on the function of speech rather than what it can tell us about their underlying thoughts (Potter and Wetherell, 1987; Parker, 1999). Conversation analysis is similarly constructionist in its focus but takes a more micro analytical approach emphasising the dynamics between people and addresses factors such as the format of questions and their answers and the role of pauses and laughter (Atkinson and Heritage, 1984; Roulston, 2006). From a more positivist perspective, Interpretative Phenomenological Analysis (IPA; Smith & Osborn, 2003) explores the meanings people ascribe to circumstances in their lives, whilst Grounded Theory (Glaser and Strauss, 1967; Glaser, 1992) employs qualitative data to develop theory and relates to both positivist and interpretative approaches.

Although deriving from different epistemological positions most qualitative approaches involve the use of interviews which can be transcribed and closely analysed to reveal patterns of interaction between interviewee and interviewer and / or patterns and divergences between interviewees (Brenner, 1985; Brenner et al, 1985; Smith, 1995; Seale et al, 2007). These different perspectives, commonly, all encourage the generation of ‘rich data’ which is also referred to as ‘thick description’, ‘textured accounts’ or simply as ‘detail’ (Geertz, 1973; Silverman, 1997; Glaser, 1992). Furthermore, although not always explicit, texts on qualitative methods provide guidelines as to how rich data
can be achieved. For example, Smith (1995) suggests planning the discussion of sensitive subjects later on in the interview and beginning with more general questions, this technique being described as ‘funnelling’. These suggestions are also recommended by DiCicci-Bloom and Crabtree (2006) and Rapley (2001). Similarly, it is argued that general questions at the beginning can ease the interviewee into the interview situation and as they become more confident more specific insights can be revealed (Smith, 1995). As much as possible, researchers are also advised to use open rather than closed questions (Burman, 1994; King, 1996; Silverman, 1997; Smith, 1995; Rapley, 2001) as expanded answers are preferable to monosyllabic responses. Although initial general questions are advised, Breakwell (2006) recommends avoiding ‘catchall’ questions, for example asking everything a participant knows about a certain topic. Such questions are too broad and might even panic respondents. Question wording is also suggested to have great influence on the outcome of an interview. Certain questions may help interviewees orient themselves in the interview or to a particular event, for example ‘how’, ‘what’, ‘when’ and ‘why’ questions (Charmaz, 2003). Question wording might also threaten the validity of the interview situation; Smith (1995) and Breakwell (2006) advise that questions should not lead or have statements of value implicit within them as participants may feel that they will be judged on their responses. Facilitating participants’ responses can also be achieved by ensuring they understand the topic and aiming questions at their level of language (Smith, 1995). In addition, several authors suggest that jargon and complex terminology should be avoided (Breakwell, 2006; Smith, 1995; Wragg, 1978). Similarly, asking more than one question at a time may also confuse the interviewee (Smith, 1995) and double negatives in question wording can be difficult for participants to interpret (Breakwell, 2006).

In sum, research within the social sciences increasingly uses qualitative methods, particularly interviews which should be carefully constructed as a means to generate rich
data and numerous guidelines are provided as to how this can be achieved. To date however, there is no existing empirical evidence base for these guidelines and the assumptions appear to be based on academic consensus and ‘common sense’. Indeed, whilst acknowledging the unique benefits of qualitative interview techniques, Collins et al. (2005) called for an increase in systematic research of the factors which might influence the outcome of such studies. One study which has attempted to systematically investigate influences on interview outcomes is that of Dijkstra, van der Veen and van der Zouwen (1985). In a double-blind experiment, interviewers were trained in either a formal task-oriented style or a socio-emotional style of interviewing (Dijkstra et al., 1985). A total of 384 interviews were conducted and analyses revealed that interviewees in the socio-emotional interviews talked more and disclosed more personal information than those in the formal task-oriented interviews (Dijkstra et al., 1985). Therefore the present study aimed to evaluate what aspects of the interview resulted in data that could be conceptualised as data that was 'rich'. In particular, in line with the guidelines in this area the study focused on the role of topic area, interviewee and framing of question.

Defining ‘richness’ however is problematic as although most qualitative texts refer to the concept of ‘richness’ (Ashworth, 2003; Breakwell, 2006; Charmaz, 2003; Coyle, 2007; DiCicco-Bloom & Crabtree, 2006; Harris & Huntington, 2001; King, 1996; Rapley, 2001; Smith, 2003; Smith, 1995; Stangor, 2004) only one could be found that provides a definition:

‘Rich data reveal participants’ thoughts, feelings, intentions, and actions as well as context and structure … (they) afford views of human experience that etiquette, social conventions, and inaccessibility hide or minimise in ordinary discourse’ (Charmaz, 2003).

The present study also, therefore aimed to develop an operationalisation of ‘rich data’ that could be tested empirically.
Method

Design

The study used a cross-sectional design to evaluate whether aspects of the predictor variables (topic, interviewee, question) predicted ‘rich data’ in health related research interviews.

Sample

A sample of 400 interview questions and their corresponding responses were identified as follows:

i) 10 university colleagues working in the area of health research who use qualitative methods were emailed and asked to supply any interviews relating to health research. In addition, the UK Data Archive [UKDA] (accessed at http://www.data-archive.ac.uk) was searched for interview studies using the term ‘qualitative health’. This generated a pool of 41 research studies relating to aspects of health research. From this a stratified sample of 10 studies was selected to reflect heterogeneity of topic, discipline (psychology, sociology, health services research) and demographics of interviewee. Studies were eligible for inclusion if they included full interview transcripts of interviews with adults and were conducted in English. Unfortunately the demographics of the interviewer were not available for those studies identified from the data archive so this variable was not included in the analysis.

ii) Five transcripts were selected at random from each study using an online random number generator.
iii) A systematic sample of eight questions and their corresponding responses were selected from each transcript: four questions from the beginning; two from the middle; and two from the end of each transcript.

iv) Microsoft WordPad was used to save the questions and responses in a text format suitable for analysis.

The transcripts

The resulting 10 sets of transcripts were from a range of studies exploring the following topics: commissioning in primary care, GPs’ behaviour, choice in Primary Care, young men and health, people with multiple problems and needs and their experiences of the labour market, experiences of nuclear medical procedures, health technology at women’s midlife, incentives in Primary Care, experiences of obesity surgery, domestic cooking and cooking skills.

Coding the transcripts

The topic, interviewee, and questions were the predictor variables and the responses the outcome variables. All variables were coded manually by two researchers through discussion concerning each code. The questions and responses were also coded electronically using Linguistic Inquiry and Word Count (LIWClite7 version 1.02, Pennebaker et al., 2007). LIWC is a text analysis program which provides a word count and also the degree to which people use words from around seventy language groups (Pennebaker et al., 2007) such as positive or negative emotional words, family and friendship-related words, self-references, and words relate to health, death and religion. The choice of coding frame was derived from the guidelines in the literature concerning strategies to generate ‘rich data’ and through an analysis of what constituted ‘rich data’ (see later for definition).
Predictor variables

The topic, interviewee and questions were coded as follows (explanations are given where not entirely clear):

**Manual coding**

**Topic:** i) Personal: on a scale of not personal (1) (eg. Commissioning in Primary Health Care) to extremely personal (5) (eg. Experiences of obesity surgery); ii) Specific: on a scale of not specific (1) (eg. Choice in primary care) and highly specific (5) (eg. Domestic cooking and cooking skills); iii) Positive: on a scale of not positive (1) (eg. Experiences of nuclear medical procedures) and very positive (5) (eg. Note no interviews were totally positive but some eg. Incentives in primary care were quite positive).

**Interviewee:** i) Patient: not patient (0) and patient (1); ii) Professional: not health professional (0) vs health professional (1); iii) Layperson: not layperson (0) or layperson with relatively little medical knowledge or experience (1); iv) Health status: the health status of the interviewee was coded not healthy (0) (eg. Those who had experienced nuclear medical procedures) or healthy (1) (eg. Doctors or those describing cooking skills).

**Question:** i) Position: each question was coded as beginning (1), middle (2), end (3); ii) Openness: coded as closed (0) or open (1). An open question was defined as one which left the interviewee to construct an answer themselves (eg. ‘What do you think about a self-help group for people who have a weight problem?’). A closed question generally indicated a shorter answer and had some implicit indication of categories, for example yes/no responses, a number or quantity, or a single noun (eg. ‘How old is she?’); iii) Leading questions: non-leading (0) (eg. ‘Why did you decide to have the operation?’)
and leading (1) (eg. ‘Do you generally try to have a healthy diet?’); iv) **More than one question**: only one question was asked (0) (eg. ‘So do you think the government’s doing enough for people like you?’) or 2 or more questions (1) (eg. ‘What do you think about the staff in the hospital when you had the operation? How do you rate the hospital and doctors? If something was not so good...what would you like to change in the way they dealt with you?’); v) **Narrative question**: non-narrative question which focused on one particular time point (0) and narrative question which requested some sort of sequence in their response (1) (eg. ‘Run through the day very roughly what you would do?’).

**LIWC coding**

**Questions**: The questions were also electronically coded in terms of the following: i) **Word count** ii) **Common verbs** – questions that ask people about things they do or have done may lead to richer responses (eg. walk, went, see) iii) **Tense** (coded separately as past / present / future); iv) **Insight** – questions that ask interviewees to make insights may lead to richer data (eg. think, know, consider); iv) **Causation** (eg. because, effect, hence); v) **Perceptual processes** (eg. observing, heard, feeling).

**Outcome variables**

**Responses** were coded electronically as follows: i) **Word count** ii) **First person singular** (eg. I, me, mine); iii) **First person plural** (eg. we, us, our); iv) **Common verbs** (eg. walk, went, see); v) **Adverbs** (eg. very, really, quickly); vi) **Quantifiers** – providing quantification in responses increases their richness (eg. few, many, much); vii) **Affective processes** – talking about affect reflects richness. (eg. happy, cried, abandon); viii) **Positive emotion** (eg. love, nice, sweet); ix) **Negative emotion** (eg. hurt, ugly, nasty); x)
Anxiety (eg. worried, fearful, nervous); xii) Anger (eg. hate, kill, annoyed); xiii) Sadness (eg. crying, grief, sad); xiv) Insight (eg. think, know, consider) xv) Causation – providing some analysis of causation reveals a greater detail of thought processes, indicating richness (eg. because, effect, hence) xiv) Perceptual processes (eg. observing, heard, feeling).

Defining and operationalising richness

Although the many texts on qualitative methods emphasise the importance of collecting ‘rich’ data only one clear definition of richness could be found relating to ‘thoughts, feelings, intentions, and actions as well as context and structure …’ (Charmaz, 2003). In line with this richness was operationalised in terms of emotional expression, behaviour, and descriptions of both emotion and behaviour in terms of adjectives, adverbs and quantifiers. In addition, the literature also implicitly describes other elements of data as contributing to richness such as length of response, the use of personal pronouns, expressions of insight and causation and expressions of certainty to reflect strength of belief. In line with this the 16 aspects of the responses were summated to provide five dimensions of ‘richness’ which are described below.

i) Length: (consisting of word count only);

ii) Descriptive: this reflected the degree of emotion and description (quantifiers, affective processes, positive emotion, negative emotion, anxiety, anger, sadness, perceptual processes, adverbs);

iii) Personal: to reflect how personal the response was (first person singular, first person plural);

iv) Analytical: this reflected the level of analysis in the response (insight, causation);
v) **Action:** included to describe descriptions of behaviour (common verbs, adverbs);

**Relationship between these dimensions**

In order to explore whether a ‘total’ richness score could be computed the five dimensions were analysed using factor analysis and Cronbach’s alpha.

**Factor analysis:** The results from the factor analysis with oblimin rotation produced a 3 factor solution which converged after 17 iterations. Factor 1 consisted ‘personal’ (loading 0.89), factor 2 consisted of ‘descriptive’ (loading 0.9) and ‘Action’ (loading -0.7) and factor 3 consisted of ‘analytical’ (loading 0.61) and length (loading 0.85). From this analysis it seemed that the five dimensions could not be grouped in any meaningful way.

**Cronbach’s alpha:** To clarify the association between the variables Cronbach’s alpha was computed. For all five variables alpha was unacceptable at 0.02. If ‘length’ was removed it remained unacceptable at 0.4.

As a result of this analysis the five dimensions for ‘richness’ were analysed individually.

**Data analysis**

The data were analysed using SPSS 16.0 to describe the data and to assess the role of predictor variables (**Topic** (personal; specific; positive); **Interviewee** (patient; professional; layperson; health status); **Question** (position; narrative; length; verbs; tense (past; present; future); insight; causation; perceptual; openness; leading; more than one
question)) in predicting the five defined areas of ‘richness’ (response length; descriptive richness; personal richness; analytical richness; and action richness).

**Results**

i) Descriptive statistics

**Topic:** The median scores for topics were i) personal score 2.50 (IQR = 2.00); ii) specific score 3.50 (IQR = 2.00); iii) positive score 3.00 (IQR = 1.00).

**Interviewee:** Half of the questions / responses analysed were conducted with lay people (n=200), 30% (n = 120) with health professionals, and 20% (n = 80) with patients. Almost three-quarters (n = 288, 72%) were healthy while the remainder (n = 112, 28%) had some form of health-related issue.

**Questions:** Half of the sampled questions (n = 200) were positioned at the beginning of the transcripts, a quarter (n = 100) from the middle, and a quarter (n = 100) from the end. Half the questions (n = 197, 49.2%) were open questions and half (n = 203, 50.8%) were closed questions. Just under a quarter of the questions were leading (n = 94, 23.5%), with the remainder being non-leading (n = 306, 76.5%). 6.5% (n = 26) of the questions were double-barrelled, containing more than one question, but the remaining 93.5% (n = 374) asked only one question. Narrative questions made up 2.2% (n = 9) of the sample, with the remaining 97.8% of questions (n = 391) not requesting a narrative account.

The data from the LIWC analysis of the questions is shown in Table 1.

- -insert table 1 about here -

**Predicting ‘richness’ of data**
Multiple regression analyses were conducted to explore the role of predictor variables in predicting the five dimensions of richness. The health professional variable violated the assumption of multi collinearity and was removed from analyses by SPSS. Descriptive statistics for the dimensions of richness are shown in Table 2. The results of the Multiple Regression analyses are shown in Table 3.

-insert tables 2 and 3 about here -

i) Richness as length

The model predicting richness as length was significant ($F(19,380) = 6.314, p<.001$) accounting for 20.2% of the variance. The results showed that an increase in ‘richness’ as defined by length was predicted by a more personal topic, a less specific topic, a more positive topic, not being a layperson, questions later on in the interview, questions being more open, and questions that contained more than one question.

ii) Richness as descriptive: The model predicting richness as descriptive was non-significant ($F(19,380) = 1.093, p>.05$) accounting for only 0.4% of variance.

iii) Richness as personal: An increase in ‘richness’ defined as personal was predicted by participants being healthy, questions being related to the past, and questions related to the present ($F(19,380) = 2.789, p<.001$) accounting for 7.8% of the variance.

iv) Richness as analytical: The model predicting richness as analytical was significant ($F(19,380) = 4.264, p<.001$) accounting for 13.5% of the variance. An increase in ‘richness’ defined as analytical was predicted by a more personal topic, a less specific
topic, questions later on in the interview, questions containing references to insight, and questions referring to causation.

v) **Richness as action:** An increase in richness defined as action was predicted by a less specific topic, not being a layperson, being a healthy interviewee, questions later on in the interview, and questions being more open \((F(19,380) = 3.347, \ p<.001)\) accounting for 10.1% of the variance.

**Discussion**

Most qualitative texts describe the importance of carrying out interviews in such a way as to produce ‘rich’ data (eg. Ashworth, 2003; Breakwell, 2006; Charmaz, 2003; Coyle, 2007; DiCicco-Bloom & Crabtree, 2006; Harris & Huntington, 2001; King, 1996; Rapley, 2001; Smith, 2003). To date, however, there is no empirical evidence as to how ‘rich data’ can be best achieved. The present study explored the role of aspects of the topic, interviewee and question in predicting ‘richness’ of response defined in terms of length, personal richness, descriptive richness, analytical richness and action. These five dimensions were deemed to be relatively statistically discrete suggesting that generating ‘rich’ data is dependent upon which component of ‘richness’ is judged most appropriate to any particular research topic.

Perhaps the most obvious definition of ‘richness’ is response length with many qualitative texts describing strategies designed to encourage interviewees to speak openly and at length about the interview topic and the analysis indicated that this definition had the greatest number of significant predictors. The results from the present study
indicated that a longer response was predicted by either a personal or a positive topic and one that was less specific, an interviewee who wasn’t a lay person and a question that was towards the end of the interview, open and double barrelled. This provides support for some of the strategies highlighted in the literature which suggest that questions should be open and that those later on in the interview may be more productive as the interviewee has had time to relax (Burman, 1994; King, 1996; Smith, 1995; Rapley, 2001). In addition, a personal topic may enable the interviewee to draw upon experience and to feel more confident and a less specific topic may encourage the interviewee to speak more broadly. Furthermore a more positive topic may enable interviewees to elaborate without having to manage any negative emotions. In addition, lay people may produce shorter responses if they feel less experienced with the interview scenario. In this study, questions which contained more than one question were also predictive of longer responses which contrasts with the instructions of Smith (1995) who suggests that asking more than one question at a time can confuse participants. In this instance, the interviewees may simply have given two answers in one response which would therefore be longer or may have been confused giving more redundant or irrelevant information. Further analysis on the content of the responses is needed to clarify whether double barrelled questions are indeed related to ‘richness’ or not.

Aspects of the question were also predictive of ‘personal richness’ involving references to the self (eg. ‘I’ and ‘we’) with more self references being related to a question that involved the past or present tense (but not future). Using a focus on the past enables participants to describe their own personal chronology and therefore might engender a more personal response as could might a focus on the present, whereas a question asking the individual to project into the future may encourage them to distance themselves from the topic and / or to produce a more limited and muted response. A more personal
response was also greater in healthy rather than unhealthy participants, which is a curious
finding as one might expect unhealthy participants to have more rather than less to say in
reference to themselves. Perhaps talking about health problems for those that are
unhealthy is easier if they objectify their condition and make fewer personal references
whereas healthy participants may be less motivated to distance themselves from the
interview topic.

Analytical richness involving both references to insight and causation was predicted by
questions which directly asked about insight and causation and those which were later on
in the interview, topics that were personal but not specific and a participant who was not
a lay person. The association between insight and causation in both the question and
response indicates that direct questioning about these components is the best strategy for
generating insightful and causal answers, particularly if such questions are asked after the
interviewee has had time to relax and settle into the interview. It also illustrates a
process of mirroring between interviewer and interviewee. The impact of a personal
topic may reflect the participants’ willingness to analyse subjects that are personal to
them and a more general topic may allow for more speculation about issues of causality.

Finally, for richness as ‘action’ it would seem that people use more action words if they
are asked open questions that are later on in the interview about a less specific topic and
if they are healthy and not a lay person. A focus on behaviour and the tendency to
elaborate descriptions of behaviour may be more common when the topic is general as
participants are less restricted in what they can talk about and non lay people may be
more focused on behaviour rather than beliefs or events. Interestingly, richness as
‘action’ was predicted by interviewees being healthy, perhaps quite literally because
healthy participants were in the position to do more, whereas unhealthy participants might be involved in less activity and dwell more on other aspects of their experiences in interviews.

The results however showed no significant predictors of richness as defined as descriptive which consisted of emotional terms and descriptions of behaviour. There are several possible explanations for this. First it is possible that descriptive responses are unrelated to aspects of the interview question. This seems unlikely as simple factual questions are predictably less likely to produce description than open questions. Second, description may be related to aspects of the question that were not measured in the present study; it may relate to the demeanour of the interviewer or the mood and emotional state of the interviewee. Third, it is possible that the degree of description in the responses analysed for the present study was too specific to the subject matter under scrutiny resulting in too little variance by question characteristics. Future research needs to explore these possibilities.

There are, however, some issues with the present study that need to be considered. The first relates to the use of an empirical and statistical approach to evaluate interviews which are predominately a tool used within less traditional qualitative methodologies. This contradiction may not sit easily with more constructionist researchers. The present study, however, is positioned within an epistemological framework whereby interviews should be carried out in the most ‘productive’ way possible and therefore provides some insights into how this may be done. Secondly, the study analysed questions and their responses as independent of each other rather than the dynamic interaction between the two which conflicts with approaches emphasised by either discourse or conversational
analysts. Although, however, in some ways at odds with these perspectives the present study does provide statistical support for the notion that responses are strongly influenced by the phrasing of questions. Finally, there are a number of other factors that were not considered within the design of the present study including the demographics of the interviewer, non verbal communication and mood which may well also influence the ‘richness’ of the data produced.

Overall, however, the results provide some evidence for a relationship between aspects of the topic, interviewee and question and the resulting richness of the response. These results have implications both for further research and for the development of strategies to maximise the richness of any data generated by an interview. In terms of further research, a more detailed micro analysis of questions could be carried out to assess the impact of individual words such as ‘why’, ‘when’ or ‘how’ or the role of pauses and silence. Further, audio or video tapes could be evaluated to assess the impact of more subtle factors such as tone of voice, smiling and body language and the environment in which the interview takes place (King, 1996; Wragg, 1978; Harris and Huntington, 2001). In addition, research could explore the structure of the interview schedule in greater depth to assess the relative of impact of questions versus prompts and the positioning of each question within the broader narrative of the dialogue between interviewer and interviewee. It would also be possible to evaluate the impact of matched or unmatched demographics between the interviewer and the interviewee (Breakwell, 2006; Boutain and Hitti, 2006).

In terms of the development of strategies the results from the present study do provide a preliminary evidence base for guidelines designed to generate rich data. In part the
results indicate fairly consistent support for the use of open questions and those later on in the interview as a means to generating rich data which provides supports for suggestions by several authors writing about qualitative methods (Burman, 1994; King, 1996; Smith, 1995; Rapley, 2001). In addition, the results indicate that questions relating to either the past or present may be more productive than those relating to the future. Furthermore, the results indicate that if the interview asks for analysis then this is what they are most likely to receive.

The results therefore showed some consistent patterns relating to open questions located later on in the interview and phrased in the present or past tense. There was also, however, some variability relating not only to the demographics of the interviewee and their health status, but also the nature of the interview topic and aspects of the question. This suggests that any defined strategies to generate richness must be dependent upon what kind of rich data the researcher wishes to generate which in turn must depend upon the specific aims of any given research study. In the present study five dimensions of richness were operationalised relating to the length and content of the response. These were shown to be statistically discrete indicating that richness may not be a uni dimensional construct. Accordingly, the guidelines for producing one dimension of richness may not be the same for producing another dimension. Therefore the design of any interview schedule should incorporate strategies to maximise the richness of the interviewees’ responses according to how richness would be defined for that specific study.

To conclude, the present study aimed to provide a preliminary empirical evidence base for the generation of rich data from qualitative interviews and illustrates some support for a role for aspects of the interviewee, topic and question. The study also laid a basis for
an operationalised definition of richness which consists of five dimensions relating to length, description, personal content, analysis and action. If there are some universal strategies concerning the generation of rich data, then the results from the present study highlight a role for open questions, positioned towards the end of the interview and the framing of questions in the past or present tense rather than the future. The results however also indicate much variability according not only to the demographics of the interviewee and the topic but also the question and suggest that any guidelines for the generation of ‘rich’ data must be specific to the aims of the research question. There may be some ways in which rich data can be consistently generated, but this also depends upon which operationalisation of rich is the focus of each particular research study.
References


Table 1: Descriptive statistics for predictor variables calculated in LIWC.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question length</td>
<td>400</td>
<td>18.73</td>
<td>15.36</td>
<td>3-121</td>
</tr>
<tr>
<td>Question verb</td>
<td>400</td>
<td>19.14</td>
<td>9.54</td>
<td>0-50</td>
</tr>
<tr>
<td>Question past</td>
<td>400</td>
<td>5.37</td>
<td>6.88</td>
<td>0-33.33</td>
</tr>
<tr>
<td>Question present</td>
<td>400</td>
<td>11.86</td>
<td>8.37</td>
<td>0-44.44</td>
</tr>
<tr>
<td>Question future</td>
<td>400</td>
<td>0.98</td>
<td>2.98</td>
<td>0-28.57</td>
</tr>
<tr>
<td>Question insight</td>
<td>400</td>
<td>3.70</td>
<td>5.10</td>
<td>0-25.00</td>
</tr>
<tr>
<td>Question causation</td>
<td>400</td>
<td>3.11</td>
<td>5.27</td>
<td>0-30.00</td>
</tr>
<tr>
<td>Question perceptual processes</td>
<td>400</td>
<td>1.97</td>
<td>3.95</td>
<td>0-28.57</td>
</tr>
</tbody>
</table>
Table 2: Descriptive statistics for richness outcome variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richness as length</td>
<td>400</td>
<td>97.43</td>
<td>146.62</td>
<td>1-1590</td>
</tr>
<tr>
<td>Richness as descriptive</td>
<td>400</td>
<td>2.04</td>
<td>1.62</td>
<td>0-20.00</td>
</tr>
<tr>
<td>Richness as personal</td>
<td>400</td>
<td>3.97</td>
<td>3.18</td>
<td>0-20.00</td>
</tr>
<tr>
<td>Richness as analytical</td>
<td>400</td>
<td>2.21</td>
<td>2.20</td>
<td>0-16.66</td>
</tr>
<tr>
<td>Richness as action</td>
<td>400</td>
<td>10.39</td>
<td>4.80</td>
<td>0-30.76</td>
</tr>
</tbody>
</table>
Table 3: Predictors of ‘richness’ (standardised Beta / p value).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Length</th>
<th>Descript.</th>
<th>Personal</th>
<th>Analysis</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal</td>
<td>.66 ***</td>
<td>.36</td>
<td>.32</td>
<td>.37*</td>
<td>.34</td>
</tr>
<tr>
<td>Specific</td>
<td>-.46**</td>
<td>-.05</td>
<td>.12</td>
<td>-.34*</td>
<td>-.4*</td>
</tr>
<tr>
<td>Positive</td>
<td>.31**</td>
<td>.19</td>
<td>.02</td>
<td>.07</td>
<td>.01</td>
</tr>
<tr>
<td>Patient</td>
<td>-.22</td>
<td>-.29</td>
<td>.11</td>
<td>-.08</td>
<td>.05</td>
</tr>
<tr>
<td>Layperson</td>
<td>-.86***</td>
<td>-.25</td>
<td>.05</td>
<td>-.35*</td>
<td>-.56**</td>
</tr>
<tr>
<td>Health status</td>
<td>.01</td>
<td>-.36</td>
<td>.41**</td>
<td>.23</td>
<td>.34*</td>
</tr>
<tr>
<td>Position</td>
<td>.16**</td>
<td>.13*</td>
<td>-.09</td>
<td>.13**</td>
<td>.15**</td>
</tr>
<tr>
<td>Length</td>
<td>.05</td>
<td>-.007</td>
<td>.04</td>
<td>.07</td>
<td>.02</td>
</tr>
<tr>
<td>Verbs</td>
<td>.05</td>
<td>.06</td>
<td>-.21</td>
<td>.16</td>
<td>-.3</td>
</tr>
<tr>
<td>Past tense</td>
<td>-.08</td>
<td>-.04</td>
<td>.32**</td>
<td>-.09</td>
<td>.01</td>
</tr>
<tr>
<td>Present</td>
<td>-.003</td>
<td>-.1</td>
<td>.29*</td>
<td>-.12</td>
<td>.06</td>
</tr>
<tr>
<td>Future</td>
<td>-.07</td>
<td>.005</td>
<td>.06</td>
<td>-.08</td>
<td>.06</td>
</tr>
<tr>
<td>Insight</td>
<td>-.002</td>
<td>-.04</td>
<td>-.02</td>
<td>.19***</td>
<td>-.04</td>
</tr>
<tr>
<td>Causation</td>
<td>.023</td>
<td>.06</td>
<td>.003</td>
<td>.12*</td>
<td>-.03</td>
</tr>
<tr>
<td>Perceptual</td>
<td>-.04</td>
<td>-.01</td>
<td>-.002</td>
<td>.09</td>
<td>-.07</td>
</tr>
<tr>
<td>Openness</td>
<td>.16**</td>
<td>.08</td>
<td>.005</td>
<td>.09</td>
<td>.17**</td>
</tr>
<tr>
<td>Leading</td>
<td>.04</td>
<td>.05</td>
<td>-.02</td>
<td>.03</td>
<td>.09</td>
</tr>
<tr>
<td>&gt; than one</td>
<td>.14**</td>
<td>.01</td>
<td>-.02</td>
<td>-.03</td>
<td>.02</td>
</tr>
<tr>
<td>Narrative</td>
<td>.09</td>
<td>-.02</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
</tr>
</tbody>
</table>

* Significant at p<.05; ** Significant at p<.01; *** Significant at p<.001