A blind spot in higher education research? A call for the explicit use of qualitative survey validation methods

Cathryn McCormack
Southern Cross University, Lismore, Australia
cathryn.mccormack@scu.edu.au

This paper draws attention to the potential for using qualitative pretesting techniques for increasing the validity of questionnaires returning quantitative data. Using a sample of all higher educational research publications from HERDSA 2008-2011, a total of 69 questionnaires for collection of quantitative data were identified. Of these, 18 (26%) were researcher developed questionnaires where no validation processes were reported, 14 (20%) used an institutional questionnaire, and 20 (29%) used an existing validated instrument. No examples of qualitative pretesting were identified. In all cases research strength could have been increased by improving survey validity through the use of documented qualitative pretesting techniques, thereby facilitating connections through the research. The sample raises particular concerns for the questionnaires where no validation of any kind was reported, and that validation studies were rarely cited for institutional surveys. The paper concludes with a call for higher education researchers to recognise this blind spot and take advantage of qualitative validation processes.

Keywords: survey, validation, pretesting

Introduction

Higher education research frequently involves the use of surveys. These require validation in order for the research findings to be treated with confidence, and thereby allow deeper exploration of the issues being investigated and connection with the research audience. Literature on validating surveys is divided into two main streams. One stream takes a distinctly quantitative approach, focusing on psychometric validation through statistical analysis. The other is based in cognitive psychology and the social sciences and focuses on a variety of qualitative pretesting techniques. Most researchers are acquainted with the quantitative approach, although may be unfamiliar with the specialist knowledge and skills involved. Knowledge of qualitative pretesting techniques, on the other hand, appears to be much less common, with the few educational researchers who have used them commenting on their infrequency of use (Desimone & Le Floch, 2004; McCormack, 2008). The division between these two streams is distinct in the literature, with those using quantitative approaches rarely undertaking more than limited qualitative pretesting, and those using qualitative pretesting rarely undertaking more than limited quantitative data analysis. Furthermore, no publications have been identified where the outcomes from qualitative and quantitative methods of survey validation are compared.

When undertaking survey research we aim to produce data that accurately reflects the concepts and behaviours we aimed to measure and that does not contain too much random variability. We also want our questions be sensitive enough to measure important real differences or changes (Collins, 2003). In other words, we want our research to be valid. Validity is being increasingly called for in higher education management, as seen in requirement for “regular, valid and
reliable feedback” in the TEQSA Provider Standards (Australian Government, 2011, Chapter 1, Section 5.6). If this is the standard required in higher education management, then should not higher education research aim somewhat higher?

Background

The survey as conversation

The concept of the ‘survey as conversation’ arose from the design of questionnaires for face to face delivery in the social sciences. However, even with paper based and online delivery, Dillman (2007) advises imitating the natural flow of a conversation as much as possible. Following this advice encourages the researcher to consider the questionnaire from the respondent’s perspective, and consider how instructions, layout, section order and question wording might impact on the types of responses received. Furthermore, a survey designed as conversation, albeit one-sided, encourages connection between the respondent and the topic as well as between the researcher and respondent. This is likely to result in a higher response rate and greater thought on the part of respondents in constructing their answers.

The most common qualitative pretesting method, the cognitive interview, provides a method for researchers to ensure their questionnaire follows as closely as possible the norms of conversation, in the process identifying both overt and covert problems with the questions. In a cognitive interview the respondent is asked to think aloud as they attempt to complete the questionnaire (Willis, 2005). The exact protocol may be highly structured where interviewers are provided with a script and specific probing questions to use throughout the questionnaire, or unstructured where interviewers ask questions as they deem appropriate. The greatest strength of cognitive interviews is their ability to assist in identifying covert problems, such as where a difference in interpretation of a question between respondents and researcher differs but is not immediately apparent. Cognitive interviews are particularly useful where a questionnaire will be used for culturally and linguistically diverse respondents.

Terminology

It is important to clarify terminology before progressing the discussion. The terms survey and questionnaire are frequently used interchangeably. To add more confusion, the term survey instrument is also used. This paper uses the conventions laid out in many introductory books (e.g., Peterson, 2000; Punch, 2003; Thomas, 2004), that the term survey refers to the entire research process from developing a questionnaire that addresses the research question, sampling, delivery, data collection and analysis; the term survey instrument refers to the questions asked; and the term questionnaire refers to the delivered instrument. Questionnaire therefore incorporates elements such as graphic design, online implementation and skip logic.

This paper is concerned with survey instruments and questionnaires, which will be referred to as questionnaires throughout, rather than the larger survey research process. It will introduce issues in questionnaire validity and cognitive interviews as a qualitative pretesting method. It will then report on the frequency of reported use of any types of qualitative pretesting methods from a sample of educational research literature. Finally, researchers are presented with recommendations for using and reporting on surveys.
Validity

In creating a questionnaire, researchers start from the position of designing questions that address the research question(s). A questionnaire that is based in the relevant literature, clearly addresses the research questions and makes sense to the researcher has *face validity*. However, a plethora of evidence shows that face validity is insufficient to claim the results are valid (e.g., Converse & Presser, 1986; Payne, 1951; Schwarz & Sudman, 1996; Sudman, Bradburn, & Schwarz, 1996; Toepoel, Vis, Das, & van Soest, 2009). Just as qualitative researchers should implement verification strategies throughout the research process for improved reliability and validity (Morse, Barrett, Mayan, Olson, & Spiers, 2002), qualitative validation of surveys can also take place during the research process and prior to delivery of the questionnaire.

Punch (2003) says, “Validity means whether the data represents what we think they represent,” and “reliability means stability of response. ... We want survey questions which produce stable responses and that depends in part on whether the questions can be consistently and straightforwardly answered using the response scales and alternatives provided.” Although the terms reliability and validity are frequently used together, it is possible to develop a questionnaire that is reliable without being valid. For instance, a questionnaire delivered for several offerings before and after an educational intervention may show an increase from consistently lower ratings to consistently higher ratings. It is therefore likely to be reliable, as it is clear that the ratings are consistent at the lower and higher levels, however it may not be valid, as the rating for any particular question may actually reflect the overall experience rather than the subject of the question. For example, Williams and Ceci (1997) report on how an intervention focused on improving presentation and communication skills resulted in increases for a range of teaching related topics, including the (unchanged) textbook.

The goal of writing a survey question for self-administration is to develop a query that every potential respondent will interpret in the same way, be able to respond to accurately, and be willing to answer. However, in practice, producing good questions is often difficult. (Dillman, 2007, p. 32)

The crux of the problem is that respondents may not think about questions or the questionnaire in the same way as the researcher. Foddy summarises the problem as “If different respondents define a topic in different ways and the researcher is not aware of this fact, the respondents’ answers will neither be properly interpreted nor meaningfully comparable with one another” (p.36). As educators, we are all too familiar with how students can misinterpret assessment tasks and test questions, so why do we assume they will do better when completing a questionnaire?

Apart from variations in understanding of specific question wording, poorly constructed questions present a number of other challenges to survey validity. Questions requiring respondents to recall information may be subject to significant recall errors, with most people overestimating the frequency of rare behaviors and underestimating the frequency of frequent behaviors (Sudman et al., 1996). Even greater errors may occur when asking respondents to recall attitudes or whether change occurred unless respondents are prompted to think deeply (Ross, 1989; Schwarz & Oyserman, 2001).

The set of responses provided for any question can also introduce bias if not designed carefully. Changing the numerical values on a scale can impact on the distribution of responses (Schwarz, Grayson, & Knauper, 1998; Schwarz, Knauper, Hippler, Noelle-Neumann, & Clark, 1991) as can providing an inappropriate set of ranges (Rockwood, Sangster, & Dillman, 1997; Toepoel et al., 2009). The use of a five point Likert scale can also create a positive bias in responses for attitude questions (Foddy, 1993), with the bias even more pronounced for satisfaction questions.
Lists of items from which respondents select one or more responses offer several potentialities for bias. Items toward the top of the list will be more frequently selected, as will items that stand out (Sudman et al., 1996), while simply being included in the list is likely to result in occasional selection (Peterson, 2000). Finally, items poorly understood appear to be allocated a meaning based on items around them (Tourangeau, Couper, & Conrad, 2004).

**Cognitive interviews**

A questionnaire is in essence a series of cognitive tasks for the respondent. A good questionnaire will therefore present itself as a conversation while simultaneously minimizing the cognitive workload for the respondent. In a conversation or interview, the researcher is able to clarify when the initial response is limited or appears inappropriate. In a self-administered questionnaire the researcher is not present and therefore must predict in advance the course of the conversation (Dillman, 2007). One of the most common methods for pretesting questions is cognitive interviewing, as they are highly effective in testing respondents’ interpretation of individual questions as well as identifying both overt and covert problems (Willis, 2004). The cognitive interview theoretical model is based on the four-stage response model of thought process (Tourangeau, Rips, & Rasinski, 2000), which involves comprehension, retrieval, judgement, and mapping to the reporting system.

The importance of cognitive interviews in improving questionnaire validity is recognised by the number of permanent cognitive interview facilities established around the world by organisations such as the US Census Bureau, the US National Centre for Health Statistics, Statistics New Zealand, and Statistics Canada (Willis, 2005). In other fields authors have argued that cognitive testing should be a standard part of the development process of any survey instrument (Collins, 2003; Conrad, Blair, & Tracy, 1999).

In a cognitive interview, the respondent is asked to think their way aloud through the question being tested. Within this basic structure there are many variations, from a protocol where the interviewer does not question at all, through to one where the interviewer actively engages in unscripted questioning (Willis, 2005). In terms of timing, the respondent may be asked to think aloud as they answer the question, or to explain their reasoning after answering each question or the entire survey. Questioning may focus on any of the four stages of the response model, for example, by probing the respondent’s understanding of terms, meaning of questions, strategies for recall, the basis of their judgement, or how they mapped their response to the scale. Dillman (2007) argues cognitive interviewing approaches can also be used to test layout as well as questions.

While cognitive interviews can reveal many problems with questions they have limitations and requirements for maximum impact. Firstly, volunteer participants are likely to be more diligent and intrinsically motivated than average. Secondly, the mental work (Conrad et al., 1999) called for in a cognitive interview is likely to be more than respondents would typically use in responding to the survey. Cognitive interviews may therefore amplify any problems with the questions. Thirdly, cognitive interview participants should be selected for diversity rather than representativeness to ensure all viewpoints are considered. Results can therefore only be considered as a series of case studies, and it is the responsibility of the researcher to consider how widespread a problem is likely to be; it may be that a single case is adequate to highlight a larger problem. The number of cognitive interviews undertaken should reflect the size of the research project, but even for small scale surveys, a single cognitive is better than none (Willis, 2005).
Reported use of qualitative pretesting methods

A sample of educational research publications were reviewed to investigate the quality and quantity of reporting about pre-testing of questionnaires returning quantitative data. Papers from HERDSA were selected for analysis as they represent a diverse range of reviewed educational research and encompass a wide range of approaches, methodology and methods. These papers are not assumed to represent all higher education research being undertaken in Australia. Convenient access to the papers was also a factor. Links to the papers listed on a single web page allowed convenient download to an Excel spreadsheet for analysis, and the papers themselves are openly accessible and in searchable pdf format. It is important to note the focus is on the reported use of qualitative pretesting methods; it is possible researchers used one or more of these approaches but decided not to include a description in their paper.

All papers from HERDSA 2008-2011 were included in the sample, except for two from 2008 and one from 2009 which were no longer available. In total 186 papers were reviewed. Papers were listed in an Excel spreadsheet, and as each was analysed the following were recorded: number of surveys used, an initial identification of validation processes reported (qualitative pre-testing, quantitative methods or other such as existing questionnaire or reference elsewhere), the scale of the survey (rounds, respondents, classes etc.), and a brief description. In the first instance, a search was made for the key words “survey” and “questionnaire” to identify whether or not a survey was used in the research. Papers containing either of these key words were then investigated further in order to complete the remaining details. A total of 82 papers were identified which reported on 92 surveys. Table 1 shows the number of papers and surveys identified.

Table 1: HERDSA papers reviewed for surveys

<table>
<thead>
<tr>
<th>Year</th>
<th>Conference Theme</th>
<th>Papers reviewed</th>
<th>Survey in method</th>
<th>Total surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>Higher Education on the Edge</td>
<td>33</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>2010</td>
<td>Reshaping Higher Education</td>
<td>56</td>
<td>27</td>
<td>32</td>
</tr>
<tr>
<td>2009</td>
<td>The Student Experience</td>
<td>64</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>2008</td>
<td>Engaging Communities</td>
<td>33</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>186</td>
<td>82</td>
<td>92</td>
</tr>
</tbody>
</table>

The 82 papers identified were then reviewed to determine the role of the survey in the research and the nature of the questionnaire used. Of the 92 surveys reported, 14 were eliminated as they included only qualitative questions, a further 3 which discussed survey theory only, and a final 6 because they had insufficient information about the nature of the questionnaire. Table 2 shows the breakdown of surveys in HERDSA papers.

Table 2: Surveys in HERDSA papers

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Surveys</th>
<th>Qualitative only</th>
<th>Theoretical only</th>
<th>No description</th>
<th>Quantitative in whole or part</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>22</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>2010</td>
<td>32</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>2009</td>
<td>31</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>2008</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>92</td>
<td>14</td>
<td>3</td>
<td>6</td>
<td>69</td>
</tr>
</tbody>
</table>
The validation methods reported for the 69 surveys were then revisited and coded into broad groups. The level of detail of descriptions of questionnaires varied significantly. At one end of the spectrum, as previously noted, 6 papers provided no description of the survey process or questionnaire and therefore could not be considered further. At the other end the full questionnaire was included as an appendix along with details of validation processes used in its development. Validation processes reported included use of institutional surveys, use of existing validated surveys, variations on existing validated surveys, approaches grounded in theory, quantitative methods, triangulation with qualitative results, and references to previously published validation processes. Table 3 lists the number and proportion of validation processes reported. Note that one survey involved multiple methods.

<table>
<thead>
<tr>
<th>Validation Method</th>
<th>Count</th>
<th>% of 69 (# surveys)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Existing Validated (includes minor variations)</td>
<td>20</td>
<td>29%</td>
</tr>
<tr>
<td>None reported</td>
<td>18</td>
<td>26%</td>
</tr>
<tr>
<td>Institutional Survey</td>
<td>14</td>
<td>20%</td>
</tr>
<tr>
<td>Theory Based</td>
<td>7</td>
<td>10%</td>
</tr>
<tr>
<td>Triangulation</td>
<td>5</td>
<td>7%</td>
</tr>
<tr>
<td>Quantitative Methods</td>
<td>4</td>
<td>6%</td>
</tr>
<tr>
<td>Reported Elsewhere</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>70</strong></td>
<td><strong>101%</strong></td>
</tr>
</tbody>
</table>

The following points were identified:

- Validated instruments, 20 (29%) were most common, sometimes with minor variations to suit local conditions. However claims of validation made in publishing the original instrument were always accepted without question.
- No validation reports 18 (26%) made up a significant proportion of cases. Note that questionnaires from mixed methods studies were counted as ‘none reported’ unless explicit reference was made to triangulation as a means of validation. This was made explicit in 5 (7%) cases.
- Institutional surveys, 14 (20%) included few references to their validity studies.
- Quantitative validation methods, 4 (6% of the surveys) were taken as sufficient without consideration of other validation methods.
- There were no reports of validation using cognitive interviews.

In seeking information about validation approaches two other points were noted. Firstly, in some cases description of the questionnaire was limited, as well as description of the survey process. Secondly, use of the 5 point Likert scale was widespread, with no explanation provided as to why this scale was selected.

**Validation reports**

The significant proportion (26%) of the surveys without report of any validation undertaken is concerning. The scale of surveys without report ranged from post workshop happy sheets...
distributed to 10-20 people, through to surveys distributed to over 1000 respondents across multiple sites. The amount of validation undertaken should be proportional to the scale of the survey, and Willis (2005) points out that even small surveys should be tested with at least one cognitive interview. A one-off workshop happy sheet, therefore, could be improved with 1 or 2 cognitive interviews, whereas a survey distributed to over 1000 respondents should be subject to much more rigorous scrutiny.

Trust in any research result is dependent upon the quality of reporting of validation processes. Where a survey is the only research method used, this trust is generated by reports of the questionnaire development process and the research protocols. In some cases where the survey description was confined to a single sentence: “… an on-line confidential survey released post-exam” or “… an in-class survey conducted by a staff member,” the reader is left with a few questions. They may then look for a copy of the questionnaire or the actual questions asked to make up their own mind as to the validity of responses. In some rare cases in this study these also were not reported and questions of validity remained unanswered. In all these cases, use and reporting of pretesting methods to improve the validity of the questionnaire would allow the reader to have greater confidence in the research findings.

**Mixed method studies**

In mixed method studies, the survey may form only a small part of a large suite of data. Within the papers reviewed, studies were identified where interviews and focus groups were used to inform questionnaire construction, or conversely to add additional data or confirm questionnaire results. Both of these methods increase the validity of questionnaire results, but further improvements are possible. Focus groups in themselves are not considered a questionnaire pretesting method (Presser & Blair, 1994) as the tendency to group think does not accurately reflect the cognitive processes for an individual responding to a questionnaire (Willis, 2004). However they, along with other forms of interviews, are a powerful tool for informing questionnaire construction. In the reverse situation, where the questionnaire informs interviews or focus groups, the subsequent information collected aids in confirming the level of validity of the questionnaire data. In all cases, using cognitive interviews to pretest the questionnaire could increase the overall validity of each study to an even higher level.

The level of detail of methodological reporting in mixed methods studies was weighted toward qualitative methods over surveys. For instance, a detailed description of qualitative methods may be followed by a description such as “… the survey was trialled and minor revisions made.” Missing is a description of the nature of the trial and the basis on which the revisions were made, as well as any mention of qualitative validation of question wordings.

**Existing surveys**

Institutional surveys were used in 14 (20%) of the surveys reviewed, generally to supplement other sources of data, or in one instance to validate a new questionnaire. However, apart from one specific university, no publications were referred to documenting how the instrument was validated. This may simply reflect the convenience of using institutional surveys, as the data is collected as a matter of course. Alternatively it may reflect either an assumption that the institutional survey has been validated or simply lack of understanding of the need for validation. Furthermore, institutional surveys are influenced by factors outside the educational experience being investigated (Hirschberg, Lye, Davies, & Johnston, 2011), are rarely psychometrically validated (Barrie, 2000), and when validation is reported cognitive interview pretesting was not used (Tucker, Oliver, & Pegden, 2008).
Existing validated instruments were used in 20 (29%) of the surveys reviewed, made up of 12 (17%) used as published, and 8 (12%) where minor variations were introduced to address local requirements. Use of an existing validated instrument is recommended as one way to ensure validity (Peterson, 2000), however questions remain about their use. One concern is that with changing language use and social norms over time, validation conducted at a single point in time can become out of date (Converse & Presser, 1986). Furthermore, given their validation is invariably quantitative, and there is no literature addressing whether cognitive interviews and other qualitative methods identify the same problems, a question remains as to whether their validity could be improved further. This same question can be addressed to the 4 (6%) of surveys identified that were validated using quantitative methods.

**Likert scale**

Whilst the focus of the analysis was on pretesting methods, it became apparent while undertaking the analysis that the use of the 5 point Likert scale for collecting responses is widespread and without explanation. As previously noted, the use of a five point Likert scale can create a positive bias in responses for attitude questions (Foddy, 1993), with the bias even more pronounced for satisfaction questions (Peterson, 2000). The widespread use of this scale raises the question of whether researchers are using this scale simply because it appears to be the de facto standard, and if indeed researchers are considering other options and making conscious and thoughtful decisions in their questionnaire design? The final stage in the four stage cognitive process model (Tourangeau et al., 2000) is mapping the response to the answer scale provided, and cognitive interviews can assist in identifying alternative scales that may provide richer and more meaningful data.

**Summary**

It is important to recognize that overall the quality of research in the reviewed papers ranged between satisfactory to excellent, with many questionnaires based soundly on theory or prior qualitative research findings. However, using qualitative pretesting methods to increase the validity of questionnaires could increase the impact of all the research reported by removing doubt about bias and uncertainty about survey results. It is important to note that validity is not an either/or proposition, it can more accurately be considered a scale where the goal is the highest practical rating possible and 100% is not possible. Time, resources and the scale of the survey will all influence the amount of pretesting possible, however a lot is better than some, and some is better than none (Willis, 2005).

Finally, to address what appears to be a blind spot in survey research and increase the validity of survey research, it is important to improve reporting by including a copy of the questionnaire or relevant section, an explanation of how the questionnaire was developed, and the validation processes undertaken. This would increase the strength of the research findings, and therefore connections with the research audience. To address the first point, publishers of educational research are encouraged to include copies of questionnaires as an appendix without impacting on word count. For the second, researchers are encouraged to not only explain the theoretical basis for their questionnaire, but to also explain how they decided upon question order and wording, layout and response scales. Cognitive interviews and other qualitative pretesting methods are simple to use and should always be used to aid in improving survey validity. A forthcoming publication by this author explains a range of qualitative pretesting methods.
References


Copyright © 2012 Cathryn McCormack. The author assigns to HERDSA and educational non-profit institutions a non-exclusive license to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive license to HERDSA to publish this document in full on the World Wide Web (prime site and mirrors) and within the portable electronic format HERDSA 2012 conference proceedings. Any other usage is prohibited without the express permission of the author.