&YUSBDU GSPN BSUJDMF

%VODBO %   /VMUZ   5IF BEFRVBDZ PG

BOE QBQFS TVSWFZT XIBU DBO CF EPOF

)JHIFS &EVDBUJPO   %0*

What is an adequate response rate?

It might be strictly more correct at this point to be asking what an adequate sample size is. However, in the context of teaching and course evaluation surveys, sampling is not likely to be in the minds of academics. It is much more likely that they will ask a question about response rates. Furthermore, if a determination is made regarding sample size, the size of the population being sampled needs to be known first and so the corresponding response rate can be readily calculated from these two figures.

Whether or not a response rate is adequate depends (in part) on the use that is being made of the data. If the data gathered from a teaching evaluation survey were to be used only to bring

completed their course units of programmes.Õ (p.409.), despite having noted earlier that this rate Ôclearly leaves ample opportunity for sampling bias to affect the resultsÕ (406).

Assertions regarding the adequacy or otherwise of a particular percentage response rate appear to be made without reference to any theoretical justificationÑor to the total number of potential respondents. Behind the assertions appears to be a balance between rational and political considerations of acceptability. It would be better if there was a theoretically justified, systematic way to calculate the response rate required.

Calculating required response rates

When academics survey their students to gather opinions on their teaching, or the quality of courses, they may either ask every student enrolled in those courses to respond, or may select only a smaller sub-set of students. If every student is surveyed, the purpose is to establish the views of the entire group of students. In this instance the population is every student enrolled on the course.

When academics elect to survey a sub-set of the enrolled students, there is one of two purposes. They might only be interested in the opinions of that particular sub-set of students because they possess some characteristic that is of particular interest. For instance, the population could consist of only the mature-age students who are enrolled in the course. In these circumstances it follows that the academics have neither the interest nor the intent to deduce anything about other students, nor subsequently to take actions that in any way relate to those students or their views.

Alternatively, an academic might be interested in the views of all students enrolled on his/her course but simply finds it more practical to survey only one sub-set. In this case, the population remains all students enrolled on the course. The sub-set which is surveyed is a sample of that that the ace mainsbekt ixtral aay rimplns from (thaeveis a s1or tw T* (ndobdents,)Tj 0 T2 Tbout i

older. The views of these people may deviate systematically from the views expressed by those who attend the daytime lecture.

Sample bias can also be introduced as a product of the survey method that is chosen. Watt et al. (2002, 329) have reported that web users are demographically different from other users. Salmon et al. (2004) reported that variance in data from web surveys was less than for paper surveys. It is reasonable to suppose that an online survey will attract responses from students who are demographically different from students who would respond to a paper survey.

Third, sample bias can be introduced because of systematic differences between respondents and non-respondents. As noted by Richardson (2005, 406), research shows that Ôdemographic characteristics of people responding to surveys are different from those who do not respond in terms of age and social classÕ (Goyder 1987, Chapter 5). While that may not matter to most academics conducting evaluations of their teaching and courses, Goyder more importantly reported that Ôrespondents differ from non-respondents in their attitudes and behaviourÕ (Goyder 1987, Chapter 7) and other research has shown that Ôstudents who respond to surveys differ from those who do not respond in terms of their study behaviour and academic attainment ÉÕ (Astin 1970; Neilsen et al. 1978; Watkins & Hattie 1985, 406).

Richardson (2005) concluded: ÔIt is therefore reasonable to assume that students who respond to feedback questionnaires will be systematically different from those who do not respond in their attitudes and experience of higher educationÕ (406, emphasis added) and furthermore, Ôit is not possible to predict attitudes or behaviours on the basis of known demographic characteristicsÕ (Goyder 1987, Chapter 7, emphasis added). This means it impossible to use demographic data concerning students to construct a sampling frame that might seek to overcome sampling bias.

Thus, not only are the expressed views of respondents likely to be different from those of non-respondents but responses gathered using web surveys are likely to be different from those gathered using paper-based surveys.

In the face of evidence of this kind, are we still prepared to accept response rates of 50%Ð 60%Ð70% as adequate? It seems reasonable to argue that despite our best efforts it will often be difficult and/or expensive to obtain response rates above 70%. Politically, it is discomforting to accept low response rates because the proportion of non-respondents may be too high for us to be sure that those who responded are representative of the others who did not. The issue becomes Ôwhat are we prepared to accept?Õ. As such, there is some degree of arbitrariness about the decision.

But there is some theory to guide us in the domain of statisticians and mathematicians beginning with a seminal paper by Neyman (1934), which discusses Ôthe method of stratified samplingÕ compared with Ôthe method of purposive selectionÕ, followed in 1955 by a paper entitled ÔA unified theory of sampling from finite populationsÕ (Godambe 1955) and more recently a paper by Smith (1983), ÔOn the validity of inferences from non-random sampleÕ. A more accessible account of the salient points has been provided in Chapter 5 of Dillman (2000, 194Ð213).

First, there is a systematic way to calculate the sample size required for a specified level of confidence in the result, in relation to a population of a specified size, with a specified 5 specpds0.0

Thueaonduiions ore :

rate. In practice, if the reader wants to calculate sample size instead, the requirement to survey all the students can be removed.)

-

For example, let us consider a hypothetical scenario. If an online survey is used, the respondents are more likely to be students who are familiar with and able to use this medium. As such, these students may also comment more favourably regarding online teaching matters than the other students would. Hypothetically, these students may also constitute a minority. The result will be a survey with a low overall response rate, made up of students who are mostly familiar with, able to use and favourably disposed toward online teaching and learning provisions of the course. If this happens, and these are the only data considered, the academic concerned could form a false view that she/he should do more to boost the use of online teaching approaches.

It should be noted that the problem here is not simply that the responses to the survey have come from a minority of students, but that the survey results suffer from systematic bias. This means that these data may also misrepresent and misinform summative judgements regarding the performance of the teacher. Unfortunately, it is not possible to determine the direction of that bias. Although (in this hypothetical case) students responding to online surveys may be more positively disposed towards online teaching approaches, this does not mean that they will also be more positively disposed towards the teacherÕs teaching.

The hypothetical scenario above serves to illustrate another problem too: imagine an online survey of all students yields a 30% response and an on-paper survey of the same students yields a 60% response. The temptation would be to regard the results of the latter as more valid and more worthy of consideration. However, as already described above, it may be that the online survey attracted responses from those who predominantly make use of online teaching and learning resources, while the respondents to the paper survey may contain few of these people. Effectively the two surveys have sampled two different sub-groups of students with systematically different views which may (or may not) be reflected in the nature of their answers to survey questions (depending on the questions). Neither survey may be a valid reflection of the whole group but each one may be a valid reflection of each sub-group.

In practice, it is likely that only one of these two surveys would be conducted Ñ the academic will not have both sets of data for comparison. The academicÕs responses to improve his/her teaching and/or his/her course might therefore be erroneous. Similarly, the data for either survey may be misleading if used for summative purposes. This is not a problem resulting from low response rate per se but, rather, a problem associated with the potential for systematic sample bias in respect of the respondents to any one survey type Ñ or, indeed, any survey.

This last point takes us into territory that is beyond the scope of this paper. Suffice it to say that the design of a survey, not only the mode of administration, may also affect who responds to it and what they say. Thus, when interpreting survey results, it is important to think about what was asked, how it was asked and how these variables may have resulted in bias in respect of who responded, what they said and how these responses may have differed if the survey itself, the mode of administration and the resultant pool of respondents had been different. The implication is that data derived from surveys are likely to be somewhat more easily and validly used if the surveys themselves are appropriately designed and used for particular targeted purposes. Given that doing this is difficult, even in the best of worlds, this observation underscores the need to evaluate courses and teachers using multiple methods, and to carefully consider the differences between the pictures that emerge from each in order to triangulate a more accurate position.

It follows from all this discussion that, although Table 3 gives us a guide for response rates which could (in a theoretically ideal world) be considered adequate, the reality is that even if the response rates suggested are achieved, great care is needed to be sure that results for a survey are representative of the whole group of students enrolled. Although this is known, current practice frequently ignores this need for caution. Generic course and teaching surveys are often used to evaluate situations they were not designed for, and response rates which are below those