Routledge
Taylor & Francis Group

The adequacy of response rates to online and paper surveys: what can be done?

Duncan D. Nulty*

Griffith University, Queensland, Australia

This article is about differences between, and the adequacy of, response rates to online and paper-based course and teaching evaluation surveys. Its aim is to provide practical guidance on these matters. The first part of the article gives an overview of online surveying in general, a review of data relating to survey response rates and practical advice to help boost response rates. The second part of the article discusses when a response rate may be considered large enough for the survey data to provide adequate evidence for accountability and improvement purposes. The article ends with suggestions for improving the effectiveness of evaluation strategy. These suggestions are: to seek to obtain the highest response rates possible to all surveys; to take account of probable effects of survey design and methods on the feedback obtained when interpreting that feedback; and to enhance this action by making use of data derived from multiple methods of gathering feedback.

Online surveying in general

There are many advantages associated with the use of information technology to support approaches to evaluation (Dommeyer et al., 2004; Salmon et al. 2004; Watt et al. 2002). As examples, Watt et al. (2002) note that 'using web-based evaluation questionnaires can bypass many of the bottlenecks in the evaluation system (e.g. data entry and administration) and move to a more "just in time" evaluation model' (327). Another advantage is avoiding the need to administer surveys in class (Dommeyer et al. 2004). Unsurprisingly, there is increasing growth in the use of web-based surveying for course and teaching evaluation (Hastie & Palmer 1997; Seal & Przasnyski 2001). This growth is happening despite concerns from students (e.g. regarding confidentiality and ease of use) (Dommeyer, Baum & Hanna 2002), and concerns from staff (e.g. about the adequacy of response rates) (Dommeyer, Baum et al. 2002).

Online surveying practice varies greatly. For example, in Australia, the University of South Australia uses a system supporting *solely* online administration of surveys, while Murdoch University and Curtin University among others are moving the same way. Griffith University and Queensland University of Technology have each developed integrated web-based systems that take a hybrid approach offering academics a choice of paper or online administration for their surveys. Respondents, however, have no choice: they receive either a paper-based survey or an online survey. Other emerging systems allow choice of response mode by combining multiple modes of administration and response (Pearson Assessments 2006), thereby allowing survey designers to better match the method of survey administration to the needs, abilities or preferences of respondents and avoid skewing the data.

*Email: d.nulty@griffith.edu.au

Despite these variations, there are some common features to online surveying practice. These have been described by (Dommeyer et al. 2004). They reported that: a typical online evaluation involves: giving students assurances that their responses will be de-identified and that aggregate reports will be made available only after the final grades are determined; providing students with the URL to access the survey—generally using their student ID number; students responding numerically to multiple response items and typing answers to open-ended questions; providing students with a receipt verifying that they have completed the evaluation; and providing at least two weeks in which the students can respond, usually near the end of term/semester (612).

Comparability of online and on-paper survey response-rate data

(McCormack 2003) reported that there are 'new expectations in relation to the evaluation of teaching, for example, expectations about the role of evaluation of teaching in promotion and probation and about the public availability of student evaluation results on institution web sites …' (2). More specifically, the expectations are that teaching evaluations should be used directly, openly and compulsorily in promotion and probation decisions, and that data on student evaluation of courses should be made available publicly to inform the public. Such expectations may be seen as an extension of the change in the focus of teaching and course evaluations from formative to summative (Ballantyne 2003).

These changes in expectations and focus are occurring at the same time that the use of online surveying is increasing. Considered together, this has raised interest in issues around response rates to these surveys. Yet, a recent review of literature regarding instruments for obtaining student feedback (Richardson 2005) claimed that 'little is known about the response rates obtained in electronic surveys, or whether different modes of administration yield similar patterns of results' (406).

Closer scrutiny of the literature, however, reveals that a good deal *is* known. Moreover, there is also a fair amount of information available in relation to the comparison between patterns of results obtained through using different modes of administration of surveys. Some of that literature is reviewed below—with the caveat that while it is strongly suggestive of what one might call a 'prevailing position', it also illustrates substantial variability.

In general, online surveys are much less likely to achieve response rates as high as surveys administered on paper—despite the use of various practices to lift them. Some literature demonstrating this follows and has been summarized in Table 1. In addition, in some cases (such as Griffith University), the reported response rate for paper-based surveys is conservative because an academic may only hand out paper surveys to one sub-group (e.g. one class) of students rather than to all that were enrolled. Given that this practice is not reported centrally, there is no way to take it into account when calculating the overall response rate.

In summary, of the eight examples cited in Table 1, most of the online surveys achieved response rates that were much lower than the paper-based ones (on average, 33% compared with 56% = 23% lower). Thus, in general, these data show that online surveys do not achieve response rates that are even close to what is achieved with paper-based surveys. There are just two exceptions which will be detailed next.

In the research by Watt et al. (2002), the overall response rate for online surveys was 32.6%, while for paper surveys it was 33.3% (333). This finding is inconsistent with the other data reported in Table 1. However, the context for the low on-paper response rate in Watt et al.'s research is that the courses surveyed were all taught in distance education mode. This means that these paper surveys were not handed out in a face-to-face environment as they were in the other studies. This finding raises a question about the impact of face-to-face administration of surveys.

Table 1.   Comparisons of response rates to paper-based and online surveys.

Who	Paper–based response rate (%)	Online response rate (%)	Difference (%)
Cook et al. (2000)	55.6	–	−16
Baruch (1999)	–	39.6	
Dommeyer et al. (2004)	75	43	−32
Ballantyne (2005)	55	47	−8
Ogier (2005)	65	30	−35
Nair et al. (2005)	56	31	−31
Griffith University (2005)	57	20	−37
Sweep (2006)*	56	23	−33
Watt et al. (2002)	32.6	33.3	<1
Overall	56	33	−23

*(T. Sweep, pers. comm.)

The data clearly show that face-to-face administration results in higher response rates. What is unknown is whether response rates to online surveys would rise to the same level if they were also conducted in a face-to-face way.

The author has not found any study reporting on this question. It seems likely that this is because one of the main benefits (and uses) of the online survey process is to avoid the need to conduct the survey in class (Dommeyer et al. 2004). Clearly, if the only way to achieve high response rates with online surveys was to administer them in a face-to-face setting it would negate these benefits. In general, such steps are not taken—and are unlikely to be taken.

Watt et al.'s (2002) research suggests that when paper surveys of courses and teaching are not administered face to face, the response rates might be as low as for non-face-to-face online surveys. It is reasonable to hypothesize that in a non-face-to-face setting it is easier to submit an online response than it is to physically mail a paper one. It follows that in non-face-to-face settings this should advantage online survey response rates. It is not, therefore, a conclusion of this article that on-paper surveys are intrinsically 'better' than online surveys.

The second exception to the data reported in Table 1 is contained within the detail of the study conducted by Dommeyer et al. (2004). These researchers conducted an experiment in which they found that response rates to online surveys were lower than for on-paper surveys in 14 cases out of 16—significantly so in 10 of these. Where response rates were not significantly different was usually when students were offered a (very) small grade incentive (respondents' grades were increased by one quarter of 1%). When the grade incentive was applied, the response rates for both online and on-paper surveys were high—and almost identical (86.67% and 86.99% respectively). This result appears to be unique: that is, I have found no other literature to demonstrate that it can be, or has been, repeated. Overall, however, Dommeyer et al. reported that online surveys achieved a 43% response rate, while on-paper achieved 75%.

Boosting online survey response rates

The most prevalent methods for boosting online survey response rates are:

(1) repeat reminder emails to non-respondents (students);
(2) repeat reminder emails to survey owners (academics);
(3) incentives to students in the form of prizes for respondents awarded through a lottery.

Table 2.   Methods used to boost online survey response rates in five universities.

University	Methods used	Online survey response rate (%)
Murdoch University (Ballantyne 2005)	1, 2 & 3	47
Canterbury University (Ogier 2005)	1 & 3	30
Monash University (Nair et al. 2005)	1	31
Griffith University	no measures taken	20
QUT (Sweep 2006)*	no measures taken	23

*(T. Sweep, pers. comm.)

Methods used in the institutions investigated, together with the response rate achieved for online surveys, are summarized in Table 2. These data suggest that, generally speaking, the greater the number of measures taken to boost online response rates, the higher those rates are.

In addition to the measures specified above, Ballantyne (2005) reported that for each survey at her university the email sent to students contained a URL which allowed them to access the survey more easily. This same URL was also embedded in the course WebCT pages and the course welcome pages. All surveys were also, by default, open for 20 days. Aside from these extensive mechanisms, Ballantyne speculated on the reasons for the relative success at Murdoch University. She noted that Murdoch had been using online surveys since 1998 and that it has had mandatory surveying since 1993. She proposed that this has helped to create a culture in which such surveys were accepted by students and staff.

Neither Griffith University nor QUT used email reminders for online surveys, nor any form of incentive scheme to potential respondents. Academics were simply advised to ensure that they encourage the students to respond. Clearly, given that these institutions achieved the lowest online response rates (20% and 23% respectively) encouragement alone appears to have little effect.

Additional approaches to boosting response rates

Two websites offer particularly succinct, credible and partly overlapping advice regarding practices that can boost response rates. These are Zúñiga (2004) from the US Teaching and Learning with Technology/Flashlight Group, and Quinn (2002) from the University of South Australia. Zúñiga offered a set of seven 'best practices for increasing response rates to online surveys'. These are:

(1) *Push the survey*. This basically means making it easy for students to access the survey by, for example, providing them with the survey URL in an email sent directly to them.
(2) *Provide frequent reminders*. Zúñiga advocated 'At least three reminders'. Others, however, point to the inevitable diminishing return on this investment coupled with the possibility of irritating the survey population (Kittleson 1995; Cook et al. 2000). In the context of surveying multiple lecturers in any one course, and multiple courses in any one semester, respondents are likely to have several surveys to complete. The potential for a barrage of reminders—and commensurately higher levels of irritation—is evident.
(3) *Involve academics*. Zúñiga contended that 'Nothing helps more than regular reminders to students from faculty'. This assertion does not appear to be entirely supported by the literature. As shown earlier in this paper, institutions that did not use direct email reminders to students—implicitly relying on academics to promote participation—achieved much lower response rates than those that did. The combination of direct reminders backed up

by encouragement from academics, however, was certainly better than either method alone. It may be particularly so if the academics also take the opportunity to demonstrate and/or convince students that their feedback has been, or will be, used to good effect (see no. 4 below).

(4) *Persuade respondents that their responses will be used.* The issue here is whether students believe that the academics will take the feedback seriously (Nulty 1992). There is a range of ways to achieve this but all involve some active demonstration to students that feedback is valued and acted upon.

(5) *Provide rewards.* Zúñiga stated that 'Many institutions have found that a drawing for a prize of general interest … [helps]'. He went on: 'even one point earned for the course also works well even though it is not enough to change any individual student's grade. Sometimes this reward is given to individuals, and sometimes to the whole class if more than a certain percentage of students responds.' However, he echoed a warning made more clearly by Ehrmann (2004) that thoughtful participation is best achieved by ensuring the survey is worth students' time, and that using extrinsic motivators may bias the sample to include more responses from those who need that form of encouragement.

(6) *Help students understand how to give constructive criticism.* When such help is given it seems likely that there will be at least two benefits. First, students will improve their ability to make points of value in ways that are unlikely to bruise academics' egos. Second, providing this kind of help to students will help convince them that their responses will be used (point no. 4).

(7) *Create surveys that seek constructive criticism.* If a survey does not demand constructive criticism—for example if all the items require a simple numerical rating—then there will probably be less engagement with the survey because the survey itself sends a message that conflicts with attempts made under no. 4.

Quinn (2002) specified eight strategies that have been used by people who have achieved high response rates to online surveys. Some of these overlapped with those already detailed above, but the following five did not:

(1) *Extend the duration of a survey's availability.* The longer it is there the higher the chance students will respond.

(2) *Involve students in the choice of optional question*s. Aside from making the survey intrinsically more interesting to students, this also addresses Zúñiga's #4.

(3) *Assure students of the anonymity of their responses.* Dommeyer, Baum & Hanna (2002) indicated that this was a concern for students, so anonymity seems likely to boost responses if it is managed effectively.

(4) *Familiarize students with online environments by using online teaching aids/methods.* Related to this point, Richardson (2005) gave the following advice:

It would be sensible to administer feedback surveys by the same mode as that used for delivering the curriculum (classroom administration for face-to-face teaching, postal surveys for correspondence courses and electronic surveys for online courses). (406)

In the context of online surveying, it seems reasonable that the more familiar students are with the medium to be used for the survey, the more likely they will use it. Consistency of mode is likely to help achieve this outcome.

(5) *Keep questionnaires brief.* The proposition here is that the less time it takes for a student to complete a survey, the more likely it is they will do so.

From the evidence available (e.g. Ballantyne 2005) it seems reasonable to suggest that the effect of these measures will be additive: those who use more of these approaches will achieve higher

response rates. Clearly, the literature and practice reviewed in the first section of this paper show that there is a long way to go before online survey response rates will match those of on-paper. There is an argument that can be made here. The two primary purposes of teaching and course evaluation surveys are for monitoring quality and for improving quality. Hence the actions of academics that relate to Zúñiga's fourth point, '*Persuade respondents that their responses will be used*', are the most critical—yet also the most difficult—to impact on.

In summary, there are many methods for boosting response rates to online surveys. Many of these would apply equally well to boosting response rates to any kind of survey. At present, few of the methods advocated above are used for on-paper surveys, yet on-paper surveys already achieve relatively high response rates—perhaps because they are administered to a captive audience, often with some dedicated class time sacrificed for the purpose. If classes were conducted in computer laboratories, online surveying done in class could possibly reap similar rewards. This suggestion may therefore be added to the lists offered by Zúñiga (2004) and Quinn (2002). Conversely, if some of the measures above were used with on-paper surveys, their response rates might be even higher than they already are.

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 about improvements by that teacher, and there is even one response that provides information which can be used in this way, the survey's purpose has, at least in part, been served and the response rate is technically irrelevant. If such a single useful response were just one from (say) a hundred or more possible respondents, that is of no consequence—unless that response is entirely at odds with what the majority of other students would have said. A more likely outcome would be that a single response would be regarded as completely inadequate in the context of a summative appraisal of the performance of the teacher. Generally, course and teaching evaluation data are used for both of these purposes, and increasingly the latter (Ballantyne 2003).

Accepting that course and teaching evaluations are rarely conducted for solely formative purposes, there is certain to be widespread concern about the adequacy of the responses to these surveys. In part, this will translate into a concern about response rates. It should be noted however, that this concern occurs without sufficient awareness of the importance of sample size and population size.

Richardson (2005) cited Babbie (1973, 165) and Kidder (1981, 150–151) when stating that 50% is regarded as an acceptable response rate in social research postal surveys. Baruch (1999) researched the response rates reported by 141 published studies and 175 surveys in five top management journals published in 1975, 1985 and 1995. He found that the overall average response rate was 55.6%. Richardson (2005), however, indicated that the Australian Vice-Chancellors' Committee & Graduate Careers Council of Australia (2001) regarded 'an overall institutional response rate for the Course Experience Questionnaire (CEQ) of at least 70% [to be] both desirable and achievable' (4). But, in concluding comments, he stated: 'Response rates of 60% or more are both desirable and achievable for students who have satisfactorily

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

Assertions regarding the adequacy or otherwise of a particular percentage response rate appear to be made without reference to any theoretical justification—or 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 population. It is common that an academic may survey those students who attend a particular class on a particular day of the week and not other students who attend on other days. In these circumstances, the academic will seek to extrapolate findings from the sample to the population. Whether it is valid to do so is the issue.

In all three scenarios outlined above, it is unlikely that every student who is asked to respond to a survey will actually do so. As a result, there are a number of matters to consider before it is possible to determine whether it is valid to extrapolate findings derived from the students who did respond to either the sample from which they came or the population to which they belong.

In the first two scenarios, every student in the population is surveyed but not all respond. The respondents represent a non-random sample of the population. An appropriate question is whether the respondents differ systematically from the non-respondents, and if so, whether these differences would cause them to respond differently to the questions asked. If the answer to both questions is 'yes', the sample is biased and simple extrapolation of findings from the sample to the population is not valid.

It is reasonable to expect that any survey that samples a population (or that achieves only a sample by way of respondents) will incur some sampling error and possibly also some sample bias. The former is the extent to which any statistical measure applied to the sample (such as the mean) gives a result that deviates from the mean of the population as a result of random variation in the membership of the sample. The latter is where a statistical measure applied to the sample deviates from the population measure because of systematic bias in the membership of the sample. In principle, both can be reduced by increasing the sample size and/or response rate—however, neither of these steps *guarantees* a reduction in either error or bias (Dillman 2000).

There are different ways in which sample bias can be introduced. In the context of course and teaching evaluation surveys, sample bias might be introduced if the academic chooses to administer a survey in a daytime lecture in preference to an evening lecture. The evening lecture might consist of a higher proportion of people who are in full employment, study part time, and are

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 'demographic characteristics of people responding to surveys are different from those who do not respond in terms of age and social class' (Goyder 1987, Chapter 5). While that may not matter to most academics conducting evaluations of their teaching and courses, Goyder more importantly reported that 'respondents differ from non-respondents in their attitudes and behaviour' (Goyder 1987, Chapter 7) and other research has shown that 'students who respond to surveys differ from those who do not respond in terms of their study behaviour and academic attainment …' (Astin 1970; Neilsen et al. 1978; Watkins & Hattie 1985, 406).

Richardson (2005) concluded: 'It 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 education' (406, emphasis added) and furthermore, 'it is not possible to *predict* attitudes or behaviours on the basis of known demographic characteristics' (Goyder 1987, Chapter7, 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%–60%–70% 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 'what are we prepared to accept?'. 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 'the method of stratified sampling' compared with 'the method of purposive selection', followed in 1955 by a paper entitled 'A unified theory of sampling from finite populations' (Godambe 1955) and more recently a paper by Smith (1983), 'On the validity of inferences from non-random sample'. A more accessible account of the salient points has been provided in Chapter 5 of Dillman (2000, 194–213).

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 degree of sampling error, given a specified level of probability for a particular answer to be provided by a respondent (Dillman 2000, 206–207).

Specifically, and in relation to the context of teaching evaluation, under the following conditions it is possible to use a formula provided by Dillman (2000) to calculate how many respondents are required (and therefore also the required response rate).

The conditions are:

- The total number of students in the population that is being surveyed is known.
- All students in the population are surveyed. (Note: It is not actually necessary to survey all the students, but this assumption is necessary for the argument being made about *response*

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

- There is a known probability of any one student providing a certain answer to a question on a survey.
- The required/desired level of accuracy of result is known or set.
- There is a known or chosen level of confidence required/desired for the same result to be obtained from other samples of the same size from the same total group of students in the course.

In order to seek to present data representing the 'best possible scenario' (i.e. one that maximizes the probability of needing the lowest response rates) the formula supplied by Dillman (2000) was initially applied with liberal conditions set. These were: to set a 10% sampling error (higher than the normal 3%), to assume a simple yes/no question is to be answered equally by respondents in 50:50 ratio (the most conservative situation), and to accept an 80% confidence level (much lower than the normal 95% used by statisticians).

However, in practice it is known that students' responses to questions on teaching and course evaluation surveys use the top ratings more frequently than the lower ones. Considering data gathered in one Australian university over an eight-year period with over 25,000 surveys using a 1 to 5 scale, actual percentages are 72% of students responding with a rating of 4 or 5, the remainder using a rating of 1, 2 or 3. Thus, the assumption of a 50:50 split on a 'yes/no' question can be altered to a (nominal) 70:30 split. Applying this more liberal condition yields lower required response rates, which are tabulated in Table 3 in the columns headed 'Liberal conditions'.

Columns under the heading 'Stringent conditions' present the required responses and response rates when more stringent (and more common) conditions are set: specifically 3% sampling error, and 95% confidence level.

Starting with the data from the liberal conditions, the table shows that for class sizes below 20 the response rate required needs to be above 58%. This is greater than the maximum achieved by all but one of the universities cited earlier when using paper-based surveys (that maximum was only a little higher at 65%). In other words, the table suggests that even the relatively good response rates obtained to paper surveys of teaching and courses are only adequate when the class size is 20 or higher—and, even then, only when liberal conditions in relation to the acceptable sampling error and required confidence level are acceptable.

Similarly, considering the response rates achieved with online surveys, the table shows that the highest response rate reported earlier (47%) is only adequate when class sizes are above (approximately) 30—and again, even then, only when liberal conditions in relation to the acceptable sampling error and required confidence level are acceptable.

In other institutions, such as Griffith University for example, class size (at best) needs to exceed 100 before its existing response rate of 20% can be considered adequate. In other words, for this institution, unless the response rate can be boosted, online surveys should not be used on classes with less than 100 students.

When the more traditional and conservative conditions are set, the best reported response rate obtained for on-paper surveys (65%) is only adequate when the class size exceeds approximately 500 students. The best reported response rates for online surveys (47%) are only adequate for class sizes above 750 students. The 20% response rate achieved for online surveys by Griffith University would not be adequate even with class sizes of 2000 students.

Table 3 is, however, only a *guide* as it is based on the application of a formula derived from a theory that has random sampling as a basic requirement. With teaching and course evaluations *this requirement is not met*. If the total enrolment of a course is sampled, it is generally a convenience sample—selecting all students who show up to the Monday daytime lecture for example.

Table 3.   Required response rates by class size.

Total no. of students on the course	'Liberal conditions' 10% sampling error; 80% confidence level; 70:30 split responses 4 or 5 compared with 1, 2, 3		'Stringent conditions' 3% sampling error; 95% confidence level; 70:30 split responses 4 or 5 compared with 1, 2, 3	
	Required no. of respondents	Response rate required (%)	Required no. of respondents	Response rate required (%)
10	7	75%	10	100%
20	12	58	19	97
30	14	48	29	96
40	16	40	38	95
50	17	35	47	93
60	18	31	55	92
70	19	28	64	91
80	20	25	72	90
90	21	23	80	88
100	21	21	87	87
150	23	15	123	82
200	23	12	155	77
250	24	10	183	73
300	24	8	209	70
500	25	5	289	58
750	25	3	358	48
1000	26	3	406	41
2000	26	1	509	25

If all students enrolled are surveyed, or if a random selection of these are surveyed, random sampling is still not achieved in practice because those who respond are not a random selection. Indeed, those who respond are systematically different from those who do not, and that those who respond will be different depending on the method of evaluation selected (Astin 1970; Neilsen et al. 1978; Watkins & Hattie 1985; Goyder 1987; Watt et al. 2002).

Discussion

What are the consequences of ignoring these facts? If the sample size is too small, results obtained will not be representative of the whole group of students. That is, the results will suffer from both sample error and sample bias. This means that the results obtained (from a sample) are not likely to be an indication of what the group as a whole (the population) would have said. Given that the respondents may be systematically different from non-respondents it is possible that the feedback provided could influence an academic to respond in ways that are counter to what they would do if they had feedback from all students. Similarly, if the data are used summatively to judge a teacher's performance, it may lead a person to make an erroneous judgement. Although academics (like the rest of us) have to make judgements all the time in the absence of useful information, it would be helpful if the parameters affecting the feedback were more transparently obvious. It would also be helpful if the information available was not itself misleading—as may be the case.

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 teacher's 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—the academic will not have both sets of data for comparison. The academic's 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—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

advocated by Table 3 are generally accepted. Despite this a high weight is simultaneously placed on student evaluation results.

Conclusion

This article has confirmed earlier research (Cook et al. 2000) which showed that response rates to online surveys of teaching and courses are nearly always very much lower than those obtained when using on-paper surveys. While a wide range of methods exists for boosting response rates, institutions do not make full use of these. The methods that are used are more likely to be applied to boosting response rates to online surveys than on-paper surveys. This is despite the fact that this article has shown that in many cases the response rates obtained for course and teaching evaluation surveys are not adequate regardless of the method of surveying used.

Given the anonymity of responses and the impossibility of using demographic data to predict attitudinal variables in students (and therefore there being no viable way to systematically target surveys at a minimal sample of students that would be representative of the whole group), appropriate paths of action that remain are to:

(1) use multiple methods to boost survey response rates as high as possible (regardless of whether on-paper or online surveys are used—but *especially* when online surveys are used);
(2) consider the probable effect that use of a particular survey design and method might have on the make-up of the respondents and take this into account when interpreting the feedback obtained;
(3) use multiple methods of evaluation to elucidate findings—so as to construct a better informed understanding of what the true picture is.

Without these actions being taken, relying heavily on student evaluations of courses and teaching is likely to be, at best, inadequate, at worst, misleading.

Notes on contributor

Duncan Nulty is senior lecturer in the Griffith Institute for Higher Education at Griffith University, Queensland Australia. He has more than a decade of experience in teaching, course and programme evaluation obtained in several large universities. He has also conducted many educational evaluation consultancies in Australia and overseas.

References

Astin, A.W. 1970. The methodology of research on college impact, part two. *Sociology of Education* 43: 437–450.
Australian Vice-Chancellors' Committee & Graduate Careers Council of Australia. 2001. Code of practice on the public disclosure of data from the Graduate Careers Council of Australia's Graduate Destination Survey, Course Experience Questionnaire and Postgraduate Research Experience Questionnaire, available online at: http://www.avcc.edu.au/documents/policies_programs/graduates/2003/COP_2001.pdf (accessed 7 December 2006).
Babbie, E.R. 1973. *Survey research methods.* Belmont, CA: Wadsworth.
Ballantyne, C. 2003. Measuring quality units: considerations in choosing mandatory questions. Paper presented at the *Evaluations and Assessment Conference: A Commitment to Quality*, University of South Australia, Adelaide, 24–25 November.
———. 2005. Moving student evaluation of teaching online: reporting pilot outcomes and issues with a focus on how to increase student response rate. Paper presented at the 2005 *Australasian Evaluations Forum: University Learning and Reaching: Evaluating and Enhancing the Experience*, UNSW, Sydney, 28–29 November.

Baruch, Y. 1999. Response rates in academic studies: a comparative analysis. *Human Relations* 52: 421–434.

Cook, C., F. Heath, and R.L. Thompson. 2000. A meta-analysis of response rates in web or internet-based surveys. *Educational and Psychological Measurement* 60, no. 6: 821–836.

Dillman, D.A. 2000. *Mail and internet surveys: the tailored design method.* Brisbane: Wiley.

Dommeyer, C.J., P., Baum, K. Chapman, and R.W. Hanna, 2002. Attitudes of business faculty towards two methods of collecting teaching evaluations: paper vs. online. *Assessment and Evaluation in Higher Education* 27, no. 5: 455–462.

Dommeyer, C.J., P. Baum, and R.W. Hanna. 2002. College students' attitudes toward methods of collecting teaching evaluation: in-class versus online. *Journal of Education for Business* 78, no. 2: 11–15.

Dommeyer, C.J., P. Baum, R.W. Hanna, and K.S. Chapman. 2004. Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education* 29, no. 5: 611–623.

Ehrmann, S. 2004. Increasing Survey Response Rates by 'Closing the Loop' from Survey Back to Respondent, available online at: http://www.tltgroup.org/resources/Flashlight/Participation.html (accessed 15 March 2006).

Godambe, V.P. 1955. A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society* 17, no. 2: 269–278.

Goyder, J. 1987. *The silent majority: non-respondents on sample surveys.* Cambridge: Polity Press.

Hastie, M., and A. Palmer. 1997. The development of online evaluation instruments to compliment web-based educational resources. Paper presented at the *Third Australian World Wide Web Conference*, Lismore, New South Wales.

Kidder, L.H. 1981. *Research methods in social relations.* New York, Holt: Rinehart & Winston.

Kittleson, M. 1995. Determining effective follow-up of e-mail surveys. *American Journal of Health Behavior* 21, no. 3: 193–196.

McCormack, C. 2003. Reconceptualising student evaluation of teaching: an ethical framework for changing times. Paper presented at the Evaluations *and Assessment Conference*, Adelaide, 24–25 November.

Nair, C.S., C. Wayland, and S. Soediro. 2005. Evaluating the student experience: a leap into the future. Paper presented at the *2005 Australasian Evaluations Forum: University Learning and Teaching: Evaluating and Enhancing the Experience*, UNSW, Sydney, 28–29 November.

Neilsen, H.D., R.H. Moos, and E.A. Lee, 1978. Response bias in follow-up studies of college students. *Research in Higher Education* 9: 97–113.

Neyman, J. 1934. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 97, no. 4: 558–625.

Nulty, D.D. 1992. Students' opinions on students' opinions. In *Research and development in higher education: academia under pressure: theory and practice for the 21st century,* ed M.S. Parer Jamieson, ACT, Australia: Higher Education Research and Development Society of Australasia Vol. 15, 498–502.

Ogier, J. 2005. The response rates for online surveys—a hit and miss affair. Paper presented at the 2005 *Australasian Evaluations Forum: University Learning and Teaching: Evaluating and Enhancing the Experience*, UNSW, Sydney. 28–29 November.

Pearson Assessments 2006. Survey Tracker Plus, available online at: http://www.pearsonncs.com/surveytracker/index.htm (accessed 3 May 2006).

Quinn, D. 2002. Improving online response rates, available online at: http://www.unisanet.unisa.edu.au/sei/website/Online-respnrates.asp (accessed 15 March 2006).

Richardson, J.T.E. 2005. Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education* 30, no. 4: 387–415.

Salmon, P., Deasy, T., and B. Garrigan. 2004. What escapes the Net? A statistical comparison of responses from paper and web surveys. Paper presented at the *2004 Evaluation Forum: Communicating Evaluation Outcomes: Issues and Approaches*, Melbourne, Australia, 24–25 November.

Seal, K.C., and Z.H. Przasnyski. 2001. Using the world wide web for teaching improvement. *Computers and Education* 36: 33–40.

Smith, T.M.F. 1983. On the validity of inferences from non-random sample. *Journal of the Royal Statistical Society* 146, no. 4: 394–403.

Watkins, D., and J. Hattie. 1985. A longitudinal study of the approaches to learning of Australian tertiary students. *Human Learning* 4: 127–141.

Watt, S., C. Simpson, C. McKillop, and V. Nunn. 2002. Electronic course surveys: does automating feedback and reporting give better results? *Assessment & Evaluation in Higher Education* 27, no. 4: 325–337.

Zúñiga, R.E. 2004, March. Increasing response rates for online surveys—a report from the Flashlight Program's BeTA Project, available online at: http://www.tltgroup.org/resources/F-LIGHT/2004/03-04.html#BeTA (accessed 15 March 2006).