

literature software for enhancing the reading skills of students with learning disabilities.

Nolinske, T. (1998). Minimizing error when developing questionnaires. In M. Kaplan (Ed.), *To Improve the Academy*, Vol. 17 (pp. 291-310). Stillwater, OK: New Forums Press and the Professional and Organizational Development Network in Higher Education. Key Words: assessment, evaluation methods, research, surveys, faculty development programs.

## Minimizing Error When Developing Questionnaires

**Terrie Nolinske**

Lincoln Park Zoo

*Questionnaires are used by faculty developers, administrators, faculty, and students in higher education to assess need, conduct research, and evaluate teaching or learning. While used often, questionnaires may be the most misused method of collecting information, due to the potential for sampling error and nonsampling error, which includes questionnaire design, sample selection, nonresponse, wording, social desirability, recall, format, order, and context effects. This article offers methods and strategies to minimize these errors during questionnaire development, discusses the importance of pilot-testing questionnaires, and underscores the importance of an ethical approach to the process. Examples relevant to higher education illustrate key points.*

Questionnaires are used by administrators, faculty, support staff, students, and faculty developers in higher education for many reasons. For example, faculty developers conduct needs assessments of faculty to determine relevant content for workshops or programs. Faculty developers also work with faculty to develop and implement questionnaires in the classroom to assess teaching and learning effectiveness

or to conduct an alumni/employer survey to determine academic readiness for the job.

Those developing the questionnaire must think carefully about its purpose. As a result of having collected this information, what is it that will be learned? What is the desired outcome? As the questionnaire is developed, it is tempting to add a few more items and get more information "just in case." Remember that each item in the questionnaire must relate directly to the intent or outcome—the stated purpose of the questionnaire. Adding "just a few more items" leads down tangential paths best left for another questionnaire or perhaps another mode of data collection.

Questionnaires may be used to explore, explain, describe, or determine satisfaction. They may collect information to measure attitudes, behaviors, and perceptions of a sample of students, faculty, or administrators. Faculty developers draw inferences about the larger population from the sample responding to the questionnaire.

Although widely used, questionnaires may be the most misused method of collecting information. Too often they are put together hurriedly with ambiguous questions crowded on each page in an attempt to keep the questionnaire short to save on postage. Questionnaires such as these will likely yield a low response rate with inaccurate, incomplete information. Effective design and use of questionnaires requires methodical planning, content expertise, time, and money. Some of the more common problems faced when developing questionnaires include errors related to sampling, nonresponse, format, order, design, wording, social desirability, recall, and context effects. This article focuses on ways in which such error can be controlled or accounted for when developing questionnaires and provides examples related to faculty development in higher education.

## Sources of Error

### *The Sample*

Faculty developers must define criteria or characteristics that will describe those included in the sample as well as criteria or characteristics that will describe those excluded from the sample. A sample

can be selected from a relatively complete list of individuals in the population to be studied. Faculty developers might want to survey faculty to determine their level of satisfaction with services provided. The director of a teaching center might randomly select names from faculty who used the center but, due to chance, end up with mostly faculty from one college. A method to assure a more equitable distribution of faculty representation could be to compile a list of faculty from each of the six colleges who used the center. If the sampling design calls for selecting a sample from one in every three faculty on the list, this arrangement ensures that faculty would be selected from each of the six colleges dependent on the length of the lists. A sample can also be selected from a group of people who go somewhere or do something relative to a certain topic. For example, an administrator might be interested in faculty from a specific school or college, including only them in the sample. Regardless of the method used to select the sample, some degree of error will be present. Sampling error, or the imprecision in a survey, occurs because only a part of an entire population is being studied. Sampling error is generally a function of three things: diversity of what is being measured, the size of the sample, and the size of the population (Lavrakas, 1993). It is calculated and described statistically as the standard error of a mean.

Myths abound in determining sample size. One myth is that a fraction of the population must be specified for the sample (i.e., data must be gathered on five percent of all faculty in an institution). Not true. A second myth is that the size of the population from which a sample is drawn has an impact on how well the sample will describe the population. Not true. In truth, if all else is equal, a sample of 150 will describe a population of 15,000 or 15 million with the same degree of accuracy (Fowler, 1992). A third myth is that national surveys should include about 1,500 people while state or local surveys should number about 500. Again, untrue (Fowler, 1992).

Sample size is determined by the purpose and content under study. A thorough analysis of the overall plan is critical to determine sample size. Will separate estimates be necessary for males and females, senior and junior faculty, or for faculty from different disciplines? Will separate regional estimates be necessary for a national sample? Those in charge of the questionnaire need to identify any subgroups among

potential respondents and estimate the sample size necessary to provide a representative sample of each subgroup.

How accurately do the characteristics of the sample describe the population? The answer depends, in part, on whether or not all people or things in the population had at least some probability of being included in the sample. Coverage errors result when everyone or everything did not have the same chance. For example, in an electronic mail survey of all students, those who have no computer have zero probability of being sampled and may have had more (fewer) problems during the past quarter than the group with computers. To the extent that the topic under study is correlated with the coverage/non-coverage issue, the accuracy of the questionnaire's findings will be lowered (Lavrakas, 1993).

### *Nonresponse*

Of concern when using mailed questionnaires is whether or not the response rate will be sufficient. Faculty developers must define an adequate response rate at the onset of the study. This is especially challenging since no standardized formula exists, and a review of the literature reveals diverse opinions on the subject. Bailey (1997) reports a 30% response rate to be reasonable if respondents are interested in the study, while Babbie (1998) suggests that a response rate of at least 50% is adequate for analysis and reporting; 60% is good; and a response rate of 70% or more is very good. Another researcher suggests that every effort should be made to obtain a response rate of 80% to 90% if mailed questionnaires are used (Kerlinger, 1986). If nonresponse is a significant problem, a new sample should be selected.

An effort should also be made to determine whether or not nonrespondents differ from respondents. If there is no difference between the two groups, study findings can be extrapolated from one group to the other. If differences between respondents and nonrespondents are found, results are limited strictly to respondents. One way to determine the characteristics of nonrespondents, and thus the costs of nonresponse to a study, is to interview or send a postcard/short followup questionnaire to a sample of nonrespondents. In this way researchers can obtain demographics relevant to the study (e.g., fac-

ulty experience, discipline, years as a mentor) and reason(s) for their original nonresponse. Be forewarned that events occurring between the time the original questionnaire was received and the time of the followup call or postcard can influence information given by nonrespondents. Comparing information from nonrespondents and respondents may not provide a definitive description about the two groups because the method used to obtain nonrespondent information may differ from the method used in the study-at-large. The comparison will, however, give some indication of how similar or dissimilar the two groups are.

How much of an attempt should be made to increase response rates and obtain information on nonrespondents? It depends on many things including time, money, the study design, purpose, and content. When using a mailed questionnaire, faculty developers can do several things to remind the sample to return it. They can send out a reminder postcard two weeks after the original mailing to those who have yet to return it. Upon receipt of the postcard, however, respondents may not be able to complete the questionnaire even if they wanted to because they probably threw it away. An ideal reminder method is to send another complete packet of materials to nonrespondents with a note emphasizing the importance of their contribution. As with this or any other reminder, it is appropriate to advise the respondent to ignore the reminder if they already responded. This procedure was followed three times over two months in a study using a self-administered questionnaire. The original mailing yielded a response rate of 48% while the second mailing of a complete packet resulted in a 78% response. The third mailing, coupled with a reminder phone call to each nonrespondent, resulted in a response rate of 95% (Nolinske, 1994).

In attempts to increase response rates, faculty developers must be assertive yet unoffensive in methods used to engage respondents. Mailing the questionnaire in an eye-catching yet professional envelope can set it apart from other mail. Respondents might also be enticed with motivators such as money, food, or a promise to send a summary of the results (Fowler, 1992).

Most importantly, people are likely to respond to a questionnaire, regardless of length, if clear, concise questions are presented in an

uncluttered format (Fowler, 1992) and accompanied by a cover letter that makes the topic of the study relevant to them.

### ***Reliability and Validity***

A reliable questionnaire yields consistent responses, suggesting the same data will be collected in repeated observations of the same phenomena. The faculty developer can then assume that differences in responses are likely due to differences in respondents' points of view, not from differences in interpreting the questionnaire. Item wording is very important in shaping reliability. Respondents should have a sense of what constitutes an acceptable response based on the item itself or by something communicated to them by the researcher (Fowler, 1992).

Reliability is not to be confused with the concept of validity. A questionnaire is generally valid to the extent that measurement procedures accurately reflect the topic under study. The concept of validity has more to do with assuring accuracy of responses than does the concept of reliability, which assures consistency of responses. There are several types of validity, each of which must be addressed by the faculty developer when still in the development phase of the questionnaire. Reliable items contribute towards a valid questionnaire. Chances of validity increase when multiple items and varied formats are used (Babbie, 1998; Fowler, 1992).

Pilot studies, in which a draft of the questionnaire is sent to a representative sample from the target population, can determine how reliable and valid items are. If possible, respondents should be interviewed to determine which items were unclear or misleading. If not, it is a good idea to include a feedback page asking specific questions about the questionnaire-draft. Was it easy to follow? Which items were confusing? Why? What sections or items were most/least difficult? Why? Modifications to items based on these responses will improve the content, reliability, validity, and flow of the questionnaire.

### ***Social Desirability***

Respondents may feel threatened by questionnaire items that are controversial or of a sensitive nature, such as race, gender, or income (Fowler, 1992; Lavrakas, 1993). In fact, issues thought to be socially undesirable are underreported significantly more than other issues (Johanson, 1993; Kerlinger, 1986; Lavrakas, 1993). Often respondents will not answer sensitive items due to mistrust that the information will not be held in confidence. For example, in writing a questionnaire to determine whether female faculty report sexual harassment more often than male faculty, care must be taken to establish trust with respondents. This may be accomplished through careful wording and ordering of items on the questionnaire. This means placing sensitive items, including demographics, towards the latter half of the questionnaire. By the time respondents get to the end they are warmed up, have a sense of what the questionnaire is about, and are more likely to respond to sensitive items. Use of a questionnaire (as opposed to an interview) to gather sensitive information can make some respondents feel invisible and increase the likelihood of their response on controversial or sensitive issues.

### ***Confidentiality and Anonymity***

Respondents are often promised that data will be kept confidential. This is stated in a cover letter, on the questionnaire, or in signed agreements between the respondent and the researcher. Each member of the research team should also sign a letter of intent stating they will maintain confidentiality. One way of doing this is to keep information such as respondents' names and addresses separate, so that one cannot be linked to the other. Assigning study numbers at the onset of the study is a good way to forget about names entirely, since the list linking a number to a name is locked away in a file. Individuals such as supervisors or colleagues who might be able to identify people with their responses (e.g., by handwriting or demographics) should not be allowed to look at the actual completed questionnaire.

When analyzing data, care should be taken to report data in aggregate, avoiding small categories in which individual respondents could be singled out and recognized. Even so, respondents may still

feel that results could be used against them as an individual or group and not respond. Or they may respond but do so in a socially desirable manner, not wanting to be the one to appear different or undesirable when results are reported. If the faculty developer feels that the content of the questionnaire is so controversial, then perhaps the promise of confidentiality is not enough.

Faculty developers can promise that information will be gathered anonymously, if indeed it will. That is, neither respondents nor those conducting the study know who has participated. Guaranteeing anonymity may help increase the response rate since respondents may no longer fear retaliation. But other rich correlations are sacrificed. When deciding what to promise respondents, one must carefully consider the purpose of the study, decide what information is most needed, and weigh the cost-benefit of maintaining confidentiality or promising anonymity.

### **Recall**

Despite good intentions, respondents may make errors because they cannot recall appropriate information when responding to questionnaire items. Respondents displace events in time, associate the behavior with the wrong time period, or forget entirely. They distort the facts, or they may not understand the question. A bias in social desirability often results in overreporting the positive and underreporting the negative. Respondents' motivation also affects how much effort they are willing to make to give an accurate report. Telescoping effects might bring the distant past nearer in memory (Fowler, 1992; Lavrakas, 1993), causing students to feel that most assignments were at the end when, in fact, they were evenly distributed throughout the semester. There are several options for reducing errors resulting from recall effects.

Respondents have a more accurate memory for recent events, especially if those events made an important or emotional impact on them. The length of time about which respondents are asked to report can be reduced by decreasing the time between the event and the questionnaire. Use holidays or special events on campus as landmarks to which respondents can relate. Ask faculty, for example, how often

between Labor Day and Thanksgiving they met with their advisees. Or apply the averaging technique. Consider the faculty member who tracks the frequency with which a particular teaching strategy is used in an academic year. The question "How much time did you spend using cooperative learning activities in class this week?" is followed by "Is that amount *typical* of the time you spent during the past month? If not, what was?" That response is then multiplied by 12 to learn the average time spent in cooperative learning activities over a year. Of course, subsequent items might want to assess the success of such activities!

Memory can be improved by asking more questions, especially if the questions relate to one another or are designed to stimulate associations. This also gives respondents more time to think about issues or events. Another way to improve recall is to approach the topic from many different angles. In a questionnaire designed to determine the effectiveness of a faculty mentoring program, items could be written about perceived benefits to the mentor, protege, and institution; types of activities in which mentors and proteges engaged; and overall satisfaction with the relationship. The more respondents perceive a relevance between questionnaire items and themselves, the greater the accuracy of their response (Fowler, 1992).

Memory can also be stimulated by asking for similar information in different formats on the questionnaire. Consider the following example in which the dean of the business school wants to know how prepared alumni are perceived to be for management positions. A questionnaire is sent to all alumni holding positions in management one year after graduation. An open-ended format might ask "How did your courses (or experiences) in the business school prepare you for your management position?" Within that same questionnaire, the same concept might also be addressed in a matrix format where alumni read statements like "The case method used throughout the curriculum taught me to problem solve and use resources," "Overall, course content was relevant to what I'm experiencing on the job," and "The fieldwork component of the academic program linked academic concepts to real-world application." Alumni mark the appropriate response alternative—strongly agree, somewhat agree, somewhat disagree, or disagree. An open-ended "why" question could be used

to follow up any item in the matrix format. Use of multiple formats offers a second chance to think about answers or to verify information reported elsewhere on the questionnaire.

### Question Wording

When writing items for questionnaires, one must consider the following questions: Will this item mean the same thing to everyone? Is there an answer to this item? Does the item demand knowledge and information respondents might not have? Does the item avoid bias and encourage respondents to draw their own conclusions? Does the item contain potentially sensitive issues? Are questions loaded with social (un)desirability? Is wording clear and precise (Groves, 1989; Schuman & Presser, 1996)? Individuals in higher education would do well to consider the parallels between designing and writing items for a questionnaire and writing an effective exam.

**Adequate Words.** When writing items, words must clearly and precisely express a thought, for the response will only be as accurate as the item is written. Rather than ask, "Age?" it is more accurate to ask, "What was your age on your last birthday?" This will generate consistent responses and avoid fractional answers. Care must be taken not to include so many words as to confuse or lead the respondent, as in "What did you like best about the course? (We're especially interested in prompt return of papers, fair tests, availability of professor to answer questions, etc.)."

**Well-Defined Words.** The same words will mean different things to different people; therefore one must clarify words or concepts that might be misconstrued. Consider the question, "How many times in the past month have you seen your mentor?" What is meant by the word seen? Does this literally mean a visual sighting? Do telephone conversations and correspondence by electronic mail count? To avoid confusion, either change the wording or provide dyads with definitions of terms at the beginning of that section of the questionnaire. This item might be reworded to read: "This next section asks about meetings you had with your mentor during the past month. We are interested in *all* communications you have had whether in person, via fax, electronic mail, or phone."

**Negation.** Using negation in a questionnaire can be easily misinterpreted, since the negative word is often missed when reading the item. Consider how new faculty might respond to the statement, "The following teaching activities may *not* be appropriate for use with classes of 300 students" followed by a list of teaching activities. In going down the list, faculty will likely respond by selecting activities that *would* be appropriate to use rather than *not* to use, even though the negation is italicized in the stem.

**Biased items.** Since respondents usually generate responses based on how an item is worded, care must be taken not to lead respondents in their thinking and bias their responses. For example, bias can result if a respondent is asked to respond to a statement attributed to a prestigious researcher or university. It may be appropriate to delete the researcher or university name accompanying the statement.

### Question Format

Using multiple formats in a questionnaire is just as important as using them on a course exam. It stimulates and maintains interest. It approaches similar content in a different way, giving respondents another chance to think about their answers in a new context (Converse & Presser, 1986; Payne, 1951).

**Double-barreled items** are ambiguous and produce inaccurate answers. Responses to the item, "Instructor displays effective pacing and stimulating thought" will only be accurate if the instructor is, in fact, effective in both his or her pacing *and* stimulating thought in students. The student may agree with part of the statement but answer in the negative, thinking that a positive response requires agreement with both parts of the statement or vice versa. The statement should be reworded or divided into two items.

**Open-ended items** allow respondents to generate their own answers. Never underestimate the feeling of control this gives each respondent. Open-ended items should not give many clues or much information about what the response should be. There should be the same number of lines or inches of white space left for responses to each open-ended item. Unequal space introduces a bias since respon-

dents will gauge the desired response accordingly. A rule of thumb is to leave enough space or blank lines for three responses to any open-ended item.

Another consideration when using open-ended items is the tremendous volume of data obtained. While interesting, data may become problematic since responses are time consuming to read and are subject to interpretation as they are read, clustered, and analyzed. Methods of addressing data volume must be considered early on.

Use of open-ended items can be effective with sensitive topics or as follow-up "why" questions to capture respondent insight. Open-ended items are useful in pilot tests to generate response sets or closed-ended items for the final questionnaire. Consider the faculty developer whose budget will only support four instructional seminars for faculty that year. She or he probably wants to determine the four topics of *most* interest to faculty. A pilot test containing the item, "Please list four topics you would like to learn more about during instructional seminars this year," could be sent to a sampling of faculty across departments and colleges in the institution. If, for example, 13 different responses are received, the faculty developer could list those 13 topics as response alternatives on the final questionnaire. Faculty then identify their top four preferences, which gives the faculty developer valuable planning information.

**Closed-ended items** should contain response alternatives that are always exhaustive and mutually exclusive (Babbie, 1998; Biemer, Groves, Lyberg, Mathiowetz, & Sudman, 1991). Use of the closed-ended format usually increases the likelihood that there will be enough responses in any one category to use in data analysis. Closed-ended items present respondents with a recognition, not a recall, task. In a questionnaire targeting use of technology in teaching, is it enough that faculty recognize and select the types of media or technology they use from a pre-established list? Or must they remember the different types of technology available to them before identifying what they use in their classroom? Closed-ended items offer response alternatives that respondents may never have considered. This may bias some respondents by putting words in their mouths. If the faculty developer needs to know what faculty really think about something, an open-ended item should be used instead. But if originality of thought is not the

issue, then closed-ended items are a convenient, standardized way to get information, because each respondent selects from the same response alternatives.

**Multiple choice items** usually have no more than six response alternatives that are exhaustive and mutually exclusive. Questionnaire developers may get carried away with the amount of material that can be obtained. For example, unless the purpose of the questionnaire requires matching faculty by age, gender, or income, collecting detailed data from each respondent may be unnecessary. Sacrificing specificity for a more general number is often less threatening to respondents and easier for those coding and analyzing. Asking anyone to reveal their income is a sensitive topic. To increase the probability of response, a range of figures can be presented, as reflected in the following:

1. The amount of gross income I received in 1998 was

- ☐ a. Below 30,000
- ☐ b. 30,000 - 39,000
- ☐ c. 40,000 - 49,000
- ☐ d. 50,000 or more

**Matrix items** require listing multiple statements that have the same set of response alternatives. Respondents are asked, for example, whether they strongly agree, slightly agree, slightly disagree, or strongly disagree with each statement. Is it undesirable to have a midpoint on a scale of response alternatives? No, although the odd numbered scale (which has a midpoint) should be used judiciously. If a numbered scale includes "5...4...3...2...1," respondents may be pulled towards either endpoint but likely select the midpoint of three (Fowler, 1992; Payne, 1951; Schwarz, Hippler, Deutsch, & Strack, 1989). If a midpoint is used it must be carefully defined. Will the midpoint be a continuation of the other categories or will it be neutral? If it is a continuation of other categories, is it different enough in meaning? For example, if the words "excellent, good, fair, average, and poor" are used, will respondents discriminate meaningfully between "fair" and "average"? A more meaningful scale might be a

three-point scale using "above average, average, and below average." A midpoint alternative is useful if it is meaningful and discriminating. It should not be a throwaway category. Another consideration is whether to include an "undecided" or "not applicable" alternative in the scale. These alternatives should be used only if they contribute meaningful information when selected. Otherwise, this neutral alternative becomes a throwaway response for the noncommittal respondent. Most questionnaire designers prefer to force the respondent to make a choice and state an opinion.

For responses to be meaningful, each statement must be worded so it is anchored at one end of a continuum (Fowler, 1992). For example, students may be asked to indicate the degree to which they agree with the statement, "Learning activities provided by this instructor are average." The descriptor "average" falls midway on a continuum between good and poor. If students agree with the statement (i.e., think that activities *are* "average"), there is no problem. If students disagree with the statement, it means they rate the instructor as either good or poor, which still does not provide an accurate, meaningful response. The statement could be rewritten as, "Learning activities used by this teacher meet my needs" or "Learning activities provided by this instructor are excellent" (or poor), keeping the descriptor anchored at either end of the continuum.

The matrix format allows for efficient use of space, although respondents tend to work through items in this format rather quickly. Many develop a pattern of agreeing with all statements without stopping to think about any of them. This is a phenomenon called *response-set* and can be minimized by alternating statements representing different orientations (Babbie, 1998). For example, including "Learning activities used by this instructor are *poor*" in a series of statements with positive wording may force respondents to read each statement more carefully.

**Don't know and contingency items** are another format option. The *don't know* option must be used judiciously, since it becomes an easy out for respondents who are unwilling to think about and commit to an answer. But many times questionnaires *do* ask respondents about something they do not know, relying on the respondent to volunteer the fact that they do not know. Due to social desirability effects,

however, most respondents would rather guess than admit to not knowing. To counter the invalidity of responses that result from this bias towards social desirability, a contingency item or question can be used (Babbie, 1998; Fowler, 1992).

A contingency question might be used to ask faculty members whether they use technology in teaching. If so, they are directed to complete that section of the questionnaire. That is, whether they respond to the next item is *contingent* upon their response to the first item in the series. If faculty do not use technology, however it has been defined in the questionnaire, there are two options. Faculty might be asked to respond to the next item to explain why they do not use technology, or they might be excused from answering that section through use of a skip pattern.

**Skip patterns.** When skip patterns are used, clear instructions must be given to respondents to let them know exactly where and how to direct their attention. Consider an example in which the purpose of the questionnaire is to determine the extent to which departmental faculty use specific cooperative learning activities. After defining cooperative learning activities, an initial question might be, "Do you use cooperative learning activities?" The options of "yes" and "no" are presented. If faculty select "yes," they proceed to the next item, which might ask them to identify and select the cooperative learning activities used. If they select "no," a skip pattern contained in brackets, printed in italics, and located on the same line directs them to the next numbered question they are to answer. This is illustrated in the following example in which respondents who select "no" as the response to question number two (Q2) are directed to skip to Q23:

2. Do you use cooperative learning activities?

a. Yes

b. No [*Skip to Q23*]

Skip patterns must be very clear. Arrows and diagrams are unnecessary. The cleanest skip pattern directs respondents to the number of the next question to be answered.

## Context Effects and Question Order

The context of any questionnaire depends upon item format and order. Once questionnaire items have been drafted, they need to be put into an order that establishes the overall context. The questionnaire might begin with a statement of purpose. The first section should contain items perceived by respondents as being easy and nonthreatening. This does *not* usually include demographic information. Until respondents trust and feel comfortable with the content and intent of the questionnaire, they will not feel safe in revealing personal information. Difficult or sensitive items, including demographic information, should be addressed in later sections of the questionnaire.

In what order should items be placed? Will a better or a different answer result if items are placed in another order? Probably, so examine the options and select the one that best reflects the purpose of the questionnaire. If it is necessary to ask similar questions, consider including a short note to respondents acknowledging that fact by saying, "Some questions may appear to be similar or the same. Please try to answer each to the best of your ability, regardless of whether or not you believe you have already answered it." This will help minimize redundancy effects, which occur when a question is asked and the response appears to be the answer to a different question. Later in the questionnaire that "different question" is asked, and the respondent either leaves it blank or repeats the answer given earlier.

In what order are general and specific items presented? Some studies have found that if specific items come first, the attention given them probably influences responses to general items (Israel & Taylor, 1990; Strack & Martin, 1991). If issues are complex or are about things that most respondents think little about, starting with specific items might provide respondents with more background on which they can base their responses to subsequent, general items (Sheatsley, 1983; Strack & Martin, 1991). It could also be argued that more thoughtful responses to specific items may be given after respondents have a chance to consider the material in a more general context.

Consider a questionnaire section that assesses overall faculty satisfaction with resources. The section might begin by asking faculty to identify resources, resource availability, how often resources are

used, and if resources are needed but unavailable. These specific items would be followed by a general item asking faculty to indicate their overall level of satisfaction with teaching resources. By responding to specific items first, faculty have time to reflect thoughtfully on their experiences and preferences before indicating their general or overall level of satisfaction with teaching resources.

The importance of pretesting or piloting any questionnaire cannot be overemphasized. Respondents in a pilot test should be selected because they have characteristics similar to potential respondents. Responses to the pilot test provide feedback on item wording and format; questionnaire context, layout, style, and flow; reliability and validity. These responses give the faculty developer invaluable information on which to base revisions that will strengthen the questionnaire and allow for the collection of reliable and valid data.

## Layout and Design

The layout and design of a questionnaire is very important. A self-administered questionnaire must be self explanatory, since there is no one present to clarify information. Separate or detailed instructions on how to complete each section of the questionnaire are unnecessary, since they are usually not read. The questionnaire must be uncluttered and pleasing to the eye. Avoid any attempt to fit more copy onto a page by using single spacing or by reducing the point size. For ease of reading, type should be no smaller than 10 points, although a larger type size is necessary if respondents are older or have impaired vision. Studies have shown that cluttered pages or pages with small print actually reduce the response rate as compared with the same number of questions spread out attractively over a greater number of pages (Fowler, 1992). Respondents are more inclined to respond to a lengthy questionnaire if it is precise, focused, and legible than they would be to respond to a short questionnaire that is cluttered and ambiguous.

Questionnaire items must appear in their entirety on the same page. That is, the stem of an item should never be split from its response alternatives. Each item and each response alternative should appear on a separate line with the response box or code to the left. If

placed to the right of the response alternative, the response box or code is often left unfilled, which lowers the rate of complete responses.

## Ethical Considerations When Using Questionnaires

An ethical approach to the process of developing and implementing questionnaires is very important. Most institutions have a research committee or board of human investigations that must review and approve questionnaires, whether part of a formal research project or not. These same committees or boards may also approve use of a questionnaire submitted by someone outside of the institution wishing to survey those within. Quality control of the design and use of questionnaires is important.

Respondents have a right to know several things before responding to any survey. They have a right to know the name of the organization implementing the questionnaire and its policy on confidentiality. Respondents should be told how the information they are about to provide will be used and protected (Babbie, 1998; Fowler, 1992; Singer, 1978). All identifying information should either be destroyed at the end of the study or stored in confidential, locked files.

Respondents should enjoy responding to and completing the questionnaire. They should even learn something in the process. If motivators such as food, money, small prizes, or gifts were promised, care must be taken to distribute them to respondents in such a way so as not to breach any promises of confidentiality or anonymity that were made (Fowler, 1992).

## Summary

Information from questionnaires is used to make important decisions in higher education. Critical decisions about faculty recruitment, promotion and retention, student satisfaction, and program effectiveness are often made, in part, based on responses to questionnaires. Yet such responses are only as reliable and valid as the words used to create each item on the questionnaire. Each item must reflect the purpose of the questionnaire, and each item must be ordered so as to reflect that

purpose in context. Multiple formats should be used to maintain interest and facilitate recall. Care should be taken to build trust and rapport with respondents before broaching items of a sensitive nature. Attention must be paid to every detail throughout the process.

Perhaps the ultimate success of questionnaire design and implementation depends on the morals and ethics of the individual or team who puts it all together, interacting with respondents and the data to ensure that all are treated with honesty and integrity.

## References

- Babbie, E. (1998). *The practice of social research* (8th ed.). Belmont, CA: Wadsworth Publishing.
- Bailey, D. (1997). *Research for the health professional* (2nd ed.). Philadelphia, PA: F. A. Davis Company.
- Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds.). (1991). *Measurement errors in surveys*. New York: John Wiley & Sons.
- Converse, J., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Thousand Oaks, CA: SAGE.
- Fowler, F. J. (1992). *Survey research methods* (2nd edition). Thousand Oaks, CA: SAGE.
- Groves, R. M. (1989). *Survey errors and survey costs*. New York: John Wiley & Sons.
- Israel, G. D., & Taylor, C. L. (1990). Can response order bias evaluations? *Evaluation & Program Planning*, 13(2), 365-371.
- Johanson, G. A. (1993). If you can't say something nice: A variation on the social desirability response set. *Evaluation Review*, 17(1), 116-122.
- Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). Fort Worth, TX: Harcourt Brace.
- Lavrakas, P. J. (1993). *Telephone survey methods: Sampling, selection and supervision*. Thousand Oaks, CA: SAGE Publications.
- Nolinske, T. L. (1994). The value of the mentor-protege relationship as perceived by the mentor, the protege, and the protege's boss. *Dissertation Abstracts International*, 55, 5508B. (University Microfilms No. 94-33, 900).
- Payne, S. L. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.
- Schwarz, N., Hippler, J., Deutsch, B., & Strack, F. (1989). Response scales: Effects of category range on reported behavior and comparative judgments. In E. Singer & S. Presser (Eds.), *Survey research methods* (pp. 165-172). Chicago: University of Chicago Press.
- Sheatsley, P. (1983). Questionnaire construction and item writing. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (Chapter 6). San Diego: Academic Press, Inc.

Shuman, H., & Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context* (second printing). Thousand Oaks, CA: SAGE Publications.

Singer, E. (1978). Informed consent: Consequences of response rate and response quality in social surveys. *American Sociological Review*, 43, 144-162.

Strack F., & Martin, L. (1991). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (Chapter 7). New York: John Wiley & Sons.

Contact:

Terrie Nolinske  
Associate Director, Education  
Lincoln Park Zoo  
2150 North Cannon Drive  
Chicago, IL 60614  
(312) 742-2070  
(312) 742-2040 FAX

Terrie Nolinske is Associate Director of Education at Lincoln Park Zoo. At the time this manuscript was accepted, Terrie was Associate Professor of Occupational Therapy and Assistant Professor of Health Systems Management at Rush University in Chicago, where she received the 1996 Outstanding Faculty Award for Excellence in Teaching. Terrie is also a freelance writer for national publications and a consultant across industries in the areas of strategic planning, presentation skills, survey research, people with disabilities, program development and evaluation, instructional design, and mentoring programs.

**TNI** Consultants in Professional Development

**Terrie Nolinske, Ph.D.**

9223 Rockrose Drive  
Tampa, FL 33647  
813. 973. 2102

tni@tniconsultants.com  
www.tniconsultants.com

Carbone, E., & Greenberg, J. (1998). Teaching large classes: Unpacking the problem and responding creatively. In M. Kaplan (Ed.), *To Improve the Academy*, Vol. 17 (pp. 311-326). Stillwater, OK: New Forums Press and the Professional and Organizational Development Network in Higher Education. Key Words: educational objectives, faculty development, large group instruction, newsletters.

## Teaching Large Classes: Unpacking the Problem and Responding Creatively

**Elisa Carbone**

**James Greenberg**

University of Maryland

*Teaching large classes well is a continuing challenge for many universities. This article looks at one university's systematic approach to the problem. It describes how faculty and administrators from all over campus were involved in a Continuous Quality Improvement (CQI) process, how the problems were clearly defined and recommendations made, and how the solutions that emerged also involved faculty from across the curriculum.*

How do we normally deal with a complex problem such as how to improve the teaching and learning in large classes? Too often—at least in our experience in academic settings—we refer it to an administrator, who can, at best, only deal with part of the problem, or we form a committee and talk it to death. This is perhaps especially true at large universities that have a tradition of decentralized academic responsibility and decision making.