



21106
117784

UNITED STATES GENERAL ACCOUNTING OFFICE
WASHINGTON, D.C. 20548

FEDERAL PERSONNEL AND
COMPENSATION DIVISION

B-205566

MARCH 11, 1982

The Honorable Herbert E. Ellingwood
Chairman, Merit Systems Protection
Board

Dear Mr. Chairman:

Subject: Questionnaire Design Problems Limit
Usefulness of Results (FPCD-82-9)

The Civil Service Reform Act of 1978 gave the Merit
Systems Protection Board a broad mandate to

"* * * conduct * * * special studies relating to
the civil service and to other merit systems in
the executive branch." (5 U.S.C. 1205(a)(3))

The Merit Systems Review and Studies (MSRS) group was estab-
lished by the Board to make these studies. Because of lim-
ited resources, MSRS chose to approach the studies' require-
ment through a series of questionnaires to Federal employees.

As part of our congressionally mandated activities, we
reviewed the design of two MSRS questionnaires: (1) "What
is the State of the Merit System and What Has OPM's [Office
of Personnel Management's] Impact on the Merit System Been
During 1980?" and (2) "Do Federal Employees Face Reprisal
for Reporting Fraud, Waste, or Mismanagement?" An MSRS re-
port based on the merit system questionnaire was released in
August 1981, and MSRS plans other uses of the data. MSRS
released a final report on the reprisal questionnaire in
October 1981.



117784

(966036)

220932

Questionnaire item design is not an exact science, and, therefore, designing good questions requires substantial knowledge of research in the area and extensive, practical experience. The purpose of this letter is to share with you our concerns with the design of these questionnaires. Our comments are based on the type of review that we have been doing on our own questionnaires for nearly 10 years and can be corroborated through questionnaire design and construction literature.

QUESTIONNAIRE OBJECTIVES
ARE NOT ACHIEVED

The objectives of both questionnaires are very broad and address complex issues. The reprisal questionnaire asks for "respondent opinions * * * concerning the reporting of illegal or wasteful practices in Government operations." The merit system questionnaire asks for "views of the health of the merit system and impact of OPM's actions on the quality of that system." We question the appropriateness of using abstract opinion data to meet descriptive and evaluative objectives. For example, the reprisal survey consistently asks questions which use the phrase "illegal or wasteful activities." As noted in the cover letter to the respondents, "this covers a variety of situations * * *." Perceptions that activities are illegal differ from activities that can be proven illegal in a court of law. Moreover, whether an activity is wasteful often cannot be objectively demonstrated and may be a matter of opinion. Objective counts of observable phenomena (e.g., time trends in the number of illegal activities per agency) are preferable to broad opinions.

Neither questionnaire adequately provides for a baseline or the necessary comparisons that would permit an assessment of the relevance of any problem identified. The relevance of respondents' reported instances of illegal and wasteful activities (including estimated dollar values) cannot be assessed without benchmarks of other institutions or other economic sectors. Without comparative data, the reader does not know if the reported values are high, normal, or low. In the merit system questionnaire, the design does not provide for comparisons against which to assess the impact, or lack thereof, of OPM's actions on the merit system--a stated objective in the questionnaire.

Additionally, in the merit system questionnaire, we question whether agency personnel specialists are the appropriate population to sample. For example, delegating

certain examining authorities to agencies was a very controversial decision. Many agency personnel specialists opposed delegated examining authorities because agency budgets were not increased to take care of the additional workload. Therefore, it would not be surprising to find that personnel specialists answered negatively when asked about delegated examining authorities, thus making it difficult to objectively determine whether delegating examining authorities is desirable or what OPM's impact was.

QUESTION CONCEPTS ARE TOO ABSTRACT

Questions in both questionnaires address very abstract concepts, though this is less of a problem in the reprisal questionnaire because several key phrases are defined on the cover. In the merit system questionnaire, respondents are asked for views on statements that have vague meanings and use ambiguous words. In question 3, for instance, respondents are asked the extent certain activities are being achieved, such as "Selections are based on 'fair and open competition'" and "Women and minorities have achieved fair representation within the organization." However, the concepts used are very vague; for example, what selections are being questioned--promotions, hiring, training, or something else? Is "fair representation" some proportion of the number qualified, the number in the work force, the number in the population, or some other group?

Using abstract concepts does not necessarily present a problem if all respondents

- understand the abstract concepts equally well,
- use the same basis for assessing and combining relevant facts into their overall judgment, and
- share the same values.

When these assumptions are not met, which we believe is the case in both questionnaires, meaningful interpretations of respondents' answers are impossible. Questions to determine whether the above assumptions held true were not asked in either questionnaire. Therefore, the effect on responses of using abstract phrases cannot be determined.

One way to minimize the effect of abstract terms and phrases is to define them, as the reprisal questionnaire

attempts to do. While respondents may disagree with the definition, at least they should understand them within the questionnaire context.

Research shows that respondents cannot be assumed to pull together the salient issues necessary to make meaningful responses. This problem is compounded when the questions are abstract or when several interpretations of the same question are possible. For example, question 3 in the merit system questionnaire asks respondents how effective OPM has been in some of its activities. The question assumes that respondents have a common understanding of what OPM's role should be in such broad areas as "Monitoring agency personnel systems," "Providing leadership and support for agency internal personnel management evaluation systems," and "Providing general management consulting services."

A way to test the respondents' understanding of OPM's role is to define the phrases and ask questions about the respondents' actual experiences with OPM. For instance, "How effective do you feel OPM has been in providing general management consulting services?" could have been broken into several questions (with appropriate response alternatives):

- Have you used the general management consulting services of OPM? (If no, skip to question __.)
- How satisfied were you with OPM's services?
- What effect, if any, did these services have on your organization?

Another problem occurs when a single question asks about two or more concepts that might be evaluated differently if they appeared as separate items. Again both questionnaires contained such "double-barreled" questions, but the most troublesome examples appeared in the merit system questionnaire. That questionnaire used the phrase "women and minorities" in asking several questions. If the respondent feels one way about "women" but another way about "minorities," how is the person supposed to answer? In question 3e, if respondents answer "to some extent," it is unclear whether there are problems with both women and minorities or major problems in one area and none in the other. Double-barreled questions can be avoided by asking separate questions; for example, question 3e would be "Women have achieved * * *" and "Minorities have achieved * * *." Responses from double-barreled questions are difficult to interpret and should not

be used without sufficient qualification or supplementary analysis to determine which part of the question is driving the answers.

QUESTIONNAIRE DESIGN IS INADEQUATE

The questionnaires have an aesthetic and professional appearance, but many questions in both questionnaires contain design flaws which limit the usefulness of the results or make the data very difficult to interpret.

Titles are biased

The titles of both questionnaires and the cover letter to the merit system questionnaire are biased. That is, the respondent is confronted with several questions and phrases that infer problems exist. The reprisal questionnaire title, "Do Federal Employees Face Reprisal for Reporting Fraud, Waste, or Mismanagement?" is not neutral. If the respondent has not considered the subject before receiving the questionnaire, the person is likely to infer that Federal employees do face reprisal and frame the answers to the questions accordingly. Similarly, the title and cover letter to the merit system questionnaire implies that OPM had an impact on the quality of the merit system and that significant improvements could result from the questionnaire results. The inference is that OPM's actions are subject to significant improvement. In both questionnaires, the respondents begin with negative inferences about the questionnaires' subjects. Bias may result in the respondents' answering the questions negatively without fully considering actual experiences or thoughts.

Bias can be minimized by being attentive to the problem, choosing words carefully, and thoroughly pretesting the instrument. At a minimum, the counter alternative can be presented. For instance, the reprisal questionnaire title would have been more neutral had it been worded "Do Federal Employees Face Reprisal * * * Mismanagement, Or Not?" The bias problem with the merit system questionnaire title and cover letter could have been lessened by minimizing the references to OPM's impact in the cover letter and by shortening the title to read "What is the State of the Merit System in 1980?"

Questions and response alternatives are also biased

Biased questions could lead some respondents to give different answers than they would give to a neutral wording of the issue. Most of the questions in the merit system questionnaire and many of those in the reprisal questionnaire were biased because the question did not pose the counter alternative to the issue. Questions that ask "How effective is * * *?" or "How adequate is * * *?" should have presented the counter alternative; for example, "How effective or ineffective is * * *?" or "How adequate or inadequate is * * *?" Without the counter alternative, a degree of effectiveness or adequacy can be inferred, which may cause the respondents to choose more positive responses than they otherwise would.

Bias may also occur because the response alternatives favor positive or negative responses. Several questions in the merit system questionnaire gave as response alternatives "Yes, more than one instance," "Yes, one instance," and "No." Unless the respondent gives the question very careful consideration, the tendency may be to choose one of the yes responses because they are given more emphasis in both number of choices and space. This problem could be avoided by asking two questions--the first requiring just a yes or no response, with a second question for those answering yes to determine frequency.

Other response alternatives should have been used

Not including certain response alternatives in both questionnaires may make the results very difficult to interpret. Without "neutral" and "escape" alternatives, the respondent may be forced to select a response alternative that does not accurately reflect the person's opinion.

Most questions in both questionnaires that asked respondents for opinions used response alternatives without a neutral alternative. Excluding the neutral alternative is useful when respondents are believed to have sufficient basis for being pro or con because respondents are forced to make a choice. We believe, however, that the neutral alternative should have been provided for most of the questions, especially in the merit system questionnaire.

Many questions in the merit system questionnaire ask about broad concepts that respondents may indeed feel neutral about. The questionnaire's use of the response alternative "no basis to judge" is not a suitable substitute for a neutral response alternative. Respondents may have had a basis to judge, but their opinions were neutral. That choice, however, was not provided. For example, the merit system questionnaire asks about the effectiveness of OPM's oversight of delegated examining authorities, the Federal Equal Opportunity Recruitment Program, and general performance appraisal systems. If respondents had a basis to judge and thought OPM was neither effective nor ineffective, they could not indicate this. Rather, they were forced to take a positive or a negative position which may distort the respondents' true opinions.

In addition, most questions in both questionnaires did not provide an "escape" alternative such as "other." When closed questions are used, respondents should generally be given an escape response alternative to test the inclusiveness of the other choices. A low number of respondents selecting the escape alternative does not mean the other response alternatives are good, but a high number of escape alternatives selected may indicate that the other response alternatives were not adequate.

Response alternatives
should be better defined

Many response alternative sets are flawed because the answer choices are not adequately defined and the range of choices is not spread out in equal appearing intervals. These problems appeared in both questionnaires but occurred most often in the merit system survey.

In opinion questions, respondents need enough information to determine the order and magnitude of the response alternatives. This is done by "anchoring" the response alternatives to words or phrases the respondents should be able to readily understand. In balanced response alternative sets--the same number of positive and negative answer choices--the answers are usually anchored at the extreme choices, for example, these choices use words that should be consistently understood. Other response alternatives should be spread between the extremes at equal appearing intervals. This practice helps respondents make meaningful choices in selecting the answers that best describe their opinions. Often, balanced response alternative sets are also anchored

at the midpoint. The "neutral response" can serve as the midpoint anchor with an equal number of positive and negative choices on either side. But as noted in the previous section, the neutral alternatives were usually omitted.

The balanced response alternative sets used in the questionnaires did not always anchor the answer choices or place them at equal appearing intervals between the extremes or from the neutral response. Not knowing the distance between answer choices prevents the use of certain statistics, such as arithmetic averages, in the data analysis and makes the meaning of the midrange response alternatives more difficult for the respondents to interpret consistently and correctly.

Question 12b of the merit system questionnaire illustrates these interrelated problems. The response alternatives were: "very helpful, somewhat helpful, no significant impact one way or another, hurt more than help." The positive extreme, very helpful, is not offset by the negative extreme, making the scale unbalanced and leaving one end not anchored. Since the scale is not balanced, the neutral response is not surrounded by an equal number of positive and negative choices, perhaps biasing the respondent to the helpful responses because they appear more often.

The answer choices presented do not appear to be separated by equal intervals. We cannot tell whether the interval between "very helpful" and "somewhat helpful" is the same as the interval between "no significant impact one way or another" and "hurt more than help." The distinction between "somewhat helpful" and "hurt more than help" is not clear and may overlap. The latter implies that the program may have been somewhat helpful though it hurt more than helped. A more balanced scale, such as "greatly helps, helps more than hinders, no impact one way or another, hinders more than helps, greatly hinders," would minimize these problems. The phrases "greatly helps" and "greatly hinders" define the extremes, and the neutral response serves as a midpoint anchor. Respondents should be able to judge, with a high degree of consistency, the order and magnitude of the response alternatives.

These flaws call into question the validity of any observations and conclusions drawn from the answers received. By not anchoring and spreading the response alternatives at equal appearing intervals from the anchors, the respondents are likely to interpret the order and magnitude of the

answer choices differently. If respondents do not interpret the response alternatives the same way among themselves and as the MSRS authors intended, the answers received will be very difficult to interpret correctly and any observations and conclusions drawn must be carefully qualified.

Response sets are a problem

Research also shows that respondents may develop a pattern while answering questions. That is, after choosing a particular choice in several questions, they see a pattern and may not read the remaining questions completely and give less thought to the answers. The list format used for many questions, in particular questions 3 through 5 in the merit system questionnaire (requiring 33 responses), may evoke a response set. Since most of the subitems are favorably worded and the response alternative closest to the question subitem is favorable, the response set problem may cause the positive responses to be overstated.

The effects of response sets are controversial and may not occur when the respondent population is highly educated and verbally facile. On the other hand, even this group may get tired and bored toward the end so items later in the questionnaire (such as question 32 in the merit system questionnaire) may be subject to a response set.

As a matter of principle, both sides of an issue should be stated in the stem of a question or both positively and negatively worded stems should be used in a subitem list. Other ways to minimize response set problems would include using a numerically relative response scale or reversing the response alternatives on half of the questionnaires so that the negative response would be closest to the subitem stem.

Respondents need more support

The reprisal questionnaire also has a problem inherent in its objective. To determine whether employees face reprisal, a number of questions asked the respondents to criticize or state negative views about themselves; other employees; and their department, agency, or office. Experience indicates and surveys have shown that people are generally less willing to criticize than they are to praise. To help minimize this tendency, greater efforts should have been made to assure the respondent that answers are confidential and anonymous. Respondents should also have been

encouraged to answer honestly by informing (reminding) them of the social acceptability and possible positive results of honest answers. The objective here is to overcome the respondents' reluctance to be identified with criticism of their workplace by reassuring the respondents that honest answers are sought and by minimizing the natural discomfort of giving critical answers.

Time frames are mixed

The time frames in the merit system questionnaire are confusing. In the title, cover letter, and most questions asking for a time reference, calendar year 1980 is asked about. Since the questionnaire was sent out in April 1981, and most respondents probably filled it out in May 1981, separating events that occurred between January 1981 through April 1981 may be difficult. To compound the problem, questions 4, 5, and 6 ask for information from the last 12 months--presumably April 1980 through April 1981. Questions that do not specify a time frame or say "now" (question 17) will get the respondents' views as of May 1981. The mix of time frames will cause problems in interpreting the data.

Other technical problems

In several questions in both questionnaires, respondents were asked to give written comments and referred to space several pages away. By not providing space with the question, some respondents will be reluctant to answer, will forget to answer, or will forget parts of their answer. To minimize this burden, space should have been left where the question was asked.

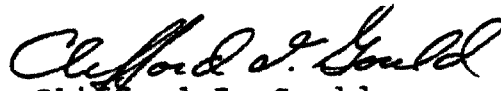
In the merit system questionnaire, acronyms are used without being defined. While the group of personnel specialists sampled should know what OPM, CSRA, and EEO mean, defining the acronym the first time it is used would help eliminate any possible confusion. Incidentally, for reasons unknown, OPM is defined in question 29 after it had been used 17 times in previous questions.

B-205566

We believe these design problems adversely affect achievement of the questionnaires' major objectives. However, if properly qualified, the questionnaires' results may help to identify general problem areas on which to focus general research.

We hope our comments will be helpful to you in your design of future questionnaires.

Sincerely yours,


Clifford I. Gould
Director