

Excerpt from “Chapter 9, Practicing the Elicitation and Training the Project Personnel” in Meyer and Booker, *Eliciting and Analyzing Expert Judgment: A Practical Guide*, Knowledge Acquisition for Knowledge-Based Systems series, vol. 5 (London, United Kingdom: Academic Press), 1991 p. 153-157. To be reprinted by the American Statistical Society, SIAM Series in March 2001.

Chapter 9 provides the last check on all aspects of the elicitation design before the elicitation is conducted (chapter 10). The logistics, ease of use, interface, and timing of the parts of the elicitation process are examined to find any remaining glitches and resolve them. The project personnel identify any remaining problems by practicing and pilot testing the different parts of the elicitation.

What Requires Pilot Testing?

The procedures that require pilot testing are those in which the experts participate (e.g., giving their judgment). The experts' understanding is critical to their participation. Therefore, anything that the expert must understand makes for a good pilot testing candidate. However, a smaller set of all those things that the expert must understand is usually pilot tested because pilot testing is time and expert intensive. As a general rule, pilot testing is done on the experts' understanding of instructions, whether these are given orally or in writing (e.g., how to fill out a survey). The following parts of the elicitation are suggested for pilot testing.

- 1. The experts' understanding of the problem or question.** In some cases, a final statement of the question or problem exists long before the experts provide their solutions. In other projects, such as the NUREG-1150 reactor risk study, the experts develop their own statement of the problem from an initial version. In the first case, the pilot test provides information on the experts' interpretation of the problem. If the project personnel did most of the question development, their interpretation is likely to differ noticeably from that of the experts. In the second case, the pilot test checks that the experts understand that they should take the basic problem area and refine it.
- 2. The experts' use of the response mode.** The pilot test allows the interviewer or knowledge engineer to make sure that the expert understands how to use the mode. The expert's response can be checked against both the instructions and logical or mathematical standards. For example, probabilities for all outcomes, previously defined as mutually exclusive events, should sum to one.
- 3. The experts' understanding of the elicitation procedures that they must follow.** For example, if verbal report is chosen, it would be important to make sure the experts knew that they must *think aloud* and that they had the capability to do so.
- 4. The experts' ability to use any documentation format.** The pilot test provides feedback on the experts' understanding of the directions on how to complete the format. (e.g., on a mail survey). We have found that experts typically do not provide as thorough a documentation of their judgment as they are requested to do.

How to Pilot Test

Sample Sizes and Selection

Ideally, pilot testing uses experts who represent the range of experts who will later participate in the actual elicitation process. For instance, if the experts have been drawn from positions in government, private industry, and academia, the pilot sample should contain members of each of these positions as well. Consider the factors used to select the expert populations and use these to choose the sample for pilot testing. In selecting an expert pilot sample, consider that this selection decreases the pool of experts whose judgments can be elicited later. In general, experts who participate in the pilot test or who have otherwise assisted in the method's development should not participate in the actual elicitation. Similarly, we would advise against using the advisory experts in the pilot sample. The advisory experts will not approach the test materials from a fresh perspective if they helped develop them.

The size of the test sample depends on the size of the expert population available for the elicitation. Test samples for expert elicitation rarely, if ever, have the large sample sizes (10% of the total) associated with traditional mail surveys. Typically, expert samples for pilot tests are five persons or less because the largest expert population does not usually exceed 50. Because of these small sample sizes, the strategy for pilot testing in expert judgment studies has been to test intensively. In other words, try to obtain as much benefit or feedback as possible from these few experts.

Two types of pilot tests are mentioned below. We recommend conducting the intensive pilot test first. The *intensive pilot test* was developed (Meyer, 1986, p. 92) to trace the expert's understanding. It consists of structured, in-depth interviews and observations. The other type of pilot test is called *limited* to distinguish it from the traditional tests of high sample size and from the intensive pilot tests mentioned above. Limited pilot tests work best after intensive pilot tests have been performed and the elicitation revised. The limited pilot test allows the project personnel to practice the elicitation procedures, time their duration, and check how these procedures fit together.

Sequence for Pilot Testing

The sequence for performing these pilot tests follows:

1. Apply part one and part two of the intensive pilot test (as described on the next page) to the elicitation.
2. Use the pilot test results to revise the tested parts.
3. Practice the entire elicitation process and do limited pilot tests on those parts that will eventually involve the experts. For example, the project personnel would rehearse their introduction to the elicitation in front of the experts for the limited pilot tests. Project personnel would then practice the elicitation procedures on these experts. The expert data from the practice elicitation would be used to check the other steps of the elicitation. Specifically, the data would be documented, aggregated, and used as input just as it would be when the elicitation is actually performed. Each of the elicitation procedures is timed during the practice.

How to Conduct Intensive Pilot Tests

To our knowledge, this type of pilot test is the only one that allows people's thinking to be tracked through information presented in written form. The intensive pilot test provides two kinds of information: how the experts progress through the information, their general impressions, and when and why they decide to respond to particular questions; and how the experts specifically interpret each direction, statement of the question, or response mode option.

First, the materials to be pilot tested are selected using the list provided under *What Requires Pilot Testing?* A typical selection would include the set of problems, any written directions on assumptions or definitions that the expert must use, and directions on what information the expert should document on the form and in what manner, such as on a continuous linear scale.

Intensive pilot testing takes place with one expert at a time in a face-to-face situation. The interviewer sits on the opposite side of the expert's handedness to allow easy viewing of the expert's writing; that is, the interviewer sits to the expert's left, if the expert is right handed.

Intensive Pilot Test—Part 1. For the first part of the intensive pilot test, the expert is instructed to complete the written materials as he or she would if no observer were present. The expert is also asked to *think aloud* (verbal report) as he or she reads through the written materials. The interviewer has a copy of the same materials given to the expert for recording the data.

While the expert pages through the written materials, the interviewer records the order in which the expert looks at the materials, his or her pauses, gestures, facial expressions, and words. For example, the expert may skim through the introduction, cover letter, or directions and then flip through the rest of the materials before returning to read each of these more thoroughly. The expert may also make a variety of comments, including *I have problems with this page and will probably let it sit on my desk for several days.*

In addition, if limited pilot tests will not follow intensive pilot tests, the interviewer can record the expert's starting and ending time on the first part of the intensive test. The limited pilot tests described in the next section provide better time estimates because the expert does not have to think aloud throughout the interview. However, gathering some data on time is better than gathering none, so this data should be gathered during the intensive pilot tests if they will not be gathered by other methods. If the intent is to obtain time data, save the questions of the expert until the second part of the intensive test. Otherwise, these questions will inflate the time estimates of how long the expert takes to respond to the written materials.

The intensive pilot test provides better time estimates than might be expected. The time estimates obtained from the first part of the intensive pilot test would be high because the expert continuously verbalized his or her thoughts (and may not do so for the actual elicitation). However, we have found that this measurement provides an adequate estimate of the upper limit of time needed. This occurs because the experts selected for the pilot test may not represent the range of experts, some of whom could take a significantly longer time to finish. For instance, we have seen pilot tests that last two hours while the actual elicitations range from 1 hour to 2 hours and 15 minutes.

The expert is frequently allowed a brief break between the first and second part of the intensive pilot test.

Intensive Pilot Test—Part 2. During the second part of the intensive pilot test, the expert is asked to paraphrase in his or her own words the meaning of each direction, question, or response option. This information allows the interviewer to track the expert's interpretation in detail. It has always amazed us that something could be interpreted in so many different ways, ways that we had not thought of because we *knew* what we meant. The interviewer can also question the expert about any reactions, such as a look of puzzlement, noted at one of the questions during the first part of the pilot test. This questioning can jog the experts' memories so that they can give a more thorough description of their impressions.

Even though the interviewer records all information on a copy of the written materials, he or she still collects the expert's hard copy. The expert is thanked and asked not to discuss the details of the pilot test

with anyone because it could influence their responses. The data from the intensive pilot test is examined and used in revising the elicitation procedures or the logistics of who performs the different tasks. If limited pilot tests are planned, they are the next step.

How to Conduct Limited Pilot Tests

In limited pilot tests, the elicitation procedures are conducted as closely as possible to the way that they will be done for the actual elicitation. There would be little point in examining the expert's responses to a form or a procedure that did not resemble the one to be administered. Thus, in the limited pilot test, the experts are given the briefings or the introductions that have been planned for the larger population of experts. Similarly, the sample experts receive whatever forms and instructions on providing their responses that the other experts will subsequently use during the elicitation sessions.

The interviewer elicits the data in the planned manner and obtains the experts' responses. In addition, data are gathered on how long each elicitation lasts so that these can be added to other time estimates to produce a total. If the elicitations are being done in person, the interviewer may also record the amount of time it takes an expert to respond to a particular problem.

The following illustrates how timing data are used. The practice briefing of the group of sample experts lasted 2 hours and one elicitation lasted 2 1/2 hours. This time included the experts' review of the documentation of their responses. Assuming that only one elicitation team performs 10 individual interviews, about 27 hours would be needed just to perform the elicitation.

As with the other pilot test, experts receive thanks for their cooperation after completing the session. The data from the pilot test are used in practicing the other stages of elicitation, including mathematical aggregation and modeling, as mentioned in the *Sequence for Pilot Testing*. In addition, the limited pilot test may prove useful in training project personnel, as will be discussed in the next section.