Q-Bank: Development of a Tested-Question Database
Kristen Miller, Ph.D., National Center for Health Statistics
ksmiller@cdc.gov

Introduction: No Place to House and the Implications for Survey Research

The concept of Q-Bank grew out of a dialog of complaints at the 2001 QUEST meeting, an international workgroup whose members evaluate survey questions primarily through cognitive interviewing. At this particular meeting, discontent was directed at the problem of vanishing cognitive test results, leading, of course, to the ongoing need to test the same questions year after year.

This is certainly the case for my agency, the National Center for Health Statistics. Our staff (the staff of the Question Design Research Laboratory) often receives requests from survey divisions within NCHS as well as from other agencies regarding the wording of specific questions along with their testing results. These requests typically leave us victim to our collective memories, laboring to recall which specific projects might have involved the question in question along with the name of the staff person who may have headed up the particular project. Over the years, as staff come and go, projects would be forgotten as well as which questions were specifically examined, and (perhaps most importantly) what was actually found.

I'm chuckling as I write this because today our staff is on a massive search for testing results to questions about income. A few months ago I was asked to provide a summary of QDRL test findings for income—one of the most common and recurring survey questions in the history of survey research and likely the most tested in the QDRL. Fully appreciating the magnitude of the request, I procrastinated as long as possible, hoping it would disappear as easily as the test results. But it didn't, and today we are all scanning our computers, going through diskettes and contacting old staff to track down work that may (or may not) have been conducted within the past 10 years.

Because findings from cognitive interviewing studies are not journal-appropriate, there is no public forum for this type of work as there is for other methodological studies. There is no warehouse to search for previous work or to draw upon the lessons of past evaluations. Chances are, at this very time, some survey researcher somewhere is writing a survey question—a question that has likely been written, tested, re-written and even fielded—from scratch.

This lack of "place" has implications at several levels. At a personal level, those of us who evaluate questions are often frustrated, testing and improving questions for specific surveys fielded in a specific year, only to be faced with the same question by another survey or even the same survey but for another year. On an administrative level, resources are routinely wasted by the repetition of tests and staff hours are lost to tracking down results from previous studies. From a survey research perspective, data quality is

compromised as surveys are unable to advance consistently with improved question design.

In my mind, however, the weightiest implication falls on the field of question evaluation, itself, and the methods for conducting evaluations. The inability to collectively house evaluation research necessarily means the inability to make comparisons across studies. Consequently, we are unable to identify where differences in methods may produce different findings. The vast majority of cognitive testing literature, for example, is dedicated to probing techniques, specifically, what types of questions the cognitive interviewer should ask and how they should be asked. Is it better to ask probe questions retrospectively (after all of the test questions are asked) or concurrently? Should cognitive interviews be fully pre-scripted—like a survey questionnaire—or is a semi-structured approach better? Despite conference papers and book chapters of debating, with no comparative methodological studies, there are no legitimate conclusions, and the dialog of "differing opinions" or "equally valid schools of thought" is destined to continue.

Even more, the lack of comparative analysis across evaluation projects hinders consideration of other, perhaps more relevant, aspects of the cognitive interview. For example, extremely little attention has addressed the cognitive interviewing sample. How do respondent characteristics impact cognitive interview findings? And, what is the most appropriate sampling strategy for cognitive interviewing studies? With little evaluation potential, it is impossible to see where the method can be improved. With no public forum, the ability to evaluate the evaluation is lost as well as the potential for verifiability. Arguably, the lack of forum produces a stagnant method.

The lack of place, additionally, has implication for the field's relationship with the larger field of survey research. I don't think it's a stretch to say that most researchers who use survey data have never seen a cognitive testing report, nor would they know what cognitive testing comprises, certainly not how it could assist them as data-users. Cognitive interview data reveals how survey questions are interpreted as well as the various types of information that respondents consider when providing an answer. Without this interpretive perspective, survey data can too easily be taken at face-value, appearing much more one-dimensional and simplistic than is actually the case. With no access to cognitive testing reports, the field of question evaluation is unnecessarily limited to the realm of front-end survey production despite its enormous potential for post-production analysis.

Database Structure

In Spring of 2002, The National Center for Health Statistics began development of a database for the QDRL. This early development comprises the elemental foundation of Q-Bank—which has since grown into a much broader, interagency-based initiative. Originally, the database was developed to meet the specific needs of the QDRL, that is, to track its own work as well as to conduct broader methodological work pertaining to

question design. The database, for example, includes question characteristics (e.g. response error type and response category type) so that, as cognitive testing reports accumulate, relationships between these characteristics can be investigated (e.g. Are yes/no questions more likely than open-ended questions to produce a particular type of response error?) The underlying design of Q-Bank, consequently, is based on this dual purpose: 1) to provide easy access to past cognitive evaluation projects and 2) to support more general methodological research in question design.

The database houses individual questions that have been tested and then links each question to the report which contains specific findings about the question as well as a methods section detailing how the evaluation was actually conducted. The coding structure, making up the essential core of the database, identifies characteristics of each question. Questions are searchable not only by topic, but also by the type of information that a question elicits, the types of errors associated with the question as well as the type of response category that the question uses. These categories can be used to search the database so that, ultimately, relationships between the question characteristics can be examined. Chart 1 outlines the question characteristics that are searchable within Q-Bank.

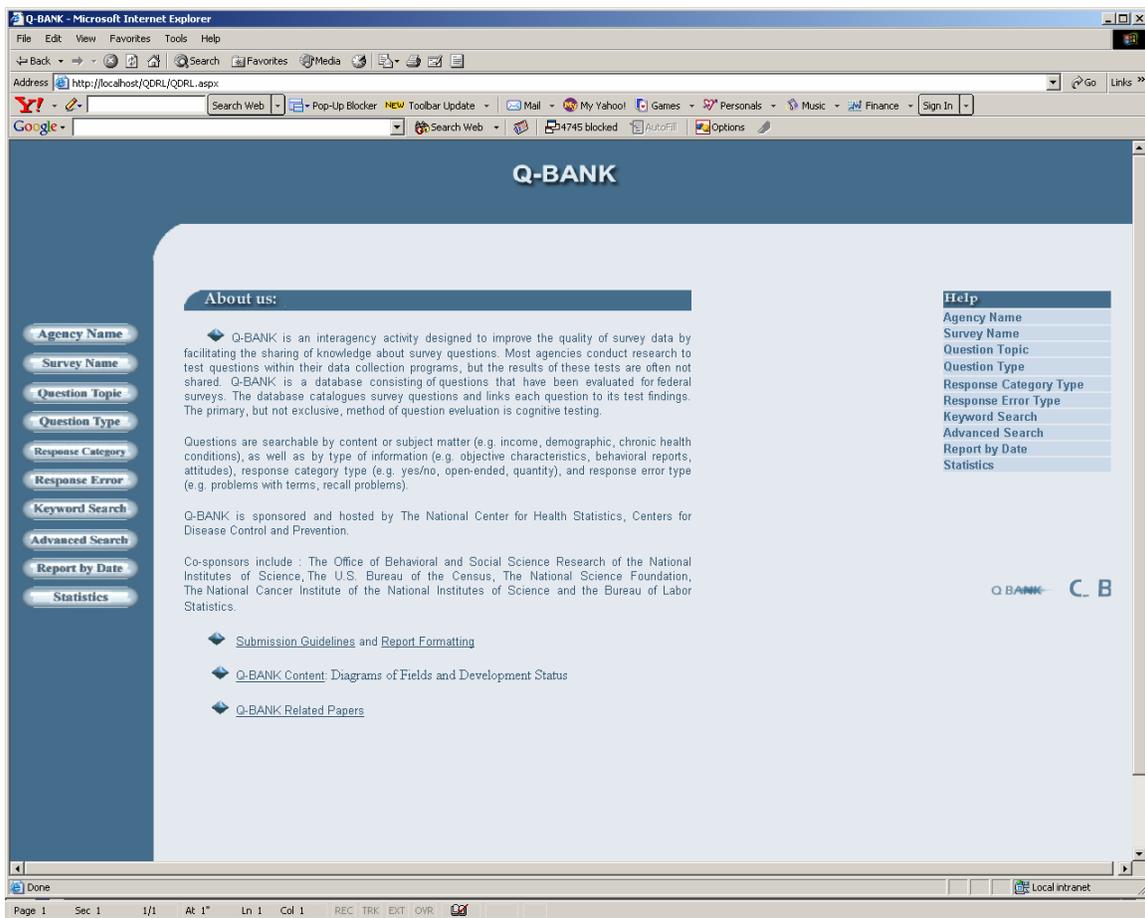Chart 1: Question characteristic fields and categories

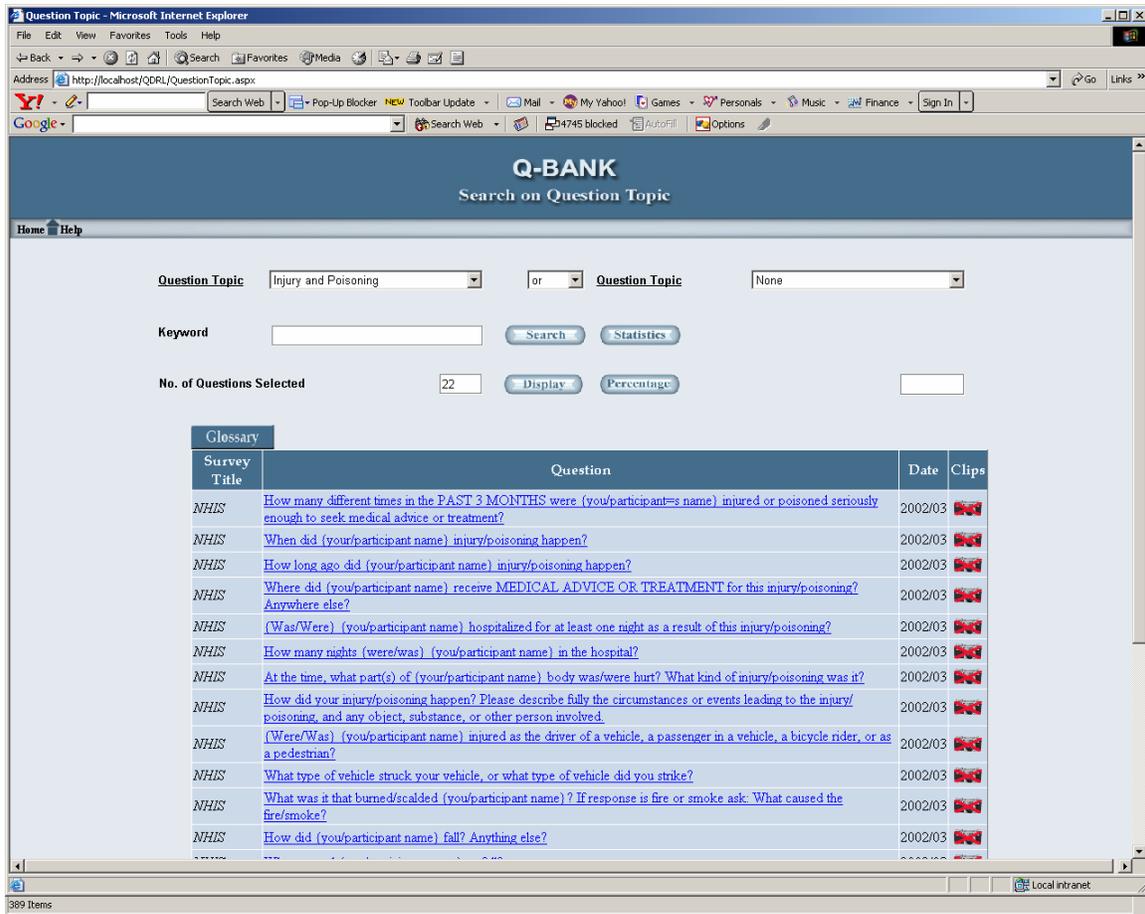| Question Topic	Question Information Type	Response Category Type	Response Error Indicator

Numerous, open categories accessed by drop-down list Examples: Demographic   Income   Race   Gender   Nationality
 	
 Health   Insurance   Conditions   Smoking
 	
 Employment   Looking status   Occupation   Job Change   Salary   Benefits
	1. Events, actions & behaviors
	 2. Objective characteristics
	 3. Subjective characteristics
	 4. Speculations
	 5. Attitudes
	 6. Knowledge tests
	 7. Other
		1. Yes/no
		 2. Select one textual
		 3. Select one numeric
		 4. Mark all that apply
		 5. Allocation
		 6. Open-ended textual
		   a. Delineated
		   b. Non-delineated
		 7. Open-ended textual
		   a. Delineated
		   b. Non-delineated
		 8. Other
			1. Interviewer difficulty**
			 2. Visual design problem*
			 3. Missed instructions*
			 4. Unknown terms
			 5. Ambiguous concepts
			 6. Overly complex
			 7. Assumption
			 8. Double barrel
			 9. Questionnaire effects
			 10. Recall & estimation error
			 11. Inadequate response options
			 12. Problematic answer space**
			 13. None as tested

* self-administered questions only
**interviewer-administered questions only

To develop the category schemas, hundreds of tested questions were categorized by QDRL staff. Categories were then refined if a question did not fit the existing schema or if they could not be consistently categorized. When Q-Bank became an interagency endeavor, fields and categories were again revised to meet the needs of the collaborating agencies. In designing the coding structure of the database, it was important to reach a balance: enough detail so that categories were useful, but not so detailed that the coding process became labor-intensive to the extent that no one would code reports—thereby leaving us with an empty database.

The picture below is the home page for Q-Bank. From this page, searches can be conducted on question topic, question information type, response category type and response error indicator.
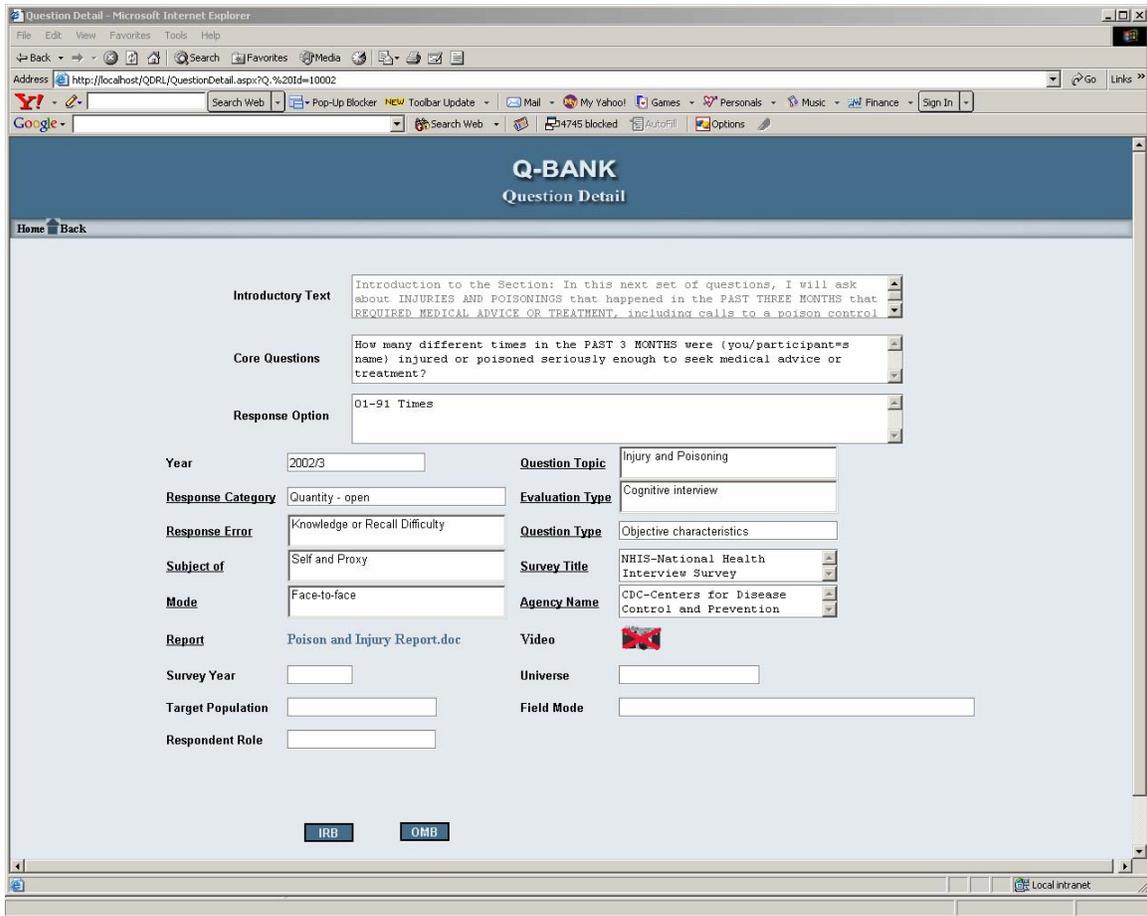

From the home page, Q-Bank can search and then report all questions related to a particular topic. For example, if a researcher entered the Question Topic "Poison and Injury" a list of questions housed in the database would appear:

The questions are listed in alphabetical order. The date that the question was tested appears next to the question along with the agency that conducted the test.[1] A question is entered into Q-Bank each time that it is tested. Therefore, questions that have been tested multiple times will be repeated in the list for as many times as it was tested. This will indicate to the Q-Bank user that more than one set of test findings are available.

From the search results page, a user may select one particular question which will then produce detailed information about the question:

---

[1] The picture of the camera to the far right is a feature that is being added to the NCHS Question Design Staff version of Q-Bank. This link will allow QDRL staff to access video clips of cognitive interviews in which the question was actually tested. Over the past 5 years, use of video has become a central tool for conducting analysis at NCHS, so adapting Q-Bank to also organize and track interviews was a particularly worthwhile development.

The information page for a specific question provides the question text, any introductory text, and the response categories. If the question has a flash card or is presented in a matrix, a link will connect this site to the additional material. Additionally, the page relates the type of evaluation that was conducted along with the error categories associated with the question, that is, the types of problems that were discovered in testing. Most importantly, a link is provided so that the user can refer to the final report and read, in detail, what was found as well as how the evaluation was actually conducted.

Value of Q-Bank

It should be emphasized that this is not a database of ready-to-use survey questions. Q-Bank users would be remiss if they chose to use a question appearing in Q-Bank without first examining what was found in the evaluation. Questions appearing in Q-Bank are the original questions—questions that were tested—and are not the revised questions. Consequently, some of the questions in the database have many problems.

In working through the underlying structure of Q-Bank, the team has concluded that such an endeavor—a database of ready-to-use questions—is more like Don Quixote's quest for the impossible dream. After all, how could a *good* survey question actually be defined? What are its characteristics? How is it possible to know that a

particular question is suitable for every potential respondent and within the context of every potential questionnaire? There is no pre-test, including cognitive testing, that can concretely determine whether a question is good or bad, but rather identify the types of problems associated with a particular question within a particular context.

In the end, Q-Bank, as a database of tested questions with links to evaluation findings, is more valuable to survey research than a repository of ready-made questions. The dream database of "good" survey questions—if it were even possible—could never function as a methodological tool serving a broader research agenda. Q-Bank contains descriptions of what actually went wrong with questions when asked in cognitive interviews (if something actually went wrong) as well as the ways in which the questions were interpreted by respondents when asked in cognitive interviews. That is, the database contains an account of the type of data that the question actually elicits; the meaning behind the statistic.

That being said, it is envisioned that Q-Bank will ultimately serve survey research in a variety of ways. Specifically, it will be a tool for:
1. Designers tasked with developing new questionnaires who would be assisted in seeing what did and did not work in previous questions and *why* they did not work,
2. Subject matter analysts seeking additional information about the data in order to help make sense of relationships that they have found or have not found
3. and, survey methodologists doing research on question design and response error.

Q-Bank Administration

In the past year and a half, Q-Bank has become a collaborative, interagency effort. Joining NCHS in the Q-Bank effort is: The Office of Behavioral and Social Science Research of NIH, the Bureau of the Census, the National Science Foundation, the National Cancer Institute of NIH and The Bureau of Labor Statistics. A steering committee of representatives from the various agencies has been formed and now oversees the direction and development of Q-Bank. NCHS will continue to manage the daily operations, hosting the database on-line and providing access to and warehousing pre-test findings of questions from all participating agencies. Our highest expectation is that the database will house a massive amount of questions on a widespread range of topics, including health-related questions as well as economic, education, justice and agricultural questions.

At the end of June, Q-Bank completed the CDC accreditation process, authorizing Q-Bank to be moved online. It is now accessible to steering committee members who will test the application by entering data and running searches. It is not yet clear when broader access to Q-Bank will be provided, but it is my hope that this will occur by the end of the year.

By far, the greatest challenge faced in the development of Q-Bank was related to the fact that the "NCHS version of Q-Bank" needed to accommodate the joining agencies. The vast majority of evaluation work conducted at NCHS is for population-based, interviewer administered surveys. Respondents almost always represent themselves, their household or are answering for another person in their household. The coding structure developed by NCHS, consequently, only pertained to these types of questions. Other agencies, however, often conduct question evaluation work for establishment surveys and self-administered surveys. As one of the papers will describe in detail, incorporating these types of surveys into Q-Bank meant re-conceptualizing the coding schema and adding many fields and categories.

In the end, the process of re-working the coding schema took over a year and tripled the size of the database structure. Though a delay, the investment is worthwhile because 1) Q-Bank will accommodate more types of surveys and is better equipped to incorporate future web-based surveys, and 2) all participating agencies now back the category schema and can support the federal standardization of cognitive testing reports according to Q-Bank requirements.

The steering committee also made decisions regarding procedures for posting reports onto Q-Bank. Specifically, reports must be submitted through one of the sponsoring agencies and will receive two tiers of review: first, from the sponsoring agency and, second, from the Q-Bank administrator located in NCHS. Inclusion into Q-Bank will require that 1) reports are coded consistently according to the Q-Bank categorization schema, 2) reports are properly formatted and include a fully developed methods section, and 3) reports provide enough detail so that conclusions are sufficiently illustrated.

Conclusion: Challenges for the Present and Future

As a centralized, multiple-search warehouse of survey questions with links to pre-test information, Q-Bank is envisioned to:
1. provide an invaluable resource to topic specialists and other non-design specialists tasked with developing new survey questions
2. Promote standardization and comparability across federal surveys
3. facilitate the comparison of cognitive testing findings over time, in different settings, with different sub-populations and across federal agencies, thereby providing insight into the reliability and validity of cognitive testing and other methods of question evaluation
4. provide ability to analyze specific types of question characteristics that may contribute to response error
5. provide an additional resource for analysts in the interpretation of survey data
6. increase awareness of response error as part of total survey error, generating a research agenda that ultimately seeks to quantify response error and improve the overall methodology of survey research

To maintain this vision, the steering committee will face ongoing challenges. The primary challenge currently faced is in the actual implementation of these procedures that were previously outlined. For Q-Bank to serve as a valuable methodological tool, it will be critical that the testing reports meet the minimal requirements set forth by the Q-Bank steering committee. After all, the quality of the research generated by Q-Bank can only be as good as the individual testing projects that it houses. Secondly, it will be critical that data, specifically the question characteristic codes, be coded and entered into the database consistently. It is hoped that the two-tier review process with one agency (i.e. NCHS) overseeing data entry will ensure data quality. Additionally, plans are underway to develop methods for testing data consistency.

Finally, there are two prominent challenges for Q-Bank of the future. First, a continual eye must be kept toward development. To thrive, Q-Bank must grow in terms of the amount of data it houses, the number of agencies it serves as well as the number of pre-test methodologies that it will accommodate. For example, a future version of Q-Bank could include questions in different languages, thereby allowing for examination of cross-cultural comparability and translation methodologies. Finally, the original goal of serving a methodological research agenda must be maintained. As only a repository of questions, it is of much less value to survey research. The value of Q-Bank lies in its potential to advance the field of question design.