

**STEPS INVOLVED IN ADDING QUESTIONS OR EXISTING SCALES
TO AN EXISTING LARGE-SCALE FEDERAL SURVEY**

Evelyn S. Mathis

Consultant

August 1, 2016

This paper was commissioned by the National Academies of Sciences, Engineering, and Medicine Standing Committee on Integrating New Behavioral Health Measures into the Substance Abuse and Mental Health Services Administration's Data Collection Programs. Opinions and statements included in the paper are solely those of the individual author, and are not necessarily adopted, endorsed or verified as accurate by the National Academies of Sciences, Engineering, and Medicine. Support for the Standing Committee was provided by a contract between the National Academy of Sciences and the U.S. Department of Health and Human Services.

Introduction and Purpose

This paper is to delineate the variety of activities that need to be conducted and assessed in order to add existing brief scales or new questions to an existing, ongoing large-scale federal survey. These activities include the identification of the most appropriate questions to solicit the desired information, testing and other activities that may be needed after the scale or questions have been identified and selected. This paper focuses on the National Survey on Drug Use and Health (NSDUH), which consists of a core set of questions and several modules devoted to different topics. The NSDUH is conducted by a contractor for the Substance Abuse and Mental Health Services Administration (SAMHSA) an agency of the Department of Health and Human Services.

SAMHSA is charged with reducing the impact of substance abuse and mental illness on America's communities; to carry out this charge, SAMHSA maintains current data on the incidence and prevalence of substance use and mental health problems in the United States. The agency is considering expanding behavioral health data collection in the following four areas:

- (1) Serious emotional disturbance in children,
- (2) Specific mental illness diagnoses with functional impairment in adults,
- (3) Trauma and,
- (4) Recovery from a substance use or mental disorder.

One of the options being considered as a mechanism for collecting new data is the addition of new modules or follow-ups to existing surveys such as the NSDUH. This paper will focus primarily on the NSDUH, but the procedures for adding questions to an existing survey are applicable to most large-scale federal surveys.

NSDUH Background

The NSDUH has been ongoing since 1971 and is the primary source of reliable statistical information on the prevalence, patterns, and consequences of the use of illicit drugs, alcohol, tobacco, nonmedical use of prescription drugs, substance abuse treatment, mental illness and mental health treatment among the U. S. civilian, noninstitutionalized population aged 12 or

older.¹ Early on, the study was conducted at various intervals, settling into a pattern of every two years. The demand for current, accurate information rose sharply by the late 1980s to early 1990s prompting SAMHSA to conduct the survey annually starting in 1990. The survey generates estimates at the national, state, and sub-state levels such as cities or counties in one state. Because of its long-running history, the NSDUH provides trend information on the use and abuse of substances to the health care community, both public and private.

NSDUH employs a stratified multistage area probability sample that is designed to be representative of both the nation as a whole and for each of the 50 states and the District of Columbia. The survey collects data through face-to-face interviews at the respondent's place of residence.²

Many sampling design and data collection changes have been made to the NSDUH over the years. Changes were designed to provide Spanish as a language option, add needed survey data, increase the efficiency of data collection and processing, and to improve the quality of the data being collected. The most recent changes made to the NSDUH were implemented for the 2014-2017 surveys. Changes were made in the sample sizes allocated to each state and to different age groups in order to increase the precision of national and many state estimates for youth as well as estimates for older adults; changes were also made to the questionnaire to improve the quality of data collected, expand the number of certain items and address changing substance use, mental health policy, and research needs.

Regardless of changes, the survey continues to track trends in specific substance use and mental illness measures, and assists stakeholders and policy makers with information for budgeting, planning, and treatment of substance use and mental illness.

Planning for the Addition of New Items

All federal agencies have access to general guidelines on adding new items or conducting major revisions to an existing survey. These guidelines are published by the Office of

¹ Center for Behavioral Health Statistics and Quality (2015). *Behavioral Health Trends in the United States: Results from the 2014 National Survey on Drug Use and Health*. Substance Abuse and Mental Health Services Administration, Department of Health and Human Services. Rockville Maryland.

² Same as above.

Management and Budget (OMB).³ The guidelines provide specific standards that contribute directly to the utility of collected information including: survey planning, survey design, and pretesting. The guidelines emphasize that adequate planning and development are critical to the identification of items that would provide needed information, in this case, to the health care community. It requires that agencies initiating a new survey or major revision of an existing survey must develop a written plan that sets forth a justification, including: goals and objectives; potential users; the decisions the survey is designed to inform; key survey estimates; the precision required of the estimates (e.g., the size of differences that need to be detected); the tabulations and analytic results that will inform decision makers and other users; related and previous surveys; steps taken to prevent unnecessary duplication with other sources of information; when and how frequently users need the data; and level of detail needed in tabulations, confidential microdata, and public-use data files.⁴

The federal agency and the survey content developers would need to remain aware of how any “new” items would affect the existing survey. For example, how would the collection of the new information affect the response rate, time to collect the data, comparability to earlier data (trend is important information), and the cost of conducting the survey with the added data items. The balance of this paper addresses the steps involved, issues that need to be considered, and activities that need to be carried out in order to add new questions or modules to an existing survey. The major steps in adding new items include the following:

- ✓ Identification of new items (also called the planning phase),
- ✓ Testing of new items,
- ✓ Questionnaire/module development,
- ✓ Obtaining OMB clearance,
- ✓ Pretesting the survey with the new items, and
- ✓ Evaluating the impact of the new items on the survey.

³ Office of Management and Budget. (2006). *Standards and Guidelines for Statistical Surveys*. Washington, DC: Author. [Retrieved on July 20, 2016 from

https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf]

⁴ Same as above.

Adding New Questions/Modules to the NSDUH

Identification/Selection of New Items

If SAMHSA were to decide to expand behavioral health data collections as part of the NSDUH, new items would need to be added to the survey. Identification of the new items is the first and most critical step in the process of adding new items to an existing survey or data system. The planning and development of new items and measurements would typically include input from many experts and stakeholders in order to ascertain what information is most useful to collect. The planning phase will serve as an opportunity to systematically gather input from individuals, organizations, federal, state and local agencies, and other stakeholders with interest in and knowledge about behavioral health.

A series of interrelated tasks would have to be carried out in order to make a complete assessment of the information to be added to the existing information collection tools and the best way to collect the additional information. Some questions would need to be answered, such as: should all or some portion of the new information be collected annually or periodically; should the new items be collected as separate modules or be merged into existing modules/questionnaires, and so on. These and other questions will need to be answered and decisions made based on research such as a literature search or an environmental scan, discussions, and recommendations from experts in the field and the agency's own experience, data needs, and anticipated uses of the new data on behavioral health issues.

Both formal and informal communications can be helpful in reaching a decision on the questions/items to add. SAMHSA can reach out to a variety of individuals in a range of related disciplines, to organizations with an interest in the topic, and to other government agencies and enlist their help through formal meetings, informal meetings, and other forms of communication to identify the questions to be developed and added.

Agencies have several options to bring interested parties together for the purpose of consensus building. For example, an interagency task force may be organized to develop a recommended list of items. Another option is to convene an expert panel to develop recommendations for the needed items. The panel could consist of leaders in the areas of substance abuse, mental health problems and treatment and other areas of behavioral health. The panel members can assist with the identification of current critical topics and issues, and with

approaches to solving them.⁵ These two options are just a small sample of the many options available to an agency to determine the most pertinent information to collect. The extent of input to develop a list of items needed to provide the desired information depends upon the amount of time and funds available to the agency for planning. After the items are selected and agreed upon, the next step is to develop questions and data collection instruments along with procedures to collect the new items. Existing data collection procedures may or may not need to be revised.

Testing of New Items

Various methods are used to develop and test new items prior to adding them to an existing survey^{6,7}. One agency uses a questionnaire design lab to develop and test questions that are new to a survey. Some agencies use focus groups or usability assessments to test new items, but one of the most effective methods of testing and evaluating new items is through cognitive testing. The primary purpose of cognitive testing is to evaluate how well the questions perform when asked of respondents similar to respondents in the survey population. Cognitive interviews provide a way to find out how well respondents understand the questions, how they arrived at their answers, and how they define the terms.⁸ Questions that are misunderstood or that are difficult to answer can be improved prior to adding them to a questionnaire. Unlike the sample for a pretest, the sample for cognitive testing is usually very small and purposive. Cognitive testing is recognized in the OMB Guidelines as a useful tool, but OMB clearance is not needed for the test if it is conducted with fewer than 10 respondents. Additionally, there are times when an agency will conduct cognitive testing after the pretest if substantive revisions are made to the questionnaires as a result of pretest findings.

⁵ National Center for Science and Engineering Statistics, National Science Foundation. (December, 2008). *NCSES Guidance for a Major Redesign of an Existing Survey, Design of a New Survey, or Improvement of a Continuing Survey*. Retrieved July 19, 2016 from https://www.nsf.gov/statistics/infoqual/ncses_suppguidelines_dec2008.pdf

⁶ Same as above

⁷ U.S. Census Bureau (Reissued July, 2013). *Statistical Quality Standards. Appendix 2. Questionnaire Testing and Evaluation Methods for Censuses and Surveys*. Washington, DC: Author. [Retrieved 7-19-2016 from https://www.census.gov/content/dam/Census/about/about-the-bureau/policies_and_notices/quality/statistical-quality-standards/Quality_Standards.pdf]

⁸ National Center for Science and Engineering Statistics, National Science Foundation. (December, 2008).

Questionnaire Development

Once the wording of the new items is finalized, the next step is typically to integrate them into the form of a hard-copy questionnaire or module, and distribute a draft of the questionnaire to a small group of experts for comments. After the comments are received and the draft questionnaire is revised as necessary, the agency may want to distribute the questionnaire for additional input by a larger group of stakeholders. This larger group of stakeholders can include staff of other federal agencies, universities, private research organizations, professional organizations, and others as determined by the agency. Reviewers may be convened at a meeting to provide additional input about the questionnaire with the new items; or the agency may convene an expert panel (this panel could be the same panel that developed the new items) to review the draft and provide a set of recommendations about the questionnaire. Feedback from the review process will inform the final draft of the questionnaire.

The final draft will be prepared for submission to an Institutional Review Board (IRB) and to OMB for clearance. IRBs must approve all non-exempt research (the NSDUH is non-exempt) before the involvement of human subjects may begin. One of the requirements of studies collecting sensitive information such as information on behavioral health is that the data collectors must obtain informed consent. Because an IRB will need to review the entire study protocol, it is efficient to submit the same materials for IRB review which are prepared for submission to OMB.

The OMB approval process is discussed in detail below. Upon receipt of OMB clearance, the survey will be ready for pretesting. Any revisions recommended by the IRB will be made prior to OMB submission. It is typically cost-effective to maintain the questionnaires in hard copy until they are finalized and ready for the pretest at which time they can be converted to electronic form.

Obtaining OMB Clearance

The addition of new items (or substantive revisions), modules, or questions to NSDUH or any other federal survey, requires clearance from OMB because the pretest sample size would need to be large enough (greater than 9) to make meaningful comparisons to the full survey. The agency is responsible for all activities necessary to obtain OMB approval. This includes preparing an Office of Management and Budget/Paperwork Reduction Act (OMB/PRA)

clearance package, including the preparation of the information collection instruments, consent forms for respondents, correspondence and instructions to support the information collection, interviewer and field staff training materials, and all other documents planned to be used for the data collection. In addition to the data collection protocol, an agency prepares and submits a draft of the *Federal Register* notice, an OMB application form, and a Supporting Statement, Parts A and B. The *Federal Register* notice must be prepared according to the procedures outlined in the most recent “Federal Register Document Drafting Handbook”.

The initial submission of the OMB package is for the *Federal Register* notice which gives the public 60 days to comment on the planned data collection. After the close of the 60-day comment period, the agency will review all submitted comments and questions, develop responses to the comments and questions, make any necessary revisions to all documents and submit a revised clearance package to OMB. The notice will again be published in the *Federal Register* but the public will have 30 days to submit comments and or questions. If there are comments or questions after the 30-day comment period, the agency will prepare and submit responses, make any necessary revisions to the package and resubmit to OMB. This submission is usually followed by OMB’s approval.

Pretesting the Survey with the New Items

For any survey, it has been found useful to test the questionnaire and the data collection procedures on a sample of respondents smaller than the sample for the actual survey. This activity is sometimes referred to as pilot-testing or field testing. In this case, a pretest with the new items added, will allow researchers to gain direct experience under the same data collection conditions as in the full survey (i.e., the same questionnaire, survey design, and survey methods). A pretest is often referred to as a “dress rehearsal”.

OMB’s guidelines require that agencies ensure that all components of a survey function as intended when implemented in the full-scale survey and that measurement error is controlled by conducting a pretest of the survey components or by having successfully fielded the survey components on a previous occasion. In addition, the guidelines stipulate that the agency develop a plan for a pretest as one of the activities in planning to substantially revise an existing data

collection system.⁹ A pretest is designed to measure the effect of the new items on a survey's procedures and resulting estimates. The following elements of pre-testing are essential for ensuring that the new items are sufficiently assessed.

Pretest Study Design

The sampling design for the pretest is typically the same as the sampling design for the national survey at the time of the pretest. Thus, if the pretest were to be conducted based on the present design of the NSDUH, the sample for the pretest would typically be representative of the United States population age 12 or older. It would employ a 50-state design with an independent, multi-stage area probability sample for each of the 50 States and the District of Columbia.

Pretest Population

The respondent universe for the pretest should ideally be the same as for the actual survey. In other words, in the case of the NSDUH, the population would be the civilian, noninstitutionalized population age 12 or older. It would include college students living in dormitories, people living in homeless shelters, and civilians living on military bases. Oversampling of particular demographic or ethnic groups that is done in the main survey should also be considered for the pretest to reflect as accurately as possible the target population, the data collection processes with this population, and the results.¹⁰ The number in the sample should be sufficient in size to provide adequate statistical power for quantitative analyses of the pretest data, with reasonable projections for non-eligibility, non-response, and skip patterns.^{11,12}

Mode of Data Collection

The pretest data collection with the new questions should be in the same mode and format as the actual survey in order to have reliable data for comparison. For example, the NSDUH is a face-to-face household interview survey. The entire NSDUH data collection process is conducted electronically which means the pretest should also be conducted electronically. The interviewer uses a small touch screen tablet to conduct a screening of the

⁹ Office of Management and Budget (2006).

¹⁰ Caspar, R., & Peytcheva, E. (2011). Guidelines for best practice in cross cultural surveys. Chapter X1 Pretesting. Survey Research Center, Institute for Social Research, University of Michigan. Retrieved on July 19, 2016 from <http://ccsg.isr.umich.edu/pdf/FullGuidelines1301.pdf>

¹¹ Same as above.

¹² U.S. Census Bureau (Reissued July, 2013).

eligible household with an adult resident (18 or older) in order to determine who should be selected for the interview. Selected respondents are interviewed using computer assisted interviewing (CAI) on a laptop computer. Portions of the interview are conducted via computer-assisted personal interviewing (CAPI) where the interviewer reads the less sensitive questions and records the answers on the laptop computer. Sensitive questions are completed using audio computer-assisted self-interviewing (ACASI), where the respondents listen to the questions and enter their own responses. With ACASI, only the respondent knows the answer provided to potentially personal and sensitive questions such as illicit drug use, mental health issues, or other sensitive conditions or behaviors. Studies show that maximizing privacy helps to encourage honest, accurate answers and produces high quality data when collecting sensitive information.

In addition to testing the survey, the pretest also provides an opportunity to prepare and test the electronic equipment for proper functioning. As part of this process, the tablets and computers would need to be configured and coded to incorporate the new information associated with the new survey items. (For electronic surveys, data are coded as they are collected.)

Because all data collection in NSDUH is done electronically, when the draft hard copy questionnaire is finalized, the final questionnaire should be given to programmers who will write the program codes so that the new questionnaire can be administered using CAI, CAPI, and ACASI. It is more cost and time effective to transfer the questionnaires into electronic form after they are finalized. The programming for the new version would need to undergo extensive testing to verify that the survey is working as intended. The electronic collection of data requires that the routing and variants in a questionnaire be specific and in full detail before interviewing begins. The proper functioning of this aspect reduces the need for most of the manual and machine editing required after data collection. Upon verification that the programs are working properly, the questionnaires are ready for pretesting.

Quality Control

Quality control is important to all stages of the data collection process. For example, staffing requirements for the pretest field team should be identical to the requirements for the field team for the main study. One activity that helps ensure the quality of the data collected is systematic training and monitoring of the field staff. The training sessions should include practice conducting interviews and role playing. All field staff should undergo the same level of

training for the pretest as for the regular survey. Another quality control mechanism is observation of the interviewers by field supervisors and agency staff during training and data collection. Quality control will also involve updating and revising survey procedures and documentation as follows:

- Documents used for locating and contacting respondents, as used in the main survey, may need to be revised (e.g., lead letters to the respondents).
- Informed consent forms and other study documents provided to respondents may need to be revised.
- Training manuals and other materials for interviewers and other field staff will need to be revised or developed to incorporate the new questions.
- Interviewer debriefing materials will need to be revised.
- Interview observations feedback documents will need to be revised.
- Editing and imputation procedures and related documents will need to be revised or developed.
- Survey receipt and control procedures may need to be updated.
- Equipment, such as tablets and computers, will need to be updated to incorporate the new items; and will need to be tested and debugged.
- Specifications for data files will need to be developed or updated.

Methods for Pretesting

In an earlier section of this paper on testing of new items to be added to the survey, methods such as focus groups and cognitive testing were discussed that are used to assess how well respondents understand the questions, how they arrive at their answers, and how they define the terms. When the final new items are inserted into the questionnaire a pretest of the new questionnaire is carried out in the context of the actual survey design, sample, data collection, and procedures as described above.

Different methods can be used to pretest how well the adapted survey performs.^{13,14} For example, *behavior coding of respondent/interviewer interactions*, which is a systematic coding of interactions from live, taped, or telephone interviews, allows for the identification of problems that arose with the questions or responses, such as respondents asking for clarification. Another

¹³ Caspar, R., & Peytcheva, E. (2011)

¹⁴ U.S. Census Bureau (Reissued July, 2013).

method of pretesting involves *debriefing of respondents or interviewers* after the interviews. Respondent debriefings are typically accomplished using a structured questionnaire to learn to what extent respondents interpreted questions in the same way as intended by the survey developers. The purpose of interviewer debriefings is to examine their perceptions and experience of questionnaire problems. Group discussions or rating forms can also be used.

A powerful pretesting method, referred to as *split panel tests*, utilizes experimental conditions to assess different questionnaire versions or different data collection modes. The U.S. Census Bureau's statistical quality standards¹⁵ point out that split panel tests, though more expensive, may offer an advantage in surveys where comparability of results over time is critical. This is because they may be able to discriminate between survey results stemming from methodological changes, such as adding survey items and changes in data collection, versus actual changes over time in the subject of interest. The US Census Bureau also advises that pretesting using multiple methods is more effective than relying on only one method.¹⁶

Evaluating the Impact of the New Items on the Survey

The primary goal of the pretest is to measure the effects of the new items on the full survey estimates. Upon completion of the pretest, study staff will typically create data files, tabulate the data, conduct necessary analyses, and summarize these findings in one or more reports. To determine the potential impact of the new questions on existing questions, measures and values from the pretest may be compared to those from the full study. If there are detectable differences for any measure, the reason for the differences will need to be evaluated and resolved prior to continuing with the planning for the implementation of the full survey.

Study staff can use a variety of indicators to assess the quality of the pretest data collection as compared to the full study. An examination of the indicators will assist in identifying potential problem areas and inform study staff about the possibility of needed revisions to procedures or data collection instruments. One of these indicators is the response rate. Response rates can be calculated for all levels of responses including screening, eligible participants, and the items in the questionnaires. An adequate number of people participating in the study is important to the reliability of the survey results. Item completion rates are important to assure the quality of the estimates. Values for missing data can also be calculated and

¹⁵ Same as above.

¹⁶ Same as above.

compared to the full study. Inconsistencies in responses to questions can also be evaluated although most of these may be resolved based on checks programmed into the electronic data collection.

In addition to evaluating collected data, other aspects of the pretest can be evaluated and provide useful information. For example, the length of time for data collection in the pretest will be compared to the length of time for data collection prior to the addition of the new items, because a substantial increase in the length of the survey can have an adverse influence on the item response rates. Information provided by the field staff about their observations and experiences can also be useful at this step in reflecting on the analyses of the pretest data.

In many cases, if sufficient research is invested in the planning stages, the pretest demonstrates that the new items perform as anticipated and only minor changes are needed before full implementation. In these cases, the pretest can be considered a “dress rehearsal”.