

2017-2020 NSDUH Methodological Field
Tests, Supporting Statement

Appendix B – 2015 National Survey on Drug
Use and Health: Text To Speech Investigation
and Report

2015 NATIONAL SURVEY ON DRUG USE AND HEALTH

TEXT TO SPEECH INVESTIGATION AND PRETEST REPORT

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland

February 2015

2015 NATIONAL SURVEY ON DRUG USE AND HEALTH: TEXT TO SPEECH INVESTIGATION AND PRETEST REPORT

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.208.006.002

RTI Authors:

Emily Geisen
Patty LeBaron
Marty Meyer
Gil Rodriguez
David Roe
Christina Touarti
Hilary Zelko

RTI Project Director:

David Hunter

SAMHSA Project Officer:

Peter Tice

For questions about this report, please email Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration,
Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

February 2015

Acknowledgments

This report was prepared for the Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International (a registered trademark and a trade name of Research Triangle Institute). Contributors to this report from CBHSQ include Joel Kennet, Grace O'Neill, and Dicy Painter. Contributors to this report at RTI include Debbie Bond, Doug Currivan, Patti Dukes, Tim Flanigan, Valerie Garner, Becky Granger, David Hunter, Georgina McAvinchey, Gretchen McHenry, Allison McKamey, Susan Myers, Rosanna Quiroz, Bonnie Shook-Sa, Margaret Smith, Richard Straw, and Kevin Wang.

Table of Contents

Chapter	Page
1. Introduction	1
2. Literature Review.....	3
3. Investigations of Text to Speech Software	7
3.1 2009 Investigation.....	7
3.2 2013 Investigation.....	14
3.3 Costs and Impact on Work Processes	22
3.4 Overall Conclusions and Recommendations	23
4. 2014 Pretest of Text to Speech in NSDUH	25
4.1 Cognitive Interview Phase	25
4.2 Pilot Test Phase.....	38
4.3 Pretest Conclusions.....	46
5. References.....	47
Appendix	
A ACASI Questions Selected for TTS Prototypes	A-1
B Recruitment Advertisements (English and Spanish)	B-1
C Cognitive Testing Protocol (English and Spanish).....	C-1
D Informed Consent Forms (English and Spanish)	D-1
E Text to Speech Pilot Test Detailed Timing Tables	E-1
F Text to Speech FI Debriefing Moderator's Guide.....	F-1

List of Tables

Table	Page
3.1 Product Scoring for Acapela.....	9
3.2 Product Scoring for NeoSpeech.....	10
3.3 Product Scoring for AT&T Natural Voices.....	10
3.4 Length of Time Required to Initially Customize Audio for a Particular Variable.....	11
3.5 Ratings of Customized Text to Speech Voices.....	13
3.6 Text to Speech Prototypes.....	17
3.7 Product Averages for Evaluated Text to Speech Prototypes.....	18
3.8 Gender of Voice Preference.....	18
3.9 Average Ratings on Quality Dimensions for NeoSpeech English Prototypes.....	18
3.10 Average Ratings on Quality Dimensions for NeoSpeech Spanish Prototypes.....	19
3.11 Average Ratings on Quality Dimensions for Microsoft Speech Platform English Prototypes.....	19
3.12 Average Ratings on Quality Dimensions for Microsoft Speech Platform Spanish Prototypes.....	19
3.13 Average Ratings on Quality Dimensions for TextSpeech Pro English Prototypes.....	20
3.14 Estimated Software Costs of Text to Speech Products.....	22
4.1 English-Speaking Participant Demographics.....	27
4.2 Spanish-Speaking Participant Demographics.....	27
4.3 Comprehension Rating of English Survey Questions, by Version.....	30
4.4 English-Speaking Participant Preference Ratings, by Voice.....	32
4.5 Comprehension Rating of Spanish Survey Questions, by Version.....	33
4.6 Spanish-Speaking Participant Preference Ratings, by Voice.....	35
4.7 Question Length and Percentage Faster, by English Voice.....	36
4.8 Question Length and Percentage Faster, by Spanish Voice.....	37
4.9 Text to Speech Interview Respondents, by Age Group.....	41
4.10 Text to Speech Audit Trail Timing Data: Mean and Median in Minutes, Comparison across Instruments, English-Speaking Respondents, by Age Group.....	42
4.11 Text to Speech Audit Trail Timing Data: Mean and Median in Minutes, Comparison across Instruments, Spanish-Speaking Respondents, by Age Group.....	43

List of Tables (continued)

Table		Page
E.1	Text to Speech Audit Trail Timing Data: Interview Overall, All Respondents	E-2
E.2	Text to Speech Audit Trail Timing Data: ACASI, All Respondents.....	E-3
E.3	Text to Speech Audit Trail Timing Data: ACASI Tutorial, All Respondents	E-4
E.4	Text to Speech Audit Trail Timing Data: ACASI Risk Availability, All Respondents	E-5
E.5	Text to Speech Audit Trail Timing Date: Interview Overall, English-Speaking Respondents	E-6
E.6	Text to Speech Audit Trail Timing Data: ACASI, English-Speaking Respondents	E-7
E.7	Text to Speech Audit Trail Timing Data: ACASI Tutorial, English-Speaking Respondents	E-8
E.8	Text to Speech Audit Trail Timing Data: ACASI Risk Availability, English-Speaking Respondents	E-9
E.9	Text to Speech Audit Trail Timing Date: Interview Overall, Spanish-Speaking Respondents	E-10
E.10	Text to Speech Audit Trail Timing Data: ACASI, Spanish-Speaking Respondents	E-11
E.11	Text to Speech Audit Trail Timing Data: ACASI Tutorial, Spanish-Speaking Respondents	E-12
E.12	Text to Speech Audit Trail Timing Data: ACASI Risk Availability, Spanish-Speaking Respondents	E-13

1. Introduction

Audio computer-assisted self-interviewing (ACASI) is a survey technology in which question text is simultaneously displayed on screen and read aloud to respondents through a headset. It has been adopted widely by survey researchers and has been used in a number of national Federal surveys because of its effectiveness in eliciting more accurate responses to highly sensitive or personal questions (Tourangeau & Smith, 1996) and for enabling participation among respondent populations with low literacy levels or sight impairments (Phillips, Edwards, & Dolbow, 2013). Typically, ACASI uses human voices recorded in the form of WAV or MP3 files that read the survey questions and response options on each screen. This requires a costly and labor-intensive effort both to record and edit high-quality audio and to programmatically integrate audio files with survey software (Phillips et al., 2013). This process can be especially difficult for questionnaires that dynamically generate question text using previous response fills or other changing components, such as dates, which need separate audio files that must be "stitched" together and can cause breaks or pauses in the flow of the audio.

Another challenge for voice-recorded ACASI questionnaires is the reliance on a particular human voice that may change over time or become unavailable, thereby requiring the rerecording of all audio files with a replacement voice in order to ensure consistency within and across interviews. A potential solution to these challenges is the integration of text to speech (TTS) software, which uses a computer-generated voice to read text displayed on screen. This text can be recorded in WAV file format (static) or run dynamically in conjunction with the survey instrument (dynamic), thereby eliminating the use of audio files altogether (Phillips et al., 2013). Compared with recording WAV files with a human voice, either approach has the potential to simplify the production of audio content associated with modifications to questionnaire text or the incorporation of new survey items.

To assess whether the use of TTS would be viable on the National Survey on Drug Use and Health (NSDUH), RTI International and the Substance Abuse and Mental Health Services Administration (SAMHSA) completed a thorough investigation into the features and quality of various TTS software packages. This investigation began in 2009 with a literature review of TTS software used for survey data collection, followed by the evaluation of available TTS software to assess which of the commercially available products were potentially viable for use on the NSDUH. At that time, the TTS products available were not deemed suitable for the NSDUH. However over the next few years, significant advances in TTS technology were made. In 2013, RTI and SAMHSA continued their investigation of TTS software for use on the NSDUH. The 2013 evaluation resulted in the selection one TTS product, which was then tested in 2014, and finally implemented on the 2015 NSDUH.

The results of the literature review and investigations are provided in the remainder of this report. Chapter 2 provides the results of the TTS literature review that was initially conducted in 2009 and updated in 2013. It summarizes the projects and studies that used TTS in data collection and the potential effects it had on data quality. Chapter 3 summarizes the results of the investigations of commercially available TTS products. In 2009, the investigation focused on researching different TTS systems and identifying products suitable for further evaluation. In 2013, given the advances in TTS technology, the investigation re-assessed the different TTS systems and conducted an evaluation of selected products.

After the 2013 investigation, a recommendation regarding which of the evaluated products was most promising for the NSDUH's ACASI modules was provided and a product was selected by SAMHSA for the 2015 NSDUH. However, the comprehensibility of TTS in the NSDUH had not been evaluated by the individuals who may rely more heavily on the audio component when completing the ACASI portion of the NSDUH interview. These individuals are likely to be the youngest and oldest respondents (i.e., those aged 12-17 and 65 or older), respondents with low levels of literacy, and non-native English speakers. Therefore, a pretest was conducted in 2014 to evaluate whether the use of TTS had any effect on comprehension of specific survey items, to determine the best TTS presentation speed for administering the survey questions, and to identify any major issues in administration time or unanticipated issues with the use of TTS. The TTS pretest also included two phases: a cognitive interview phase and a pilot test phase. The findings from the 2014 TTS pretest are reported in Chapter 4. Based on the findings in the cognitive interview phase, the decision was made to implement TTS for the 2015 NSDUH.

2. Literature Review

For years, the survey research industry has studied and documented the benefits of audio computer-assisted self-interviewing (ACASI) when compared with other data collection modes, from paper and pencil to computer-assisted personal interviewing (CAPI) (Tourangeau & Smith, 1996, Turner et al. 1998, Dykema, Basson, & Schaeffer, 2007). As time went on, ACASI became even more popular, with Couper, Tourangeau, & Marvin (2009) furthering the conversation by noting that modest gains found through the use of ACASI called for continued research and experimentation.

As ACASI grew, the methods of audio delivery to respondents also began to evolve. Researchers began experimenting and examining the use of text to speech (TTS) technology, which converts entered or keyed (programmed) text into speech (Kraft & Taylor, 2006; Couper, Kirgis, Buageila, & Berglund, 2012; Phillips et al., 2013). In survey data collection, TTS is most commonly used in interactive voice response (IVR) or telephone audio computer-assisted self-interviewing (T-ACASI) modes. Surveys deployed in these modes are conducted by telephone, and the survey instructions and questions are delivered via prerecorded scripts (Couper, Singer, & Tourangeau, 2004). Using TTS systems to provide audio for ACASI in-person interviews is not yet as common as in telephone data collection modes. As Couper (2005, p. 488) noted, continued improvement to the quality of TTS technology "... opens the way to increased use of TTS systems for replacing the recording of interviewers [i.e., human voices]," including ACASI. Kraft and Taylor (2006) provided an early example of investigating TTS for use with an ACASI protocol, in which they suggested that the use of TTS provided greater flexibility in making edits and changes and reduced development time and costs while providing respondents with an impersonal solicitor for sensitive questions. In recent years, advances in TTS synthesis technology have been enabling the generation of more realistic, accurate, and human-sounding "voices" in multiple languages that may be customized with respect to pacing, volume, pitch, and pronunciation. Concurrent with these advances, the use of computer-generated voices has gone from a few, highly technical applications to the point at which many people encounter synthesized speech regularly in their daily lives and are likely becoming more accustomed to it.

In response to this improved technology, survey researchers have begun experimenting with implementing TTS. In September 2011, the developers of the National Survey of Family Growth (NSFG) replaced human voice recordings with static TTS recordings for the ACASI portion of the Cycle 8 interview, and they compared Cycle 8 data with Cycle 7 data that used human voice recordings (Couper et al., 2012). In addition to supporting past literature on the overall cost and time efficiencies provided by TTS in preparation for data collection, the NSFG researchers found that more respondents used ACASI but took less time (in terms of overall interview time) to do so, suggesting that smaller audio files, less downtime or "space" between files and fills, and the pace of TTS led to better overall interview efficiency (and, one could assume, reduced burden) while in the field. In 2012, the Population Assessment of Tobacco and Health (PATH) Study also turned to TTS for a field test. Using a dynamic implementation mode, in which the TTS engine reads question text in real time without prerecorded audio WAV files, Phillips and colleagues (2013) supported the aforementioned research on the efficiency of using TTS, especially when it comes to eliminating the sheer size and volume of files that must be "stitched" together to provide smooth audio delivery.

They noted that in this mode, TTS consumes far fewer resources (server space, storage, etc.) than live voice recordings.¹

Because many TTS products offer both female and male voices, a review of the survey research methods literature examining how the gender of an ACASI voice might affect survey response was also conducted. The literature remains inconclusive regarding the impact of the gender of the voice on survey response. Outside of the survey research industry, scholars and technology experts have suggested that computer voices are mostly female because of biology (people finding female voices more pleasing than male voices) or history (e.g., telephone operators) (Griggs, 2011). That said, gender experts have suggested that gender stereotypes, if present in an individual, can in fact extend to machines, suggesting that voice selection (in terms of gender) could be highly consequential (Nass, Moon, & Greene, 1997).

In the survey research literature, information is somewhat scarce, although Dykema, Diloreto, Price, White, and Schaeffer (2012) presented an excellent review of the literature at the beginning of their investigation into ACASI gender-of-voice effects on the reporting of sensitive behaviors among young adults. They explained that, depending on a question's topic, respondents may refer to gender-based stereotypes, conversational norms, or identities when responding (Tannen, 1996; see also Schaeffer, 2000). According to self-disclosure theory, individuals are expected to be more honest and disclose more to someone they trust and with whom they feel comfortable (Jourard, 1971). Insofar as respondents hold stereotypes that women are more sympathetic (Pollner, 1998) or nonjudgmental (Nass, Robles, Heenan, Bienstock, & Treinen, 2003), respondents may disclose or report higher levels of sensitive behaviors to female interviewers (Dindia & Allen, 1992). In contrast to self-disclosure theory, other researchers have offered "explanations of exaggeration," which hold that higher levels of reporting may be less valid. For example, the "macho hypothesis" of Catania and colleagues (1996, p. 371) explains the higher levels of some sexual behaviors that males report to male interviewers as an effort to seem more virile and manly. In a related vein but predicting a different outcome, Weisel (2002, p. 102) argued in her study of contemporary gangs that "a female interviewer may [have] inadvertently encourage[d] male interviewees to put on a macho bravado and exaggerate some points." However, Dykema and colleagues (2012) found higher levels of engagement in sensitive behaviors and more consistent reporting among males when responding to a female voice. They did not find any evidence that female respondents were influenced by the voice's gender.

Couper and colleagues (2004) also found no significant difference between male and female voices on reporting sensitive behaviors or gender attitudes in a telephone survey. They compared question administration by live telephone interviewers, an IVR system that used human voice recordings, and an IVR system that used synthesized speech generated from TTS software. Equal numbers of male and female respondents were enrolled in the study. Although the authors expected to find gender-of-voice effects for gender attitudes and some sensitive questions (i.e., sexual behavior, weight) in both the live interview and IVR conditions, surprisingly, they found no significant differences due to the gender of the voices used. However, they did find effects for differences in responses to the live interviewers as compared with the automated IVR system, which consistently revealed greater disclosure of sensitive behaviors in the IVR conditions. This finding is consistent with the research methods literature that suggests that ACASI facilitates more accurate and candid reporting of sensitive behaviors, although it does not indicate that the gender of the voice used in

¹ The PATH Study did not report any results associated with the impact on survey response because a full implementation of TTS had not been completed at that time.

ACASI always has a significant impact on survey responses. One other study, conducted exclusively with male respondents, directly tested the effect of the ACASI voice's gender on survey reports; Fahrney, Uhrig, and Kuo (2010) explored the impact of a male versus a female voice on reports of sexual activity among men who have sex with men. Their results were consistent with more accurate reporting among the males who heard questions read by a female voice. Although Dykema and colleagues (2012) concluded that a female voice should be the convention for ACASI studies, they also noted that more experimentation and investigation are required to ensure the elimination of underreporting across specialized topics and populations.

Based on the review of the TTS literature, RTI and the Substance Abuse and Mental Health Services Administration proceeded with an investigation of TTS software products as detailed in Chapter 3. Given the inconclusive literature on voice gender, both male and female voices were examined in the TTS investigation for preference.

3. Investigations of Text to Speech Software

3.1 2009 Investigation

In July 2009, survey research staff at RTI completed an investigation into the features and quality of various text to speech (TTS) software packages to determine which, if any, software is best suited for use with the National Survey on Drug Use and Health (NSDUH). The objective was to determine whether the NSDUH could transition from using a human voice for the English and Spanish audio computer-assisted self-interviewing (ACASI) portions of the survey to using an automated voice created through TTS software. RTI reviewed relevant literature and presentations evaluating TTS systems for surveys or similar purposes, as described in Chapter 2. This research and review of articles, such as Couper (2005) and Couper et al. (2004), provided guidance in the decision to further investigate and test a carefully chosen subset of currently available packages.

3.1.1 Static versus Dynamic Text to Speech Implementation

The use of TTS software in the NSDUH interview could take one of two forms. In the dynamic alternative, each field laptop would be equipped with TTS software. The computer-assisted interviewing (CAI) instrument would use a Blaise alien router to call the TTS run-time software passing the question text as input. The TTS software would then generate the audio in real time during the survey. The TTS software would theoretically have the pronunciations for hard-to-pronounce words defined ahead of time, and identical versions of the TTS software, along with the pronunciation definitions, would be loaded on all laptops.

The static alternative would largely mirror the approach that is currently used for creating audio files and loading them on the field laptops. A centralized TTS application would be used to create audio files for the ACASI section of the interview. This central bank of audio files would then be loaded onto the field laptops for use during the interview.

In discussing the merits of these two alternatives, a decision was reached to proceed with the latter. The Substance Abuse and Mental Health Services Administration (SAMHSA) expressed concerns about standardization across field laptops with respect to pronunciations of the same words from question to question if the TTS software were to be installed in each field laptop individually. The level of effort for creating the customized pronunciations across the two alternatives would be quite similar. RTI noted that the dynamic alternative would allow the programming staff to reduce hours needed to maintain and store the WAV files needed for the interview from year to year. However, it was decided that the risk of violating standardization would outweigh any additional time required to maintain the bank of audio files. Therefore, this investigation identified packages that would be suitable for creating customized audio files for storage and dissemination to the field.

3.1.2 Methods

RTI initially evaluated six TTS software packages in order to identify those with the capability to provide English and Spanish audio for NSDUH instrumentation. The packages

evaluated were Acapela, AT&T Natural Voices, Cepstral, Loquendo, NeoSpeech, and Real Speech. This evaluation yielded three TTS products that RTI recommended for further testing.

Creating a standardized audio environment across field interviewer (FI) laptops was a critical requirement of the TTS package. Neither Acapela nor NeoSpeech offered a male Spanish TTS voice. Therefore, RTI recommended focusing investigations on female voices offered by the companies.

RTI provided SAMHSA with the sample audio files that were created as part of the initial investigation, using freely available demo versions of the six TTS software packages. The audio files were created using the text from the NSDUH's lifetime alcohol use question (AL01). The wording of this question was simple, and RTI expected the TTS audio of this question to be high quality. However, that was not the reality.

RTI survey research staff's initial reactions to the quality of the "off-the-shelf" products included observations that none of the packages provided voices that were free from problems. Many of the voices possessed robotic qualities. Others stuttered when pronouncing certain words (e.g., "alcohol"). Rankings of the Spanish versions of these TTS products were considerably lower than their English counterparts. Evaluations of the Spanish voices elicited comments regarding the presence of mispronunciations, heavy Spanish accents, an inappropriate tone that may sound rude to respondents, and trembling voices.

The first round of review indicated that real difficulties might be found when the products were subject to further testing. First, RTI staff suspected that it could be entirely possible that the speech quality offered by even the best and most expensive TTS software would not be good enough for NSDUH interviewing. Second, staff believed that the manual "tweaking" of text, which would be required to produce understandable pronunciations of drug names and other nonstandard words, might require extensive effort. This could perhaps result in expending several times as much effort as would be required to simply re-record all of the WAV files using a human voice. RTI recommended testing this customization process with the goal of estimating the amount of effort that it would entail.

3.1.3 Results of the 2009 Investigation

3.1.2.1 Off-the-Shelf Testing

The next steps in the investigation included focusing on the three packages: Acapela, NeoSpeech, and AT&T Natural Voices. Tables 3.1–3.3 show average subjective rankings of each product on a number of dimensions. These evaluations ranked the off-the-shelf audio produced by the individual products. These files were not customized. The English voices were ranked by six RTI evaluators: Chuchun Chien, Doug Currivan, Patty LeBaron, Hyunjoo Park, Gil Rodriguez, and Mai Wickelgren. Each evaluator was asked to rank the TTS voice on each dimension using a scale from 1 to 5. The points on the scale were labeled as follows:

Numeric Value	Label
1	Poor
2	Fair
3	Good
4	Very Good
5	Excellent

Spanish rankings used the same scale and were scored by three RTI evaluators: Georgina McAvinchey, Rosanna Quiroz, and Gil Rodriguez.

Initial investigations of each of the three products listed as follows indicated that they were compatible with existing systems, were priced competitively, and had potential to be customized to correctly pronounce and emphasize words contained within the NSDUH interview. Further testing of these products can provide more comprehensive insight into the process of customization and deployment.

3.1.2.2 Product 1: Acapela

As shown in Table 3.1, the product average for Acapela's English female voice tied with NeoSpeech's for the highest ranking. Acapela's Spanish female voice also was ranked the highest among Spanish female voices. However, this ranking was almost a full point lower than comparable English rankings. Acapela's Spanish female voice was described as "trembling." Acapela software possessed the capability to edit pronunciations for abbreviations and exceptions. It was batch processing-supported and had an adjustable speaking rate and an adjustable voice tone. Acapela included speech enhancements and control tags.

Table 3.1 Product Scoring for Acapela

English Female	Average Score	Spanish Female	Average Score
Clarity	2.83	Clarity	2.00
Inflection	2.67	Inflection	2.00
Tone	3.00	Tone	2.00
Humanness	2.83	Humanness	2.00
Pace	3.83	Pace	2.33
<i>Product Average</i>	<i>3.03</i>	<i>Product Average</i>	<i>2.07</i>

<http://www.acapela-group.com/text-to-speech-interactive-demo.html>

In 2009, Acapela's product was priced according to the number of hours of audio that were produced. A 5-hour pack cost \$2,250, a 10-hour pack cost \$3,750, and a 20-hour pack cost \$5,250. If one bought a 5-hour pack, for instance, up to 5 hours of back-to-back sound files could be created. Sound files were license free and could be duplicated and distributed without having to pay an additional license fee. For NSDUH's purposes, a 20-hour pack would have been needed to record the entire ACASI portion of the English and Spanish interviews, along with all fills, and would therefore have cost \$5,250.

3.1.2.3 Product 2: NeoSpeech

As shown in Table 3.2, NeoSpeech's female Spanish voice was ranked relatively low, with a score of 1.4. This was the third highest ranking of a female Spanish voice, with Acapela and AT&T Natural Voices receiving higher scores. One evaluator commented that the voice sounded "too loud or too exclamative" and "may sound rude to respondents." A comment about NeoSpeech's English female voice described it as "assertive."

Table 3.2 Product Scoring for NeoSpeech

English Female	Average Score	Spanish Female	Average Score
Clarity	2.67	Clarity	1.67
Inflection	3.33	Inflection	1.33
Tone	3.17	Tone	1.33
Humanness	3.00	Humanness	1.33
Pace	3.00	Pace	1.33
<i>Product Average</i>	<i>3.03</i>	<i>Product Average</i>	<i>1.40</i>

<http://www.neospeech.com/>

NeoSpeech software supported customization of a dictionary so that developers could adjust pronunciations of symbols, abbreviations, and new terms. In 2009, this product cost \$3,996 *per year* and included two developer licenses and an additional Spanish voice. There was no additional cost for audio file distribution.

3.1.2.4 Product 3: AT&T Natural Voices

As shown in Table 3.3., the clarity score for the AT&T Natural Voices English female voice was below the average of all of the product's clarity scores (1.83 vs. 2.28). AT&T's Spanish voice was described as having a heavy Spanish accent and included mispronunciations.

Table 3.3 Product Scoring for AT&T Natural Voices

English Female	Average Score	Spanish Female	Average Score
Clarity	1.83	Clarity	1.67
Inflection	2.33	Inflection	1.67
Tone	2.17	Tone	1.67
Humanness	2.50	Humanness	1.33
Pace	3.83	Pace	1.33
<i>Product Average</i>	<i>2.53</i>	<i>Product Average</i>	<i>1.53</i>

<http://www.naturalvoices.att.com/>

AT&T Natural Voices included custom dictionaries that could be used to define pronunciations phonetically. It could run the application to read text from a file and output a WAV file. Control tags could be added to mark up text to affect pronunciation, such as controlling the emphasis on a word.

In 2009, a single AT&T software development kit cost \$295, and an extra Spanish voice font cost \$50. This purchase provided a perpetual license with no need to renew. Deployment licenses for the software cost \$5 each. The overall cost included deployment licenses for 800 laptops (\$4,000). Fewer licenses would have been needed for testing the product, resulting in significant cost savings. The estimated cost for NSDUH's purposes was \$4,345.

AT&T also offered a prescription drug module containing English pronunciation definitions of certain drug names to RTI for \$1,000. Of the 99 drugs used in the four psychotherapeutic modules and in the special drugs modules of the 2009 NSDUH, 49 were an exact match. Many were partial matches. For example, AT&T's Wizzard Module contained

Darvocet, Tylenol, and codeine, whereas the NSDUH used Darvocet-N and Tylenol with codeine.

Each software Web site listed earlier provided the opportunity to listen to the TTS voices speak customized phrases, often in a variety of voices. After providing audio files to SAMHSA for further evaluation, RTI conducted further testing of available software packages listed earlier to gain more insight into their ability to provide quality, standardized audio for the NSDUH interview.

3.1.2.5 Customized Testing

To assess the quality of customized audio and the level of effort involved in creating such customizations, RTI requested evaluation versions of the three software packages and tested them off the shelf. Acapela and NeoSpeech provided RTI with trial evaluation versions of their software at no charge. These software packages allowed for the customization of pace, pronunciation, and tone for English voices only. Representatives from both companies indicated that there were no differences in capabilities between the trial version and the full version of the software. A trial evaluation version of the AT&T Natural Voices product was not available.

As a next step, RTI customized the tone of the audio files, the pronunciations of the words contained within a select number of variables, and the pace of the voice. A number of questions within the NSDUH's tranquilizer module were selected for this customization. These questions contained both standard text and more complicated drug names. Throughout the customization process, programming staff tracked the level of effort that was required to achieve the highest possible quality in the end product.

To customize the audio files, pronunciations of the words were defined either phonetically or using pronunciation symbols within the lexicon/pronunciation editor of the particular application. RTI found that the same phonetic pronunciation definition of a word did not always yield the same pronunciation in the two software packages that were tested. Pronunciations of drug names were particularly laborious to customize, as the default pronunciation of these words differed dramatically from the correct pronunciation. Because some questions contained more than one drug name, the time required to customize each question varied. It is estimated that each drug name required between 10 and 20 minutes to define.

Additional time was required to customize other words and pronunciations in the question. Table 3.4 illustrates the length of time it took to *initially* program each WAV file. It was estimated that after the iterative rounds of review and revision needed to achieve an acceptable level of quality, these lengths of time could increase threefold.

Table 3.4 Length of Time Required to Initially Customize Audio for a Particular Variable

Variable	Time
INTROTR2	15 minutes
TR01	45 minutes
TR02	1 hour, 15 minutes
TR03	45 minutes
TR04a question	3 hours, 30 minutes
TR04a responses	15 minutes

After defining pronunciations and customizing the results, RTI reviewed and rated the resulting Acapela and NeoSpeech audio files through a process similar to the off-shelf evaluation process (see Table 3.5 for the results). The review and rating process elicited some negative comments about the quality of the audio files. The pronunciation of most drug names by both products presented a problem. Comments about the customized Acapela voice noted difficulty with pronouncing most prescription drug names, problems with inflection at the end of a question, and a strange accent when pronouncing the word "caused." Other comments noted that the voice seemed "laborious" while speaking. Comments about the NeoSpeech product were more favorable. However, the pronunciations of the drug names were not accurate, and in one particular WAV file, the voice sounded as though she talked with a lisp (TR04a question).

3.1.4 Conclusions from the 2009 Investigation

In response to a lack of acceptable quality and to the extensive level of effort that would be needed to create the WAV files for the English and Spanish NSDUH interviews, RTI compiled observations about the packages. After performing a fairly thorough initial investigation using free demo versions of two of the popular representative commercial TTS packages, RTI observed the following:

- It was fairly easy to quickly produce audio files from question text. However, these files sounded robotic, and pronunciations of many words were clumsy or incorrect. Customization of pronunciation was possible, so RTI investigated that in some detail. Some observations on the customized products were noted as follows.
- Experimentation with some simple examples of questions from the NSDUH CAI instrument clearly showed that it might be difficult, if not impossible, to produce high-quality and natural-sounding audio files using any of the TTS packages RTI examined. This was true even after fairly extensive tweaking and customization of pronunciations using phonetic spellings and pronunciation enhancements.
- Customization of simple question text proved quite time consuming. On the basis of what was seen, RTI estimated that it could take *between 1 and 2 hours per question* to optimize the pronunciations and cadence of the questions. This was an optimistic estimate because it neglected to factor in the effect of "fills," in which the question audio was dynamically produced based on the answers to prior questions. It also did not factor in much time for the iterative review process, in which RTI sent audio to SAMHSA to review and comment on, RTI modified the audio, and so on. This would undoubtedly be a time-consuming process.
- If RTI estimated that it would take 1.5 hours to optimize each audio file, for approximately 3,500 audio files in the questionnaire, then twice that to account for Spanish as well, RTI concluded that it could take 10,500 person hours (5 people working full time for more than 1 year) to finish the work. Even if a 1.5-hour per question estimate was a significant overestimate, the amount of effort involved still could be prohibitive and would undoubtedly be many times the effort involved in completely re-recording all the audio files from scratch using a human voice.

Table 3.5 Ratings of Customized Text to Speech Voices

Voice	Acapela	NeoSpeech
Intro TR02		
Clarity	1.8	3.8
Inflection	2.0	3.6
Tone	3.0	4.0
Humanness	1.8	3.6
Pace	2.6	3.2
<i>Product Average</i>	2.2	3.6
TR01		
Clarity	1.4	2.8
Inflection	2.0	3.0
Tone	2.8	3.4
Humanness	1.8	3.2
Pace	2.8	3.0
<i>Product Average</i>	2.2	3.1
TR02		
Clarity	1.8	3.0
Inflection	2.0	2.8
Tone	3.0	3.4
Humanness	1.6	3.2
Pace	3.2	3.0
<i>Product Average</i>	2.3	3.1
TR03		
Clarity	2.2	3.0
Inflection	2.4	3.0
Tone	3.2	3.6
Humanness	1.8	3.2
Pace	3.0	3.4
<i>Product Average</i>	2.5	3.2
TR04a Question		
Clarity	3.0	3.8
Inflection	2.4	3.4
Tone	3.2	3.8
Humanness	2.2	3.4
Pace	2.8	3.4
<i>Product Average</i>	2.7	3.56
TR04a Responses		
Clarity	1.8	2.4
Inflection	2.2	3.4
Tone	2.6	3.2
Humanness	1.8	3.4
Pace	3.0	3.8
<i>Product Average</i>	2.3	3.2

NOTE: Scale: 5=Excellent, 4=Very good, 3=Good, 2=Fair, 1=Poor. A trial evaluation version of the AT&T Natural Voices product was not available.

- The whole issue of an audio Spanish version seemed extremely challenging. The effort involved, the ability of the TTS software packages to produce accurate Spanish pronunciations, and the limited availability of Spanish speakers who could tweak those have caused RTI to further question the viability of the proposed approach.
- The option to purchase AT&T Wizzard's prescription drug module might be a time saver. However, this module would help with the pronunciation of drug names in English only. In cases where the Spanish pronunciation differed from the English, RTI would have been charged with the task of creating custom pronunciations for each drug name in Spanish.

3.1.5 Recommendations from the 2009 Investigation

At the conclusion of the 2009 investigation, it was determined that TTS technology was not appropriate for the NSDUH. At that time, very high levels of effort would have been needed to customize audio files so that they would be of a quality appropriate for use with the NSDUH. As a result of the level of effort that would be involved in this process, RTI did not recommend the use of NSDUH resources to continue further down this path at that time. However, RTI and SAMHSA understood the importance of reevaluating TTS options as technology improved with time. A reevaluation was conducted in 2013 as part of the TTS investigation and is described in the following section.

3.2 2013 Investigation

In 2013, SAMHSA asked RTI to assess changes in the TTS software landscape and reexamine the possibility of implementing TTS on the NSDUH. RTI conducted a broad review of the TTS landscape at the time to identify software products that could be considered for further hands-on assessment. A wide variety of TTS software products was available in 2013, ranging from small Web plug-ins or mobile applications, to stand-alone desktop software, to server-based solutions for streaming speech output. The sound and quality of the speech output depended largely on the speech synthesizer software that generated the "voice" and on the methods that were available to vary the pacing, volume, emphasis, or pronunciation of terms. Products varied with respect to the synthesizer technology used and the functionality provided for customization. For example, some products provided a "studio" editing interface that allowed for easier manual editing and manipulation of the pronunciation, volume, and pacing of the speech output, whereas others required customization using programming code, such as Speech Synthesis Markup Language (SSML). In 2013, Microsoft, Apple, and Android all had a TTS synthesizer built into their operating systems. In addition, Microsoft's Speech Platform Software Development Kit (SDK) allowed developers to build applications that interacted with the synthesizer programmatically. Other commercial providers, such as AT&T Natural Voices, Acapela, and NeoSpeech, had developed their own TTS synthesizers that could run on the Windows platform and be integrated with a variety of applications. Some commercial providers also included editing or "studio" software that could be used for customizing the speech output. Other products typically provided a user interface for inputting text or selecting documents/web pages to be read and used either the operating system's built-in synthesizer or one of the commercial third-party synthesizers, such as AT&T or NeoSpeech.

To identify feasible options for the evaluation, RTI staff conducted online research and contacted a variety of vendors to learn about different TTS products. The most important factors

in selecting products were the sound and quality of the voice, the ability to customize speech output (pronunciation/pacing), and the ability to run on a laptop in an offline environment. Also, the search was limited to vendors that had a track record in the industry and were likely to remain key players with well-supported products. As requested by SAMHSA, the research began with the three products evaluated as part of the 2009 TTS investigation (i.e., Acapela, AT&T Natural Voices, and NeoSpeech) to determine whether they remained feasible options for the 2013 evaluations. These products evolved since 2009 and were considered among the top commercially available TTS products in 2013. RTI staff examined software features, listened to online demos of the voices, and spoke with company representatives to discuss the software's capability to provide both a dynamic speech solution and a static, prerecorded audio file-based solution. The conclusion was that all three products should remain viable options for further investigation. Next, RTI staff extended the search to include other products that other survey groups had used or had come on the market or gained prominence since 2009. TextSpeech Pro emerged as an additional product for consideration because it was used on both the National Survey of Family Growth (NSFG) and Population Assessment of Tobacco and Health (PATH) Study (see the discussion in Chapter 2).

In addition to the four products mentioned earlier, RTI staff researched a wide variety of new TTS products. Several turned out to be less suitable with limited capabilities and were marketed mainly for personal use, so the team focused on seven mainstream and commercially available TTS packages, each of which supported integration with other applications, as follows:

- TextSpeech Pro
- AT&T Natural Voices
- NeoSpeech
- Microsoft Speech Platform
- Acapela
- IVONA
- Loquendo

The 2013 evaluation plan required an assessment of the two core modes for implementing TTS software in the NSDUH ACASI module: static and dynamic implementations. As mentioned earlier, "static" implementation used TTS software to record all of the audio files needed by the ACASI instrument. These recordings were saved as WAV files, stored on the laptop, and played by the ACASI program as the respondent moved through the interview. This was similar to the NSDUH's existing ACASI approach, with the primary difference being that the TTS "voice" was recorded rather than a human voice. "Dynamic" implementation installed the TTS engine on each laptop, and the ACASI program interfaced with the TTS engine to construct the audio in real time as each question appeared on screen.

The features of the TTS packages were examined with respect to both implementation modes. The dynamic implementation required a Windows-based desktop SDK or an application programming interface (API) to interact with the NSDUH's Blaise survey software. The static audio file-based approach required that TTS speech be output as WAV audio files. In summary, the TTS software features required to support the NSDUH's needs were as follows:

- For dynamic implementation, Windows-compatible SDK or API to allow integration with the NSDUH interview software and the ability to run offline
- For static implementation, the ability to output speech to WAV file format
- Customizable pronunciation and extensible dictionaries to accommodate specialized terminology
- A graphical user interface (GUI) that enabled customization of volume, pausing, and speaking rate
- Availability of male and female voices
- Availability of English and Spanish voices

Three of the seven products listed earlier satisfied the requirements: TextSpeech Pro, NeoSpeech, and AT&T Natural Voices. The Microsoft Speech Platform met most of the requirements but did not include any male voices. The requirements excluded three of the seven TTS products listed earlier: Acapela, IVONA, and Loquendo.

Sections 3.2.1 and 3.2.2 describe the 2013 evaluation methods and results. Additionally, RTI's technical team evaluated the advantages and disadvantages of the static and dynamic implementation modes, which are described in Section 3.2.3.

3.2.1 Methods

After discussions with the NSDUH team at SAMHSA, three products were selected for purchase: TextSpeech Pro, NeoSpeech, and AT&T Natural Voices.² Shortly before the purchase was made, AT&T unexpectedly pulled its Natural Voices product off the market without providing any information on possible future availability. As a result, SAMHSA and RTI staff decided to eliminate the AT&T product from further consideration, and the Microsoft Speech Platform (which met most of the requirements to support the NSDUH) was used to replace the AT&T product in the 2013 evaluation.

An initial set of prototypes using a subset of 12 ACASI questions was developed using the static and dynamic modes, English and Spanish languages, and male and female voices. The CAI questions were selected to include a mix of specialized terms, such as the names of alcoholic beverages or prescription drug names, long sections of text, and question fills. These questions are listed in Appendix A. Because many of the default pronunciations, particularly for specialized terms and prescription drug names, were incorrect, the technical team worked to fine-tune and enhance these pronunciations and customize the pausing, speed, and pitch of voices to maximize the probability that respondents could understand the voice. The team spent approximately 10 minutes per screen to enhance the quality, which provided a realistic test of the quality that could be achieved in a full-scale implementation.

A total of 12 prototypes—8 in English and 4 in Spanish—were developed and are listed in Table 3.6. These included male, female, and Spanish versions that were available from each product. Although both female and male voices were originally required for testing, no male

² The Microsoft Speech Platform product was not included initially because it did not need to be purchased, and an initial review indicated that it did not meet all of the requirements. Before RTI conducted the evaluation, a newer and higher quality voice option offered by Microsoft was identified and included in the 2013 evaluation.

Spanish voices were available for any of the products, so only female Spanish voices were tested. Also, no male English voice was available for the Microsoft Speech Platform.

Table 3.6 Text to Speech Prototypes

English Prototypes	Spanish Prototypes
1. NeoSpeech male static	1. NeoSpeech female static
2. NeoSpeech female static	2. Microsoft Speech Platform female static
3. Microsoft Speech Platform female static	3. NeoSpeech female dynamic
4. TextSpeechPro female static	4. Microsoft Speech Platform female dynamic
5. TextSpeechPro male static	
6. Microsoft Speech Platform female dynamic	
7. NeoSpeech female dynamic	
8. NeoSpeech male dynamic	

A team of seven evaluators (five English speakers and two Spanish speakers), including one male and six females who had experience in survey research were recruited to assess the prototypes. Both Spanish-speaking evaluators had experience in designing Spanish-language surveys.

Each evaluator was assigned to assess four prototypes in a specific order. Assignments were made so that evaluators were presented with a mix of static, dynamic, male, and female prototypes in different orders to minimize any potential bias due to the order in which evaluators heard the prototypes. Each prototype was assessed by at least two evaluators. The evaluators were instructed to navigate through the CAI prototype and rate each screen along the following dimensions using a rating scale from 1 to 5: clarity, inflection, tone, humanness, and pace. The points on the scale were labeled as follows:

Numeric Value	Label
1	Poor
2	Fair
3	Good
4	Very Good
5	Excellent

Ratings were then averaged across all screens, and the evaluators generated an average product score for each prototype. After completing the ratings, evaluators were asked to report a gender preference and which, if any, of the prototypes they preferred.

3.2.2 Results of the 2013 Investigation

This section presents the results of the evaluation. Ratings on all screens for each dimension were averaged across all evaluators to provide an overall product average for each prototype (see Table 3.7). Table 3.8 presents the gender of voice preferences. The following sections include the detailed results for each product.

Table 3.7 Product Averages for Evaluated Text to Speech Prototypes

Product	English Static Female	English Dynamic Female	English Static Male	English Dynamic Male	Spanish Static Female	Spanish Dynamic Female
NeoSpeech	2.6	2.9	2.8	3.0	2.1	2.5
Microsoft Speech Platform	2.6	2.9	--	--	2.1	2.8
TextSpeech Pro	2.1	--	2.0	--	--	--

-- Not available.

When asked if they had a preference for a male or female voice, most evaluators were indifferent or preferred female, as shown in Table 3.8.

Table 3.8 Gender of Voice Preference

Gender of Voice Preference	Male	Female	Indifferent
Number of Evaluators	1	2	2

3.2.2.1 Product 1: NeoSpeech

Table 3.9 shows the audio quality ratings for the NeoSpeech English prototypes, and Table 3.10 shows the ratings for the Spanish prototypes. The product average scores for NeoSpeech's female English voice tied with Microsoft Speech Platform's female English voice in both the static and dynamic versions (see Section 3.2.2.2 for the Microsoft product results). NeoSpeech's static prototypes were ranked higher than both TextSpeech Pro's female and male prototypes (see Section 3.2.2.3 for the TextSpeech Pro product results). The dynamic implementations were higher than the static versions and were the highest of all female prototypes. NeoSpeech's male English voice was ranked slightly higher than the female voice. Despite these higher rankings, only one English evaluator preferred a male voice over a female one.

The product average for NeoSpeech's female Spanish voice was higher for the dynamic version than the static version, but it was lower than the product average for Microsoft Speech Platform's female Spanish voice. No male Spanish voices were available from NeoSpeech or Microsoft Speech Platform.

Table 3.9 Average Ratings on Quality Dimensions for NeoSpeech English Prototypes

Dimension	Static Female	Dynamic Female	Static Male	Dynamic Male
Clarity	2.8	3.3	2.9	3.6
Inflection	2.5	2.7	2.6	2.8
Tone	2.9	2.9	3.4	2.9
Humanness	2.0	2.5	2.2	2.6
Pace	2.8	3.2	3.0	3.1
<i>Product Average</i>	2.6	2.9	2.8	3.0

Table 3.10 Average Ratings on Quality Dimensions for NeoSpeech Spanish Prototypes

Dimension	Static Female	Dynamic Female
Clarity	2.4	2.5
Inflection	1.9	2.3
Tone	1.9	2.4
Humanness	2.1	2.3
Pace	2.3	3.1
<i>Product Average</i>	<i>2.1</i>	<i>2.5</i>

3.2.2.2 Product 2: Microsoft Speech Platform

Table 3.11 shows the audio quality ratings for the Microsoft Speech Platform English prototypes, and Table 3.12 shows the ratings for the Spanish prototypes. Microsoft Speech Platform's female English voice tied with NeoSpeech's female English voice for the highest ranking of the female voices and outranked the TextSpeech Pro female prototype by one-half rating point. As with NeoSpeech, the product average for the dynamic version was higher than that of the static version. No male English voices were available with Microsoft Speech Platform at the time of evaluation.

Table 3.11 Average Ratings on Quality Dimensions for Microsoft Speech Platform English Prototypes

Dimension	Static Female	Dynamic Female
Clarity	3.5	3.9
Inflection	2.2	2.8
Tone	2.7	2.9
Humanness	1.7	2.5
Pace	2.8	2.7
<i>Product Average</i>	<i>2.6</i>	<i>2.9</i>

Table 3.12 Average Ratings on Quality Dimensions for Microsoft Speech Platform Spanish Prototypes

Dimension	Static Female	Dynamic Female
Clarity	2.1	2.8
Inflection	2.0	2.7
Tone	2.0	2.8
Humanness	2.4	2.8
Pace	2.1	3.0
<i>Product Average</i>	<i>2.1</i>	<i>2.8</i>

Microsoft Speech Platform's dynamic female Spanish voice was ranked the highest of all Spanish voices. The static version had a product average equivalent to NeoSpeech's female Spanish static version.

3.2.2.3 Product 3: TextSpeech Pro

The product averages for the TextSpeech Pro female and male English prototypes were the lowest of all English prototypes (see Table 3.13). On almost every dimension, they scored

lower than both NeoSpeech's female and male static prototypes as well as Microsoft Speech Platform's female static prototype. A dynamic version of TextSpeech Pro was unavailable, and no Spanish voices were available with the TextSpeech Pro software.

Table 3.13 Average Ratings on Quality Dimensions for TextSpeech Pro English Prototypes

Dimension	Static Female	Static Male
Clarity	2.7	2.4
Inflection	1.8	1.8
Tone	1.9	2.0
Humanness	1.5	1.4
Pace	2.8	2.5
<i>Product Average</i>	<i>2.1</i>	<i>2.0</i>

3.2.3 Static versus Dynamic Implementation

As part of the activity associated with developing the prototypes, the NSDUH's technical team at RTI also assessed requirements and tradeoffs associated with the static versus dynamic TTS implementations. This section summarizes the key advantages and disadvantages of each approach. Overall, the team concluded that the advantages of the dynamic approach, which allowed greater flexibility, efficiency, and audio quality, outweighed those of a static approach. In addition, the dynamic prototypes were consistently rated higher than the static prototypes.

Using the static TTS approach to prerecord audio files would have many advantages associated with eliminating the process for human voice recording and minimizing the impact on the development process:

- Audio files could be generated in a batch from text descriptions of the questions. This would require significantly less time than human voice recording and would not require the use of a recording studio.
- Also, in a static TTS implementation, the Blaise program on the laptop would remain largely the same as the current version, with only minimal changes required.
- Installing TTS software on the laptops would be unnecessary, and the configuration of the NSDUH's field laptops would remain largely the same, with audio files installed locally on each laptop.
- Testing would be similar to the current approach and somewhat easier than a dynamic implementation because no TTS software would be required to listen to audio files, which could be played on any computer or even a mobile device.

One of the main drawbacks of the static approach was that audio files still had to be "stitched" together, which would not eliminate some of the breaks or pauses that could occur when multiple files must be played for a single question. Also, the static implementation would require the production of approximately 10,000 audio files (5,000 for ACASI in English and 5,000 for ACASI in Spanish) to replace the human voice recordings currently used on the NSDUH. This work would require purchasing or developing software to generate and edit the audio files on RTI development computers, and it would also require developing a database for storing and maintaining all of the audio files and text scripts for each audio segment.

A dynamic implementation would offer many advantages largely associated with eliminating the use of audio files:

- Audio files would not be prerecorded, thereby eliminating the audio file production process and the effort associated with editing, prerecording, storing, and maintaining a database of audio files.
- The dynamic approach also offered a smoother audio delivery on questions with fills because breaks or pauses that could occur when audio files have to be "stitched" together would no longer be present.
- Eliminating audio files would simplify the process of distributing patches to in-field laptops because there would be no audio files to distribute as part of a patch. This would make patches smaller and easier to deploy, providing increased flexibility to make needed changes to the questionnaire during ongoing data collection.
- There would be no need to maintain audio files and scripts used for recording.

Among the disadvantages of a dynamic implementation was that more sophisticated programming would be necessary, requiring familiarity with specific tools and features of the TTS product and the Microsoft Speech API. Although many changes would be required in the Blaise programming code to implement a dynamic approach, the experience acquired during the development of the prototypes allowed the team to establish methods that could be implemented in the entire set of ACASI modules fairly easily. Additionally, the dynamic approach would tend to make the interview software testing process somewhat more cumbersome because each testing laptop would require installation and configuration of TTS software, whereas in a static implementation, a finite set of audio files could be easily shared and played on any laptop. Also of interest is that the dynamic implementation required that TTS software be installed and configured on every field laptop, which could potentially introduce complications associated with software licensing agreements. In particular, the NeoSpeech TTS product required each laptop to be configured with a unique license activation key, permanently tied to the media access control (MAC) address (the hardware ID number) of that laptop. As such, a dynamic implementation based on NeoSpeech would severely complicate the duplication and maintenance of NSDUH field laptops and would require that a NeoSpeech license be purchased for every laptop. It is certainly worth noting here that the Microsoft TTS product, which was free, did not impose any such licensing requirements or restrictions. Therefore, a dynamic implementation based on the Microsoft TTS product would not have the same drawbacks associated with a dynamic implementation based on NeoSpeech.

Finally, it should be noted that regardless of mode, TTS implementation in the first year would require a significant effort with respect to customizing pronunciations and testing the audio components. Many of the default pronunciations for specialized terms and drug names would need to be corrected using alternate, phonetically based spellings. Furthermore, even with customization, the pronunciations for some terms or phrases would likely be suboptimal, which was true for both the static and dynamic implementations. The technical team estimated that the level of effort to customize and test the audio components would be similar in either the static or dynamic approaches.

3.3 Costs and Impact on Work Processes

Cost is a key factor in adopting any new technology. The implementation of TTS software offered the potential for work process efficiencies and long-term cost savings associated with developing and updating the NSDUH ACASI program. Startup and ongoing costs were associated with purchasing and using the TTS software and the level of effort required to implement the new technology. This section discusses both the software costs and the impact on the level of effort in key work processes involved in TTS implementation.

Each product evaluated in 2013 is summarized in Table 3.14. Estimated costs associated with the purchase and ongoing use of the product are provided in both static and dynamic implementations based on vendors' estimates in 2013. Licensing requirements were different for each product. NeoSpeech required an annual license agreement with the purchase of their products, whereas Microsoft did not. TextSpeech Pro licenses were bundled in the purchase prices of the software on a one-time, per-laptop basis, and no renewal was required.

Table 3.14 Estimated Software Costs of Text to Speech Products

Text to Speech Product	License Required	Cost: Static Implementation¹	Cost: Dynamic Implementation¹
NeoSpeech	Yes, annual renewal	\$5,300/year (700 laptops)	\$9,800/year (700 laptops)
Microsoft Speech Platform	No	\$0	\$0
TextSpeech Pro	Yes, one time	\$1,500 (10 development computers)	N/A ²

N/A = Not applicable.

¹ Costs presented in the table are based on estimates gathered from vendors in 2013.

² Although RTI staff were initially told that a software development kit (SDK) was available for implementing a dynamic version, only a demo version could be obtained, which appeared to be last updated in 2009. RTI staff were later told that the voice provider (i.e., NeoSpeech, AT&T) would need to be contacted to obtain the SDK needed to implement a dynamic version.

RTI staff also reviewed key work processes involved in the NSDUH's current ACASI development effort that would be affected by TTS implementation to assess whether an increase, decrease, or similar level of effort would be required. These work processes were in four areas:

1. Production of audio files for ACASI
2. Customization/editing and testing of audio components
3. ACASI programming
4. Field laptop configuration and mass duplication

The production of audio files in the current approach involves activities associated with human voice recording using a recording studio and sound engineer and maintaining a database for storing and maintaining audio files. A reduced level of effort and greater efficiency over time would be expected in a static implementation because audio files could be generated from text scripts in batch using purchased or developed software, thereby eliminating costs associated with human voice recording and rerecording. Clearly, in the first year of implementation, a learning curve would be associated with using the new software, implementing new processes, and

generating scripts for approximately 10,000 new audio files. In a dynamic implementation, this process would be eliminated altogether.

The second area (i.e., customization/editing and testing of audio components) would require the greatest increase in effort, regardless of implementation mode, especially in the first year of implementation. The RTI staff's experience in developing the TTS prototypes revealed that default pronunciations for specialized terms (e.g., alcohol types, prescription drugs, health conditions) had to be customized using alternate phonetic spellings. The speaking rate, pitch, and volume also were modified to improve the quality. As a result, it was expected that a greater number of audio components would require editing and quality control checks than in the current approach, where pronunciations, pacing, and pitch are more easily adjusted by the human voice in the studio at the time of initial recording. Therefore, TTS implementation was expected to require an increased level of effort associated with editing pronunciations compared with the current approach. However, the same level of effort for customization/editing and testing was expected, regardless of whether a static or dynamic implementation is adopted.

In the third area, the ACASI programming effort, a static implementation would require essentially the same effort as in the current approach, and minimal changes would be required in the Blaise code to accommodate the TTS audio files. A dynamic implementation would require more complex programming to integrate the TTS engine with the Blaise software. However, because RTI staff developed the programming approach and an initial set of code to achieve this while building the prototypes, a small increase was anticipated in programming effort to implement this for the full set of ACASI modules.

Similarly, the process for field laptop configuration and mass duplication (the fourth area) would be similar to the current effort in a static implementation and somewhat more complicated for a dynamic implementation, which would require TTS software to be installed and configured on each field laptop. As noted earlier, this process would be significantly more complex if the NeoSpeech product were deployed because of its hardware-based annual licensing structure, which would complicate not only the initial deployment, but also annual updates and ongoing replacements of field laptops. Because Microsoft Speech Platform did not require a license, the configuration process would be significantly simplified.

3.4 Overall Conclusions and Recommendations

Implementing TTS technology in the ACASI module of the NSDUH interview offered an opportunity for work process efficiencies and cost savings in the NSDUH's ACASI software development. Although TTS software did not match the audio quality of human voice recording, the evaluation presented in this report, along with the experience of both the NSFG and the PATH Study, indicate that it was of sufficient quality to replace human voice recordings in ACASI. The goal of implementation was to create audio recordings that both English- and Spanish-speaking respondents of all ages could easily understand. None of the evaluators had difficulty understanding any words or phrases produced by the TTS voices. A significant advantage of TTS, regardless of implementation mode, was the elimination of the effort and costs associated with the use of a recording studio and human voices for generating ACASI audio. TTS also eliminated relying on a particular human voice, which could change or become unavailable over time.

The evaluation results indicated that a dynamic implementation offered a higher quality audio experience than the static implementation largely because of the elimination of audio files. The dynamic approach also eliminated processes for producing, maintaining, and storing audio files and would simplify the process for updating or modifying the NSDUH's ACASI modules. The dynamic approach outperformed the static approach in the evaluation, and all of the evaluators indicated that they preferred the dynamic prototypes. Therefore, a dynamic implementation mode was recommended over a static mode. With respect to the TTS products evaluated in both English and Spanish, the evaluation teams ranked Microsoft Speech Platform the highest and therefore recommended use of this product. Also, the Microsoft product offered a significant advantage over NeoSpeech because it was freely available and required no licensing agreement or user fees.

As a result of the 2013 evaluation, SAMHSA asked RTI to develop two versions of the 2015 NSDUH CAI. One version used the traditional human voice recordings, and the other version included TTS using the Microsoft Speech Platform female voice. Developing the two instruments simultaneously allowed RTI to customize and test the TTS pronunciations in the NSDUH but provided an alternative in case the TTS version was deemed unacceptable. The TTS instrument was then evaluated in a pretest, which was used by SAMHSA in its final decision to use TTS in the 2015 NSDUH.

As mentioned above, after the 2013 evaluation, SAMHSA asked RTI to develop a full version of the 2015 NSDUH CAI instrument using the Microsoft Speech Platform's female English and Spanish voices for further testing and potential full-scale implementation in 2015. A team of survey methodologists was recruited to review and edit the TTS pronunciations for all 36 ACASI modules in both the English and Spanish versions of the CAI instrument. The NSDUH programming team developed a TTS database application that would allow the editors to listen to and adjust the pronunciations of any problematic terms or phrases. All editors participated in a 1-day training to learn how to use the application. Edited terms or phrases were saved in the database along with the module and screen names. The database was then used by the programming team to generate pronunciation files that were incorporated into the survey instrument to ensure that the TTS audio was accurate. Once the customized pronunciations were incorporated into the CAI, the instrument underwent additional reviews by the editing team and the NSDUH survey methodologists. This process was conducted over 4 months and involved multiple iterations of review and customization. All modules were reviewed on a flow basis by SAMHSA staff. The TTS version of the instrument was then configured for use in the pretest described in Chapter 4.

4. 2014 Pretest of Text to Speech in NSDUH

The text to speech (TTS) investigation determined that the National Survey on Drug Use and Health (NSDUH) could transition from using a human voice for the English and Spanish audio computer-assisted self-interviewing (ACASI) portions of the survey to using an automated voice created through TTS software. This investigation concluded that advances in TTS technology have enabled realistic, accurate, and clear-sounding voices with a reduced level of effort to program and customize. However, the comprehensibility of TTS in the NSDUH had not been evaluated by the individuals who may rely more heavily on the audio component when completing the ACASI portion of the NSDUH interview. These individuals are likely to be the youngest and oldest respondents (i.e., those aged 12-17 and 65 or older), respondents with low levels of literacy, and non-native English speakers.

The purpose of the TTS pretest was to explore the use of TTS on the NSDUH with these populations to (1) evaluate whether the use of TTS had any effect on comprehension of specific survey items or administration of those items compared with the human voice, (2) determine the best TTS presentation speed for administering the survey questions, (3) evaluate participants' preferences and opinions regarding the voices (human and computerized) used for ACASI, and (4) identify any major issues in administration time or any unanticipated issues with the use of TTS.

The TTS pretest included two phases: a cognitive interview phase and a pilot test phase. The findings from the cognitive interviews assisted the Substance and Mental Health Services Administration (SAMHSA) in determining that TTS should be used on the 2015 NSDUH and in determining the preferred speaking rate. The cognitive interview findings were also used to refine the TTS audio to improve comprehensibility (e.g., altering the speed or modifying pronunciations). These refinements were then tested in the pilot test phase to identify any major issues with administration time or any unanticipated issues with the use of TTS.

4.1 Cognitive Interview Phase

4.1.1 Methods

RTI conducted 36 cognitive interviews in August 2014 to evaluate the use of TTS software. Details on this process are outlined in the following sections.

4.1.1.1 Participant Selection and Recruitment

RTI conducted cognitive testing with the type of participants who are believed to be the most likely to rely on the audio component of the ACASI portion of the interview. RTI recruited a convenience sample of 24 English-speaking participants with at least 6 interviews in each of the following groups:

- Group 1. Participants aged 12-17
- Group 2. Participants aged 65 or older
- Group 3. Participants with low levels of literacy

- Group 4. Non-native English speakers who would complete the interview in English

Across all groups, RTI tried to recruit a diverse population with respect to gender, ethnicity, and race. Participants were recruited from the Research Triangle Park, North Carolina, area; the Washington, DC, metro area; and Chicago, Illinois.

To recruit English-speaking participants with low education or low literacy levels, RTI worked with the Literacy Volunteers and Advocates organization in Washington, DC. RTI recruited the remaining English-speaking participants using online advertisements posted on Craigslist (www.craigslist.org) and word of mouth. Online advertisements included information about the nature of the study, the incentive amount, and next steps if the individual was eligible for the study. All individuals were screened for eligibility over the telephone using a phone number provided on the flyer. Advertisements are provided in Appendix B.

RTI was contacted by 62 individuals, completed 45 screenings, and conducted 24 interviews. Screenings were conducted with all individuals who could be contacted to ensure that there was a sufficient pool of eligible participants in each recruitment group. Demographic characteristics for the English-speaking participants are provided in Table 4.1.

RTI also recruited 12 Spanish speakers with limited or no English-speaking ability to complete the cognitive interview in Spanish. The 12 Spanish-speaking participants were recruited by word of mouth and administered a screening. All participants were eligible and completed the interview, so no additional individuals were screened. Five of the participants were considered low education (i.e., they did not complete high school). Three of these participants had very low literacy, had not attended high school, and were not familiar with computers. All participants were screened for basic demographic information. Table 4.2 provides the demographics of the Spanish-speaking participants.

Table 4.1 English-Speaking Participant Demographics

Participant ID	Age	Sex	Low Education	Race/Ethnicity	Non-Native Speaker
1-006-100	14	M	No	Black	No
1-009-100	65	F	No	White	No
1-010-100	25	F	No	Asian	Yes
1-015-100	13	F	No	Black	No
1-017-100	22	M	No	Asian	Yes
1-019-100	42	F	Yes	Black	No
1-025-100	34	F	No	Hispanic/Am. Indian	Yes
1-027-100	53	M	Yes	Black	No
2-001-100	15	M	No	White	No
2-005-100	81	F	No	White	No
2-011-100	67	M	No	White	No
2-013-100	16	F	No	Black	No
2-016-100	65	M	No	White	No
2-020-100	63	F	Yes	Black	No
2-023-100	77	F	Yes	Black	No
2-026-100	32	F	No	Hispanic	Yes
3-002-100	68	F	No	White	No
3-004-100	13	M	No	Black	No
3-008-100	23	F	No	Asian	Yes
3-014-100	16	F	No	Hispanic	No
3-018-100	56	F	Yes	Black	No
3-021-100	81	F	Yes	Black	No
3-024-100	13	M	No	White	No
3-028-100	46	F	No	White	Yes

Table 4.2 Spanish-Speaking Participant Demographics

Participant ID	Age	Sex	Low Education	Country of Origin
1-101-500	50	M	Yes	Honduras
1-102-500	51	F	Yes	Mexico
1-106-500	30	F	Yes	Mexico
1-109-500	26	M	No	Dominican Republic
2-201-500	35	F	Yes	Honduras
2-202-500	26	F	No	Mexico
2-107-500	49	M	Yes	Mexico
2-110-500	46	F	No	Colombia
3-301-500	54	M	No	Mexico
3-302-500	50+	M	No	Peru
3-108-500	33	F	No	Mexico
3-111-500	36	M	No	Mexico

4.1.1.2 Description of Procedures

RTI staff conducted cognitive interviews with English-speaking participants in private locations in RTI office buildings. RTI staff conducted cognitive interviews with Spanish-

speaking participants in private locations in RTI office buildings or in participants' homes if they were unable or unwilling to travel to RTI offices.³ All participants were played audio recordings of survey questions using three voices: human voice, TTS moderate speed, and TTS slow speed. Participants only heard the questions and were not able to see the questions.⁴ Although this differs from the main study protocol, the purpose was to test the clarity of the audio without helping the participants by showing them the question text. Participants were not asked to answer the recorded survey questions; rather, they were asked follow-up probes about the questions to assess how well they understood the words or concepts presented in the questions. For example, participants were asked to repeat some of the survey questions or explain what the questions meant in their own words. As needed, spontaneous probes were asked to further evaluate participants' understanding of the questions.

After each set of audio recordings, interviewers asked participants a set of debriefing questions to assess participants' thoughts on speed, inflection or cadence, pronunciation, comprehensibility, sound quality, and overall affective response (how well participants liked the voice). A copy of the protocol is available in Appendix C.

All participants received the same questions and the same probes in the same order, but the order of the voices presented varied. Participants from each of the recruitment categories were randomly assigned to one of three versions for order of the voice tested:

- Version 1. Human voice, TTS slow, TTS moderate
- Version 2. TTS moderate, human voice, TTS slow
- Version 3. TTS slow, TTS moderate, human voice

For example, a participant assigned to Version 1 heard the first third of the interview with a human voice, then answered the debriefing questions. Next, the participant listened to the middle third of the interview using the TTS slow voice, followed by a second debriefing. Finally, the participant heard the final third of the interview in the TTS moderate voice and answered the last set of questions and debriefing. Because there were 24 English-speaking participants and 3 voices, each question was tested in each voice with 8 participants. Each Spanish question was tested in each voice with four participants.

Before starting the interviews, the cognitive interviewers administered the informed consent form (Appendix D). Participants were compensated \$40 for completing the interview, which lasted approximately 45-60 minutes.

4.1.1.3 Analysis and Reporting

RTI survey methodologists analyzed the results of the pretest by comparing how well participants understood the questions read using the TTS voice compared with the questions read

³ A total of seven interviews were conducted in participants' homes. Interviews were conducted in participants' homes when participants did not have reliable transportation and were unable to travel to RTI offices easily or when participants did not feel comfortable coming to RTI facilities, which required going through security and presenting a form of ID.

⁴ For six of the Spanish-speaking participants (1-106-500, 1-109-500, 2-107-500, 2-110-500, 3-108-500, 3-111-500), the cognitive interviewer placed the screen so participants could also read the questions if they wanted to.

using the human voice. In addition, RTI analyzed whether there were differences in comprehension for TTS slow compared with TTS moderate.

For each survey question, the cognitive interviewer reviewed the participants' responses to the probes and assigned a code based on how well the participant appeared to understand the question as intended. The code assigned was verified by another methodologist during analysis. For any codes on which the methodologists disagreed, they met to discuss the codes and came to a consensus. Codes were assigned as follows:

1. **Poor understanding.** Participant did not understand the question or instructions as intended.
2. **Moderate understanding.** Participant mostly understood the question or instructions as intended.
3. **Good understanding.** Participants clearly understood the question or instructions as intended.

In the sections 4.1.2 and 4.1.3, RTI provides a general assessment of whether differences in comprehension appeared to be due to the voice or other factors (question wording, cognitive ability, etc.).

In addition, RTI documented how many times participants asked for a question to be repeated in each voice. It may be that participants need to listen to one of the voices several times before they fully understand it. This has implications for general comprehension and for the timing of administration.

RTI also analyzed participant preference based on the responses to the debriefing questions. These results were used to assess participants' overall thoughts and preferences regarding TTS and the human voice.

It is important to note that because of the limited number of participants and qualitative nature of the study, the results of this study are not generalizable. The intent of this study is only to provide SAMHSA with a general assessment of the comprehensibility of the TTS for subpopulations that may rely most on the audio component of ACASI. SAMHSA can use these findings in evaluating whether to use TTS in the 2015 NSDUH, but the findings cannot be used to determine the potential impact of TTS on survey estimates.

4.1.2 English-Speaking Participant Results

4.1.2.1 Participant Comprehension

4.1.2.1.1 Overall Findings

As discussed in Section 4.1, participants' comprehension of the survey questions was assigned a value of 1, 2, or 3, based on whether their understanding was poor, moderate, or good, respectively. The average score across all questions and all 24 participants was 2.5, indicating that participants understood most of the questions well. As shown in Table 4.3, this score varied somewhat by question set. All participants received the same three sets of questions in the same order, but the voice played for each set of questions differed depending on which version of the

pretest the participant was assigned. The second set of questions appeared to be the hardest for participants to understand. However, the comprehension rating did not vary by voice.

Table 4.3 Comprehension Rating of English Survey Questions, by Version

Voice	Comprehension (Higher=Better)			
	Set 1	Set 2	Set 3	Overall
Human	2.3	2.5	2.7	2.5
TTS moderate	2.7	2.5	2.4	2.5
TTS slow	2.6	2.3	2.6	2.5
All voices	2.5	2.4	2.6	2.5

TTS = Text to Speech.

NOTE: This study consisted of a convenience sample, and results are not generalizable.

Examination of the qualitative comments from participants confirmed that participant understanding did not appear to differ by voice. Participants did not tend to repeat the questions word for word but tended to understand the general meaning or the intent of the question. When participants misunderstood the question, this appeared to be consistent across voices. For example, three participants (one from each group) thought that the very first question they listened to mentioned marijuana. However, it mentioned only tobacco products: "These questions are about your use of tobacco products. This includes cigarettes, chewing tobacco, snuff, cigars, and pipe tobacco. The first questions are about cigarettes only."

Although there did not appear to be much difference in comprehension overall among the voices, a few survey questions appeared more problematic in one voice compared with the other voices. These are described as follows:

- **Human.** The human voice performed the worst in the two questions about alcohol (ALCINTR2 and AL01). Understanding of these questions was low across all voices—most likely because they were the first long, detailed introductions played for participants. When summarizing the introductions, participants tended to miss the part about not including a sip or two or thought it was asking only about a sip or two. However, participants hearing the human voice tended to misunderstand the question more severely. For example, one participant said that AL01 was asking, "If I had a sip of alcohol, did it get me drunk?" Another participant said, "She's asking: when you have a glass of wine if you drink the whole glass or just have a sip of it."
- **TTS slow.** Question CG06 asks participants when they last smoked "part or all of a cigarette." Two participants who heard the question in TTS slow had a hard time understanding the word "part." They initially thought it said "heart" or "hard."
- **TTS moderate.** SP03a asks participants if they were arrested for "motor vehicle theft" in the past 12 months. Three participants who heard the question in TTS moderate had difficulty. One participant understood the phrase but said it sounded like the words were running together. Another participant thought it asked about "identity theft" instead of "motor vehicle theft." A non-native English speaker could not understand the word "booked" and thought it was asking if he voted. However, one participant who heard the human voice also had some difficulty with the question. He thought it asked about "motorcycle theft" instead of "motor vehicle theft." Across all voices, at least one participant misunderstood the specific nature of

the crime and thought it was asking about breaking into a motor vehicle or some sort of driving offense.

It also appears that some words are hard to hear in all voices. For example, METHINTRO provides an introduction for methamphetamine. In all three voices, at least one participant misheard "crank" for "crack."

4.1.2.1.2 Repeating of Bolded Words

One of the main differences between the human voice and the TTS voices is that the TTS voices cannot be used to emphasize certain words, such as words that are bolded on the computer screen. To evaluate whether this was problematic for comprehension, RTI examined how often participants used the bolded words from the survey question when repeating or summarizing the question in their own words.

Despite the fact that the bolded words and phrases were emphasized when read in the human voice, RTI found no difference in the number of bolded words repeated by participants between questions read with the human voice and TTS slow. In questions read with the TTS moderate voice, participants did repeat slightly fewer of the bolded words. However, it did not appear that repeating the bolded words was necessarily associated with a better understanding of the survey question. Participants could have repeated the bolded phrase and still misinterpreted the question. For example, RK01a asks, "How much do people risk harming themselves physically and in other ways when they smoke **one or more packs of cigarettes per day**?" One participant said this question was asking, "How do people try to make themselves feel better if they smoke more than one pack of cigarettes per day?" Conversely, most other participants did not mention smoking one or more packs of cigarettes but correctly understood that the question was asking about the risks associated with smoking.

4.1.2.1.3 Number of Times the Question Was Repeated

Another metric that RTI examined was the number of times participants asked for a question to be repeated with each voice. On average, participants asked for a question to be repeated about 12 percent of the time. This varied slightly by voice. Participants asked for questions to be repeated about 10 percent of the time with the human voice, 11 percent of the time with TTS moderate, and 14 percent of the time with TTS slow.

4.1.2.1.4 Differences by Recruitment Characteristics

RTI also examined differences across recruitment characteristics and found that participants with low education or low literacy had the most difficulty understanding the survey questions, and non-native speakers had the least difficulty (average comprehension score of 2.2 compared with 2.7, respectively). However, the non-native speakers in our sample were highly educated. For each recruitment group, comprehension did not appear to differ based on the voice heard.

4.1.2.2 Participant Preference

After questions were played for each voice, participants were asked a series of debriefing questions (see cognitive interview protocol in Appendix C) rating the voice on a number of

characteristics. The specific ratings and the number of participants who preferred each voice are shown in Table 4.4. It should be noted that although participants were asked about which voice they preferred, the ultimate goal of this exercise was to determine whether any characteristics of the TTS voices would render them unusable in the survey.

Table 4.4 English-Speaking Participant Preference Ratings, by Voice

Voice Characteristic	Preference Rating (Lower=Better)		
	Human	TTS Slow	TTS Moderate
Speed (1=Much too slow, 5=Much too fast)	2.9	3.1	3.5
Cadence (1=Excellent, 5=Very poor)	2.0	2.4	2.5
Pronunciation (1=Excellent, 5=Very poor)	1.7	2.0	2.0
Comprehension (1=Not at all difficult, 5=Extremely difficult)	1.3	1.7	1.8
Pleasantness ¹ (1=Extremely pleasant, 5=Not at all pleasant)	2.3	3.4	2.9
Overall quality (1=Excellent, 2=Good, 3=Fair, 4=Poor, 5=Very poor)	1.5	2.2	2.2
Comfort ¹ (1=Extremely comfortable, 5=Not at all comfortable)	2.1	3.0	2.7
Average (1=Best, 5=Worst)	1.8	2.4	2.4
Number of participants preferring voice ²	12	7	4

TTS = Text to Speech.

¹Reverse coded so that lower scores are better.

²One participant selected "no preference."

NOTE: It is important to note that this study consisted of a convenience sample, and results are not generalizable.

Participants rated the human voice slightly better than the two TTS voices, which were rated the same. However, participants rated all voices as more positive than negative. The cognitive interviews asked participants to elaborate on what they liked or disliked about each voice.

Participants tended to like the human voice because it was natural sounding, easy to understand, and had a good pace. Eight participants described it as nice, pleasant, or soothing. One non-native English speaker commented that the human voice sounded "authoritative" and that it was easier to give a real answer to the human voice compared with the TTS voices. However, some participants commented that the human voice was monotonous, sounded sad or depressed, and was a little too slow. Only one participant was particularly negative about the human voice. He said, "The voice is distracting. It sounds like she's going to cough or is on the edge of being sick."

Participants' opinions varied on which of the TTS voices sounded better. One participant would say that the TTS slow voice was less robotic than the TTS moderate voice, and another participant would say the opposite. In fact, although many participants commented that the TTS moderate voice was faster than the TTS slow voice, three participants thought that the TTS moderate voice was slower than the TTS slow voice. Interestingly, all three of these participants were in Group 1, which means that TTS slow was played before TTS moderate. It may be that they became more accustomed to the TTS voice by the third set of questions, and it was easier for them to understand.

When commenting on the comprehension of the TTS voices, participants tended to be less effusive in their praise of the human voice compared with the TTS voices. For example, in reference to one or both of the TTS voices, eight participants commented, "I could understand it." In contrast, no participants made that comment about the human voice. Instead, seven participants noted that the human voice was "easy" or "clear" to understand.

When remarking on which of the two TTS voices was easier to understand, participants were divided. Some participants thought the TTS slow voice was harder to understand because it was choppy and the cadence was not as good. One participant commented that the TTS slow voice was difficult to understand because it was hard to tell which words "fell together" when being listed. He provided the example of drugs that included the word generic after them. He said that it sounded like "generic" was the name of a drug and not being used to modify another drug. For example, it sounded like "zolpidem, generic, extended-release zolpidem, generic" instead of "zolpidem generic, extended-release zolpidem generic." He said that this was less problematic in the TTS moderate voice. This appears to be related to the fact that "generic" is always listed in parentheses, which causes TTS to add a slight pause. This slight pause is less noticeable in TTS moderate.

Three participants noted that the faster voice was harder for them to understand, and they would have to concentrate more or listen to the question multiple times in order to understand.

4.1.3 Spanish-Speaking Participant Results

4.1.3.1 Participant Comprehension

Of the 12 Spanish-speaking participants, half read the question text and listened to the audio. The other half of the participants listened to the audio only.

As discussed in Section 4.1, participants' comprehension of the survey questions was assigned a value of 1, 2, or 3, based on whether their understanding was poor, moderate, or good, respectively. The average score across all questions and all participants was 2.7, indicating that overall, participants understood the questions quite well. As shown in Table 4.5, this varied little by voice.

Table 4.5 Comprehension Rating of Spanish Survey Questions, by Version

Voice	Comprehension (Higher=Better)			
	Set 1	Set 2	Set 3	Overall
Human	2.7	2.4	2.9	2.7
TTS moderate	2.7	2.6	2.9	2.7
TTS slow	2.6	2.5	2.9	2.7
All voices	2.7	2.5	2.9	2.7

TTS = Text to Speech.

NOTE: It is important to note that this study consisted of a convenience sample, and results are not generalizable.

For Spanish-speaking participants, most differences in comprehension appeared to be based on the question rather than on the voice. For example, on the introduction question ALCINTR2, many participants did not repeat or recall that "a sip or two" was not considered a

drink. Following is a summary of survey questions in which comprehension may have differed by voice:

- Cognitive interviewers asked participants to repeat the following question as best as they could: "How often do you get a real kick out of doing things that are a little dangerous?" Two TTS slow participants could not repeat or recall this question correctly. One person interpreted it vaguely as, "What things give you pleasure when you do them?" Another said, "How satisfied do you feel when doing dangerous things?" The other participants appeared to understand that the question was asking about the frequency of doing dangerous things for the pleasure of doing so.
- Participants received the introduction question INHINTRO. This is a rather long introduction listing a variety of inhalants. Interviewers asked participants if the introduction was asking about times when substances were inhaled accidentally, on purpose, or both. Only two participants (one TTS moderate and one TTS slow) correctly answered that it was asking about using them on purpose.
- Several questions included the instructions to press 95 if the respondent had not used any of the drugs listed. In two separate instances on two different questions, participants misheard or misunderstood these instructions. In both cases, the participants heard the question in the human voice. In one case, the participant thought the question was asking if he had used a drug 95 percent of the time. The other participant thought it was asking if he had used the drugs in 1995. However, neither participant had ever used computers before, which was likely responsible for their confusion.

4.1.3.2 Participant Preference

After questions were played for each voice, participants were asked a series of debriefing questions rating the voice on a number of characteristics. The specific ratings and the number of participants who preferred each voice are shown in Table 4.6.

Although the human voice was rated slightly better than the two TTS voices, the ratings were fairly similar across all categories. Slightly more than half of the participants indicated that they preferred the human voice, and the remaining participants said they preferred the TTS slow voice. None of the participants preferred the TTS moderate voice.

When participants commented on the human voice, they tended to say that it was pleasant and easy to understand, but a little too slow. One participant commented, "It was perfect, very clear," whereas another participant commented, "She sounds like she is about to go to sleep; it makes me sleepy."

Three participants commented that the TTS voice has a slight Spanish accent (i.e., from Spain), which they did not care much for. The accent for the human voice is a neutral Latin American accent. However, only one participant said it was harder to understand the TTS voice and that he had to pay more attention. Although most participants thought the TTS voices were okay, one participant particularly did not like them. With respect to the slow voice, he said, "The voice is shrilling, piercing, uncomfortable, not understood." For the moderate voice, he said, "You can't understand it at all. It's extremely fast; it doesn't get the message across. It talks like a machine without conveying the message. It doesn't care." Despite these comments, this

participant had the same comprehension score for TTS moderate and the human voice. The comprehension score was slightly lower for the TTS slow voice, but that was also the voice that was heard first.

Table 4.6 Spanish-Speaking Participant Preference Ratings, by Voice

Voice Characteristic	Preference Rating (Lower=Better)		
	Human	TTS Slow	TTS Moderate
Speed (1=Much too slow, 5=Much too fast)	2.3	3.3	3.6
Cadence (1=Excellent, 5=Very poor)	2.2	2.3	2.3
Pronunciation (1=Excellent, 5=Very poor)	1.9	2.3	2.1
Comprehension (1=Not at all difficult, 5=Extremely difficult)	1.6	1.8	1.8
Pleasantness ¹ (1=Extremely pleasant, 5=Not at all pleasant)	2.8	2.9	3.2
Overall quality (1=Excellent, 2=Good, 3=Fair, 4=Poor, 5=Very poor)	2.0	2.4	2.3
Comfort ¹ (1=Extremely comfortable, 5=Not at all comfortable)	2.7	3.3	2.9
Average (1=Best, 5=Worst)	2.2	2.5	2.4
Number of participants preferring voice	7	5	0

TTS = Text to Speech.

¹ Reverse coded so that lower scores are better.

NOTE: It is important to note that this study consisted of a convenience sample, and results are not generalizable.

Two participants commented that the moderate voice was too fast, but one other participant thought the slow voice was faster than the moderate voice. However, this person was in Group 2 and heard the TTS slow voice immediately after hearing the human voice.

4.1.4 Timing

RTI conducted a brief timing test to compare the length of time required to listen to a question being read with the human voice compared with TTS slow or TTS moderate. For the comparison, RTI selected the 14 items in the ACASI tutorial section. These items include a variety of question types that may affect speed, including introductions with no response options, yes/no response options, and relatively long lists of response options.

As shown in Table 4.7, the timing for the English TTS slow voice was about 10 percent faster than the English human voice across all 14 items, and the English TTS moderate voice was about 21 percent faster. However, for some items, timing differences among the three voices varied somewhat compared with the overall pattern.

Two variables (INTRO1 and RANGEERR) had very little difference between the TTS slow voice and the human voice. RTI retested the timing to verify the lengths and found that the human voice recordings on those questions had practically no pause between paragraphs, enabling them to be read more quickly than other items of the same length. RTI also found that the TTS moderate voice is about 12 percent faster than the TTS slow voice.

Table 4.7 Question Length and Percentage Faster, by English Voice

Item	Length (seconds)			Percentage Faster		
	Human	TTS Slow	TTS Moderate	TSS Slow vs. Human	TTS Moderate vs. Human	TTS Moderate vs. TTS Slow
HEADPHONE	20	18	16	10.0	20.0	11.1
INTRO1	29	27	26	6.9	10.3	3.7
INTRO2	15	14	12	6.7	20.0	14.3
GOTDOG	26	22	19	15.4	26.9	13.6
EYECOLOR	35	29	26	17.1	25.7	10.3
ALLAPPLY	58	52	46	10.3	20.7	11.5
NUMBER	19	17	15	10.5	21.1	11.8
BACKUP	26	22	19	15.4	26.9	13.6
PLAYINFO	28	23	20	17.9	28.6	13.0
RANGEERR	45	44	39	2.2	13.3	11.4
CALENDAR	41	38	33	7.3	19.5	13.2
CALENDR2	25	23	20	8.0	20.0	13.0
CALENDR3	27	25	21	7.4	22.2	16.0
ANYQUES	8	7	6	12.5	25.0	14.3
Total	402	363	318	9.7	20.9	12.4

TTS = Text to Speech.

All three Spanish voices are slower than the respective English voices, but the differences between the TTS voices and the human voice show a similar pattern in Spanish and English. As shown in Table 4.8, the timing for the Spanish TTS slow voice was about 13.5 percent faster than the Spanish human voice across all 14 items, and the Spanish TTS moderate voice was about 21 percent faster. However, for some items, timing differences among the three voices varied somewhat compared with the overall pattern.

Although the TTS voices are faster than the human voice, it does not necessarily mean that administration times will be shorter. For example, if use of TTS makes respondents more likely to listen to the survey questions being read, that could increase administration time. On the other hand, if use of TTS makes respondents less likely to listen to the survey questions, it could decrease time of administration even more than what is shown in the table. The timing was evaluated further in the TTS pilot test phase of this project (see Section 4.2).

Table 4.8 Question Length and Percentage Faster, by Spanish Voice

Item	Length (seconds)			Percentage Faster		
	Human	TTS Slow	TTS Moderate	TSS Slow vs. Human	TTS Moderate vs. Human	TTS Moderate vs. TTS Slow
HEADPHONE	22	22	19	0.0	13.6	13.6
INTRO1	38	36	32	5.3	15.8	11.1
INTRO2	19	17	16	10.5	15.8	5.9
GOTDOG	32	26	25	18.8	21.9	3.8
EYECOLOR	46	37	34	19.6	26.1	8.1
ALLAPPLY	82	67	61	18.3	25.6	9.0
NUMBER	24	19	17	20.8	29.2	10.5
BACKUP	28	26	24	7.1	14.3	7.7
PLAYINFO	36	26	25	27.8	30.6	3.8
RANGEERR	50	54	49	-8.0	2.0	9.3
CALENDAR	53	45	41	15.1	22.6	8.9
CALENDR2	32	27	25	15.6	21.9	7.4
CALENDR3	38	30	27	21.1	28.9	10.0
ANYQUES	11	10	9	9.1	18.2	10.0
Total	511	442	404	13.5	20.9	8.6

TTS = Text to Speech.

4.1.5 Summary

NSDUH interviewers encourage all survey respondents to wear the headphones when completing the survey, even if they choose to turn off the volume. As a result, it is unclear exactly what percentage of NSDUH respondents rely on the audio when completing the survey questions, and thus, how much change in the voice used to read the questions could have on their comprehension of the questions. As a result, the cognitive interview phase was designed to evaluate differences in comprehension between the human voice and TTS for the most difficult situations—relying on audio only—and for participants who might have the hardest time understanding the questions; that is, the youngest and oldest participants (12-17 and 65 or older), non-native English speakers, and participants with low education or low literacy levels. Any issues that were identified were likely exacerbated by not allowing respondents to read the screen along with the voice.

Although both English and Spanish-speaking participants seemed to prefer the human voice slightly, most participants thought the TTS voice was pleasant and understandable. It is important to recognize that this cognitive interview included only static questions and did not include any dynamic questions that apply fills. It is believed that TTS sounds better compared with human voice recordings that must be stitched together on these types of questions. Furthermore, there were no differences in comprehension ratings among the three voices for either English- or Spanish-speaking participants. As a result, RTI recommended proceeding with the use of TTS for the 2015 NSDUH.

Although it appeared that there were very few differences between TTS slow and TTS moderate, RTI recommended using TTS slow for the NSDUH. Several participants noted that the TTS moderate voice required them to concentrate harder to understand it. Therefore, the slower voice might be less cognitively demanding when used for the entire ACASI portion of the survey.

4.2 Pilot Test Phase

Upon deciding that TTS software will be implemented on the 2015 NSDUH, the cognitive interview findings were used to refine the TTS audio to improve comprehensibility (e.g., altering the speed or modifying pronunciations), and then these refinements were tested in the pilot test phase to identify any major issues with administration time or any unanticipated issues with the use of TTS. In October and November 2014, RTI conducted 43 field interviews to evaluate the use of TTS. These field interviews mirrored the 2015 main study protocol in order to identify any major issues in administration time or any unanticipated issues with the use of TTS. Details on the pilot test and the results are outlined in the following sections.

4.2.1 Methods

4.2.1.1 Sampling

The respondent universe for the pilot test was the civilian, noninstitutionalized population aged 12 or older residing in the selected areas. Eligibility for the pilot test was determined based on where the occupants of the sampled dwelling units (DUs) reside for most of October, November, and December 2014. Data collection took place in the fourth quarter of 2014.

The pilot test goal was to include at least 20 interviews completed in English and at least 10 interviews completed in Spanish. The sample was selected in Los Angeles, California, and Miami, Florida, to meet staffing needs and ensure a sufficient number of Spanish interviews.

Retired quarter 1 2014 segments were used for selection. Based on past experience with these segments, three segments were selected to yield the desired number of interviews. After accounting for eligibility, nonresponse, and the person-level sample selection procedures, RTI estimated that approximately 107 selected DUs would yield at least 30 completed interviews. As discussed in the data collection section, RTI did not return to convert refusals for the pilot test. Refusal rates were taken into account when selecting the sample, and an additional reserve sample of 37 DUs was also sampled in case the 107 selected DUs did not yield a sufficient number of interviews.

To sufficiently evaluate the impacts of TTS among youths and the older population, the target respondent sample by age group was as follows:

Age Group	Allocation
12-17	10
18-25	5
26-34	4
35-49	4
50+	7

4.2.1.2 Staffing and Training

The pilot test involved purposefully selecting 3 segments to yield at least 30 interviews in total, including a minimum of 10 Spanish interviews. Nine field interviewers (FIs) were needed, including seven bilingual FIs. FIs were selected based on their performance in the 2013 Dress Rehearsal, location, data quality, dependability, availability to travel, and availability to attend training and complete data collection.

The FI training for the pilot test included at-home and in-person components. The training was a brief refresher training because all of the FIs already participated in the Dress Rehearsal, which used similar equipment, questionnaires, and procedures. For the at-home training, the pilot test FIs were sent an FI handbook, comprising relevant sections from the draft 2015 FI Manual and FI Computer Manual on the tablet and laptop computers and screening and interview procedures and materials, as well as a memorandum outlining the pilot test schedule, training, and related details. The FIs carefully read the FI handbook before attending the in-person refresher training session.

The in-person FI refresher training was a 1-day session held at RTI's Research Triangle Park, North Carolina, office. The morning of training was spent covering the equipment, instruments, and procedures, and the afternoon included practice exercises and mock interviews under trainer observation. Bilingual FIs completed one mock interview in Spanish under the observation of a bilingual trainer.

During training, trainers carefully observed FIs as they completed the practice exercises and mock interviews, and they provided specific feedback and retraining on any items completed incorrectly. Similar to the Dress Rehearsal, this process was used to reinforce the proper procedures before FIs began their fieldwork.

4.2.1.3 Description of Procedures

The pilot test FIs administered the screener instrument using 7-inch touchscreen Samsung Galaxy tablets that will be used for the 2015 NSDUH. The tablet contained the 2015 screening program, as well as tablet tools such as the parental introductory script and tablet video. FIs were provided with instruction on how to use these new tools at training.

The FIs administered the computer-assisted interview (CAI) instrument using lightweight Samsung Ultrabook laptops that will be used for the 2015 NSDUH. The laptops contained both the 2015 NSDUH CAI instrument and the TTS program. Customized TTS pronunciations were used in both languages. Based on the results from the TTS cognitive interview phase, TTS slow was used for both the English and Spanish instruments.

Data collection occurred immediately after training and lasted approximately 2 weeks. FIs reported to a survey specialist for case management and field issues during data collection. The programmers modified the case management system to accommodate case assignment and transfer requirements for the pilot test.

All data collection software and procedures were very similar to those planned for use in the 2015 NSDUH data collection. All materials mirrored those for 2015, with some minor wording changes to references regarding sample size, as well as data collection and project dates

and Office of Management and Budget approval numbers. Respondents received a \$30 incentive for their participation, which is the same incentive amount used in the NSDUH main data collection. Finally, case management procedures mirrored those for main study data collection with a few exceptions:

- Lead letters were sent to selected DUs with valid mailing addresses, but no follow-up contact occurred through mail. Refusal and unable-to-contact letters were not sent for the pilot test.
- Efforts were made to successfully complete each case within the data collection period. However, given that data collection lasted approximately 2 weeks, all screening or interview refusals were finalized at the initial refusal without any refusal conversion attempts.

4.2.1.4 Analysis and Reporting

RTI staff analyzed the results of the pilot test by comparing the timing data from the pilot test against the timing data from the Dress Rehearsal and those from the 2013 main study. Results were evaluated to assess whether the implementation of TTS affected overall instrument timing. However, it should be noted that the comparison to the Dress Rehearsal and the 2013 main study are limited because of the different questionnaires used. This is particularly true when comparing the results with the 2013 main study results.

Because of the short turnaround required for this pilot test, a full range of timing estimates was not produced. Timing data were provided by age (12-17, 18-64, and 65 or older) and by language (English vs. Spanish) for the following:

- Interview overall
- ACASI portion of the interview
- ACASI tutorial
- ACASI risk availability

The ACASI tutorial and risk availability sections were chosen because they are offered to all respondents, and all respondents are asked identical questions (i.e., there are no differences due to routing or skip logic), which made the timing estimates more comparable to those of the Dress Rehearsal and the 2013 main study.

Of the 107 selected DUs, 74 screenings were completed and 63 interviews were yielded. Of these 63 interviews, 43 interviews were completed. Of the interviews, 22 (51 percent) were completed in English, and 21 (49 percent) were completed in Spanish. Of the 43 interview respondents, 19 (44 percent) were aged 12-17, 23 (54 percent) were aged 18-64, and 1 (2 percent) was aged 65 or older. The respondent sample by the target age groups is provided in Table 4.9.

It is important to note that these distributions differ from those of the Dress Rehearsal and the 2013 main study, particularly regarding the proportion of interviews that were completed in Spanish, which was much higher in the pilot test (at 49 percent) compared with the Dress Rehearsal (at 9 percent) and the 2013 main study (at 3 percent).

Table 4.9 Text to Speech Interview Respondents, by Age Group

Age Group	Completed Interviews	Percentage	Cumulative Percentage
12-17	19	44	44
18-25	5	12	56
26-34	8	17	73
35-49	6	14	87
50+	5	12	99

Comparing the TTS pilot test mean in minutes for all respondents with that of the Dress Rehearsal and the 2013 main study showed that the timings were longest for the pilot test across the entire interview, the entire ACASI, and certain portions of the ACASI. As noted previously, this is primarily due to the high proportion of Spanish respondents, who tend to have longer interview times. Therefore, Sections 4.1.2 and 4.1.3 detail timing results and analysis by English and Spanish respondents. Detailed timing tables for all respondents who completed the pilot test can be found in Appendix E.

4.2.2 English-Speaking Participant Results

Of the pilot test interviews, 22 were completed in English. Of these 22 interviews, 12 (55 percent) were completed with respondents aged 12-17, 9 (41 percent) with respondents aged 18-64, and 1 (5 percent) with a respondent aged 65 or older.

Table 4.10 compares the TTS pilot test mean and median in minutes with that of the Dress Rehearsal and the 2013 main study for English respondents across the entire interview, the entire ACASI, and certain portions of the ACASI. The average overall pilot test interview time was slightly longer compared with the Dress Rehearsal, but the median overall pilot test interview time was slightly shorter. This suggests the presence of outliers, which can have a large effect when sample sizes are small. The average and median ACASI times for the pilot were slightly shorter compared with the Dress Rehearsal. This is consistent with the fact that the number of questions administered via ACASI in the Dress Rehearsal and the 2015 instrument were very similar.

The overall pilot test interview time was about the same (or shorter) as the 2013 main study, but the ACASI time was longer. This difference is likely explained by the addition of some questions to ACASI in the pilot test, such as questions on disability, language, sexual orientation, and military family members. Some questions from the 2013 computer-assisted personal interview were also moved to the ACASI for the pilot test, including questions on education, employment, moving, and country in which the respondent was born.

The two modules that had roughly the same content among the pilot test, the Dress Rehearsal, and the 2013 main study were the ACASI tutorial and ACASI risk availability. The ACASI tutorial time was shorter in the pilot test than the Dress Rehearsal or the 2013 main study, but the pilot test ACASI risk availability time was slightly longer compared with both the Dress Rehearsal and the 2013 main study. The cause of this difference is unclear.

Table 4.10 also details the pilot test data by age group. Pilot test timings were the same as or faster than those of the Dress Rehearsal and the 2013 main study for participants aged 18 or older. However, timings for the 12-17 age group were longer on the pilot compared with the

Dress Rehearsal and the 2013 main study, with the exception of the ACASI tutorial. The 12-17 age group represented a much larger portion of the English interview respondents in the pilot test than in the Dress Rehearsal and the 2013 main study (55 percent compared with 25 percent and 33 percent, respectively). Consequently, the difference in the overall timings is almost entirely explained by the differences in the 12-17 age group. It is unclear why the younger age group had slower timings on the pilot test compared with the Dress Rehearsal or the 2013 main study. Increased timings may be due to differences in respondents among the three studies. Another possible explanation is that respondents in this age group are more likely to listen to the TTS slow voice reading the question compared with the human voice.

Detailed timing tables for respondents who completed the pilot test in English can be found in Appendix E.

Table 4.10 Text to Speech Audit Trail Timing Data: Mean and Median in Minutes, Comparison across Instruments, English-Speaking Respondents, by Age Group

	Mean			Median		
	Text to Speech	Dress Rehearsal	2013 Main Study	Text to Speech	Dress Rehearsal	2013 Main Study
Interview Overall						
Overall	61.40	59.56	61.95	54.86	55.88	58.87
12-17	69.63	59.55	61.37	68.62	57.13	58.97
18-64	50.38	57.16	61.38	51.22	53.53	58.15
65+	61.80	84.98	74.38	61.80	79.75	70.40
ACASI¹						
Overall	43.27	44.72	41.01	40.65	41.20	38.30
12-17	49.17	42.09	41.31	51.01	40.37	39.37
18-64	35.21	43.56	39.94	35.15	40.30	37.03
65+	45.02	67.11	53.87	45.02	61.48	50.37
ACASI Tutorial²						
Overall	2.92	3.43	3.44	2.67	3.23	3.25
12-17	3.37	3.56	3.69	3.17	3.47	3.58
18-64	2.53	3.19	3.20	2.62	2.95	3.00
65+	1.02	5.42	4.88	1.02	4.85	4.72
ACASI Risk Availability						
Overall	3.14	2.89	2.94	2.83	2.62	2.65
12-17	3.38	2.76	3.00	2.99	2.63	2.78
18-64	2.80	2.77	2.80	2.75	2.50	2.53
65+	3.40	4.59	4.48	3.40	4.15	3.93

ACASI = audio computer-assisted self-interview.

¹Timing for the ACASI section began with the INTROACASI1 variable and ended with the ENDAUDIO variable. The ACASI section for the 2013 main study did not include several questions that were included in the Text to Speech and Dress Rehearsal instruments.

²Timing for the ACASI tutorial began with the INTRO1 variable for all instruments. The end variable for the Text to Speech and Dress Rehearsal was RANGEERR, and the end variable was ANYQUES for the 2013 main study.

4.2.3 Spanish-Speaking Participant Results

Of the pilot test interviews, 21 were completed in Spanish. Of these 21 interviews, 7 (33 percent) were completed with respondents aged 12-17, and 14 (67 percent) were completed with

respondents aged 18-64. No interviews were completed in Spanish with respondents aged 65 or older.

Table 4.11 compares the TTS pilot test mean and median in minutes for all Spanish respondents with that of respondents who completed the Dress Rehearsal and the 2013 main study in Spanish across the entire interview, the entire ACASI, and portions of the ACASI. All timings were longest for the pilot test, including the tutorials and risk availability portions for the ACASI. The reason for the longer times in the pilot test is unclear because the TTS Spanish voice is faster than the Spanish human voice. The mean and median times for the pilot are similar; therefore, the difference does not appear to be due to outliers. Again, increased timings may be due to differences in respondents among the three studies or because Spanish-speaking respondents are more likely to listen to the TTS voice reading the question compared with the human voice.

Table 4.11 Text to Speech Audit Trail Timing Data: Mean and Median in Minutes, Comparison across Instruments, Spanish-Speaking Respondents, by Age Group

	Mean			Median		
	Text to Speech	Dress Rehearsal	2013 Main Study	Text to Speech	Dress Rehearsal	2013 Main Study
Interview Overall						
Overall	89.45	83.94	83.43	89.67	79.32	79.83
12-17	82.05	65.57	71.24	74.95	63.40	67.87
18-64	93.15	86.70	86.54	92.73	83.44	84.07
65+	N/A	109.71	92.85	N/A	100.19	93.38
ACASI¹						
Overall	69.90	63.61	57.70	66.98	59.47	55.25
12-17	60.88	40.48	48.09	56.68	40.27	46.42
18-64	74.41	67.87	60.03	73.77	62.89	58.35
65+	N/A	87.45	67.37	N/A	78.98	66.82
ACASI Tutorial²						
Overall	5.36	4.94	5.07	5.53	4.92	5.10
12-17	4.94	3.63	4.44	4.25	3.20	4.33
18-64	5.57	5.16	5.26	6.00	5.18	5.32
65+	N/A	6.48	5.16	N/A	6.50	5.57
ACASI Risk Availability						
Overall	4.88	4.38	4.85	4.60	4.20	4.40
12-17	3.71	2.92	3.79	3.47	2.82	3.53
18-64	5.47	4.59	5.10	4.77	4.47	4.63
65+	N/A	6.56	6.02	N/A	7.58	5.97

N/A = Not applicable; ACASI = audio computer-assisted self-interview.

¹ Timing for the ACASI section began with the IntroAcasi1 variable and ended with the ENDAUDIO variable. The ACASI section for the 2013 main study did not include several questions that were included in the TTS and Dress Rehearsal instruments.

² Timing for the ACASI tutorial began with the INTRO1 variable for all instruments. The end variable for the TTS and Dress Rehearsal was RANGEERR, and the end variable was ANYQUES for the 2013 main study.

Table 4.11 also details the pilot test data by age group. Pilot test timings for the 12-17 age group were longer compared with the Dress Rehearsal and the 2013 main study, except when looking at the ACASI risk availability timing in which the pilot test timing was shorter than the

2013 main study timing but longer than the Dress Rehearsal timing. Pilot test timings for the 18-64 age group were longer compared with the Dress Rehearsal and the 2013 main study across all interview sections. Detailed timing tables for respondents who completed the pilot test in Spanish can be found in Appendix E.

4.2.4 FI Debriefing Call Results

The purpose of the pilot test FI debriefing call was to obtain direct feedback from FIs on their experiences collecting data using the 2015 NSDUH questionnaire with the TTS voice. The debriefing also provided the opportunity to gather additional feedback on other 2015 changes, such as completing interviews on the new Samsung Ultrabook laptop and completing screenings on the Samsung Galaxy tablet in both English and Spanish. The goal of the debriefing call was to gather feedback from FIs (including bilingual FIs) on topics including

- significant questions or concerns raised by interview respondents about the computerized ACASI voice;
- significant questions or concerns raised by members of sample households about the tablet video, which is a new tool for 2015;
- challenges encountered using the tablet to conduct household screenings; and
- challenges encountered using the laptop to conduct interviews.

The results of the pilot test FI debriefing call will be used to inform preparations for the 2015 NSDUH or the 2016 NSDUH.

One FI debriefing call, which lasted approximately 45 minutes, was held with 8 pilot test FIs attending. The call included a moderator and a note taker, along with several observers including SAMHSA staff, and was recorded. The remaining pilot test FI who was unable to join the call provided feedback on the debriefing questions on an individual call.

The moderator began the call with a brief introduction, and the remainder of the call focused on specific section topics:

- Screening and Using the Tablet
- TTS Questionnaire and ACASI Voice
- Administering the TTS Interview and Using the Laptop

The TTS Debriefing Moderator's Guide can be found in Appendix F and includes the specific questions covered in each section.

Feedback on the tablet video was positive. Two FIs used the video with respondents, and in these situations, the respondents did not make any comments about the video. However, these respondents also did not have any further questions after viewing the video, and the FIs thought this was due to the content of the video. Neither FI experienced issues with bringing up the video for the respondents. All FIs liked the content of the video and did not have any suggested changes. It should be noted that refusal conversion was not attempted for the pilot test, and some FIs thought the video would be best used with difficult respondents.

Regarding the tablet and the screening program, the FIs mentioned only a few issues, outlined here:

- Some FIs mentioned that they were concerned about losing the stylus because it did not fit tightly into the holder. Some FIs chose not to use the stylus to avoid losing it, and one FI did lose the stylus while in the field. However, this did not interrupt the screening process because FIs could use their fingers to navigate through the program. The stylus has a clip that can be used to secure it in the holder.
- Two FIs mentioned that they had some difficulty connecting to the Internet in public locations, mainly because they needed to open the Internet browser to accept the business's user agreement before connecting. All FIs will practice connecting to WiFi at the 2015 Veteran FI Training.
- Within the screening program, some FIs noted that they could not enter certain symbols such as apostrophes in the Record of Call comments. Although most standard symbols may be entered into comment fields in the screening program, the use of single and double apostrophes was disabled in the 2015 screener program to prevent problems that these symbols could cause in the data transmission process.
- A couple of FIs mentioned that unlike the iPAQ, the tablet screening program did not allow persistent highlighting of cases on the Select Case screen. This functionality is not available because of constraints imposed by the tablet's operating system.
- A few FIs thought that the tablet goes into sleep mode too quickly, resulting in having to enter the password frequently. One FI expressed concern about entering the password in the presence of respondents. The tablet has been configured to allow for the maximum time of inactivity (10 minutes) before entering sleep mode.
- One FI thought the tablet battery drained too quickly, which may have been due to having the WiFi setting enabled. Training and practice have been incorporated into the 2015 Veteran FI Training program to ensure that FIs know how to turn WiFi on and off. Also, previous testing on the tablet battery has indicated an expected battery life of approximately 6 hours, and all FIs will be provided with a car charger and a standard charger to use with the tablet.

The debriefing call uncovered no issues with the TTS ACASI voice or the questionnaire. No respondents commented on the voice or showed nonverbal signs of confusion, frustration, or changing the volume during the ACASI section of the interview. One FI commented that two youth respondents made comments about the questionnaire being repetitive. Two FIs had respondents (one English-speaking and one Spanish-speaking) who commented that the interview was long. However, the FIs also mentioned that these comments are not different from what they hear with the main study data collection.

Regarding the laptop and the interview program, the FIs mentioned only a few issues, outlined as follows:

- No FI reported respondent confusion over the function key labels, and all FIs were fine with using the same labels for 2015. One FI did comment that the function keys themselves are very small, and a couple of respondents had to look closely to find

them. This FI thought it may be harder for some respondents (particularly older respondents) to use and select the keys accurately.

- Two FIs had respondents say that the laptop screen was too bright. However, both of these respondents were completing the interview in rooms with dim lighting.
- One FI commented that the Windows lock screen appeared between interviews when completing two interviews in the same household. The FI needed to call technical support for the password the first time this happened and was able to use the laptop without technical support on future occasions.

Aside from the minor issues detailed previously, the FIs liked the new equipment and the changes to the screening and interview programs and look forward to using the equipment in 2015.

4.3 Pretest Conclusions

Based on the TTS cognitive interviews, SAMHSA decided to use the TTS slow voice for the 2015 NSDUH. After the cognitive interviews, RTI continued to test and modify the TTS pronunciations to improve the quality and cadence before the pilot test. For example, drugs that include "generic" in parentheses were modified so that the TTS did not add the slight pause associated with parentheses for drug lists. The updated 2015 CAI instrument with TTS software was used to complete 43 interviews in the pilot test.

The overall results of the pilot test did not reveal any issues with administering the 2015 CAI instrument with the TTS software. Respondents did not make any comments about the voice. The pilot test timing data show only slight variations among all English respondents compared with the Dress Rehearsal and the 2013 main study, although timings were longer for adolescents and shorter for other age groups. It is inconclusive at this point whether the 2015 administration times will be longer or shorter for English respondents. For Spanish respondents, the pilot test timing data show that TTS was longer when compared with the Dress Rehearsal and the 2013 main study. However, it is important to note that with a small sample size, it is difficult to know if these longer times can be attributed to the instrument, the respondents, or both. The early data review conducted in January 2015 will provide another opportunity to review overall timing data across all respondents compared with the Dress Rehearsal and the 2013 main study.

The FI feedback provided on the pilot test data collection will be taken into consideration to refine the screening and interviewing programs as feasible. Only small suggestions were made, and overall, the data collection did not uncover any major issues with the 2015 instrument, equipment, or protocols.

5. References

- Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., Hauck, W., & Coates, T. J. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly*, *60*, 345-375. doi:10.1086/297758
- Couper, M. P. (2005). Technology trends in survey data collection. *Social Science Computer Review*, *23*, 486-501. doi:10.1177/0894439305278972
- Couper, M. P., Kirgis, N., Buageila, S., & Berglund, P. (2012). *Using text-to-speech (TTS) for audio-CASI*. Presented at the American Association for Public Opinion Research 67th Annual Conference, Orlando, FL.
- Couper, M. P., Singer, E., & Tourangeau, R. (2004). Does voice matter? An interactive voice response (IVR) experiment. *Journal of Official Statistics*, *20*, 551-570.
- Couper, M., Tourangeau, R., & Marvin, T. (2009). Taking the audio out of audio CASI. *Public Opinion Quarterly*, *73*, 281-303.
- Dindia, K., & Allen, M. (1992). Sex differences in self-disclosure: A meta-analysis. *Psychological Bulletin*, *112*, 106-124. doi:10.1037/0033-2909.112.1.106
- Dykema, J., Basson, D., & Schaeffer, N. C. (2007). Face-to-face surveys. In W. Donsbach & M. W. Traugott (Eds.), *The SAGE handbook of public opinion research* (pp. 240-248). London: Sage Publications.
- Dykema, J., Diloreto, K., Price, J. L., White, E., & Schaeffer, N. C. (2012). ACASI gender-of-interviewer voice effects on reports to questions about sensitive behaviors among young adults. *Public Opinion Quarterly*, *76*, 311-325. doi:10.1093/poq/nfs021
- Fahrney, K. M., Uhrig, J., & Kuo, T. M. (2010, April). *Gender-of-voice effects in an ACASI study of same-sex behavior* (RTI Press Methods Report MR-0017-100). Retrieved from <http://www.rti.org/publications/rtipress.cfm?pubid=14766>
- Griggs, B. (2011, October 21). *Why computer voices are mostly female*. Retrieved from <http://www.cnn.com/2011/10/21/tech/innovation/female-computer-voices/>
- Jourard, S. M. (1971). *Self-disclosure: An experimental analysis of the transparent self*. New York, NY: Wiley-Interscience.
- Kraft, J., & Taylor, W. (2006, May). *Text-to-speech application in audio CASI: Evaluation of implementation and deployment*. Presented at the International Field Directors & Technologies Conference, Montreal, Quebec, CA.
- Nass, C., Moon, Y., & Greene, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*, 864-876. doi:10.1111/j.1559-1816.1997.tb00275.x

Nass, C., Robles, E., Heenan, C., Bienstock, H., & Treinen, M. (2003). Speech-based disclosure systems: Effects of modality, gender of prompt, and gender of user. *International Journal of Speech Technology*, 6, 113-121. doi:10.1023/A:1022378312670

Phillips, J., Edwards, B., & Dolbow, E. (2013). *Using text-to-speech software for ACASI*. Presented at the Federal CASIC Workshops, Washington, DC.

Pollner, M. (1998). The effects of interviewer gender in mental health interviews. *Journal of Nervous and Mental Disease*, 186, 369-373.

Schaeffer, N. C. (2000). Asking questions about threatening topics: A selective overview. In A.A. Stone, J.S. Turkkan, C.A. Bachrach, J.B. Jobe, H.S. Kurtzman, & V.S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 105-121). Mahwah, NJ: Lawrence Erlbaum Associates.

Tannen, D. (1996). *Gender and discourse*. New York, NY: Oxford University Press.

Tourangeau, R., & Smith, T.W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60, 275-304.

Turner, C.F., Forsyth, B. H., O'Reilly, J., Cooley, P. C., Smith, T. K., Rogers, S. M., & Miller, H. G. (1998). Automated self-interviewing and the survey measurement of sensitive behaviors. In M. P. Couper, et al. (Eds.), *Computer-assisted survey information collection*. New York: Wiley.

Weisel, D. L. (2002). *Contemporary gangs: An organizational analysis*. New York: LFB Scholarly.

Appendix A: ACASI Questions Selected for TTS Prototypes

A.1. Summary of Proposed CAI Items for Prototypes

The full question text and response options from the 2013 Dress Rehearsal computer-assisted interviewing (CAI) specifications are provided in Section A.2.

Variable Name	Description
1. Card 3a	Long list of types of alcoholic beverages
2. AL 01	Asks whether R has ever drank alcohol
3. HALINTRO	Introduction to hallucinogen module – lists various hallucinogens
4. INHINTRO	Introduction to inhalant module – lists various inhalants
5. TR03	Asks about past use of specific tranquilizers
6. TR05	Asks about past use of other tranquilizers
7. ST04	Asks about past use of specific stimulants
8. SV03	Asks about past use of specific sedatives
9. PRINTROYR2	Pre-fills a previous response
10. HLTH25	Asks whether R has had a specific set of health conditions
11. AD19	Includes pre-fills from previous response about mood problems
12. INTROINC	Asks about family income, includes fills and wording changes depending on family relationships and whether proxy is answering

R = respondent.

A.2. Selected Items from CAI Specifications—Full Question Text and Response Options (2013 Dress Rehearsal CAI Specifications)

A. English Versions

1. CARD3a Types of Alcoholic Beverages

Beer

Regular beer

Lite or light beer

Low-alcohol (LA) beer

Malt liquor

Ale

Stout

Lager

Wine

Red, white, blush wine

Wine coolers

Champagne

Sherry

Homemade wines, such as muscadine, scuppernong, or fruit wines

Fortified wines, such as Cisco

Liquor

Bourbon

Gin

Rum
Scotch
Tequila
Vodka
Homemade liquor, such as moonshine

Liqueurs, Cordials, and Brandy

Brandy
Cassis
Cognac
Creme de menthe
Drambuie
Grand Marnier
Kahlua
Port
Schnapps
Tia Maria
Triple sec
Vermouth

Mixed Drinks and Cocktails

Bloody Mary
Bourbon and water
Daiquiri
Gin and tonic
Manhattan
Margarita
Martini
Piña colada
Rob Roy
Rum and cola
Scotch and soda
Whiskey sour

Press [ENTER] to continue.

2. AL01

Have you **ever**, even once, had a drink of any type of alcoholic beverage? Please do not include times when you only had a sip or two from a drink.

- 1 Yes
- 2 No
- DK/REF

3. HALINTRO

The next questions are about substances called hallucinogens. These drugs often cause people to see or experience things that are not real.

A list of some common hallucinogens is shown below. These and many other substances that people use as hallucinogens are often known by street names, and we can't list them all. Please take a moment to look at the substances listed below so you know what kind of drugs the next questions are about.

- LSD, also called "acid"
- PCP, also called "angel dust" or phencyclidine
- Peyote
- Mescaline
- Psilocybin
- "Ecstasy," also called MDMA
- Ketamine, also called "Special K" or "Super K"
- DMT, also called dimethyltryptamine
- AMT, also called alpha-methyltryptamine
- Foxy, also called 5-MeO-DIPT
- Salvia divinorum

Press [ENTER] to continue

4. INHINTRO

These next questions are about liquids, sprays, and gases that people sniff or inhale to get high or to make them feel good.

We are not interested in times when you inhaled a substance accidentally—such as when painting, cleaning an oven, or filling a car with gasoline. The questions use the word "inhalant" to include all the things listed below, as well as any other substances that people sniff or inhale for kicks or to get high.

Take a moment to look at the substances listed below so you know what kinds of liquids, sprays, and gases these questions are about.

- Amyl nitrite, "poppers," locker room odorizers, or "rush"
- Correction fluid, degreaser, or cleaning fluid
- Gasoline or lighter fluid
- Glue, shoe polish, or toluene
- Halothane, ether, or other anesthetics
- Lacquer thinner, or other paint solvents
- Lighter gases, such as butane or propane
- Nitrous oxide or "whippits"
- Felt-tip pens, felt-tip markers, or magic markers
- Spray paints
- Computer keyboard cleaner, also known as air duster
- Other aerosol sprays

Press [ENTER] to continue.

5. TR03

Please look at the names and pictures of the tranquilizers shown below.

PROGRAMMER: DISPLAY PILLS HERE FOR VALIUM, DIAZPEPAM, TRANXENE, AND OXAZEPAM.

In the **past 12 months**, which, if any, of these tranquilizers have you used?

To select more than one drug from the list, press the space bar between each number you have typed. When you have finished, press [ENTER].

- 1 Valium
 - 2 Librium
 - 3 Tranxene
 - 4 Diazepam (generic)
 - 5 Oxazepam (generic), also known as Serax
 - 95 I have not used any of these tranquilizers in the past 12 months
- DK/REF

6. TR05

Please look at the names and pictures of the tranquilizers shown below.

PROGRAMMER: DISPLAY PILLS FOR BUSPIRONE, HYDROXYZINE, AND MEPROBAMATE.

In the **past 12 months**, which, if any, of these tranquilizers have you used?

To select more than one drug from the list, press the space bar between each number you have typed.

When you have finished, press [ENTER].

- 1 Buspirone (generic), also known as BuSpar
 - 2 Hydroxyzine (generic), also known as Atarax or Vistaril
 - 3 Meprobamate (generic), also known as Equanil or Miltown
 - 95 I have not used any of these tranquilizers in the past 12 months
- DK/REF

7. ST04

Please look at the names and pictures of the stimulants shown below.

PROGRAMMER: DISPLAY PILLS FOR BENZPHETAMINE, DIDREX, DIETHYLPROPION, PHENDIMETRAZINE, AND PHENTERMINE.

In the **past 12 months**, which, if any, of these stimulants have you used?

To select more than one drug from the list, press the space bar between each number you have typed. When you have finished, press [ENTER].

- 1 Benzphetamine
- 2 Didrex
- 3 Diethylpropion
- 4 Phendimetrazine
- 5 Phentermine
- 95 I have not used any of these stimulants in the past 12 months

DK/REF

8. SV03

Please look at the names and pictures of the sedatives shown below.

PROGRAMMER: DISPLAY PILLS FOR DALMANE, HALCION, FLURAZEPAM AND TRIAZOLAM.

In the **past 12 months**, which, if any, of these sedatives have you used?

To select more than one drug from the list, press the space bar between each number you have typed. When you have finished, press [ENTER].

- 1 Dalmane
- 2 Halcion
- 3 Flurazepam (generic)
- 4 Triazolam (generic)
- 95 I have not used any of these sedatives in the past 12 months

DK/REF

9. PRINTROYR2

NOTE: For this question, we will assume a fill of 4 drug names.

[IF PR12MON=1 AND (PR11 NE 1 OR (PR11=1 AND PRYRCOUNT > 1))]

Earlier, the computer recorded that, in the **past 12 months**, you used [PRFILL].

Press Enter to continue.

PROGRAMMER: SHOW CALENDAR WITH 12-MONTH REFERENCE DATE FOR THE INTRO SCREEN

10. HLTH25

Below is a list of health conditions that you may have had during your lifetime.

Please read the list and type in the numbers of all of the conditions that a doctor or other health care professional has ever told you that you had.

To select more than one condition, press the space bar between each number you type. When you have finished, press [ENTER].

- 1 Any kind of heart condition or heart disease
- 2 Diabetes or sugar diabetes
- 3 Chronic bronchitis, emphysema, chronic obstructive pulmonary disease, also called COPD
- 4 Cirrhosis of the liver
- 5 Hepatitis B or C
- 6 Kidney disease, not including bladder infection or incontinence
- 7 Asthma
- 8 HIV or AIDS

- 9 Cancer or a malignancy of any kind
- 10 Hypertension, also called high blood pressure
- 95 None of the above - I have never had any of these conditions
- DK/REF

11. AD19

[IF AD16 = 2, 3, 4, OR DK/REF] Once again, please think of times lasting two weeks or longer when [NUMPROBS] with your mood [WASWERE] most **severe and frequent**.

How often, during those times, was your emotional distress so severe that you **could not carry out your daily activities**?

- 1 Often
- 2 Sometimes
- 3 Rarely
- 4 Never
- DK/REF

12. INTROINC

NOTE: For this question, we will assume a 4 person family with father, mother, son, daughter

[IF NO FAMILY MEMBERS IN ROSTER]

These next questions are about the kinds and amounts of income that you receive.

[IF ONE FAMILY MEMBER IN ROSTER AND HASJOIN NE 1]

These next questions are about the kinds and amounts of income received by you and your

[FAMILY RELATIONSHIP FILL].

[IF ONE FAMILY MEMBER IN ROSTER AND HASJOIN=1]

These next questions are about the kinds and amounts of income received by [SAMPLE MEMBER] and you.

[IF AT LEAST TWO FAMILY MEMBERS IN ROSTER AND HASJOIN NE 1]

These next questions are about the kinds and amounts of income received by your family living here, including you, your [FAMILY RELATIONSHIP FILLS].

[IF AT LEAST TWO FAMILY MEMBERS IN ROSTER AND HASJOIN=1]

These next questions are about the kinds and amounts of income received by [SAMPLE MEMBER] and [IF QD01=5 FILL his, QD01 = 9 FILL her] family living here, including you,

[IF QD01=5 FILL his, QD01 = 9 FILL her] [FAMILY RELATIONSHIP FILLS].

[PROGRAMMER NOTE: THE PROXY SHOULD NOT APPEAR IN [FAMILY RELATIONSHIP FILLS]. ALSO, USE 'other' AS A MODIFIER TO THE FAMILY RELATIONSHIP FILL WHEN THE RELATIONSHIP TYPE IS EQUAL TO PROXY

RELATIONSHIP TYPE AND ONE OF THESE RELATIONSHIP TYPES IS STILL IN THE LIST. PLEASE PRECEDE EACH RELATIONSHIP WITH 'HIS/HER'.]

[IF HASJOIN NE 1] These questions refer to the calendar year [CURRENT YEAR - 1] rather than to the past 12 months that were referred to in some earlier questions. The calendar year [CURRENT YEAR - 1] would be from January 1st, [CURRENT YEAR - 1], through December 31st, [CURRENT YEAR - 1].

Press [ENTER] to continue.

A.3. Spanish Versions

1. CARD3a Tipos de Bebidas Alcohólicas

Cerveza

Cerveza

Cerveza ligera o "lite"

Cerveza con poco alcohol (LA)

Malta con alcohol

"Ale"

Cerveza negra

Lager o Cerveza dorada

Vino

Vino tinto, blanco, rosado

"Wine coolers"

Champaña

Jerez

Vinos caseros, tales como uva moscatel, scuppernong o vinos frutales

Vinos fortificados tales como Cisco

Licor

Whisky americano

Ginebra o "Gin"

Ron

Whisky escocés

Tequila

Vodka

Alcohol casero destilado ("Moonshine")

Licores de esencias, Cordiales y Brandy

Brandy

Cassis

Coñac

Crema de menta

Drambuie

Grand Marnier

Kahlua

Oporto

Schnapps
Tía María
Triple seco
Vermut

Bebidas mezcladas y Cocteles

Bloody Mary
Whisky con agua
Daiquiri
"Gin y tónico"
Manhattan
Margarita
Martini
Piña colada
Rob Roy
Ron con Coca Cola
Whisky escocés con soda
Whisky sour

Presione [ENTER] para continuar.

2. AL01

¿**Alguna vez** ha tomado una bebida alcohólica, aunque haya sido solo una vez? Por favor no incluya las ocasiones en que usted haya tomado solo uno o dos sorbos de una bebida.

1 Sí
2 No
DK/REF

3. HALINTRO

Las siguientes preguntas se tratan de las sustancias que se llaman alucinógenos. Estas drogas muchas veces hacen que las personas vean o experimenten cosas que no son reales.

A continuación hay una lista de algunos alucinógenos populares. Estas y muchas otras sustancias que la gente usa como alucinógenos se conocen frecuentemente por su nombre popular o de la calle. No podemos enumerarlos todos. Por favor preste atención al leer la lista de sustancias que sigue para saber a qué drogas se refieren las próximas preguntas.

LSD, también llamado 'ácido'
PCP, también llamado 'polvo de ángel' o fenciclidina
Peyote
Mescalina
Psilocibina
'Éxtasis,' también llamado MDMA
Ketamina, conocida en inglés como "Special K" o "Super K" y en español se le llama "Ketalar," "Hoyo K" o "vitamina K"
DMT, también llamado dimetiltriptamina
AMT, también llamado alfa-metiltriptamina

Foxy, también llamado "metoxi foxy" y cuyo nombre químico es 5-metoxi-N o 5-MeO-DIPT

Salvia divinorum, también llamada "Salvia de los adivinadores," "San Pedro," "planta sagrada" o "hierba pastora"

Presione [ENTER] para continuar.

4. INHINTRO

Las siguientes preguntas son acerca de líquidos, aerosoles o esprays y gases que las personas aspiran o inhalan para drogarse o para sentirse alegres.

No estamos interesados en ocasiones en que usted inhaló alguna sustancia accidentalmente como en el caso de pintar, limpiar un horno o echarle gasolina al automóvil. Las preguntas usan el término 'inhalante' para incluir todas las cosas mencionadas a continuación, así como cualquier otra sustancia que las personas aspiran o inhalan para divertirse o para drogarse. Por favor mire con atención la lista de sustancias a continuación, para saber a qué clases de líquidos, aerosoles o esprays y gases se refieren las próximas preguntas.

Nitrato de amilo, 'bombitas,' desodorante ambiental, o 'rush'

Líquido de corrección o 'liquid paper', desengrasador o líquido de limpieza

Gasolina o líquido para encendedores

Pegamento, crema o betún para limpiar zapatos, o tolueno

Halotano, éter u otros anestésicos

'Tiner' u otros solventes para pintura

Gases para encendedores, tales como butano o propano

Óxido nitroso o 'whippits'

Marcadores de punta fina, plumones o plumones mágicos

Pintura en aerosol

Limpiador para teclado de computadora, también llamado aire comprimido removedor de polvo

Otros aerosoles o esprays

Presione [ENTER] para continuar.

5. TR03

Por favor mire los nombres y las fotos de los tranquilizantes que se muestran a continuación.

PROGRAMMER: DISPLAY PILLS HERE FOR VALIUM, DIAZPEPAM, TRANXENE, AND OXAZEPAM.

En los **últimos 12 meses**, ¿cuál de estos tranquilizantes ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

- 1 Valium
- 2 Librium
- 3 Tranxene

- 4 Diazepam (genérico)
 - 5 Oxazepam (genérico), también conocido como Serax
 - 95 No he usado ninguno de estos tranquilizantes en los últimos 12 meses
- DK/REF

6. TR05

Por favor mire los nombres y las fotos de los tranquilizantes que se muestran a continuación.

PROGRAMMER: DISPLAY PILLS HERE FOR BUSPIRONE, HYDROXYZINE, AND MEPROBAMATE.

En los **últimos 12 meses**, ¿cuál de estos tranquilizantes ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

- 1 Buspirona (genérico), también conocido como BuSpar
 - 2 Hidroxizina (genérico), también conocido como Atarax o Vistaril
 - 3 Meprobamato, (genérico) también conocido como Equanil o Miltown
 - 95 No he usado ninguno de estos tranquilizantes en los últimos 12 meses
- DK/REF

7. ST04

Por favor mire los nombres y las fotos de los estimulantes que se muestran a continuación.

PROGRAMMER: DISPLAY PILLS FOR BENZPHETAMINE, DIDREX, DIETHYLPROPION, PHENDIMETRAZINE, AND PHENTERMINE.

En los **últimos 12 meses**, ¿cuál de estos estimulantes ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

- 1 Benzfetamina
 - 2 Didrex
 - 3 Dietilpropión
 - 4 Fendimetracina
 - 5 Fentermina
 - 95 No he usado ninguno de estos estimulantes en los últimos 12 meses
- DK/REF

8. SV03

Por favor mire los nombres y las fotos de los sedantes que se muestran a continuación.

PROGRAMMER: DISPLAY PILLS FOR DALMANE, HALCION, FLURAZEPAM AND TRIAZOLAM.

En los **últimos 12 meses**, ¿cuál de estos sedantes ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

- 1 Dalmane
 - 2 Halcion
 - 3 Flurazepam (genérico)
 - 4 Triazolam (genérico)
 - 95 No he usado ninguno de estos sedantes en los últimos 12 meses
- DK/REF

9. PRINTROYR2

NOTE: For this question, we will assume a fill of 4 drug names

[IF PR12MON=1 AND (PR11 NE 1 OR (PR11=1 AND PRYRCOUNT > 1))]

Anteriormente, la computadora registró que usted usó [PRFILL] en los últimos 12 meses.

Presione [Enter] para continuar.

PROGRAMMER: SHOW CALENDAR WITH 12-MONTH REFERENCE DATE FOR THE INTRO SCREEN

10. HLTH25

A continuación se muestra una lista de trastornos de la salud que usted pudiera haber tenido en el **transcurso de su vida**.

Por favor, lea la lista y escriba los números correspondientes a las enfermedades que **alguna vez** un doctor u otro profesional médico le dijo que tuvo.

Para seleccionar más de una enfermedad, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

- 1 Algún tipo de enfermedad o trastorno del corazón
- 2 Diabetes o diabetes del azúcar
- 3 Bronquitis crónica, enfisema, enfermedad pulmonar obstructiva crónica, también llamada COPD en inglés
- 4 Cirrosis del hígado
- 5 Hepatitis B o C
- 6 Enfermedad de los riñones, sin incluir infección a la vejiga o incontinencia urinaria
- 7 Asma
- 8 VIH o SIDA
- 9 Cáncer o algún tipo de tumor maligno
- 10 Hipertensión, también llamada presión sanguínea alta
- 95 Ninguna enfermedad arriba mencionada - Nunca he tenido ninguna de estas enfermedades

DK/REF

11. AD19

[IF AD16 = 2, 3, 4, OR DK/REF] Una vez más, por favor piense en las veces que [NUMPROBS] con su estado de ánimo [WASWERE] por dos semanas o más.

Durante esas ocasiones, ¿con qué frecuencia era su malestar emocional tan grave que **no podía realizar sus actividades diarias**?

- 1 Muchas veces
- 2 Algunas veces
- 3 Casi nunca
- 4 Nunca
- DK/REF

12. INTROINC

NOTE: For this question, we will assume a 4 person family with father, mother, son, daughter

[IF NO FAMILY MEMBERS IN ROSTER]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que usted recibe.

[IF ONE FAMILY MEMBER IN ROSTER AND HASJOIN NE 1]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que reciben usted y su [FAMILY RELATIONSHIP FILL].

[IF ONE FAMILY MEMBER IN ROSTER AND HASJOIN=1]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que reciben [SAMPLE MEMBER] y usted.

[IF AT LEAST TWO FAMILY MEMBERS IN ROSTER AND HASJOIN NE 1]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que reciben los miembros de su familia que viven aquí, incluyéndose usted, su [FAMILY RELATIONSHIP FILL].

[IF AT LEAST TWO FAMILY MEMBERS IN ROSTER AND HASJOIN=1]

Las siguientes preguntas se tratan de los tipos de ingreso y las cantidades que reciben [SAMPLE MEMBER] y los miembros de su familia que viven aquí, incluyéndose usted, [FAMILY RELATIONSHIP FILL] de su [SAMPLE MEMBER].

[PROGRAMMER NOTE: THE PROXY SHOULD NOT APPEAR IN [FAMILY RELATIONSHIP FILLS]. ALSO, USE 'otro' AS A MODIFIER TO THE FAMILY RELATIONSHIP FILL WHEN THE RELATIONSHIP TYPE IS EQUAL TO PROXY RELATIONSHIP TYPE AND ONE OF THESE RELATIONSHIP TYPES IS STILL IN THE LIST. PLEASE PRECEDE EACH RELATIONSHIP WITH 'SU'.]

[IF HASJOIN NE 1]Estas preguntas se refieren al año calendario [CURRENT YEAR-1] en vez de los últimos 12 meses a los que se refirieron algunas preguntas anteriores. El año calendario [CURRENT YEAR-1] es del 1 de enero de [CURRENT YEAR-1] hasta el 31 de diciembre del año [CURRENT YEAR-1].

Presione [ENTER] para continuar.

Appendix B: Recruitment Advertisements (English and Spanish)

Adolescents

Ages 12 to 17 Needed for Study

RTI International, a not-for-profit research organization, is looking for adolescents aged 12 to 17 to provide input on questions for a national study on alcohol use, drug use, and other health-related issues. All responses will be kept confidential under federal law. No medical tests or examinations are involved. Requires about 60 minutes. A parent or guardian must accompany the adolescent to the interview. Parents will not observe the interview or find out answers to any questions. The private interview will be conducted at our offices in [LOCATION]. Eligible participants who complete the interview will receive **\$40**.

For eligibility, call:

XXXX

or

1-800-334-8571 ext. XXXX

Adults Aged 65 or Older

Ages 65 and Older Needed for Study

RTI International, a not-for-profit research organization, is looking for persons aged 65 and older to provide input on questions for a national study on alcohol use, drug use, and other health-related issues. All responses will be kept confidential under federal law. No medical tests or examinations are involved. Requires about 60 minutes. The private interview will be conducted at our offices in [LOCATION]. Eligible participants who complete the interview will receive **\$40**.

For eligibility, call:

XXXX

or

1-800-334-8571 ext. XXXX

Non-Native English Speakers

Research Opportunity for Non-native English Speakers

RTI International, a not-for-profit research organization, is looking for adults who speak English as a second language to provide input on questions for a national study on alcohol use, drug use, and other health-related issues. All responses will be kept confidential under federal law. No medical tests or examinations are involved. Requires about 60 minutes. The private interview will be conducted at our offices in [LOCATION]. Eligible participants who complete the interview will receive **\$40**.

For eligibility, call:

XXXX

or

1-800-334-8571 ext. XXXX

Low Education/Low Literacy

Research Opportunity for Qualified Participants

RTI International, a not-for-profit research organization, is looking for respondents to provide input on questions for a national study on alcohol use, drug use, and other health-related issues. We are interested in interviewing adults who are illiterate or have limited reading skills. All responses will be kept confidential under federal law. No medical tests or examinations are involved. Requires about 60 minutes. The private interview will be conducted at our offices near Metro Center. Eligible participants who complete the interview will receive **\$40**.

For eligibility, call:

919-485-7743

or

1-800-334-8571 ext. 27743

Master Advertisement (Print or Online)

Research Opportunity for Qualified Participants

RTI International, a not-for-profit research organization, is looking for respondents to provide input on questions for a national study on alcohol use, drug use, and other health-related issues. We are interested in interviewing adults who are over 65 years old or speak English as a second language. We are also interested in interviewing adolescents aged 12 to 17. All responses will be kept confidential under federal law. No medical tests or examinations are involved. Requires about 60 minutes. The private interview will be conducted at our offices in [LOCATION]. Eligible participants who complete the interview will receive **\$40**.

For eligibility, call:

XXXX

or

1-800-334-8571 ext. XXXX

Spanish Speakers

Oportunidad de participar en un estudio para personas que hablan español

RTI International, una organizacion sin fines de lucro que realiza estudios sobre la salud, está buscando a personas de 18 años de edad o más que hablan español como su idioma principal para dar sus opiniones sobre las preguntas de un estudio nacional sobre el uso de alcohol, el uso de drogas y otros temas relacionados a la salud. Todas las respuestas se mantendrán en forma confidencial de acuerdo a la ley federal. No se hará ninguna prueba ni examen médico. No haremos preguntas sobre la situación legal o de inmigración. La entrevista se realizará en privado en una de nuestras oficinas locales y tomará aproximadamente 60 minutos. Los participantes que sean elegibles y que completen la entrevista recibirán **\$40** dólares.

Para determinar si es elegible, llame al:

XXXX

o al

1-800-334-8571 ext. XXXX

Appendix C: Cognitive Testing Protocol (English and Spanish)

NSDUH Text-to-Speech Cognitive Testing Protocol

CASEID __ - ____ - ____

DATE: __ __ / __ __ / ____

SELECT VERSION

- Group 1 HUMAN, SLOW, MODERATE
- Group 2 MODERATE, HUMAN, SLOW
- Group 3 SLOW, MODERATE, HUMAN

SET UP CASE ON LAPTOP

DEMOGRAPHICS (INTERVIEWER READ)

Thank you for participating in our study. My job is to take a lot of notes and to figure out what we can do to make the questions easier to understand and to determine which voice works best for everyone who takes the survey. If a survey question doesn't make sense or you don't understand a certain word used, tell me that. If you need me to replay any question, just let me know. When we're done, I'll ask you a few overall questions, and then you'll receive \$40 in cash as a token of our appreciation.

First I'll ask a few demographic questions to help us analyze the results of the study.

age How are old are you?

YEARS: _____
DK/REF

QD03 Are you of Hispanic, Latino, or Spanish origin or descent?

- 1 YES
- 2 NO
- DK/REF

QD05 Which of these groups describes you? You may select all that apply.

- 1 White
- 2 Black or African American
- 3 American Indian or Alaska Native
- 4 Native Hawaiian
- 5 Guamanian or Chamorro
- 6 Samoan
- 7 Other Pacific Islander
- 8 Asian
- 9 OTHER (SPECIFY)
- DK/REF

QD11 What is the highest grade or year of school you have **completed**?

INCLUDE JUNIOR OR COMMUNITY COLLEGE ATTENDANCE; DO NOT INCLUDE TECHNICAL SCHOOLS (BEAUTICIAN, MECHANIC, ETC.).

- 0 NO SCHOOLING COMPLETED
- 1 1ST GRADE COMPLETED
- 2 2ND GRADE COMPLETED
- 3 3RD GRADE COMPLETED
- 4 4TH GRADE COMPLETED
- 5 5TH GRADE COMPLETED
- 6 6TH GRADE COMPLETED
- 7 7TH GRADE COMPLETED
- 8 8TH GRADE COMPLETED
- 9 9TH GRADE COMPLETED
- 10 10TH GRADE COMPLETED
- 11 11TH GRADE COMPLETED

- 12 REGULAR HIGH SCHOOL DIPLOMA
- 13 12TH GRADE, NO DIPLOMA

- 14 GED CERTIFICATE OF HIGH SCHOOL COMPLETION
- 15 SOME COLLEGE CREDIT, BUT NO DEGREE
- 16 ASSOCIATE'S DEGREE (FOR EXAMPLE, AA, AS)
- 17 BACHELOR'S DEGREE (FOR EXAMPLE, BA, BS)

- 18 MASTER'S DEGREE (FOR EXAMPLE, MA, MS, MENG, M. ED, MSW, MBA)
- 19 DOCTORATE DEGREE (FOR EXAMPLE, PHD, EDD)
- 20 PROFESSIONAL DEGREE BEYOND A BACHELOR'S DEGREE (FOR EXAMPLE, MD, DDS, DVM, LLB, JD)
- DK/REF

QD14 Were you born in the United States?

- 1 Yes
- 2 No
- DK/REF

QD15 [IF QD14 = NO] In what country or U.S. territory were you born?

COUNTRY OR US TERRITORY: _____
DK/REF

QD55 How well would you say you speak English?

- 1 Very well
- 2 Well
- 3 Not well
- 4 Not at all
- DK/REF

QD56 Are you deaf or do you have any difficulty hearing?

- 1 Yes
- 2 No
- DK/REF

QD57 Are you blind or do you have serious difficulty seeing, even when wearing glasses?

- 1 Yes
- 2 No
- DK/REF

QD58 Do you have any difficulty concentrating, remembering, or making decisions?

- 1 Yes
- 2 No
- DK/REF

INTRODUCTION:

As I described earlier, I will play survey questions that have been recorded using different voices. You do not have to answer the recorded questions. For several of the questions, your task will be to simply repeat the question. After some of the questions, I will instead ask you some follow up items about the question you just heard. For example, I might ask you to try to put the question into your own words. That will help me understand how you have interpreted it. Let me give you an example.

[INTROVOICE1: INTERVIEWER: TYPE '1' AND PRESS ENTER WHEN YOU ARE READY TO PROCEED WITH THE SCREENS FOR VOICE 1.]

HLTH19 During the past 12 months, how many times have you visited a doctor, nurse, physician assistant or nurse practitioner about your **own** health at a doctor's office, a clinic, or some other place?

[INTERVIEWER: ADJUST ACASI VOLUME AS NEEDED.] Putting this in my own words, I might say, "They want to know if I saw a doctor, nurse, or other health care person in the past 12 months for my own health in a doctor's office, clinic, or some place."

Do you have any questions before we begin?

ACASI VOICE 1: _____

LEADCIG These questions are about your use of tobacco products. This includes cigarettes, chewing tobacco, snuff, cigars, and pipe tobacco. The first questions are about cigarettes only.

- P1. Can you tell me what this introduction is telling you?
- P2. What are some of the tobacco products that were listed in this question?

RK01a How much do people risk harming themselves physically and in other ways when they smoke **one or more packs of cigarettes per day**?

- 1 No risk
- 2 Slight risk
- 3 Moderate risk
- 4 Great risk

- P3. In your own words, what is this question asking?

ALCINTR2 These questions are about drinks of alcoholic beverages. Throughout these questions, by a "drink," we mean a can or bottle of beer, a glass of wine or a wine cooler, a shot of liquor, or a mixed drink with liquor in it. We are not asking about times when you only had a sip or two from a drink.

- P4. What would be considered a "drink" of alcohol according to this question? What would not be considered a "drink" of alcohol?

AL01 Have you **ever**, even once, had a drink of any type of alcoholic beverage? Please do not include times when you only had a sip or two from a drink.

- 1 Yes
- 2 No

P5. Please repeat this question as best you can.

MRJINTRO The next questions are about marijuana and hashish. Marijuana is also called pot or grass. Marijuana is usually smoked, either in cigarettes, called joints, or in a pipe. It is sometimes cooked in food. Hashish is a form of marijuana that is also called "hash." It is usually smoked in a pipe. Another form of hashish is hash oil.

P6. Can you tell me in your own words what this introduction is telling you?

P7. Can you recall any of the examples of marijuana or hashish that were mentioned?
[PROBE FOR AS MANY EXAMPLES AS THEY RECALL.]

MJ01 Have you **ever**, even once, used marijuana or hashish?

- 1 Yes
- 2 No

P8. Please repeat this question as best you can.

RK04a How often do you get a real kick out of doing things that are a little dangerous?

- 1 Never
- 2 Seldom
- 3 Sometimes
- 4 Always

P9. Please repeat this question as best you can.

PR01 Please look at the names and pictures of the pain relievers shown below. Please note that some forms of these pain relievers may look different from the pictures, but you should include any form that you have used.

In the **past 12 months**, which, if any, of these pain relievers have you used?

To select more than one drug from the list, press the space bar between each number you have typed. When you have finished, press [ENTER].

- 1 Vicodin

2 Lortab
3 Hydrocodone (generic)
95 I have not used any of these pain relievers in the past 12 months
DK/REF

P10. Based on this question, how would you select more than one drug at a time?

DEBRIEFING1:

The next few questions ask for your opinions about the voice that read the questions you just heard. Please think carefully about this voice as you answer these questions.

1. How would you rate the speed or pace of the voice that read the interview questions? Would you say the pace of the voice was much too slow, a little too slow, just right, a little too fast, or much too fast?

- 1 MUCH TOO SLOW
- 2 A LITTLE TOO SLOW
- 3 JUST RIGHT (NEITHER TOO FAST NOR TOO SLOW)
- 4 A LITTLE TOO FAST
- 5 MUCH TOO FAST

2. Cadence is the way a voice changes by gently rising and falling when speaking. How would you rate the cadence of the voice? Would you say the cadence of the voice was excellent, good, fair, poor, or very poor?

- 1 EXCELLENT
- 2 GOOD
- 3 FAIR
- 4 POOR
- 5 VERY POOR

3. How would you rate the pronunciation of the questions read by the voice? Would you say the pronunciation of the voice was excellent, good, fair, poor, or very poor?

- 1 EXCELLENT
- 2 GOOD
- 3 FAIR
- 4 POOR
- 5 VERY POOR

4. How difficult was it to understand the voice? Would you say not at all difficult, slightly difficult, moderately difficult, very difficult or extremely difficult?

- 1 NOT AT ALL DIFFICULT
- 2 SLIGHTLY DIFFICULT
- 3 MODERATELY DIFFICULT
- 4 VERY DIFFICULT
- 5 EXTREMELY DIFFICULT

5. How pleasant was the voice? Would you say not at all pleasant, somewhat pleasant, moderately pleasant, very pleasant or extremely pleasant?

- 1 NOT AT ALL PLEASANT
- 2 SOMEWHAT PLEASANT
- 3 MODERATELY PLEASANT
- 4 VERY PLEASANT
- 5 EXTREMELY PLEASANT

6. How do you rate the overall quality of the voice? Would you say the quality of the voice was excellent, good, fair, poor, or very poor?

- 1 EXCELLENT
- 2 GOOD
- 3 FAIR
- 4 POOR
- 5 VERY POOR

7. This voice might be used in an interview for about 30 minutes. How comfortable would you be listening to that voice for 30 minutes? Would you say not at all comfortable, slightly comfortable, moderately comfortable, very comfortable or completely comfortable?

- 1 NOT AT ALL COMFORTABLE
- 2 SLIGHTLY COMFORTABLE
- 3 MODERATELY COMFORTABLE
- 4 VERY COMFORTABLE
- 5 COMPLETELY COMFORTABLE

8. You said [ANSWER FROM Q7]. Tell me more about how you chose your answer.

9. Was there anything else about the voice that you liked or disliked?

ACASI VOICE 2: _____

AD12 Think about the times when you were sad, discouraged, or lost interest in most things. Did you ever have a period of time like this that lasted **most of the day, nearly every day, for two weeks or longer?**

- 1 Yes
- 2 No
- DK/REF

P11. Please repeat this question as best you can. [ASK AFTER EACH SENTENCE IN THE QUESTION.]

INHINTRO These next questions are about liquids, sprays, and gases that people sniff or inhale to get high or to make them feel good.

We are not interested in times when you inhaled a substance accidentally — such as when painting, cleaning an oven, or filling a car with gasoline. The questions use the word 'inhalant' to include all the things listed below, as well as any other substances that people sniff or inhale for kicks or to get high. Take a moment to look at the substances listed below so you know what kinds of liquids, sprays, and gases these questions are about.

Amyl nitrite, 'poppers,' locker room odorizers, or 'rush'
Correction fluid, degreaser, or cleaning fluid
Gasoline or lighter fluid
Glue, shoe polish, or toluene
Halothane, ether, or other anesthetics
Lacquer thinner, or other paint solvents
Lighter gases, such as butane or propane
Nitrous oxide or 'whippits'
Felt-tip pens, felt-tip markers, or magic markers
Spray paints
Computer keyboard cleaner, also known as air duster
Other aerosol sprays

P12. According to this introduction, do they want to know about times when substances were inhaled accidentally, on purpose, or both?

IN01c Have you **ever**, even once, inhaled **gasoline or lighter fluid** for kicks or to get high?

- 1 Yes
- 2 No
- DK/REF

P13. Please repeat this question as best you can. [IF NEEDED: Can you remember what time period the question was asking about?]

METHINTRO Methamphetamine, also known as crank, ice, crystal meth, speed, glass, and many other names, is a stimulant that usually comes in crystal or powder forms. It can be smoked, "snorted," swallowed or injected.

P14. What is this introduction telling you?

INTROST These next questions are about **any** use of **prescription stimulants**. People sometimes take these drugs for attention deficit disorders, to lose weight, or to stay awake. Please do **not** include "over-the-counter" stimulants such as Dexatrim, No-Doz, Hydroxycut, or 5-Hour Energy.

P15. Can you tell me in your own words what this introduction is saying?

P16. This introduction said not to include certain types of stimulants. What stimulants should not be included?

INTROSV These next questions ask about **any** use of **prescription sedatives or barbiturates**. These drugs are also called "downers" or "sleeping pills." People take these drugs to help them relax or help them sleep. Please do **not** include "over-the-counter" sedatives such as Sominex, Unisom, Nytol, or Benadryl.

P17. Can you tell me in your own words what this introduction is saying?

P18. According to this introduction, what are some of the reasons that people take these drugs?

PRINTROYR1 Earlier you reported having used certain **prescription pain relievers** during the past year. Now please think about whether you used any of these pain relievers in any way **a doctor did not direct you to use them**.

When you answer these questions, please think only about your use of the drug in any way **a doctor did not direct you to use it**, including:

0. Using it without a prescription of your own
1. Using it in greater amounts, more often, or longer than you were told to take it
2. Using it in **any other way** a doctor did not direct you to use it

P19. This introduction provided examples of using prescription drugs in a way a doctor did not direct you to use. Do you recall what those were? [PROBE FOR ALL THREE RESPONSES]

SV01

Please look at the names and pictures of the sedatives shown below. Please note that some forms of these sedatives may look different from the pictures, but you should include any form that you have used.

In the **past 12 months**, which, if any, of these sedatives have you used?

To select more than one drug from the list, press the space bar between each number you have typed. When you have finished, press [ENTER].

- 1 Ambien
- 2 Ambien CR
- 3 Zolpidem (generic)
- 4 Extended-release zolpidem (generic)
- 95 I have not used any of these sedatives in the past 12 months
- DK/REF

P20. Can you recall what time period the question was asking about?

DEBRIEFING2:

The next few questions ask for your opinions about the voice that read the questions you just heard. Please think carefully about this voice as you answer these questions.

1. How would you rate the speed or pace of the voice that read the interview questions? Would you say the pace of the voice was much too slow, a little too slow, just right, a little too fast, or much too fast?

- 1 MUCH TOO SLOW
- 2 A LITTLE TOO SLOW
- 3 JUST RIGHT (NEITHER TOO FAST NOR TOO SLOW)
- 4 A LITTLE TOO FAST
- 5 MUCH TOO FAST

2. **Cadence is the way a voice changes by gently rising and falling when speaking. How would you rate the cadence of the voice? Would you say the cadence of the voice was excellent, good, fair, poor, or very poor?**

- 1 EXCELLENT
- 2 GOOD
- 3 FAIR
- 4 POOR
- 5 VERY POOR

3. **How would you rate the pronunciation of the questions read by the voice? Would you say the pronunciation of the voice was excellent, good, fair, poor, or very poor?**

- 1 EXCELLENT
- 2 GOOD
- 3 FAIR
- 4 POOR
- 5 VERY POOR

4. **How difficult was it to understand the voice? Would you say not at all difficult, slightly difficult, moderately difficult, very difficult or extremely difficult?**

- 1 NOT AT ALL DIFFICULT
- 2 SLIGHTLY DIFFICULT
- 3 MODERATELY DIFFICULT
- 4 VERY DIFFICULT
- 5 EXTREMELY DIFFICULT

5. **How pleasant was the voice? Would you say not at all pleasant, somewhat pleasant, moderately pleasant, very pleasant or extremely pleasant?**

- 1 NOT AT ALL PLEASANT
- 2 SOMEWHAT PLEASANT
- 3 MODERATELY PLEASANT
- 4 VERY PLEASANT
- 5 EXTREMELY PLEASANT

6. How do you rate the overall quality of the voice? Would you say the quality of the voice was excellent, good, fair, poor, or very poor?

- 1 EXCELLENT
- 2 GOOD
- 3 FAIR
- 4 POOR
- 5 VERY POOR

7. This voice might be used in an interview for about 30 minutes. How comfortable would you be listening to that voice for 30 minutes? Would you say not at all comfortable, slightly comfortable, moderately comfortable, very comfortable or completely comfortable?

- 1 NOT AT ALL COMFORTABLE
- 2 SLIGHTLY COMFORTABLE
- 3 MODERATELY COMFORTABLE
- 4 VERY COMFORTABLE
- 5 COMPLETELY COMFORTABLE

8. You said [ANSWER FROM Q7]. Tell me more about how you chose your answer.

9. Was there anything else about the voice that you liked or disliked?

ACASI Voice 3: _____

SD01 The last questions were about prescription drugs. The next question is about **non-prescription** cough or cold medicines, also known as "**over-the-counter**" medicines. Have you ever, even once, taken a non-prescription cough or cold medicine just to get high?

- 1 Yes
- 2 No
- DK/REF

P21. In your own words, what is this question asking?

CG06 How long has it been since you **last** smoked part or all of a cigarette?

- 1 More than 30 days ago but within the past 12 months
- 2 More than 12 months ago but within the past 3 years
- 3 More than 3 years ago

P22. Can you tell me in your own words what that question was asking? [And what were the response categories provided?]

INTROBK The next questions are about offenses that are against the law. As you read each question, please answer whether you were arrested and booked for that offense during the **past 12 months**.

SP03a In the **past 12 months**, were you arrested and booked for **motor vehicle theft**?

- 1 Yes
- 2 No

P23. In your own words, what is this question asking?

senrelat **During the past 12 months**, how many times did you attend religious services? Please do **not** include special occasions such as weddings, funerals, or other special events in your answer.

- 1 0 times
- 2 1 to 2 times
- 3 3 to 5 times
- 4 6 to 24 times
- 5 25 to 52 times
- 6 More than 52 times

P24. Can you tell me in your own words what that question was asking?

P25. Can you recall what sort of occasions should be excluded when answering this question?

YE09 Have you attended any type of school at any time **during the past 12 months**?

- 1 Yes
- 2 No

P26. Please repeat this question as best you can.

YDS23 Have you ever had a period of time lasting several days or longer when you **lost interest and became bored** with most things you usually enjoy, like work, hobbies, and personal relationships?

- 1 Yes
- 2 No
- DK/REF

P27. In your own words, what is this question asking?

P28. Can you remember what time period the question was asking about?

QD17 The next questions are about school. Are you now attending or are you currently enrolled in school? By "school," we mean an elementary school, a junior high or middle school, a high school, or a college or university. Please include home schooling as well.

- 1 Yes
- 2 No
- DK/REF

P29. In your own words, what is this question asking?

TR01 Please look at the names and pictures of the tranquilizers shown below. Please note that some forms of these tranquilizers may look different from the pictures, but you should include any form that you have used.

In the **past 12 months**, which, if any, of these tranquilizers have you used?

To select more than one drug from the list, press the space bar between each number you have typed. When you have finished, press [ENTER].

- 1 Xanax
- 2 Xanax XR
- 3 Alprazolam (generic)
- 4 Extended-release alprazolam (generic)
- 95 I have not used any of these tranquilizers in the past 12 months
- DK/REF

P30. Can you tell me in your own words what that question was asking?

DEBRIEFING1:

The next few questions ask for your opinions about the voice that read the questions you just heard. Please think carefully about this voice as you answer these questions.

1. How would you rate the speed or pace of the voice that read the interview questions? Would you say the pace of the voice was much too slow, a little too slow, just right, a little too fast, or much too fast?

- 1 MUCH TOO SLOW
- 2 A LITTLE TOO SLOW
- 3 JUST RIGHT (NEITHER TOO FAST NOR TOO SLOW)
- 4 A LITTLE TOO FAST
- 5 MUCH TOO FAST

2. Cadence is the way a voice changes by gently rising and falling when speaking. How would you rate the cadence of the voice? Would you say the cadence of the voice was excellent, good, fair, poor, or very poor?

- 1 EXCELLENT
- 2 GOOD
- 3 FAIR
- 4 POOR
- 5 VERY POOR

3. How would you rate the pronunciation of the questions read by the voice? Would you say the pronunciation of the voice was excellent, good, fair, poor, or very poor?

- 1 EXCELLENT
- 2 GOOD
- 3 FAIR
- 4 POOR
- 5 VERY POOR

4. How difficult was it to understand the voice? Would you say not at all difficult, slightly difficult, moderately difficult, very difficult or extremely difficult?

- 1 NOT AT ALL DIFFICULT
- 2 SLIGHTLY DIFFICULT
- 3 MODERATELY DIFFICULT
- 4 VERY DIFFICULT
- 5 EXTREMELY DIFFICULT

5. How pleasant was the voice? Would you say not at all pleasant, somewhat pleasant, moderately pleasant, very pleasant or extremely pleasant?

- 1 NOT AT ALL PLEASANT
- 2 SOMEWHAT PLEASANT
- 3 MODERATELY PLEASANT
- 4 VERY PLEASANT
- 5 EXTREMELY PLEASANT

6. How do you rate the overall quality of the voice? Would you say the quality of the voice was excellent, good, fair, poor, or very poor?

- 1 EXCELLENT
- 2 GOOD
- 3 FAIR
- 4 POOR
- 5 VERY POOR

7. This voice might be used in an interview for about 30 minutes. How comfortable would you be listening to that voice for 30 minutes? Would you say not at all comfortable, slightly comfortable, moderately comfortable, very comfortable or completely comfortable?

- 1 NOT AT ALL COMFORTABLE
- 2 SLIGHTLY COMFORTABLE
- 3 MODERATELY COMFORTABLE
- 4 VERY COMFORTABLE
- 5 COMPLETELY COMFORTABLE

8. You said [ANSWER FROM Q7]. Tell me more about how you chose your answer.

9. Was there anything else about the voice that you liked or disliked?

FINAL DEBRIEFING (TO BE READ AFTER ALL THREE VOICES)

1. Of the three voices you heard, which voice did you prefer most?

- 1 HUMAN VOICE (VOICE ___)
- 2 SLOWER COMPUTERIZED VOICE (VOICE ___)
- 3 FASTER COMPUTERIZED VOICE (VOICE ___)
- 4 NO PREFERENCE

2. Tell me more about why you prefer that voice.

**NSDUH Text-to-Speech Cognitive Testing Protocol
SPANISH VERSION**

CASEID __ - ____ - ____

DATE: ____ / ____ / ____

SELECT VERSION

Group 1	HUMAN, SLOW, MODERATE
Group 2	MODERATE, HUMAN, SLOW
Group 3	SLOW, MODERATE, HUMAN

SET UP CASE ON LAPTOP

DEMOGRAPHICS (INTERVIEWER READ) [NO RECORDING]

Gracias por participar en nuestro estudio. Mi trabajo es tomar muchas notas y averiguar lo que podamos hacer para que las preguntas del cuestionario sean fáciles de entender y para determinar qué tipo de voz se comprende mejor para todas las personas que contestan la encuesta. Si una pregunta de la encuesta no tiene sentido o usted no entiende cierta palabra que se usó, entonces dígamelo por favor. Si necesita volver a escuchar cualquier pregunta, simplemente hágamelo saber. Cuando acabemos, le haré unas preguntas en general y entonces usted recibirá \$40 dólares en efectivo como muestra de nuestro agradecimiento.

Primero, le haré algunas preguntas demográficas para ayudarnos a analizar los resultados del estudio.

age ¿Qué edad tiene usted?

YEARS: _____

DK/REF

QD03 ¿Es usted de origen o descendencia hispana, latina o española?

1 YES

2 NO

DK/REF

QD05 ¿Cuál de estos grupos lo describe mejor a usted? Puede seleccionar todos los que correspondan.

- 1 Blanco
- 2 Negro o afroamericano
- 3 Indígena americano o nativo de Alaska
- 4 Nativo de Hawaii
- 5 Guameño o Chamorro
- 6 Samoano
- 7 De otra isla del Pacífico
- 8 Asiático
- 9 OTRO (ESPECIFIQUE)
- DK/REF

QD11 . ¿Cuál es el último grado o año de escuela o universidad que usted ha **completado**?

INCLUYA LA ASISTENCIA A UN 'JUNIOR' O 'COMMUNITY COLLEGE'; NO INCLUYA LAS ESCUELAS TECNICAS (COSMETOLOGÍA, MECÁNICA, ETC.)

- 12 NO COMPLETÓ NINGÚN GRADO EN LA ESCUELA
- 13 COMPLETÓ EL 1er. GRADO
- 14 COMPLETÓ EL 2o. GRADO
- 15 COMPLETÓ EL 3er. GRADO
- 16 COMPLETÓ EL 4o. GRADO
- 17 COMPLETÓ EL 5o. GRADO
- 18 COMPLETÓ EL 6o. GRADO
- 19 COMPLETÓ EL 7o. GRADO
- 20 COMPLETÓ EL 8o. GRADO
- 21 COMPLETÓ EL 9o. GRADO
- 22 COMPLETÓ EL 10o. GRADO
- 23 COMPLETÓ EL 11o. GRADO
- 12 DIPLOMA DE ESCUELA PREPARATORIA O 'HIGH SCHOOL'
- 13 GRADO 12, SIN DIPLOMA
- 14 CERTIFICADO 'GED' POR COMPLETAR 'HIGH SCHOOL'
- 15 ALGUNOS CRÉDITOS UNIVERSITARIOS, PERO SIN TÍTULO
- 16 TÍTULO ASOCIADO UNIVERSITARIO (POR EJEMPLO, AA, AS)
- 17 TÍTULO DE LICENCIATURA UNIVERSITARIA (POR EJEMPLO, BA, BS)
- 18 TÍTULO DE MAESTRÍA (POR EJEMPLO, MA, MS, MENG, M. ED, MSW, MBA)
- 21 TÍTULO DE DOCTORADO (POR EJEMPLO, PHD, EDD)
- 22 TÍTULO PROFESIONAL MÁS ALLÁ DE UN TÍTULO DE LICENCIATURA UNIVERSITARIA (POR EJEMPLO, MD, DDS, DVM, LLB, JD)
- DK/REF

QD14 ¿Nació usted en los Estados Unidos?

- 1 Sí
- 2 No
- DK/REF

QD15 [IF QD14 = NO] ¿En qué país o en qué territorio de los Estados Unidos nació usted?

COUNTRY OR US TERRITORY: _____
DK/REF

QD55 ¿Qué tan bien habla inglés?

- 1 Muy bien
- 2 Bien
- 3 No lo hablo bien
- 4 No lo hablo en lo absoluto

QD56 ¿Es usted sordo o tiene mucha dificultad para oír?

- 1 Sí
- 2 No
- DK/REF

QD57 ¿Es usted ciego o tiene mucha dificultad para ver, aún cuando usa lentes o anteojos?

- 1 Sí
- 2 No
- DK/REF

QD58 Debido a una condición física, mental o emocional, ¿tiene usted mucha dificultad para concentrarse, recordar o tomar decisiones?

- 1 Sí
- 2 No
- DK/REF

Gracias.

[SET UP AUDIO RECORDER IF RESPONDENT HAS AGREED TO HAVE THE INTERVIEW RECORDED]

INTRODUCTION:

Como le expliqué anteriormente, yo haré que usted escuche el audio de las preguntas de la encuesta que han sido grabadas usando diferentes voces. Usted no tiene que contestar las preguntas grabadas. Para varias preguntas, su tarea será simplemente repetir la pregunta. Después de algunas de las preguntas, yo le haré algunas preguntas de seguimiento acerca de la pregunta que acaba de escuchar. Por ejemplo, pueda ser que le pida que trate de hacer la pregunta en sus propias palabras. Eso me ayudará a comprender cómo interpretó la pregunta. Permítame darle un ejemplo.

[INTROVOICE1: INTERVIEWER: TYPE '1' AND PRESS ENTER WHEN YOU ARE READY TO PROCEED WITH THE SCREENS FOR VOICE 1.]

HLTH19 Durante los últimos 12 meses, ¿cuántas veces ha ido a ver a un doctor, una enfermera, un asistente médico o una enfermera especializada acerca de su propia salud en el consultorio de un doctor, una clínica o algún otro lugar?

[INTERVIEWER: ADJUST ACASI VOLUME AS NEEDED.] Usando mis propias palabras, yo pudiera decir, "Ellos quieren saber si yo ví a un doctor, una enfermera u otra persona que da atención médica en los últimos 12 meses para mi propia salud en el consultorio de un doctor, una clínica o algún otro lugar".

ACASI VOICE 1:

LEADCIG Estas preguntas se tratan del uso de productos de tabaco. Esto incluye cigarrillos, tabaco de mascar, tabaco en polvo (rapé o "snuff"), cigarros (puros) y tabaco en pipa. Las primeras preguntas se tratan solamente de cigarrillos.

P31. ¿Me puede decir que le están diciendo en esta introducción?

P32. ¿Cuáles son algunos de los productos de tabaco que se mencionaron en esta pregunta?

RK01a ¿Cuánto riesgo corren las personas de hacerse daño físico y de otro tipo cuando fuman **una o más cajetillas de cigarrillos al día?**

- 1 Ningún riesgo
- 2 Poco riesgo
- 3 Riesgo moderado
- 4 Mucho riesgo

P33. En sus propias palabras, ¿qué le está diciendo esta pregunta?

ALCINTR2 Estas preguntas se tratan del consumo de las bebidas alcohólicas. Para las siguientes preguntas, una 'bebida' quiere decir una lata o botella de cerveza, una copa de vino o 'wine cooler,' un trago de alcohol o un coctel que contiene alcohol. No queremos saber de las ocasiones en que usted haya tomado solo uno o dos sorbos de una bebida.

P34. ¿Qué es lo que consideraría usted como una "bebida" alcohólica de acuerdo a esta pregunta? ¿Qué es lo que no consideraría una "bebida" alcohólica?

AL01 ¿**Alguna vez** ha tomado una bebida alcohólica, aunque haya sido solo una vez? Por favor no incluya las ocasiones en que usted haya tomado solo uno o dos sorbos de una bebida.

- 1 Sí
- 2 No

P35. Por favor, repita esta pregunta de la mejor manera que pueda.

MRJINTRO Las siguientes preguntas se tratan de marihuana y hachís. Marihuana también se llama mota, pasto y hierba. Usualmente, la marihuana se fuma en forma de cigarrillos llamados pitillos o 'joints' o también en una pipa. Algunas veces se cocina en alimentos. El hachís es una forma de marihuana que también se llama 'hash.' Usualmente, se fuma en una pipa. Aceite de hachís ('hash oil') es otra forma de hachís.

P36. ¿Me puede decir en sus propias palabras qué es lo que le está diciendo esta introducción?

P37. ¿Puede recordar alguno de los ejemplos de marihuana o hachís que se mencionaron?
[PROBE FOR AS MANY EXAMPLES AS THEY RECALL.]

MJ01 ¿Alguna vez ha usado marihuana o hachís, aunque haya sido solo una vez?

- 1 Yes
- 2 No

P38. Por favor, repita esta pregunta de la mejor manera que pueda.

RK04a ¿Con qué frecuencia le da placer hacer cosas que son un poco peligrosas?

- 1 Nunca
- 2 Rara vez
- 3 Algunas veces
- 4 Siempre

P39. Por favor, repita esta pregunta de la mejor manera que pueda.

PR01 Por favor, mire los nombres y las fotos de los analgésicos que se muestran a continuación. Por favor, tome en cuenta que la forma de algunos analgésicos puede parecer diferente a la de las fotos, pero usted debe incluir los analgésicos que haya usado aunque hayan tenido otra forma.

En los **últimos 12 meses**, ¿cuál de estos analgésicos ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

- 1 Vicodin
- 2 Lortab
- 3 Hidrocodona (genérico)
- 95 No he usado ninguno de estos analgésicos en los últimos 12 meses
- DK/REF

P40. De acuerdo con esta pregunta, ¿cómo seleccionaría más de un medicamento a la vez?

DEBRIEFING1:

Las siguientes preguntas le piden sus opiniones acerca de la voz que leyó las preguntas que acaba de escuchar. Por favor, piense cuidadosamente acerca de esta voz a medida que responde estas preguntas.

10. ¿Cómo calificaría la velocidad de la voz que leyó las preguntas de la entrevista? ¿Diría usted que la velocidad de la voz era demasiada lenta, un poco lenta, tenía la velocidad adecuada, era un poco rápida o demasiada rápida?

- DEMASIADA LENTA
- UN POCO LENTA
- VELOCIDAD ADECUADA (NI MUY RÁPIDA NI MUY LENTA)
- UN POCO RÁPIDA
- DEMASIADO RÁPIDA

11. Cadencia es la manera en que una voz cambia al subir y bajar de tono suavemente al hablar. ¿Cómo calificaría la cadencia de la voz? ¿Diría que la cadencia de la voz fue excelente, muy buena, buena, regular, mala o muy mala?

- EXCELENTE
- BUENA
- REGULAR
- MALA
- MUY MALA

12. ¿Cómo calificaría la pronunciación de las preguntas que leyó la voz? ¿Diría que la pronunciación fue excelente, buena, regular, mala o muy mala?

- EXCELENTE
- BUENA
- REGULAR
- MALA
- MUY MALA

13. ¿Qué tan difícil era entender lo que decía la voz? ¿Diría que nada difícil, un poco difícil, moderadamente difícil, muy difícil o sumamente difícil de entender?

- NADA DIFÍCIL
- UN POCO DIFÍCIL
- MODERADAMENTE DIFÍCIL
- MUY DIFÍCIL
- SUMAMENTE DIFÍCIL

14. ¿Qué tan agradable era la voz? ¿Diría que nada agradable, un poco agradable, moderadamente agradable, muy agradable o sumamente agradable?

- NADA AGRADABLE
- UN POCO AGRADABLE
- MODERADAMENTE AGRADABLE
- MUY AGRADABLE
- SUMAMENTE AGRADABLE

15. ¿Cómo calificaría la calidad de la voz en general? ¿Diría que la calidad de la voz era excelente, buena, regular, mala o muy mala?

- EXCELENTE
- BUENA
- REGULAR
- MALA
- MUY MALA

16. Puede que esta voz se use en una entrevista de aproximadamente 30 minutos. ¿Qué tan a gusto se sentiría usted escuchando esa voz por 30 minutos? ¿Diría que nada a gusto, un poco a gusto, moderadamente a gusto, muy a gusto o sumamente a gusto?

- NADA A GUSTO
- UN POCO A GUSTO
- MODERADAMENTE A GUSTO
- MUY A GUSTO
- SUMAMENTE A GUSTO

17. Uste dijo [ANSWER FROM Q7]. Hábleme más sobre cómo decidió esa respuesta.

18. ¿Hubo algo más acerca de la voz que le gustó o que no le gustó?

ACASI VOICE 2:

AD12 Piense en la veces cuando se sintió triste, desanimado o perdió interés en la mayoría de las cosas. ¿Pasó alguna vez por un periodo de tiempo como este, el cual duró **la mayor parte del día, casi todos los días, por dos semanas o más?**

- 1 Sí
- 2 No
- DK/REF

P41. Por favor, repita esta pregunta de la mejor manera que pueda. [ASK AFTER EACH SENTENCE IN THE QUESTION.]

INHINTRO Las siguientes preguntas son acerca de líquidos, aerosoles o esprays y gases que las personas aspiran o inhalan para drogarse o para sentirse alegres.

No estamos interesados en ocasiones en que usted inhaló alguna sustancia accidentalmente como en el caso de pintar, limpiar un horno o echarle gasolina al automóvil. Las preguntas usan el término 'inhalante' para incluir todas las cosas mencionadas a continuación, así como cualquier otra sustancia que las personas aspiran o inhalan para divertirse o para drogarse. Por favor mire con atención la lista de sustancias a continuación, para saber a qué clases de líquidos, aerosoles o esprays y gases se refieren las próximas preguntas.

- Nitrato de amilo, 'bombitas,' desodorante ambiental, o 'rush'
- Líquido de corrección o 'liquid paper', desengrasador o líquido de limpieza
- Gasolina o líquido para encendedores
- Pegamento, crema o betún para limpiar zapatos, o tolueno
- Halotano, éter u otros anestésicos
- 'Tiner' u otros solventes para pintura
- Gases para encendedores, tales como butano o propano
- Óxido nítrico o 'whippits'
- Marcadores de punta fina, plumones o plumones mágicos
- Pintura en aerosol
- Limpiador para teclado de computadora, también llamado aire comprimido
- removedor de polvo
- Otros aerosoles o esprays

P42. De acuerdo con esta introducción, ¿quieren saber ellos acerca de las veces que las sustancias se inhalaron en forma accidental, a propósito o ambos?

IN01c ¿Alguna vez ha inhalado **gasolina o líquido para encendedores** para divertirse o para drogarse, aunque haya sido solo una vez?

- 1 Sí
- 2 No
- DK/REF

P43. Por favor, repita esta pregunta de la mejor manera que pueda. [IF NEEDED: ¿Puede recordar a qué periodo de tiempo se estaba refiriendo la pregunta?]

METHINTRO La metanfetamina, también conocida como "arranque" o "crank", "hielo", "cristal", "crystal meth", "velocidad" o "speed", "vidrio" y muchos otros nombres, es un estimulante que normalmente consiste en pedazos de cristales o se presenta en forma de polvo. La metanfetamina se puede fumar, inhalar por la nariz, tomar por vía oral o se puede inyectar.

P44. ¿Qué le están diciendo en esta introducción?

INTROST Las siguientes preguntas se refieren a **cualquier** uso de **estimulantes que se normalmente se venden con una receta médica**. Algunas veces las personas toman estos medicamentos para trastornos por déficit de la atención, para bajar de peso o permanecer despiertas. Por favor **no** incluya estimulantes de "venta libre" tales como Dexatrim, No-Doz, Hydroxycut o 5-Hour Energy.

P45. ¿Me puede decir en sus propias palabras lo que está diciendo esta introducción?

P46. Esta introducción dijo que no se incluyan ciertos tipos de estimulantes. ¿Qué estimulantes no se deberían incluir?

INTROSV Las siguientes preguntas se refieren a **cualquier** uso de **sedantes o barbitúricos que normalmente se venden con una receta médica**. Estos medicamentos también se llaman "downers" o "pastillas para dormir". Las personas toman estos medicamentos para poder relajarse o para poder dormir. Por favor **no** incluya sedantes de "venta libre" tales como Sominex, Unisom, Nytol o Benadryl.

P47. ¿Me puede decir en sus propias palabras qué le está diciendo esta introducción?

P48. De acuerdo a esta introducción, ¿cuáles son algunas de las razones por las que las personas toman estos medicamentos?

PRINTROYR1 Anteriormente, usted reportó haber usado ciertos **analgésicos que normalmente se venden con una receta médica** durante los últimos 12 meses. Ahora, por favor piense si usted usó alguno de estos analgésicos de alguna manera **que un doctor no le haya indicado**.

Cuando responda estas preguntas, por favor, piense solamente en el uso del medicamento de alguna manera **que un doctor no le haya indicado**, incluyendo:

3. Usarlo sin tener su propia receta médica
4. Usarlo en mayor cantidad, con más frecuencia o durante más tiempo del que le dijeron
5. Usarlo de **alguna otra manera** que un doctor no le haya indicado

P49. Esta introducción le dio ejemplos de cómo usar medicamentos que normalmente se venden con una receta médica de una manera que un doctor no le haya indicado. ¿Recuerda cuáles fueron esas maneras? [PROBE FOR ALL THREE RESPONSES]

SV01 Por favor mire los nombres y las fotos de los sedantes que se muestran a continuación. Por favor, tome en cuenta que la forma de algunos sedantes puede ser diferente a la de las fotos, pero usted debe incluir los sedantes que haya tomado aunque hayan tenido otra forma.

En los últimos 12 meses, ¿cuál de estos sedantes ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

- 5 Ambien
 - 6 Ambien CR
 - 7 Zolpidem (genérico)
 - 8 Zolpidem de liberación prolongada (genérico)
 - 95 No he usado ninguno de estos sedantes en los últimos 12 meses
- DK/REF

P50. ¿Puede recordar a qué periodo de tiempo se estaba refiriendo esta pregunta?

DEBRIEFING2:

Las siguientes preguntas le piden sus opiniones acerca de la voz que leyó las preguntas que acaba de escuchar. Por favor, piense cuidadosamente acerca de esta voz a medida que responde estas preguntas.

1. **¿Cómo calificaría la velocidad de la voz que leyó las preguntas de la entrevista? ¿Diría usted que la velocidad de la voz era demasiada lenta, un poco lenta, tenía la velocidad adecuada, era un poco rápida o demasiada rápida?**

- DEMASIADA LENTA
- UN POCO LENTA
- VELOCIDAD ADECUADA (NI MUY RÁPIDA NI MUY LENTA)
- UN POCO RÁPIDA
- DEMASIADO RÁPIDA

2. **Cadencia es la manera en que una voz cambia al subir y bajar de tono suavemente al hablar. ¿Cómo calificaría la cadencia de la voz? ¿Diría que la cadencia de la voz fue excelente, muy buena, buena, regular, mala o muy mala?**

- EXCELENTE
- BUENA
- REGULAR
- MALA
- MUY MALA

3. **¿Cómo calificaría la pronunciación de las preguntas que leyó la voz? ¿Diría que la pronunciación fue excelente, buena, regular, mala o muy mala?**

- EXCELENTE
- BUENA
- REGULAR
- MALA
- MUY MALA

4. **¿Qué tan difícil era entender lo que decía la voz? ¿Diría que nada difícil, un poco difícil, moderadamente difícil, muy difícil o sumamente difícil de entender?**

- NADA DIFÍCIL
- UN POCO DIFÍCIL
- MODERADAMENTE DIFÍCIL
- MUY DIFÍCIL
- SUMAMENTE DIFÍCIL

5. **¿Qué tan agradable era la voz? ¿Diría que nada agradable, un poco agradable, moderadamente agradable, muy agradable o sumamente agradable?**

- NADA AGRADABLE
- UN POCO AGRADABLE
- MODERADAMENTE AGRADABLE
- MUY AGRADABLE
- SUMAMENTE AGRADABLE

6. **¿Cómo calificaría la calidad de la voz en general? ¿Diría que la calidad de la voz era excelente, buena, regular, mala o muy mala?**

- EXCELENTE
- BUENA
- REGULAR
- MALA
- MUY MALA

7. **Puede que esta voz se use en una entrevista de aproximadamente 30 minutos. ¿Qué tan a gusto se sentiría usted escuchando esa voz por 30 minutos? ¿Diría que nada a gusto, un poco a gusto, moderadamente a gusto, muy a gusto o sumamente a gusto?**

- NADA A GUSTO
- UN POCO A GUSTO
- MODERADAMENTE A GUSTO
- MUY A GUSTO
- SUMAMENTE A GUSTO

8. **Uste dijo [ANSWER FROM Q7]. Hábleme más sobre cómo decidió esa respuesta.**

9. ¿Hubo algo más acerca de la voz que le gustó o que no le gustó?

ACASI Voice 3

SD01 Las últimas preguntas fueron acerca de medicamentos que normalmente se venden con una receta médica. La siguiente pregunta es acerca de medicamentos para la tos o el resfrío **que se venden sin receta médica**, también conocidos como medicamentos "**de venta libre**". ¿Alguna vez tomó un medicamento para la tos o el resfrío que se vende sin una receta médica, solo para drogarse, aunque sea solo una vez?

- 3 Sí
- 4 No
- DK/REF

P51. En sus propias palabras, ¿qué le están pidiendo en esta pregunta?

CG06 ¿Cuánto tiempo hace desde la **última** vez que fumó un cigarrillo entero o parte de uno?

- 1 Hace más de 30 días pero dentro de los últimos 12 meses
- 2 Hace más de 12 meses pero dentro de los últimos 3 años
- 3 Hace más de 3 años

P52. ¿Me puede decir en sus propias palabras qué le estaba diciendo esa pregunta? [¿Y cuáles fueron las opciones de respuestas que se proporcionaron?]

INTROBK Las siguientes preguntas se tratan de delitos contra la ley. Al leer cada pregunta, por favor conteste si lo arrestaron y ficharon por ese delito en los **últimos 12 meses**.

SP03a En los **últimos 12 meses**, ¿lo arrestaron y ficharon por robar un vehículo?

- 1 Sí
- 2 No

[INTERVIEWER: REPEAT THE PROBE QUESTION FOR EACH SCREEN]

P53. En sus propias palabras, ¿qué le están diciendo en esta pregunta?

INTROBK: _____

SP03a: _____

senrelat **En los últimos 12 meses**, ¿cuántas veces fue a servicios religiosos? Por favor **no** incluya en su respuesta ocasiones especiales tales como matrimonios, entierros u otros eventos especiales.

- 1 Ninguna vez
- 2 1 a 2 veces
- 3 3 a 5 veces
- 4 6 a 24 veces
- 5 25 a 52 veces
- 6 Más de 52 veces

P54. ¿Me puede decir en sus propias palabras qué le están diciendo en esta pregunta?

P55. ¿Puede recordar qué tipo de ocasiones no deberían incluirse al responder esta pregunta?

YE09 ¿Has estado inscrito o matriculado en algún tipo de escuela en algún momento **durante los últimos 12 meses?**

- 1 Sí
- 2 No

P56. Por favor, repita esta pregunta de la mejor manera que pueda.

YDS23 ¿Alguna vez pasaste por un periodo de tiempo, el cual duró varios días o más, cuando **perdiste el interés y te sentiste aburrido** de la mayoría de las cosas que generalmente disfrutas hacer, como tu trabajo, tus pasatiempos y tus relaciones con otras personas?

- 1 Sí
- 2 No
- DK/REF

P57. En sus propias palabras, ¿qué le están diciendo en esta pregunta?

P58. ¿Puede recordar a qué periodo de tiempo se estaba refiriendo esta pregunta?

QD17 Las siguientes preguntas se tratan de la escuela. ¿Actualmente asiste a la escuela o está inscrito o matriculado en la escuela? Por 'escuela', nos referimos a la escuela primaria, media ('junior high' o 'middle school'), preparatoria o 'high school', o un 'college' o universidad. Por favor incluya también el programa de educación en el hogar llamado 'home schooling'.

- 1 Sí
- 2 No
- DK/REF

P59. En sus propias palabras, ¿qué le están diciendo en esta pregunta?

TR01 Por favor mire los nombres y las fotos de los tranquilizantes que se muestran a continuación. Por favor, tome en cuenta que la forma de algunos tranquilizantes puede parecer diferente a la de las fotos, pero usted debe incluir los tranquilizantes que haya usado aunque hayan tenido otra forma.

En los **últimos 12 meses**, ¿cuál de estos tranquilizantes ha usado, si es que ha usado alguno?

Para seleccionar más de un medicamento en la lista, presione la barra espaciadora entre cada número que haya registrado. Cuando haya terminado, presione la tecla [ENTER].

- 4 Xanax
- 5 Xanax XR
- 6 Alprazolam (genérico)
- 4 Alprazolam de liberación prolongada (genérico)
- 95 No he usado ninguno de estos tranquilizantes en los últimos 12 meses
- DK/REF

P60. ¿Me puede decir en sus propias palabras qué le están diciendo en esta pregunta?

DEBRIEFING3:

Las siguientes preguntas le piden sus opiniones acerca de la voz que leyó las preguntas que acaba de escuchar. Por favor, piense cuidadosamente acerca de esta voz a medida que responde estas preguntas.

1. **¿Cómo calificaría la velocidad de la voz que leyó las preguntas de la entrevista? ¿Diría usted que la velocidad de la voz era demasiada lenta, un poco lenta, tenía la velocidad adecuada, era un poco rápida o demasiada rápida?**

- DEMASIADA LENTA
- UN POCO LENTA
- VELOCIDAD ADECUADA (NI MUY RÁPIDA NI MUY LENTA)
- UN POCO RÁPIDA
- DEMASIADO RÁPIDA

2. **Cadencia es la manera en que una voz cambia al subir y bajar de tono suavemente al hablar. ¿Cómo calificaría la cadencia de la voz? ¿Diría que la cadencia de la voz fue excelente, muy buena, buena, regular, mala o muy mala?**

- EXCELENTE
- BUENA
- REGULAR
- MALA
- MUY MALA

3. **¿Cómo calificaría la pronunciación de las preguntas que leyó la voz? ¿Diría que la pronunciación fue excelente, buena, regular, mala o muy mala?**

- EXCELENTE
- BUENA
- REGULAR
- MALA
- MUY MALA

4. **¿Qué tan difícil era entender lo que decía la voz? ¿Diría que nada difícil, un poco difícil, moderadamente difícil, muy difícil o sumamente difícil de entender?**

- NADA DIFÍCIL
- UN POCO DIFÍCIL
- MODERADAMENTE DIFÍCIL
- MUY DIFÍCIL
- SUMAMENTE DIFÍCIL

5. **¿Qué tan agradable era la voz? ¿Diría que nada agradable, un poco agradable, moderadamente agradable, muy agradable o sumamente agradable?**

- NADA AGRADABLE
- UN POCO AGRADABLE
- MODERADAMENTE AGRADABLE
- MUY AGRADABLE
- SUMAMENTE AGRADABLE

6. **¿Cómo calificaría la calidad de la voz en general? ¿Diría que la calidad de la voz era excelente, buena, regular, mala o muy mala?**

- EXCELENTE
- BUENA
- REGULAR
- MALA
- MUY MALA

7. **Puede que esta voz se use en una entrevista de aproximadamente 30 minutos. ¿Qué tan a gusto se sentiría usted escuchando esa voz por 30 minutos? ¿Diría que nada a gusto, un poco a gusto, moderadamente a gusto, muy a gusto o sumamente a gusto?**

- NADA A GUSTO
- UN POCO A GUSTO
- MODERADAMENTE A GUSTO
- MUY A GUSTO
- SUMAMENTE A GUSTO

8. **Uste dijo [ANSWER FROM Q7]. Hábleme más sobre cómo decidió esa respuesta.**

9. **¿Hubo algo más acerca de la voz que le gustó o que no le gustó?**

FINAL DEBRIEFING (TO BE READ AFTER ALL THREE VOICES)

1. De las tres voces que escuchó, ¿qué voz prefirió mejor?

- VOZ HUMANA (VOZ ___)
- VOZ COMPUTARIZADA MÁS LENTA (VOZ ___)
- VOZ COMPUTARIZADA MÁS RÁPIDA (VOZ ___)
- NO TENGO PREFERENCIA

2. Hábleme más sobre por qué prefirió esa voz.

Appendix D: Informed Consent Forms (English and Spanish)

Text to Speech Testing
Adult Cognitive Interview Participant Informed Consent Form
National Survey on Drug Use and Health (NSDUH)

Introduction

I am going to explain this study to you. You can stop me at any time if you have questions about anything I tell you.

The purpose of this study is to test some questions that will be used in the National Survey on Drug Use and Health, or NSDUH. The NSDUH is a large survey given to about 70,000 people across the country each year. It collects information on many health-related issues. The aim is to better serve all people throughout the United States. Right now, we're interested in evaluating the voice that will read some questions in the study. We want to see how well people understand these questions and how they might go about answering them. RTI is carrying out this research study for the Substance Abuse and Mental Health Services Administration, or SAMHSA, which is part of the US Department of Health and Human Services. You are one of 36 participants at least 12 years old (including about 24 adults) who will review the survey questions for this study.

Description of the Interview

Your participation in this interview will involve listening to survey questions being played using different voices. The survey includes questions about the use of tobacco, alcohol, drugs such as marijuana, and other health issues. However, I will not be asking you to answer these questions. Instead, I will ask you follow up questions to determine whether the voices used to read the questions were clear and easy for you to understand. For some questions, I may ask you to put the questions in your own words. The interview will last approximately 60 minutes. Your participation in this study will end after you finish the interview.

We also would like to audio record what you say during the interview. Only the people who work on this study will hear the recording. It will help us make sure we have understood your answers. If you don't want us to audio record you, that's okay.

Confidentiality/Your Rights

Taking part in the interview is completely voluntary. You can skip any interview questions you do not wish to answer. Your personal information will not be connected to your answers in any way. Federal law requires us to keep your answers confidential and to use these answers only for statistical purposes (the Confidential Information Protection and Statistical Efficiency Act of 2002). With your agreement, we will audio record your interview. You can ask us to pause or stop the recording at any time. Only RTI and SAMHSA research team members will be able to listen to the recordings. The recordings will be destroyed within 60 days of the end of this study. Comments from all interviews will be combined in a report that will not identify who made the comments.

[Read only if observer is present: A member of the RTI research team or representative(s) of SAMHSA is here with us today and would like to observe this interview from a separate observation room. If you do not want anyone else to observe your interview, we will simply ask this person (these people) to leave the observation room and then do the interview.]

Possible Risks and Benefits

You can ask me to stop the interview at any time. If you want to take a break at any time during the interview, please tell me. It is possible some of the survey questions may make you feel uncomfortable or upset. If this happens, I can tell you how to contact a counselor.

There are no direct benefits to you from participating in this interview. However the answers you give will help us to improve the quality of questions for the NSDUH.

Payment for Participation

You will be given \$40 in cash for completing the interview.

Your Questions

If you have any other questions about the study, you can call Ms. Emily Geisen at 1-800-334-8571 ext. 26566. If you have any questions about your rights as a study participant, you can call RTI's Office of Research Protection at 1-866-214-2043 (a toll-free number).

I will sign my name here to indicate that I have explained this information to you and that you have agreed to be interviewed.

Signature of Interviewer

Date

Read only if observer is present: I also will sign my name here to indicate that you have given your consent for a member of the RTI research team or representative of SAMHSA to observe the interview. [INTERVIEWER, PLEASE WRITE "NA" ON THE SIGNATURE LINE IF THE INTERVIEW IS NOT BEING OBSERVED.]

Signature of Interviewer

Date

Finally, I will sign my name here to indicate that you have agreed for the interview to be audio recorded.

Signature of Interviewer

Date

Parental Permission and Informed Consent

The National Survey on Drug Use and Health is a large survey given to about 70,000 people across the country every year. RTI International conducts the National Survey on Drug Use and Health. It collects information on many health-related issues. We ask about a lot of health issues, so that we can better help everyone in the United States. Right now we're interested in testing a new computerized voice that will read some questions in the study. Before we do this, we want to see how well people understand these questions and how they might go about answering them. We are under contract with the Substance Abuse and Mental Health Services Administration to carry out this survey. You or your child responded to an advertisement that we placed for research subjects. At present, we are seeking the help of young people like your child to see how our new questions work.

Your child is one of six adolescent respondents in Washington, DC, and Research Triangle Park, NC, who are participating in this study. Taking part in the interview is strictly voluntary. Your child can skip any portion of the interview he/she does not wish to be involved with. There is no penalty if he/she chooses to skip any part of the interview. The interview will be conducted in private to ensure nobody else overhears his/her answers. All answers will be kept private and confidential. We will not share the information given to us with any person outside the project staff, and your child's name will never be connected to the answers he/she provides. Federal law requires us to keep your child's answers confidential and to use his/her answers only for statistical purposes (the Confidential Information Protection and Statistical Efficiency Act of 2002). The only exception to this promise of confidentiality is if your child tells me that he/she intends to seriously harm him/herself or someone else or if he/she has been abused or if your child identifies a person who has given him/her drugs; in this situation I may need to notify a mental health professional or other authorities.

The interview will take about one hour. During the interview, your child will listen to survey questions being played using different voices. The survey includes questions about the use of tobacco, alcohol, drugs such as marijuana, and other health issues. However, we will not be asking your child to answer these questions. Instead, we will ask follow up items about the survey questions to determine whether the voice used to read the questions was clear and easy to understand. For example, we may ask your child to repeat the question in his or her own words.

He/She will receive \$40 in cash in appreciation for the interview.

We would like to audio record the interactions between your child and the interviewer. The recording will be heard only by members of the research team to help us make sure we have all the information from your child about how these questions work. To protect his/her privacy, the recording will remain on the laptop computer, which will be protected by a password. The recording will be destroyed soon after the study ends. However, having the interactions recorded is voluntary and you can decline for your child.

If you have any questions about this study, you can contact Emily Geisen at RTI at 1-800-334-8571 X. 26566. If you have any questions about your rights as a parent or legal guardian or your child's rights as a study participant, you can call RTI's Office of Research Protection at 1-866-214-2043 (a toll-free number).

Do we have your permission for [CHILD'S NAME] to participate?

As Parent/Guardian, I give my permission for my child to participate in this interview.

Yes **No**

As Parent/Guardian, I give my permission for my child's interview to be audio recorded:

Yes **No**

Signature of Interviewer: _____

Date: _____

Participant Informed Assent (ADOLESCENT)

Introduction

I am going to explain this study to you. You can stop me at any time if you have questions about anything I tell you.

The purpose of this study is to test some questions that will be used in the National Survey on Drug Use and Health, or NSDUH. The NSDUH is a large survey given to about 70,000 people across the country each year. It collects information on many health-related issues, to better help everyone in the United States. We're interested in evaluating the voice that will read some questions in the study. We want to see how well people understand these questions. We also want to know how people go about answering the questions. RTI is doing this study for the Substance Abuse and Mental Health Services Administration, or SAMHSA. You are one of six participants between the ages of 12 to 17 who will help us test these questions.

Description of the Interview

Your participation in this interview will involve listening to survey questions being played using different voices. The survey includes questions about the use of tobacco, alcohol, drugs such as marijuana, and other health issues. However, I will not be asking you to answer these questions. Instead, I will ask you follow up questions to determine whether the voices used to read the questions were clear and easy for you to understand. For some questions, I may ask you to put the questions in your own words. The interview will last approximately 60 minutes. Your participation in this study will end after you finish the interview.

We also would like to audio record what you say during the interview. Only the people who work on this study will hear the recording. It will help us make sure we have understood your answers. If you don't want us to audio record you, that's okay.

Confidentiality/Your Rights

You don't have to answer a question if you don't want to. If you want to take a break at any time, just tell me. Your name will be kept private. No one else will see your answers to these questions. Your parents will not find out about your answers to questions. The only exceptions to this promise of confidentiality are if you tell me that you intend to seriously harm yourself or someone else or if you have been abused or if you identify an adult who has given you drugs; in these situations I may need to notify a mental health professional or other authorities.

Possible Risks and Benefits

Some of the the questions we ask may make you feel uncomfortable or upset. If this happens, let me know right away, and we can either take a break or I can give you information about talking with a counselor.

We are required by law to keep your answers private. The law also requires the study to use your answers only to learn how the questions work. The name of this law is the Confidential Information Protection and Statistical Efficiency Act of 2002.

There are no direct benefits to you from doing this interview. Your involvement in this study will help us improve the questions for the NSDUH.

When we finish, I will give you \$40 in cash to thank you for taking time to talk to me.

If you or your parent/guardian have any other questions about the study, you can call Ms. Emily Geisen at 1-800-334-8571 ext. 26566. If you or your parent/guardian have any questions about your rights as a participant in this study, you can call RTI's Office of Research Protection at 1-866-214-2043.

I will sign my name here to indicate that I have explained this information to you and that you have agreed to be interviewed. You will be given a copy of this form.

Signature of Interviewer

Date

I will sign my name here to indicate that you have agreed for the interview to be audio recorded.

Signature of Interviewer

Date

Participant Assent to Be Observed (ADOLESCENT)

[Another person who works on the study/A person or people who work(s) with the sponsor of this study] also is here with us today. This person (These people) would like to watch your interview in a separate observation room. We have already talked with your parent or guardian about this, and they have said it is okay to have this person (these people) watch the interview. What you say will still be kept private. It's okay if you don't want this person (these people) to watch your interview. We will simply ask that person(them) to leave the observation room.

Is it OK for this person (them) to watch your interview?

CHECK ONE OF THE BOXES BELOW. SIGN AND DATE FORM

Other study team member or sponsor representative may observe the interview.

Other study team member or sponsor representative may not observe the interview.

Signature of Interviewer

Date

Prueba sobre la conversión de texto a voz
Formulario de consentimiento informado del participante adulto
para la entrevista cognitiva

Encuesta Nacional Sobre la Salud y el Consumo de Drogas (NSDUH, por sus siglas en inglés)

Introducción

Le voy a explicar el estudio. Usted me puede detener en cualquier momento si tiene preguntas sobre cualquier cosa que le diga.

El objetivo de este estudio es realizar la prueba de algunas preguntas que se usarán en la Encuesta Nacional Sobre la Salud y el Consumo de Drogas (NSDUH), la cual es una encuesta grande que se realiza con aproximadamente 70,000 personas en todo el país. La encuesta recopila información sobre muchos temas relacionados con la salud. El objetivo es servir mejor a las personas en los Estados Unidos. En este momento, nos interesa evaluar la voz que se utilizará para leer algunas preguntas en el estudio. Deseamos saber qué tan bien entienden las personas estas preguntas y cómo las contestarían. RTI esta realizando un estudio para la Administración de Salud Mental y Abuso de Sustancias (SAMHSA, por sus siglas en inglés), que forma parte del Departamento de Salud y Servicios Humanos de los Estados Unidos. Usted es uno de 12 participantes de por lo menos 18 años de edad que van a revisar las preguntas de la encuesta para este estudio.

Descripción de la entrevista

Su participación en esta entrevista se tratará de escuchar las preguntas de la encuesta usando varias voces. La encuesta incluye preguntas sobre el uso de tabaco, alcohol, drogas como marihuana y otros temas de salud. Sin embargo, no se le pedirá que responda a estas preguntas. Por el contrario, yo le haré preguntas de seguimiento para determinar si las voces que se usaron para leer las preguntas fueron claras y fáciles de entender para usted. Para algunas preguntas, quizás le pida que me diga la pregunta en sus propias palabras. La entrevista tomará aproximadamente 60 minutos. Su participación en este estudio terminará al finalizar la entrevista. No haremos preguntas sobre su situación legal o de inmigración.

También nos gustaría grabar en audio lo que usted diga durante la entrevista. Solo las personas que trabajan en este estudio escucharán la grabación. Esto nos ayudará asegurar que entendimos sus respuestas. Está bien si usted no desea que se haga la grabación.

Confidencialidad y sus derechos

Tomar parte en esta entrevista es completamente voluntario. Usted puede dejar de contestar cualquier pregunta de la entrevista que no desee contestar. Su información personal no se asociará con sus respuestas de ninguna manera. La ley federal requiere que mantengamos sus respuestas en forma confidencial y que las usemos solo para propósitos estadísticos (Ley de Protección de la Información Confidencial y Eficiencia Estadística del año 2002). Con su permiso, grabaremos su entrevista. Nos puede pedir hacer una pausa o detener la grabación en cualquier momento. Solo los miembros del personal del estudio de RTI y SAMHSA podrán

escuchar las grabaciones. Las grabaciones se destruirán dentro de los siguientes 60 días de haber terminado este estudio. Los comentarios de todas las entrevistas se combinarán en un reporte que no identificará quién hizo los comentarios.

[Leer solo si un observador está presente: Un miembro del personal del estudio de RTI o un(os) representante(s) de SAMHSA está(n) aquí con nosotros y le gustaría observar la entrevista desde otro salón de observación. Si usted no desea que nadie más observe su entrevista, simplemente le pediremos a esta persona o personas que salga(n) del salón de observación y luego haremos la entrevista.

Posibles riesgos y beneficios

Usted me puede pedir que detenga la entrevista en cualquier momento. Si desea tomar un descanso en cualquier momento, solo tiene que avisarme. Es posible que algunas de las preguntas de la encuesta le puedan hacer sentirse incómodo(a) o le molesten. Si esto sucede, le puedo decir cómo comunicarse con un consejero(a).

No hay beneficios directos para usted por participar en esta entrevista. Sin embargo, las respuestas que usted nos dé nos ayudarán a mejorar la calidad de las preguntas de la encuesta NSDUH.

Pago por participación

Usted recibirá \$40 dólares en efectivo por completar la entrevista.

Sus preguntas

Si tiene cualquier otra pregunta sobre el estudio, puede llamar a la Sra. Rosanna Quiroz al 1-919-541-7172. Si tiene preguntas sobre sus derechos como participante en un estudio, puede llamar a la Oficina de RTI para la Protección de Participantes en Estudios al 1-866-214-2043 (número de teléfono gratuito).

Voy a firmar mi nombre aquí para indicar que le expliqué esta información y que usted estuvo de acuerdo en ser entrevistado(a).

Firma del entrevistador

Fecha

Leer solo si un observador está presente: También firmaré aquí para indicar que usted ha dado su consentimiento para que un miembro del personal de RTI o un representante de SAMHSA observe la entrevista. [ENTREVISTADOR, POR FAVOR ESCRIBA "NA" EN LA LÍNEA DE LA FIRMA SI LA ENTREVISTA NO SERÁ OBSERVADA.]

Firma del entrevistador

Fecha

Por último, yo voy a firmar mi nombre aquí para indicar que usted está de acuerdo en que la entrevista sea grabada en audio.

Firma del entrevistador

Fecha

Appendix E: Text to Speech Pilot Test Detailed Timing Tables

Table E.1 Text to Speech Audit Trail Timing Data: Interview Overall, All Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	43	19	23	1
Extreme Records*	0	0	0	0
Summary Statistics (Minutes) ¹				
Mean	75.10	74.20	76.41	61.80
Variance	652.55	400.64	907.45	.
Standard Deviation	25.55	20.02	30.12	.
Quartiles				
Maximum	130.25	113.87	130.25	61.80
Q3	90.42	84.75	111.57	61.80
Median	66.48	72.70	63.73	61.80
Q1	52.72	57.93	51.40	61.80
Minimum	41.80	45.18	41.80	61.80
Mode	.	.	.	61.80
Range	88.45	68.68	88.45	0.00
Percentiles				
99%	130.25	113.87	130.25	61.80
95%	122.43	113.87	126.38	61.80
90%	114.08	107.92	122.43	61.80
10%	46.43	45.70	46.43	61.80
5%	45.18	45.18	43.93	61.80
1%	41.80	45.18	41.80	61.80
Extremes				
5 Highest (Highest)	130.25	113.87	130.25	.
	126.38	107.92	126.38	.
	122.43	103.38	122.43	.
	117.03	90.42	117.03	.
	114.08	84.75	114.08	61.80
5 Lowest	46.43	57.93	51.22	.
	45.70	55.67	51.07	.
	45.18	50.65	46.43	.
	43.93	45.70	43.93	.
(Lowest)	41.80	45.18	41.80	61.80

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.2 Text to Speech Audit Trail Timing Data: ACASI, All Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	43	19	23	1
Extreme Records*	0	0	0	0
Summary Statistics (Minutes) ¹				
Mean	56.28	53.48	59.07	45.02
Variance	562.68	310.77	799.26	.
Standard Deviation	23.72	17.63	28.27	.
Quartiles				
Maximum	107.75	83.63	107.75	45.02
Q3	73.35	65.05	92.85	45.02
Median	49.45	53.05	46.95	45.02
Q1	37.07	39.67	35.15	45.02
Minimum	21.87	26.45	21.87	45.02
Mode	.	.	.	45.02
Range	85.88	57.18	85.88	0.00
Percentiles				
99%	107.75	83.63	107.75	45.02
95%	102.05	83.63	106.43	45.02
90%	92.87	81.13	102.05	45.02
10%	31.53	29.15	31.53	45.02
5%	29.15	26.45	30.40	45.02
1%	21.87	26.45	21.87	45.02
Extremes				
5 Highest (Highest)	107.75	83.63	107.75	.
	106.43	81.13	106.43	.
	102.05	80.23	102.05	.
	98.52	73.35	98.52	.
	92.87	65.05	92.87	45.02
5 Lowest	31.53	39.67	33.85	.
	30.40	33.37	32.63	.
	29.15	32.18	31.53	.
	26.45	29.15	30.40	.
(Lowest)	21.87	26.45	21.87	45.02

ACASI = audio computer-assisted self-interview.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.3 Text to Speech Audit Trail Timing Data: ACASI Tutorial, All Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	43	19	23	1
Extreme Records*	0	0	0	0
Summary Statistics (Minutes) ¹				
Mean	4.11	3.95	4.38	1.02
Variance	4.29	4.04	4.35	.
Standard Deviation	2.07	2.01	2.08	.
Quartiles				
Maximum	10.32	10.32	7.87	1.02
Q3	5.58	5.20	6.30	1.02
Median	3.67	3.88	3.67	1.02
Q1	2.48	2.48	2.62	1.02
Minimum	1.02	1.45	1.17	1.02
Mode	2.62	.	2.62	1.02
Range	9.30	8.87	6.70	0.00
Percentiles				
99%	10.32	10.32	7.87	1.02
95%	7.37	10.32	7.37	1.02
90%	6.55	5.70	7.17	1.02
10%	1.88	1.83	1.97	1.02
5%	1.45	1.45	1.88	1.02
1%	1.02	1.45	1.17	1.02
Extremes				
5 Highest (Highest)	10.32	10.32	7.87	.
	7.87	5.70	7.37	.
	7.37	5.58	7.17	.
	7.17	5.22	6.55	.
	6.55	5.20	6.45	1.02
5 Lowest	1.88	2.48	2.25	.
	1.83	2.12	2.10	.
	1.45	1.98	1.97	.
	1.17	1.83	1.88	.
(Lowest)	1.02	1.45	1.17	1.02

ACASI = audio computer-assisted self-interview.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.4 Text to Speech Audit Trail Timing Data: ACASI Risk Availability, All Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	43	19	23	1
Extreme Records*	0	0	0	0
Summary Statistics (Minutes) ¹				
Mean	3.99	3.50	4.42	3.40
Variance	4.96	2.63	6.89	.
Standard Deviation	2.23	1.62	2.62	.
Quartiles				
Maximum	9.78	8.37	9.78	3.40
Q3	4.78	4.60	4.83	3.40
Median	3.47	3.47	3.80	3.40
Q1	2.45	2.27	2.72	3.40
Minimum	1.33	1.65	1.33	3.40
Mode	3.40	.	.	3.40
Range	8.45	6.72	8.45	0.00
Percentiles				
99%	9.78	8.37	9.78	3.40
95%	9.47	8.37	9.62	3.40
90%	8.37	5.40	9.47	3.40
10%	1.88	1.77	1.88	3.40
5%	1.65	1.65	1.43	3.40
1%	1.33	1.65	1.33	3.40
Extremes				
5 Highest (Highest)	9.78	8.37	9.78	.
	9.62	5.40	9.62	.
	9.47	5.15	9.47	.
	8.87	4.62	8.87	.
	8.37	4.60	5.65	3.40
5 Lowest	1.88	2.27	2.33	.
	1.77	2.23	1.97	.
	1.65	1.98	1.88	.
	1.43	1.77	1.43	.
(Lowest)	1.33	1.65	1.33	3.40

ACASI = audio computer-assisted self-interview.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.5 Text to Speech Audit Trail Timing Data: Interview Overall, English-Speaking Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	22	12	9	1
Extreme Records*	0	0	0	0
Summary Statistics (Minutes) ¹				
Mean	61.40	69.63	50.38	61.80
Variance	286.59	348.50	34.86	.
Standard Deviation	16.93	18.67	5.90	.
Quartiles				
Maximum	107.92	107.92	61.82	61.80
Q3	70.77	82.07	51.67	61.80
Median	54.86	68.62	51.22	61.80
Q1	50.65	53.16	46.43	61.80
Minimum	41.80	45.18	41.80	61.80
Mode	.	.	.	61.80
Range	66.12	62.73	20.02	0.00
Percentiles				
99%	107.92	107.92	61.82	61.80
95%	84.75	107.92	61.82	61.80
90%	82.57	84.75	61.82	61.80
10%	45.18	45.70	41.80	61.80
5%	43.93	45.18	41.80	61.80
1%	41.80	45.18	41.80	61.80
Extremes				
5 Highest (Highest)	107.92	107.92	61.82	.
	84.75	84.75	54.05	.
	82.57	82.57	51.67	.
	81.58	81.58	51.40	.
	78.43	78.43	51.22	61.80
5 Lowest	46.43	65.82	51.22	.
	45.70	55.67	51.07	.
	45.18	50.65	46.43	.
	43.93	45.70	43.93	.
(Lowest)	41.80	45.18	41.80	61.80

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.6 Text to Speech Audit Trail Timing Data: ACASI, English-Speaking Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	22	12	9	1
Extreme Records*	0	0	0	0
Summary Statistics (Minutes) ¹				
Mean	43.27	49.17	35.21	45.02
Variance	218.58	287.53	52.80	.
Standard Deviation	14.78	16.96	7.27	.
Quartiles				
Maximum	83.63	83.63	46.95	45.02
Q3	52.57	58.90	40.12	45.02
Median	40.65	51.01	35.15	45.02
Q1	32.18	32.77	31.53	45.02
Minimum	21.87	26.45	21.87	45.02
Mode	.	.	.	45.02
Range	61.77	57.18	25.08	0.00
Percentiles				
99%	83.63	83.63	46.95	45.02
95%	65.05	83.63	46.95	45.02
90%	60.82	65.05	46.95	45.02
10%	29.15	29.15	21.87	45.02
5%	26.45	26.45	21.87	45.02
1%	21.87	26.45	21.87	45.02
Extremes				
5 Highest (Highest)	83.63	83.63	46.95	.
	65.05	65.05	41.18	.
	60.82	60.82	40.12	.
	56.98	56.98	37.07	.
	55.53	55.53	35.15	45.02
5 Lowest	31.53	44.83	35.15	.
	30.40	33.37	32.63	.
	29.15	32.18	31.53	.
	26.45	29.15	30.40	.
(Lowest)	21.87	26.45	21.87	45.02

ACASI = audio computer-assisted self-interview.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.7 Text to Speech Audit Trail Timing Data: ACASI Tutorial, English-Speaking Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	22	12	9	1
Extreme Records*	0	0	0	0
Summary Statistics (Minutes) ¹				
Mean	2.92	3.37	2.53	1.02
Variance	1.71	2.12	0.64	.
Standard Deviation	1.31	1.45	0.80	.
Quartiles				
Maximum	5.58	5.58	3.67	1.02
Q3	3.67	4.69	2.93	1.02
Median	2.67	3.17	2.62	1.02
Q1	1.97	2.05	1.97	1.02
Minimum	1.02	1.45	1.17	1.02
Mode	.	.	.	1.02
Range	4.57	4.13	2.50	0.00
Percentiles				
99%	5.58	5.58	3.67	1.02
95%	5.22	5.58	3.67	1.02
90%	5.20	5.22	3.67	1.02
10%	1.45	1.83	1.17	1.02
5%	1.17	1.45	1.17	1.02
1%	1.02	1.45	1.17	1.02
Extremes				
5 Highest (Highest)	5.58	5.58	3.67	.
	5.22	5.22	3.52	.
	5.20	5.20	2.93	.
	4.18	4.18	2.73	.
	4.08	4.08	2.62	1.02
5 Lowest	1.88	2.48	2.62	.
	1.83	2.12	2.25	.
	1.45	1.98	1.97	.
	1.17	1.83	1.88	.
(Lowest)	1.02	1.45	1.17	1.02

ACASI = audio computer-assisted self-interview.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.8 Text to Speech Audit Trail Timing Data: ACASI Risk Availability, English-Speaking Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	22	12	9	1
Extreme Records*	0	0	0	0
Summary Statistics (Minutes) ¹				
Mean	3.14	3.38	2.80	3.40
Variance	2.46	3.60	1.28	.
Standard Deviation	1.57	1.90	1.13	.
Quartiles				
Maximum	8.37	8.37	4.83	3.40
Q3	3.57	3.57	3.43	3.40
Median	2.83	2.99	2.75	3.40
Q1	1.98	2.11	1.97	3.40
Minimum	1.33	1.65	1.33	3.40
Mode	.	.	.	3.40
Range	7.03	6.72	3.50	0.00
Percentiles				
99%	8.37	8.37	4.83	3.40
95%	5.40	8.37	4.83	3.40
90%	4.83	5.40	4.83	3.40
10%	1.65	1.77	1.33	3.40
5%	1.43	1.65	1.33	3.40
1%	1.33	1.65	1.33	3.40
Extremes				
5 Highest (Highest)	8.37	8.37	4.83	.
	5.40	5.40	3.80	.
	4.83	3.58	3.43	.
	3.80	3.57	2.92	.
	3.58	3.53	2.75	3.40
5 Lowest	1.97	2.45	2.75	.
	1.77	2.23	2.72	.
	1.65	1.98	1.97	.
	1.43	1.77	1.43	.
(Lowest)	1.33	1.65	1.33	3.40

ACASI = audio computer-assisted self-interview.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.9 Text to Speech Audit Trail Timing Date: Interview Overall, Spanish-Speaking Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	21	7	14	0
Extreme Records*	0	0	0	.
Summary Statistics (Minutes) ¹				
Mean	89.45	82.05	93.15	.
Variance	646.68	449.29	743.31	.
Standard Deviation	25.43	21.20	27.26	.
Quartiles				
Maximum	130.25	113.87	130.25	.
Q3	113.87	103.38	117.03	.
Median	89.67	74.95	92.73	.
Q1	63.85	61.10	63.85	.
Minimum	52.72	57.93	52.72	.
Mode
Range	77.53	55.93	77.53	.
Percentiles				
99%	130.25	113.87	130.25	.
95%	126.38	113.87	130.25	.
90%	122.43	113.87	126.38	.
10%	58.77	57.93	58.77	.
5%	57.93	57.93	52.72	.
1%	52.72	57.93	52.72	.
Extremes				
5 Highest (Highest)	130.25	113.87	130.25	.
	126.38	103.38	126.38	.
	122.43	90.42	122.43	.
	117.03	74.95	117.03	.
	114.08	72.70	114.08	.
5 Lowest	63.73	90.42	76.63	.
	61.10	74.95	63.85	.
	58.77	72.70	63.73	.
	57.93	61.10	58.77	.
(Lowest)	52.72	57.93	52.72	.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.10 Text to Speech Audit Trail Timing Data: ACASI, Spanish-Speaking Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	21	7	14	0
Extreme Records*	0	0	0	.
Summary Statistics (Minutes) ¹				
Mean	69.90	60.88	74.41	.
Variance	571.14	304.17	672.53	.
Standard Deviation	23.90	17.44	25.93	.
Quartiles				
Maximum	107.75	81.13	107.75	.
Q3	92.85	80.23	98.52	.
Median	66.98	56.68	73.77	.
Q1	48.28	42.02	48.28	.
Minimum	33.85	39.67	33.85	.
Mode
Range	73.90	41.47	73.90	.
Percentiles				
99%	107.75	81.13	107.75	.
95%	106.43	81.13	107.75	.
90%	102.05	81.13	106.43	.
10%	42.02	39.67	44.68	.
5%	39.67	39.67	33.85	.
1%	33.85	39.67	33.85	.
Extremes				
5 Highest (Highest)	107.75	81.13	107.75	.
	106.43	80.23	106.43	.
	102.05	73.35	102.05	.
	98.52	56.68	98.52	.
	92.87	53.05	92.87	.
5 Lowest	44.93	73.35	56.15	.
	44.68	56.68	48.28	.
	42.02	53.05	44.93	.
	39.67	42.02	44.68	.
(Lowest)	33.85	39.67	33.85	.

ACASI = audio computer-assisted self-interview.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.11 Text to Speech Audit Trail Timing Data: ACASI Tutorial, Spanish-Speaking Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	21	7	14	0
Extreme Records*	0	0	0	.
Summary Statistics (Minutes) ¹				
Mean	5.36	4.94	5.57	.
Variance	4.00	6.44	3.04	.
Standard Deviation	2.00	2.54	1.74	.
Quartiles				
Maximum	10.32	10.32	7.87	.
Q3	6.45	5.70	6.55	.
Median	5.53	4.25	6.00	.
Q1	3.88	3.27	5.23	.
Minimum	2.10	2.85	2.10	.
Mode	6.00	.	6.00	.
Range	8.22	7.47	5.77	.
Percentiles				
99%	10.32	10.32	7.87	.
95%	7.87	10.32	7.87	.
90%	7.37	10.32	7.37	.
10%	2.85	2.85	2.62	.
5%	2.62	2.85	2.10	.
1%	2.10	2.85	2.10	.
Extremes				
5 Highest (Highest)	10.32	10.32	7.87	.
	7.87	5.70	7.37	.
	7.37	4.32	7.17	.
	7.17	4.25	6.55	.
	6.55	3.88	6.45	.
5 Lowest	3.33	4.32	5.52	.
	3.27	4.25	5.23	.
	2.85	3.88	3.33	.
	2.62	3.27	2.62	.
(Lowest)	2.10	2.85	2.10	.

ACASI = audio computer-assisted self-interview.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Table E.12 Text to Speech Audit Trail Timing Data: ACASI Risk Availability, Spanish-Speaking Respondents

Age Group	Overall	12-17	18-64	65+
Sample Used in Analysis	21	7	14	0
Extreme Records*	0	0	0	.
Summary Statistics (Minutes) ¹				
Mean	4.88	3.71	5.47	.
Variance	6.19	1.21	7.86	.
Standard Deviation	2.49	1.10	2.80	.
Quartiles				
Maximum	9.78	5.15	9.78	.
Q3	5.15	4.62	8.87	.
Median	4.60	3.47	4.77	.
Q1	3.25	2.62	3.40	.
Minimum	1.88	2.27	1.88	.
Mode
Range	7.90	2.88	7.90	.
Percentiles				
99%	9.78	5.15	9.78	.
95%	9.62	5.15	9.78	.
90%	9.47	5.15	9.62	.
10%	2.33	2.27	2.33	.
5%	2.27	2.27	1.88	.
1%	1.88	2.27	1.88	.
Extremes				
5 Highest (Highest)	9.78	5.15	9.78	.
	9.62	4.62	9.62	.
	9.47	4.60	9.47	.
	8.87	3.47	8.87	.
	5.65	3.25	5.65	.
5 Lowest	2.83	4.60	3.83	.
	2.62	3.47	3.40	.
	2.33	3.25	2.83	.
	2.27	2.62	2.33	.
(Lowest)	1.88	2.27	1.88	.

ACASI = audio computer-assisted self-interviewing.

* Extreme records have an interview length, without field interviewer debriefing, of shorter than 30 minutes or longer than 240 minutes.

Appendix F: Text to Speech FI Debriefing Moderator's Guide

NSDUH Text-to-Speech Pilot Test Field Interviewer Debriefing Call Moderator's Guide – FINAL

SECTION I: Introduction (5 minutes)

Hello and thank you for attending today's debriefing call to discuss your experiences during the Text-to-Speech Pilot Test.

My name is [MODERATOR'S NAME] from RTI. Also on the call today from RTI are [NOTE TAKER'S NAME], as well as [OTHER RTI OBSERVERS]. I will be leading today's discussion with help from [NOTE TAKER'S NAME], who will be taking notes.

[IF SAMHSA STAFF ON CALL] In addition, on the call with us today from our client, SAMHSA, are [STAFF NAMES].

Before we get started, I want to remind everyone to have your TTS FI Feedback Worksheet in front of you as we talk, so you can reference your notes.

This discussion is intended to gather feedback on your experiences completing data collection for the TTS Pilot Test. As you know, a new computer-generated ACASI voice was tested during this effort, but we cannot gather all of the information we need just by analyzing survey data.

Therefore, we are hoping you can share your experiences, including feedback you received from respondents, and any issues you encountered that could be improved in the future.

A summary of your feedback from today's discussion will be provided to SAMHSA to help inform potential changes in the future.

A couple of notes about our discussion today:

- We are recording this call and have a note taker so we can capture all of your comments.
- Please be respectful of everyone on this call, so only one person should speak at a time. Doing so allows the whole group to hear each person and ensures the recording will be clear.
- Also, if you have not done so already, move to a location with minimal background noise.
- If I haven't heard from you, I may call on you. If I do call on you, but you'd rather not answer a particular question or if you don't have anything to add, you can just tell me that you would like to "pass."
- Since we are on the phone, each time you speak, I would like you to begin your comments by saying your name, such as, "This is [MODERATOR'S NAME], and I think..."
- Please know there is no right or wrong answer to the questions I will be asking. Everyone's input is important and helpful.
- Also, for the bilingual FIs on this call, when providing feedback, please clarify if any issues you encountered were experienced with only Spanish-speaking respondents, only English-speaking respondents, or both.

Any general questions before we get started?

SECTION I: Screenings and Using the Tablet (10 minutes)

NOTE: BILINGUAL FIs SHOULD ALSO INCLUDE RESPONSES SPECIFIC TO THEIR ADMINISTRATION OF SPANISH SCREENINGS.

For this first section of the call, we are going to discuss the tablet video you showed to respondents.

1. How often did you use the tablet video with respondents and what were their general reactions to it?

What do you think about the content of the tablet video? What changes would you make to the video to improve it?

2. Did you have any difficulty accessing or playing the tablet video? How was the volume level of the video?
3. Based on your experience using the tablet video during this pilot test, do you think the video would be effective for use with respondents during regular main study data collection?
4. What other feedback do you have about the tablet video that we have not already discussed?

For this next set of questions, I will be asking about your experience with other aspects of the tablet and screening program both in the field and at home.

5. What problems did you experience, if any, with the tablet case or stylus when using the tablet at the door with respondents?
6. Did you have any issues transmitting from the tablet at your home? If so, what were they?
7. How often did you connect your tablet to a wireless Internet connection outside of your home to transmit? [FOR THOSE WHO DID: Where did you connect to wireless Internet? Were there any issues with transmitting from this location outside your home?]
8. What features or capabilities of the tablet did you have difficulty with? If any, please describe in detail.
9. Please describe any issues you experienced while using the screening program.
10. Did you ever call Technical Support **for assistance** with the tablet at any point during data collection? [FOR ANY WHO INDICATE CALLING TECHNICAL SUPPORT, ASK: What happened that required you to call TSG and how was the issue resolved?]

SECTION II: TTS Questionnaire and ACASI Voice (15 minutes)

NOTE: BILINGUAL FIs SHOULD INCLUDE RESPONSES SPECIFIC TO THEIR ADMINISTRATION OF SPANISH INTERVIEWS.

Now I am going to ask a series of questions about the Text-to-Speech software and the interview questionnaire.

1. Did respondents comment on whether the ACASI voice was easy or difficult to understand? If difficult, what specific questions or aspects of the survey were difficult to understand?
2. Did respondents comment on whether or not they liked the voice? If so, what did they like or dislike?
3. Did respondents comment on the speed of the voice? If so, did they think it was too fast, too slow, or just right?
4. Did respondents comment specifically on the fact that the voice was computerized (negative, positive, neutral)?
5. Did any respondents exhibit nonverbal feedback during the ACASI portion, such as showing signs of being confused or frustrated, having difficulty hearing and turning the volume up, or removing the headphones or turning the volume off? If so, did respondents exhibit these actions more or less frequently on the TTS Pilot Test compared to the Main Study?
6. Did respondents make any other comments about the ACASI voice? If so, what particular comments did they make?

FOR BILINGUAL FIs ONLY:

7. We are interested in detecting any issues in the translation of the Spanish questionnaire. Did respondents indicate they were confused or unsure about any Spanish text in the questionnaire? [PROBE: Please provide examples of questions or wording that caused confusion.]
8. Do you personally have any feedback on questions in the Spanish interview where the Spanish translation may be problematic? [PROBE: What part of the translation seems to be problematic? Can you tell me more about that?]
9. Are there any other comments that you would like to make about the Spanish questionnaire?

SECTION III: Administering the TTS Interview and Using the Laptop (10 minutes)

NOTE: BILINGUAL FIs SHOULD INCLUDE RESPONSES SPECIFIC TO THEIR ADMINISTRATION OF SPANISH INTERVIEWS.

In this section, we are going to discuss the new laptop, including how respondents reacted to it.

1. What are your reactions to the function key label for the laptop keyboard? Did you have any issues with the label?
2. How did respondents react to the function key label for the laptop keyboard? Were there any issues finding or using the function keys?
3. Did you ever call Technical Support for assistance with the laptop at any point during data collection? [FOR ANY WHO INDICATE REQUESTING TECHNICAL SUPPORT, ASK: Can you tell me why you called?]
4. Did respondents raise **any other specific concerns** when completing the ACASI portion or the questions you administered? [PROBE: Please provide examples of any concerns that you can recall.]

SECTION IV: Conclusion (5 minutes)

Are there any final comments on any of the topics we discussed, or other feedback you would like to provide at this time about your experience with this pilot test?

I want to thank you all again for your participation on this call.

NOTE TAKER WILL NOW STOP THE AUDIO RECORDING.

PROVIDE INSTRUCTIONS FOR CHARGING TIME FOR THIS CALL; GIVE CODE AND AMOUNT OF TIME TO CHARGE.

PROVIDE REMINDERS ON DESTROYING ALL TTS PT MATERIALS, PROCEDURES FOR RETURN OF EQUIPMENT (IF NOT ALREADY DONE SO), ETC.