



Published in final edited form as:

*Qual Health Res.* 2015 June ; 25(6): 820–830. doi:10.1177/1049732315582010.

## Establishing Survey Validity and Reliability for American Indians Through “Think Aloud” and Test–Retest Methods

Cindy Horst Hauge<sup>1</sup>, Jacque Jacobs-Knight<sup>2</sup>, Jamie L. Jensen<sup>1</sup>, Katherine M. Burgess<sup>1</sup>, Susan E. Puumala<sup>1</sup>, Georgiana Wilton<sup>3</sup>, and Jessica D. Hanson<sup>1</sup>

<sup>1</sup>Sanford Research, Sioux Falls, South Dakota, USA

<sup>2</sup>Oglala Sioux Tribe CHOICES Program, Pine Ridge, South Dakota, USA

<sup>3</sup>University of Wisconsin, Madison, Wisconsin, USA

### Abstract

The purpose of this study was to use a mixed-methods approach to determine the validity and reliability of measurements used within an alcohol-exposed pregnancy prevention program for American Indian women. To develop validity, content experts provided input into the survey measures, and a “think aloud” methodology was conducted with 23 American Indian women. After revising the measurements based on this input, a test–retest was conducted with 79 American Indian women who were randomized to complete either the original measurements or the new, modified measurements. The test–retest revealed that some of the questions performed better for the modified version, whereas others appeared to be more reliable for the original version. The mixed-methods approach was a useful methodology for gathering feedback on survey measurements from American Indian participants and in indicating specific survey questions that needed to be modified for this population.

### Keywords

aboriginal people, North America; alcohol / alcoholism; contraception; reliability; validity; qualitative

---

The negative health risks associated with alcohol consumption during pregnancy represent a leading preventable cause of disabilities in the United States, including fetal alcohol syndrome (FAS; Bailey & Sokol, 2008). Even moderate alcohol use during pregnancy has been associated with adverse effects, including an increased risk of low birth weight and pre-term delivery, and represents a major public health concern (Chen, 2012; Kesmodel, Wisborg, Olsen, Henriksen, & Secher, 2002). In addition, approximately half of pregnancies

---

© The Author(s) 2015

**Corresponding Author:** Jessica D. Hanson, Center for Health Outcomes & Prevention Research, Sanford Research, 2301 E. 60th St North, Sioux Falls, SD 57104, USA. [Jessica.d.hanson@sanfordhealth.org](mailto:Jessica.d.hanson@sanfordhealth.org).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Minority Health and Health Disparities or the National Institutes of Health (NIH).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

in the United States are unintended, some of which occur in women using contraception, thus delaying pregnancy recognition (Trussell, 2011). Women who experience an unplanned pregnancy are at an increased risk of adverse maternal behaviors (Finer & Zolna, 2011). Although many women stop drinking when they find out they are pregnant, an estimated 130,000 pregnancies in the United States are exposed to high levels of alcohol each year (Lupton, Burd, & Harwood, 2004).

Overall, between 10% and 26% of women are at risk for alcohol-exposed pregnancy (AEP; Project CHOICES Research Group, 2002). This risk is even higher in many American Indian tribal communities (Hanson, Miller, Winberg, & Elliott, 2013). A previous project indicated that the number of drinks consumed in an average week by non-pregnant American Indian women was nearly 18, and 30% were sexually active but not using any birth control to protect against pregnancy (Hanson et al., 2013). One American Indian tribe in the Northern Plains, de-identified for privacy reasons, conducted an adapted program, Project CHOICES (Changing High-risk alcohol use and Increasing Contraception Effectiveness Study), that focuses on the prevention of AEP in non-pregnant American Indian women. The tribe's CHOICES Program is centered on reducing the risk for AEP through alcohol reduction and/or pregnancy prevention using in-person motivational interventions (Floyd, Ebrahim, & Boyle, 1999; Floyd et al., 2007; Ingersoll, Floyd, Sobell, Velasquez, & Project CHOICES Research Group, 2003; Project CHOICES Research Group, 2002; Sobell et al., 2003).

Unfortunately, there is a lack of established validity and reliability of the CHOICES measures with American Indian women. In fact, the tribe's CHOICES Program is one of the first of its kind with American Indian women, which is laudable in terms of inclusion of a high-risk population but also concerning in that some of the measurements might not include linguistic (i.e., readability or clarity) or cultural norms for this population. While the CHOICES measurements have been previously validated with a variety of populations (Floyd et al., 2007; Ingersoll et al., 2003), there is still a great need for this type of research in American Indian communities. In particular, while this tribe's CHOICES intervention has been successfully implemented with American Indian women, there are still validity concerns. For example, interventionists report some confusion from participants regarding key measurements within the intervention, such as not understanding the meaning behind certain questions, which may affect the responses.

Therefore, the goal of the project was to establish validity (accuracy in measurement) and reliability (reproducibility) of critical measurements of the evidence-based CHOICES protocol for American Indian women. These measures included behavioral health screening measurements, temptation and confidence measurements, and readiness rulers, which are described in greater detail in the "Method" section. Using a mixed-methods approach, this reliability and validity project enriches an innovative existing program on AEP prevention with non-pregnant American Indian women and also adds to the reliability and validity literature on American Indian health behavior.

## Method

Before beginning data collection, the project was reviewed and approved by the tribe's research review board, as well as by the principal investigator's internal institutional review board. All participants in the “think aloud” and test–retest signed informed consent documents before beginning their participation in this study.

We did not seek informed consent from those involved in the content validity process because the review boards deemed this as “non-research.”

The original CHOICES authors cited several critical core theoretical components of the intervention that informed the focus of this validity/reliability study (Floyd et al., 2007). CHOICES is theory-based, drawing on social learning theories that underpin its cognitive-behavioral therapy activities and assessments (Velasquez et al., 2010). First is an initial behavioral (alcohol and contraception) assessment, collected via use of 4 of the 10 AUDIT alcohol screening questions (Babor, Higgins-Biddle, Saunders, & Monteiro, 2001) and questions about vaginal intercourse (yes/no), use of contraception (yes/no), and type of contraception used. Other key measurements include temptation and confidence related to alcohol consumption and birth control utilization, based on the Brief Situational Confidence Questionnaire (Breslin, Sobell, Sobell, & Agrawal, 2000) and the Self-Efficacy for Contraception Scale (Grimley et al., 1996). Third are “readiness to change” questions, which are based on Readiness Rulers (Nieman, Velasquez, Groff, Cheng, & Foxhall, 2005). To determine validity and reliability, we instigated a three-step, mixed-methods approach (see Figure 1), where input from extensive qualitative data collection led to changes in the CHOICES questionnaire, with the original and modified versions both tested for reliability via a test–retest method.

First, to establish *content validity*, or the content representativeness and relevance of the CHOICES measures for American Indian women, we solicited input from community members (including American Indian elders) and from content experts in American Indian health. They were recruited through current contacts, collaborations, and community advisory boards and asked via email about their interest in participating. These individuals were sent paper copies of key CHOICES measures and asked to thoroughly review and note questions, problems, and ideas for changes, and send this information back to the principal investigator via email or regular mail. Each participant received a US\$25 gift card to thank them for their time and was also asked if they were willing to “re-review” the modified CHOICES questionnaire.

Second, we utilized a “*think aloud*” methodology with American Indian women, collecting information about response reasoning and decision making by having participants verbalize their thoughts when completing the CHOICES measurements (Fonteyn & Fisher, 1995). Uses of the think-aloud method are well documented in the medical research literature (Chase, Reicks, Smith, Henry, & Reimer, 2003; Gardin, 2010; Gillam, Fargo, & Robertson, 2009; Göransson, Ehnfors, Fonteyn, & Ehrenberg, 2008; Holmstrup, Stearns-Bruening, & Rozelle, 2013; Jaspers, Steen, van den Bos, & Geenen, 2004; MacNeela et al., 2010; Van Den Haak, De Jong, & Schellens, 2003). Benefits of the think-aloud method include

capturing decision making and critical thinking in context (Gardin, 2010). The think aloud indicates validity by highlighting overarching problems when responding to measurements, and therefore issues with the accuracy of the measurement.

Based on previous literature, the goal was to engage 20 to 30 participants from the tribal community. Eligible participants for the think-aloud methodology were non-pregnant, adult American Indian women. They were recruited by posting flyers at clinics and community centers and sponsoring radio announcements, inviting interested individuals to contact the project assistant. Participants were scheduled for the think-aloud session at a later time point and were sent a reminder appointment card. Participants met with the project assistant at a private office or area. After signing an informed consent, participants were instructed to think aloud as they deliberated answering the CHOICES measures, and were asked to say whatever came into their mind while reading. If participants paused for longer than a few seconds while reading the survey, the project assistant quietly reminded them to “keep thinking aloud” and “please keep talking.” Aside from these reminders, all interactions between the participant and the project assistant were kept to a minimum to not interfere with the participants’ flow of thoughts. If participants seemed hesitant to “think aloud,” they were asked their thoughts at the end of each question. Each participant received a US\$25 gift card. The entire think-aloud session was audiotaped, and thorough notes were taken by the project assistant throughout the interaction.

To analyze the think-aloud data, the team conducted a type of “script analysis” that aims to “provide an overall description of the reasoning processes” and “illustrate what information subjects attended to during problem solving,” and the “rationale for the choices and decisions they made” (Fonteyn, Kuipers, & Grobe, 1993). To complete this, the project assistant listened to each audiotape and made notes on problem areas (i.e., confusion with a question, need for an example) for each question. These notes were given to the principal investigator, who created a spreadsheet that noted changes, problems, and confusion resulting from the content validity and think-aloud methodologies. This spreadsheet was shared on a conference call with a larger group, including the principal investigator, research assistants, and staff from the tribe's CHOICES Program. As a group, alterations to the CHOICES survey were made based on the content validity and think-aloud suggestions. The group evaluated each individual question to see what possible modifications were suggested, and changes were made based on group consensus. After amendments were made, a modified version of the CHOICES measurements was shown to community members and content experts who agreed to “re-review” the modified version.

After identifying survey modification needs via qualitative methodology, the team focused on determining reliability of both the original CHOICES survey and the modified version via a *test-retest methodology*. Based on a power analysis, the goal was to recruit at least 20 respondents in each of the groups (Donner, 1987; Fleiss, 1986), and the team aimed to oversample to reach these numbers. To recruit, the project assistant staffed tables at three health care facilities. Women who approached the table were asked to complete an eligibility screen (i.e., adult, non-pregnant American Indian women). Eligible participants were given a subject identifier and randomly assigned to complete either the modified CHOICES measurements or the original survey. Randomization occurred by asking

participants to choose a sealed envelope from a box; each envelope stated either “original” or “modified.”

After completing the survey, participants were asked for contact information (phone number and alternate numbers) for future contact for the retest process. The project assistant called each participant 2 weeks after the initial “test” to convey the date and time of the retest. About 75% of the women were able to be contacted using the numbers and information they provided. Many of those who were unable to be reached had disconnected numbers or were not at the number they had listed. The retest involved a similar procedure of setting up a table at the same location; participants were free to come to the site to fill out the survey when it was convenient for them on the designated days in their area. Participants received a US\$10 gift card for completing the “test” and a US\$15 gift card for completing the retest. The survey took 20 to 45 minutes to complete.

Test–retest reliability of the original Project CHOICES questionnaire (OPCQ) and modified Project CHOICES questionnaire (MPCQ) was analyzed using percent agreement, weighted kappa with Cicchetti–Allison weights, Gwet's AC1 and AC2 (with varying beta values; Gwet, 2001). Gwet's statistics are thought to be more stable for unusual marginal distributions when the kappa statistics does not perform as expected (Cicchetti & Feinstein, 1990). Test–retest questions were analyzed on their original scale, except for those relating to thinking about behavior for both alcohol and contraception. For these questions, the 10-point scale was reduced to a 5-point scale by collapsing ratings with the closest values (e.g., 1 and 2 were grouped together). Agreement measures were not formally compared between questionnaires, given the small sample sizes and large variances. Qualitative comparisons were made for hypothesis generation. Statistical analysis was performed using SAS software, version 9.3, SAS Institute Inc., Cary, North Carolina, United States. The AC1AC2 Macro by Blood and Spratt was used to calculate AC1 and AC2 (Blood & Spratt, 2007).

## Results

### Content Analysis

Feedback on the CHOICES measures was received by 17 native and non-native individuals, including 9 individuals working locally at tribal health centers, tribal non-profit organizations, and research projects occurring on tribal land, as well as 8 research and health professionals from universities and research institutions across the country who have previously worked with American Indians and AEP. Reviewers suggested *wording changes* (i.e., using “feeling down” instead of “depressed” and changing “experience side effects” to “have problems with” in regard to birth control), *giving examples* (for instance, adding pictures of alcoholic beverages to ease answering questions about drinking), *adding additional context to questions* (adding “feeling stressed” to a statement on being “physically tense”), and *emphasizing the role of partners* (for instance, clarifying if birth control questions address partners’ role in discouraging contraception, such as pressure to not use birth control or becoming angry if a woman utilizes birth control). Of those who agreed to re-review the modified measurements, none had additional comments or concerns.

## Think Alouds

Twenty-three non-pregnant American Indian women from various communities across the reservation and in a local non-reservation city participated in the think-aloud methodology. The average age of participants was 40.4 (range = 23–66), and all but one of the participants had children. Marital status, education level, and employment status varied among the participants as highlighted in Table 1.

When responding to questions about the number of standard drinks, participants appeared confused on how to respond, especially because *examples of standard drinks* (i.e., a glass of wine) are not a typical drink for women in this area, who tend to drink alcohol-based energy drinks, such as Joose. Therefore, appropriate types and pictures of local drinks were added to clarify the standard drinks questions. When asked about contraception, some were unsure as to the different *types of birth control*, and many also had problems understanding “effective use,” therefore a description and a definition of effective use were added for each method. As well, more description was provided on surgical methods (i.e., hysterectomy) and menopause. Local slang for certain methods was deemed necessary, such as using “morning-after pill” as opposed to “emergency contraception.” Also, self-reported questions about drinking and sexual activity behaviors used a *time frame* of “the past three months,” which appeared too specific; therefore, the behavioral time frame was adjusted to “the past few months.”

In the section on “Temptation” (or certain situations that might make an individual to drink at risky levels or *not* use contraception) and “Confidence” (how confident these individuals would be to *not* drink and to use birth control in the same situations), several major thematic challenges stood out. In general, *examples and relatable terminologies* appeared lacking. For example, participants seemed to need examples or additional clarifications on questions regarding “testing control,” “urges,” and “conflict with others” when it came to temptation to drink alcohol, and also were unsure about “side effects” of birth control. Certain words also appeared to be inappropriate for this population, such as “unpleasant feelings,” which respondents felt was confusing and might be misleading; “depression,” which was viewed as too extreme and negative; and “self-efficacy,” which was read as “self-efficiency,” “self-effective,” or skipped completely. Examples and wording changes were provided in these sections. Finally, because the overall formatting of the Temptation/Confidence questions appeared confusing, *precursors* were viewed as necessary to provide a context for each measurement. For example, “How tempted would you be to drink alcohol if . . .” was added before each alcohol temptation statement/question. Table 2 highlights many of these changes using the original alcohol Temptation/Confidence measurements and how they were modified.

The last major section of the survey was readiness rulers, where participants were asked to rank on a scale of 1 to 10 how (a) important, (b) sure, and (c) ready they were to change drinking and birth control behaviors. One area that was modified was in the alcohol readiness ruler section, as “*risky drinking*” was *defined* in an opening paragraph but not within each ruler. Therefore, the ruler was altered to say, “drink less than four drinks on any

one occasion or less than eight drinks per week” for all three (important, sure, and ready) alcohol rulers. The contraception rulers remained the same.

### Test–Retests

As stated, based on the input from qualitative means via content experts and the think-aloud methodology, several modifications were made to the existing CHOICES survey as shown in Table 2. For the test–retest to evaluate the reliability of the original versus modified measures, 79 non-pregnant American Indian women were enrolled at one of the three recruitment sites. Of these, 44 were randomly assigned to receive the MPCQ, and 35 completed the OPCQ. The average age of participants was 27.9 (range = 18–44), and the majority (90.6%) of participants had children. Marital status, education level, and employment status varied among the participants as highlighted in Table 3.

At the time of “retest,” which occurred approximately 2 weeks after the initial recruitment, 23 participants completed the MPCQ and 16 completed the OPCQ, for a retest response rate of 52.3% and 45.7%, respectively. Data for alcohol questions are presented in Table 4. Agreement statistics for alcohol questions regarding temptation and confidence appeared slightly better for the MPCQ compared with the OPCQ. Measures of agreement were consistent across the different statistical approaches. For the alcohol questions relating to thinking about behavior (i.e., readiness rulers), reliability was quite varied for both OPCQ and MPCQ, possibly because of the wide range of values included.

Data for questions regarding contraception are provided in Table 5. The weighted kappa statistic performed poorly in the contraception analysis due to an imbalance in marginal totals (e.g., for some questions only one or two response categories were used). When examining the AC statistics, unlike the alcohol questions, the OPCQ appeared to have slightly higher agreement compared with the MPCQ for both the temptation and confidence questions for the contraception analysis. The contraception readiness rulers had similar agreement measures for both the OPCQ and MPCQ.

### Discussion

While the theoretical cores of the CHOICES intervention have been utilized with American Indian women (Canales & Rakowski, 2006; Fahrenwald & Shangreaux, 2006; May et al., 2008), previous use of the CHOICES survey measures with American Indian women is lacking. Therefore, the ultimate aim of this project was to establish the validity and reliability of the CHOICES measures, and therefore to improve this AEP prevention program for American Indian women.

Our team utilized mixed-methods approach, beginning with a content analysis and think-aloud methodology to derive community input into survey measurements. Using the think-aloud qualitative methodology, we found definite “themes”—or problems with specific wording or questions that are similar for the majority of participants. This tells the team that use of a think-aloud methodology in gathering community input on survey measurements can be a fairly uncomplicated process and one that can provide a wealth of input on how the questionnaire and even the overall project can be made community-appropriate. Previous

use of the think-aloud methodology has focused primarily on nutrition and diet behaviors (Chase et al., 2003; Henry et al., 2003; Holmstrup et al., 2013; Reicks et al., 2003), as well as with understanding decision-making processes of health care providers (Fonteyn & Fisher, 1995; Göransson et al., 2008; MacNeela et al., 2010) and students (Gardin, 2010; Gillam et al., 2009; Pottier et al., 2010; Van Den Haak et al., 2003), making this particular think-aloud study with American Indian women and AEP prevention unique.

The results from the think aloud were vital in informing how questions are asked of the American Indian women. The think-aloud participant variance in comprehension of survey questions resulted in the revision of some key pieces of the survey to fit the community being served. Repeated reports by participants acknowledged the low literacy rates and lack of comprehension, especially in young women of the community and those in the more rural areas. It was clear that future use of surveys with American Indian populations needs to be reviewed by community members *before* implementation to have appropriate language and other visual aids addressed. The think aloud is a method that can gather community information while evaluating the validity of survey measurements. Examinations of verbal think-aloud protocols can provide useful information about comprehension that is not readily available from traditional measures of comprehension performance, such as question and answer and recall tasks (Gillam et al., 2009).

Based on these qualitative efforts, modifications to the original CHOICES measures were made, including refining questions using clarifying statements, giving examples, and rewording certain questions. The test–retest revealed that for the alcohol temptation/confidence questions, the modified questions were more reliable, but for the contraception temptation/confidence measures, the original questions appeared more reliable. One possible reason for this is that alcohol consumption is fairly pervasive in our country and within this community specifically. In addition, many CHOICES participants appear less aware and have less knowledge of contraception and the varying methods to prevent pregnancy when compared with their awareness and knowledge of alcohol consumption, which may indicate that birth control is less relevant to their daily lives when compared with alcohol.

These findings provide novel evidence that certain CHOICES components should be modified for use with American Indian women to increase the efficacy of the brief intervention protocol. The next step in this process is to work with the tribe's CHOICES sites to determine what survey measurements will be implemented and what measures should remain the same within the current program. Overall, this type of mixed-methods approach is significant as there is a dearth of validity and reliability studies conducted with American Indian communities, particularly in regard to alcohol or contraception behaviors (Abbott, 2011; Leonardson et al., 2005).

## Limitations

A potential limitation to the think-aloud methodology is the presence of the project assistant, who might influence verbalizations. However, most participants from previous think-aloud projects indicated that “the investigator's presence was not influential regarding what they thought, said, or did” (Chase et al., 2003, p. 503). In addition, unlike the think-aloud analysis described in this article, think-aloud data are sometimes analyzed using traditional content

analysis, including the use of data transcripts (Göransson et al., 2008). This team utilized recommendations made by other research (Fonteyn et al., 1993), focusing on script analysis as described in the method section, which provided important input into the validity of the CHOICES measures. A final limitation is that the project was conducted with one tribal area, limiting the generalizability of this study. However, this points to the potential to replicate this methodological model with other communities as CHOICES is expanded to other tribes.

## Conclusion

The use of the “think alouds” was a unique and novel way to gather input on the existing measures for a prevention program with American Indian women. The use of qualitative methodology is important when working with American Indian participants, as cultural elements are best revealed and understood through open-ended methodologies (Denzin & Lincoln, 2000; Israel et al., 2005). Through the “think aloud” methodology, we were able to better understand overarching problems and response reasoning to the CHOICES questionnaire among American Indian women, and therefore possible issues with the accuracy of the measurements. Without this qualitative piece, the test–retest could not have been conducted, highlighting the important nature of qualitative methodology, especially when working with American Indian communities.

Overall, the findings in this article are significant as they close knowledge gaps to understanding American Indian health and measuring certain health behaviors, specifically to the prevention of AEP with American Indian women. In particular, the article adds to the current literature on the CHOICES intervention, as well as informs validity and reliability studies with American Indian communities. In addition to expanding current knowledge, the results also indicate educational needs regarding AEP, such as clarifying “standard drink” measures and telling researchers the types of alcohol being consumed in such rural communities. For instance, the inexpensive energy-type drinks typically being consumed in this community are those with high alcohol content, equivalent to five to six standard drinks.

In conclusion, the mixed-methods approach utilized within this article highlights the modifications necessary for the CHOICES measurements for this particular community. As stated, the next step is to implement some of these suggested changes into current practice. This type of research is important in highlighting the need to better educate professionals interested in AEP prevention and health disparities research about cultural and linguistic differences that must be taken into account when developing and implementing intervention research. There is a definite need to continue straying from the “one size fits all” mentality by establishing validity and reliability for survey measurements with subpopulations before implementing within health research.

## Acknowledgment

Special thanks to Susan Pourier and Katana Jackson, who are both involved in the OST CHOICES Program.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and publication of this article. This project is supported by the National Center on Minority Health and Health Disparities of the National Institutes of Health under Award U54MD008164.

## Biography

**Cindy Horst Hauge**, RN, CNOR, is a research nurse at the Center for Health Outcomes and Prevention Research at Sanford Research in Sioux Falls, South Dakota, USA.

**Jacque Jacobs-Knight**, AA, is an interventionist for the Oglala Sioux Tribe CHOICES Program in Pine Ridge, South Dakota, United States.

**Jamie L. Jensen**, MS, is a research associate at the Center for Health Outcomes and Prevention Research at Sanford Research in Sioux Falls, South Dakota, United States.

**Katherine M. Burgess**, MPH, is a senior epidemiology analyst at the Center for Health Outcomes and Prevention Research at Sanford Research in Sioux Falls, South Dakota, USA.

**Susan E. Puumala**, PhD, is an associate scientist at the Center for Health Outcomes and Prevention Research at Sanford Research, Sioux Falls, South Dakota, United States and an associate professor in the Department of Pediatrics at the Sanford School of Medicine of the University of South Dakota, Sioux Falls, South Dakota, USA.

**Georgiana Wilton**, PhD, is a senior scientist at the University of Wisconsin Department of Family Medicine in Madison, Wisconsin, USA.

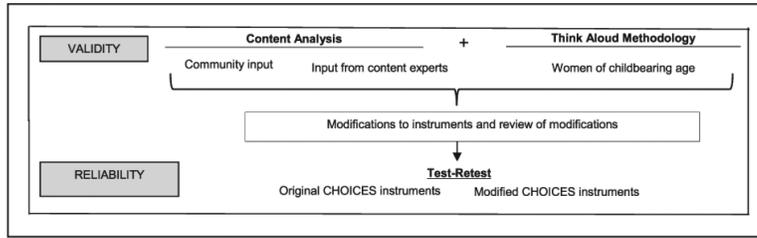
**Jessica D. Hanson**, PhD, is an associate scientist at the Center for Health Outcomes and Prevention Research, Sanford Research, in Sioux Falls, South Dakota, USA, and an assistant professor in the Department of Obstetrics and Gynecology at the Sanford School of Medicine of the University of South Dakota, Sioux Falls, South Dakota, USA.

## References

- Abbott PJ. Screening American Indian/Alaska natives for alcohol abuse and dependence in medical settings. *Current Drug Abuse Reviews*. 2011; 4:210–214. [PubMed: 21999696]
- Babor, T.; Higgins-Biddle, JC.; Saunders, JB.; Monteiro, MG. AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for use in primary care. 2nd ed.. World Health Organization; Geneva, Switzerland: 2001.
- Bailey BA, Sokol RJ. Pregnancy and alcohol use: Evidence and recommendations for prenatal care. *Clinical Obstetrics and Gynecology*. 2008; 51:436–444. [PubMed: 18463472]
- Blood, E.; Spratt, KF. Disagreement on agreement: Two alternative agreement coefficients.. Paper presented at the SAS Global Forum; Orlando, Florida. Apr. 2007
- Breslin FC, Sobell LC, Sobell MB, Agrawal S. A comparison of a brief and long version of the Situational Confidence Questionnaire. *Behaviour Research and Therapy*. 2000; 38:1211–1220. [PubMed: 11104185]
- Canales M, Rakowski W. Development of a culturally specific instrument for mammography screening: An example with American Indian women in Vermont. *Journal of Nursing Measurement*. 2006; 14:99–115. [PubMed: 17086783]
- Chase K, Reicks M, Smith C, Henry H, Reimer K. Use of the think-aloud method to identify factors influencing purchase of bread and cereals by low-income African American women and implications for whole-grain education. *Journal of the American Dietetic Association*. 2003; 103:501–504. [PubMed: 12669015]

- Chen JH. Maternal alcohol use during pregnancy, birth weight and early behavioral outcomes. *Alcohol and Alcoholism*. 2012; 47:649–656. [PubMed: 22893227]
- Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*. 1990; 43:551–558. [PubMed: 2189948]
- Denzin, NK.; Lincoln, YS. *Handbook of qualitative research*. 2nd ed.. SAGE; Thousand Oaks, CA: 2000.
- Donner AEM. Sample size requirements for reliability studies. *Statistics in Medicine*. 1987; 6:441–448. [PubMed: 3629046]
- Fahrenwald N, Shangreux P. Physical activity behavior of American Indian mothers. *Orthopaedic Nursing*. 2006; 25:22–29. [PubMed: 16465109]
- Finer LB, Zolna MR. Unintended pregnancy in the United States: Incidence and disparities, 2006. *Contraception*. 2011; 84:478–485. [PubMed: 22018121]
- Fleiss, J. *The design and analysis of clinical experiments*. John Wiley; New York: 1986.
- Floyd RL, Ebrahim SH, Boyle CA. Preventing alcohol-exposed pregnancies among women of childbearing age: The necessity of a preconceptional approach. *Journal of Women's Health & Gender-Based Medicine*. 1999; 8:733–736.
- Floyd RL, Sobell M, Velasquez MM, Ingersoll KS, Nettleman MD, Sobell L, Nagaraja J. Preventing alcohol-exposed pregnancies: A randomized controlled trial. *American Journal of Preventive Medicine*. 2007; 32:1–10. [PubMed: 17218187]
- Fonteyn M, Fisher A. Research corner. Use of think aloud method to study nurses' reasoning and decision making in clinical practice settings. *Journal of Neuroscience Nursing*. 1995; 27:124–128. [PubMed: 7622950]
- Fonteyn M, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. *Qualitative Health Research*. 1993; 3:430–441.
- Gardin FA. The “think-aloud” method to promote student modeling of expert thinking. *Athletic Therapy Today*. 2010; 15(4):18–21.
- Gillam SL, Fargo J, Robertson KC. Comprehension of expository text: Insights gained from think-aloud data. *American Journal of Speech-Language Pathology*. 2009; 18:82–94. [PubMed: 18930910]
- Göransson KE, Ehnfors M, Fonteyn ME, Ehrenberg A. Thinking strategies used by registered nurses during emergency department triage. *Journal of Advanced Nursing*. 2008; 61:163–172. [PubMed: 18186908]
- Grimley DM, Prochaska GE, Prochaska JO, Velicer WF, Galavotti C, Cabral RJ, Lansky A. Cross-validation of measures assessing decisional balance and self-efficacy for condom use. *American Journal of Health Behavior*. 1996; 20:406–416.
- Gwet, K. *Handbook of inter-rater reliability: How to measure the level of agreement between two or multiple raters*. StatAxis; Gaithersburg, MD: 2001.
- Hanson JD, Miller AL, Winberg A, Elliott AJ. Prevention of alcohol exposed pregnancies with non-pregnant American Indian women. *American Journal of Health Promotion*. 2013; 27:S66–S73. [PubMed: 23286666]
- Henry H, Reicks M, Smith C, Reimer K, Atwell J, Thomas R. Identification of factors affecting purchasing and preparation of fruit and vegetables by stage of change for low-income African American mothers using the think-aloud method. *Journal of the American Dietetic Association*. 2003; 103:1643–1646. [PubMed: 14647093]
- Holmstrup ME, Stearns-Bruening K, Rozelle J. Quantifying accurate calorie estimation using the “think aloud” method. *Journal of Nutrition Education & Behavior*. 2013; 45:77–81. [PubMed: 23141076]
- Ingersoll KS, Floyd R, Sobell M, Velasquez MM, Project CHOICES Research Group. Reducing the risk of alcohol-exposed pregnancies: A study of a motivational intervention in community settings. *Pediatrics*. 2003; 111:1131–1135. [PubMed: 12728125]
- Israel BA, Parker EA, Rowe Z, Salvatore A, Minkler M, Lopez J, Halstead S. Community-based participatory research: Lessons learned from the Centers for Children's Environmental Health and Disease Prevention Research. *Environmental Health Perspectives*. 2005; 113:1463–1471. [PubMed: 16203263]

- Jaspers MWM, Steen T, van den Bos C, Geenen M. The think aloud method: A guide to user interface design. *International Journal of Medical Informatics*. 2004; 73:781–795. [PubMed: 15491929]
- Kesmodel U, Wisborg K, Olsen SF, Henriksen TB, Secher NJ. Moderate alcohol intake during pregnancy and the risk of stillbirth and death in the first year of life. *American Journal of Epidemiology*. 2002; 155:305–312. [PubMed: 11836194]
- Leonardson GR, Kemper E, Ness FK, Koplin BA, Daniels MC, Leonardson GA. Validity and reliability of the AUDIT and CAGE-AID in Northern Plains American Indians. *Psychological Reports*. 2005; 97:161–166. [PubMed: 16279320]
- Lupton C, Burd L, Harwood R. Cost of fetal alcohol spectrum disorders. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*. 2004; 127C:42–50.
- MacNeela P, Clinton G, Place C, Scott A, Treacy P, Hyde A, Dowd H. Psychosocial care in mental health nursing: A think aloud study. *Journal of Advanced Nursing*. 2010; 66:1297–1307. [PubMed: 20384643]
- May PA, Miller JH, Goodhart KA, Maestas OR, Buckley D, Trujillo PM, Gossage JP. Enhanced case management to prevent fetal alcohol spectrum disorders in Northern Plains communities. *Maternal and Child Health Journal*. 2008; 12:747–759. [PubMed: 18026824]
- Nieman LZ, Velasquez MM, Groff JY, Cheng L, Foxhall LE. Implementation of a smoking cessation counseling module in a preceptorship program. *Family Medicine*. 2005; 37:105–111. [PubMed: 15690250]
- Pottier P, Hardouin J, Hodges BD, Pistorius M, Connault J, Durant C, Planchon B. Exploring how students think: A new method combining think-aloud and concept mapping protocols. *Medical Education*. 2010; 44:926–935. [PubMed: 20716103]
- Project CHOICES Research Group. Alcohol-exposed pregnancy: Characteristics associated with risk. *American Journal of Preventive Medicine*. 2002; 23:166–173. [PubMed: 12350448]
- Reicks M, Smith C, Henry H, Reimer K, Atwell J, Thomas R. Use of the think aloud method to examine fruit and vegetable purchasing behaviors among low-income African American women. *Journal of Nutrition Education & Behavior*. 2003; 35:154–160. [PubMed: 12773287]
- Sobell M, Sobell L, Johnson K, Velasquez MM, von Sternberg K, Nettleman MD, Sidhu J. Motivational intervention to reduce alcohol-exposed pregnancies—Florida, Texas, and Virginia, 1997-2001. *Morbidity and Mortality Weekly Report*. 2003; 52:441–444. [PubMed: 12807086]
- Trussell J. Contraceptive failure in the United States. *Contraception*. 2011; 83:397–404. [PubMed: 21477680]
- Van Den Haak MJ, De Jong MDT, Schellens PJ. Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*. 2003; 22:339–351.
- Velasquez MM, Ingersoll KS, Sobell MB, Floyd RL, Sobell LC, Von Sternberg K. A dual-focus motivational intervention to reduce the risk of alcohol-exposed pregnancy. *Cognitive and Behavioral Practice*. 2010; 17:203–212. [PubMed: 20473352]



**Figure 1.** Validity and reliability methods.  
*Note.* CHOICES = Changing High-risk alcohol use and Increasing Contraception Effectiveness Study.

**Table 1**Demographics of “Think Aloud” Participants ( $n = 23$ ).

	<i>n</i>	%
Marital status		
Single, never married	9	39.1
Divorced or separated	8	34.8
Married or living with a partner	6	26.1
Education		
High school diploma or GED	9	39.1
College degree (associates, 4-year, or graduate degree)	8	34.8
Some college	4	17.4
Less than a high school diploma	1	4.3
Master's degree	1	4.3
Employment		
Employed	10	43.5
Student	6	26.1
Unemployed	4	17.4
Homemaker	2	8.7
Retired	1	4.3

*Note.* GED = general education development.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

## Comparison of Original and Modified Temptation/Confidence Alcohol Measurements.

<b>OLD Temptation: Alcohol</b>	<b>NEW Temptation: Alcohol</b>
Unpleasant emotions: If I were depressed in general; if everything were going badly for me.	How tempted would you be to drink alcohol if you were feeling stressed, upset, or down in general?
Physical discomfort: If I were having trouble sleeping; if I felt jumpy and physically tense.	How tempted would you be to drink alcohol if you were having trouble sleeping; if you were in pain or physical discomfort?
Pleasant emotions: If something good happened and I felt like celebrating; if things were going well.	How tempted would you be to drink alcohol if something good happened and you felt like celebrating; if things were going well?
Testing control over my use of alcohol: If I started to believe that alcohol was no longer a problem for me; if I felt confident I could handle a few drinks.	How tempted would you be to continue drinking alcohol if you needed to stop drinking, such as after three to four drinks or if you had to drive home?
Urges and temptations: If I suddenly had an urge to drink; if I were in a situation in which I was in the habit of having a drink.	How tempted would you be to drink alcohol if you were in the habit of having a drink; for example, every Friday night?
Conflict with others: If I had an argument with a friend; if I weren't getting along with others at work.	How tempted would you be to drink alcohol if you had an argument with a friend, partner, or family member; if you weren't getting along with others at work or school?
Social pressure to drink: If someone pressured me to be a good sport and have a drink if I were invited to someone's home and they offered me a drink.	How tempted would you be to drink alcohol if someone pressured you to drink; if you were invited to someone's home and they offered you a drink?
Pleasant times with others: If I wanted to celebrate with a friend; if I were enjoying myself at a party and wanted to feel even better.	How tempted would you be to drink alcohol if you were enjoying yourself at a party; if everyone around you was drinking?
<b>OLD Confidence (Self-Efficacy): Alcohol</b>	<b>NEW Confidence: Alcohol</b>
Unpleasant emotions: If I were depressed in general; if everything were going badly for me.	How sure are you that you would NOT drink alcohol if you were feeling stressed, upset, or down in general?
Physical discomfort: If I were having trouble sleeping; if I felt jumpy and physically tense.	How sure are you that you would NOT drink alcohol if you were having trouble sleeping; if you were in pain or had physical discomfort?
Pleasant emotions: If something good happened and I felt like celebrating; if things were going well.	How sure are you that you would NOT drink alcohol if something good happened and you felt like celebrating; if things were going well?
Testing control over my use of alcohol: If I started to believe that alcohol was no longer a problem for me; if I felt confident I could handle a few drinks.	How sure are you that you would NOT drink alcohol if you needed to stop drinking, such as after three to four drinks or if you had to drive home?
Urges and temptations: If I suddenly had an urge to drink; if I were in a situation in which I was in the habit of having a drink.	How sure are you that you would NOT drink alcohol if you were in a situation in which you were in the habit of having a drink; for example, every Friday night?
Conflict with others: If I had an argument with a friend; if I weren't getting along with others at work.	How sure are you that you would NOT drink alcohol if you had an argument with a friend, partner, or family member; if you weren't getting along with others at work or school?
Social pressure to drink: If someone pressured me to be a good sport and have a drink; if I were invited to someone's home and they offered me a drink.	How sure are you that you would NOT drink alcohol if someone pressured you to drink; if you were invited to someone's home and they offered you a drink?
Pleasant times with others: If I wanted to celebrate with a friend; if I were enjoying myself at a party and wanted to feel even better.	How sure are you that you would NOT drink alcohol if you were enjoying yourself at a party; if everyone around you was drinking?

**Table 3**Demographics of “Test–Retest” Participants ( $n = 79$ ).

	<i>n</i>	%
Marital status <sup>a</sup>		
Married or partnered	42	53.2
Single, never married	27	34.2
Divorced or separated	8	10.1
Education <sup>b</sup>		
High school diploma or GED	30	38.0
Some college	23	29.1
College degree (associates, 4-year, or graduate degree)	12	15.2
Less than a high school diploma	11	13.9
Employment <sup>a</sup>		
Employed	31	39.2
Unemployed	21	26.6
Homemaker	14	17.7
Student	7	8.9
Other	4	5.1

Note. GED = general education development.

<sup>a</sup>Missing  $n = 2$ .

<sup>b</sup>Missing  $n = 3$ .

**Table 4**

Agreement Statistics for Alcohol Questions.

Question	Agree (n)	Observation (n)	% Agree	Weighted Kappa	AC1	AC2a	AC2b
Original questionnaire							
Temptation							
Q1	10	16	62.5	.69	.54	.42	.48
Q2	11	16	68.8	.54	.64	.52	.59
Q3	6	16	37.5	.42	.23	.20	.22
Q4	7	15	46.7	.39	.36	.29	.33
Q5	10	16	62.5	.56	.55	.42	.48
Q6	11	16	68.8	.63	.64	.51	.58
Q7	12	16	75.0	.58	.70	.52	.60
Q8	7	16	43.8	.38	.31	.25	.28
Confidence							
Q1	11	16	68.8	.53	.62	.46	.52
Q2	8	16	50.0	.31	.40	.30	.35
Q3	5	16	31.3	.25	.16	.13	.15
Q4	9	16	56.3	.40	.47	.35	.39
Q5	9	16	56.3	.50	.46	.35	.40
Q6	8	15	53.3	.39	.43	.33	.37
Q7	8	16	50.0	.36	.38	.29	.33
Q8	8	16	50.0	.36	.38	.29	.32
Thinking							
Q1	11	16	68.8	.75	.61	.47	.53
Q2	5	16	31.3	.33	.32	.25	.29
Q3	8	16	50.0	.47	.46	.35	.40
Modified questionnaire							
Temptation							
Q1	16	23	69.6	.74	.63	.49	.56
Q2	16	22	72.7	.69	.68	.54	.61
Q3	17	23	73.9	.75	.68	.52	.59
Q4	14	23	60.9	.43	.55	.44	.50
Q5	12	23	52.2	.46	.42	.33	.37
Q6	13	22	59.1	.65	.51	.41	.47
Q7	17	23	73.9	.74	.69	.53	.60
Q8	15	23	65.2	.76	.57	.45	.51
Confidence							
Q1	14	23	60.9	.48	.52	.38	.44
Q2	13	23	56.5	.58	.46	.37	.42
Q3	11	23	47.8	.34	.35	.26	.29
Q4	11	23	47.8	.37	.36	.28	.32
Q5	15	23	65.2	.56	.57	.43	.48

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Question	Agree (n)	Observation (n)	% Agree	Weighted Kappa	AC1	AC2a	AC2b
Q6	10	23	43.5	.34	.30	.22	.25
Q7	12	23	52.2	.41	.41	.31	.35
Q8	13	23	56.5	.39	.46	.34	.38
Thinking							
Q1	8	20	40.0	.41	.39	.29	.33
Q2	12	21	57.1	.61	.54	.42	.47
Q3	11	21	52.4	.40	.43	.33	.38

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5**

Agreement Statistics for Contraception Questions.

Question	Agree ( <i>n</i> )	Observation ( <i>n</i> )	% Agree	Weighted Kappa	AC1	AC2a	AC2b
Original questionnaire							
Temptation							
Q1	11	15	73.3	.62	.68	.52	.60
Q2	12	15	80.0	.61	.76	.59	.67
Q3	9	15	60.0	-.01	.56	.45	.51
Q4	13	15	86.7	.15	.86	.68	.77
Q5	14	15	93.3	—	.93	.75	.84
Confidence							
Q1	10	15	66.7	.64	.59	.44	.50
Q2	11	15	73.3	.77	.67	.51	.59
Q3	8	14	57.1	.52	.49	.38	.43
Q4	10	15	66.7	.56	.60	.46	.53
Q5	12	15	80.0	.71	.76	.59	.67
Thinking							
Q1	9	16	56.3	.71	.79	.63	.71
Q2	10	16	62.5	.63	.63	.50	.57
Q3	9	16	56.3	.72	.62	.49	.56
Modified questionnaire							
Temptation							
Q1	13	20	65.0	.34	.60	.48	.55
Q2	13	21	61.9	-.14	.59	.48	.55
Q3	14	21	66.7	.11	.63	.52	.59
Q4	13	21	61.9	.09	.58	.48	.54
Q5	16	21	76.2	.38	.75	.60	.68
Confidence							
Q1	11	21	52.4	.28	.43	.35	.40
Q2	8	21	38.1	.13	.27	.22	.26
Q3	9	21	42.9	.22	.31	.24	.28
Q4	11	21	52.4	.20	.43	.33	.38
Q5	14	21	66.7	.19	.62	.48	.55
Thinking							
Q1	15	20	75.0	.52	.78	.62	.70
Q2	13	21	61.9	.71	.73	.59	.67
Q3	17	21	81.0	.78	.84	.66	.75