THINK-ALOUD PROTOCOLS: DOES AGE MAKE A DIFFERENCE?

Erica Olmsted-Hawala[1] *and Jennifer Romano Bergstrom*[2]

The think-aloud protocol is one of the main techniques used by usability professionals when conducting usability studies. Two of the most common think-aloud protocols that usability practitioners use today are Concurrent Think-Aloud (CTA) and Retrospective Think Aloud (RTA). In this study, we analyzed whether think-aloud condition and age influenced usability measures of accuracy, efficiency, and satisfaction. We also assessed differences and similarities in verbalizations by age and by think-aloud condition. Results show that of the three age groups (young, middle-age, and older adults), only the middle-age adults in the CTA condition showed effects: they were more accurate for the difficult task and took longer to complete the task. Age did not affect satisfaction ratings. CTA led to a higher percentage of present tense, positive, and affective utterances, whereas RTA led to a higher percentage of past tense, insight and cognitive utterances. The differences in performance by age and think-aloud protocol highlight the importance of including demographic characteristics such as age when reporting results of usability testing. Depending on the goals of the usability study, a researcher might opt to use CTA when interested in obtaining a real sense of the users' experience with the interface. Conversely, if the researcher is interested in gaining a user's insight into what the issues are with a user interface, he/she might opt to do an RTA study.

Introduction: The think-aloud protocol is one of the main techniques used by usability professionals when conducting usability studies. During a think-aloud study, participants talk about what they are thinking as they work on a Web site, complete a survey, or interact with another type of interface. The benefit of using a think-aloud protocol is that once a researcher understands the participant's thoughts pertaining to the task and the specific screen (e.g., Web page), the researcher can identify what usability issues participants have with the interface and can subsequently work on potential solutions. Two of the most common think-aloud protocols that usability practitioners engage in today are:

- Concurrent Think Aloud (CTA), where the participant is encouraged to "think out loud" while working on a task
- Retrospective Think Aloud (RTA), where the participant talks only *after* the session is completed, typically while watching a video replay of his or her session.

Much of the research on think-aloud protocols in usability studies has compared the effects of CTA and RTA on verbal data quality and participant performance as measured by typical usability metrics (i.e., accuracy, efficiency, satisfaction), and the research conclusions vary. Some practitioners have found that a heavily-cued RTA (i.e., with video playback) gives insight into participants' thoughts, but session length doubles (Capra, 2002; Murphy & Norman, 2004; Van Den Haak, De Jong, & Schellens, 2003 and 2004).

Other studies show no performance or satisfaction differences between the two protocols, although there is a difference in the type of usability problems that participants talk about. CTA participants comment more on procedures, and RTA participants give explanatory or design-related comments (Bowers & Snyder 1990; Ohnemus & Biers 1993; Page & Rahimi, 1995). Van Den Haak et al. (2003) found accuracy and efficiency suffered in the CTA compared to RTA; however the usability problems identified were the same. Others have found that there are no (or minimal) differences in performance between participants using RTA and CTA

[1] This report is released to inform interested parties of research and to encourage discussion. Any views expressed on the methodological issues are those of the authors and not necessarily those of the US Census Bureau.

[2] At the time this work was performed, all authors were associated with the Human Factors and Usability Research Group of the U.S. Census Bureau. Jennifer Romano Bergstrom is now at Fors Marsh Group, LLC.

(Capra, 2002; Olmsted-Hawala, Murphy, Hawala, & Ashenfelter, 2010; Van den Haak et al., 2004). Krahmer and Ummelen (2004) summarize that most CTA leads to longer performance time on a specific task than RTA, while the accuracy of the task is not affected. Thus, there are many differences that emerge across the existing studies, and there is no clear indication which protocol is more desirable (that is, which protocol leads to a rich insight of the participants' thoughts and does not impact accuracy, efficiency, and satisfaction).

In our work at the Census Bureau's Human Factors and Usability Research Group Laboratory, we conduct usability studies with potential users of our products so we can identify what works well and what needs to be modified to better communicate with our users. During usability testing, as the participants work on their tasks, we typically have them follow a communicative CTA protocol (Boren & Ramey, 2000; Olmsted-Hawala et al., 2010). We often recruit older users to participate in our studies, so that we can ensure our sites are usable for people of all ages. This is particularly important because the US population is aging (US Census Bureau, 2009), and the use of the Internet by older adults is rapidly increasing (Madden, 2010). However, it is well known that performance for older adults decreases when they attempt to complete dual tasks (e.g., Hartley, 1992). Because of this, we were concerned that the dual task of thinking aloud in the CTA protocol while searching for information would lead to increased errors in performance. No studies to date have compared the different think-aloud protocols and the effects of aging. Thus, in this study, we examined the effects of think aloud and age on usability metrics (i.e., accuracy, efficiency, satisfaction), measures of performance that are common in usability testing. Using a linguistic analysis and word count tool, we also examined the quality of the verbal report given during the think-aloud protocol. It has been documented that the words people use often give insight into their mental and physical health (Stiles, 1992). We believed the analysis of the verbal reports could provide some understanding of the differences in the two conditions and highlight age-related differences. This analysis may be of particular interest to technical communicators who seek to understand what users are thinking and feeling as they work with a user interface. An outcome of such analysis may give insight to a technical communicator to identify pain points and potentially gain insight into how to address the design problems of a user interface while at the same time identify ways to increase the positive sense of the users' experience with the application.

METHODOLOGY

To test whether the think-aloud mode affected usability metrics differently by age group, we designed a 2 (Think Aloud) x 3 (Age) between-subjects study in which participants thought they were participating in a usability study but were, in fact, participating in an experiment (i.e., a putative usability study). We analyzed how age and think-aloud protocol were related to usability performance measures (accuracy: task completed correctly or not, efficiency: time on task, and subjective satisfaction ratings). In addition, we examined the effects of age and think-aloud protocol on the quality of verbal reports. In particular, when conducting usability tests with older adults, we wanted to know whether it is more beneficial to have older adults communicate what they are thinking about retrospectively or concurrently.

Ninety-five users from the metropolitan Washington, DC area participated. Young adults (ages 18-28), middle-age adults (ages 40-50) and older adults (ages 64-76) were assigned to a CTA condition or to a RTA condition. We identified the age groups prior to the study, intentionally selecting age ranges that were far enough apart to highlight any differences. All participants reported being experienced with computers and the Internet, and all reported being unfamiliar with the Web site tested. See Table 1 for demographic information.

	Young adults		Middle-age adults		Older adults	
	CTA	RTA	CTA	RTA	CTA	RTA
Gender	3M / 8F	6M / 15F	5M / 7F	9M / 10F	5M / 6F	9M / 12F
Age	22	22	45	46	67	68
Years of education	15	15	15	14	16	16
Ease of learning a new Web site*	1.36	1.47	1.83	1.42	2.36	2.38
Ease of navigating the Internet*	1.09	1.05	1.33	1.16	1.72	1.57

Table 1. Participants' Mean Characteristics

*Scale: 1 (Not difficult at all) – 5 (Extremely difficult).

Participants worked on five information-gathering tasks on the legacy[3] version of the US Census Bureau's American FactFinder (AFF) Web site. AFF is the Census Bureau's primary data dissemination Web site about the population, housing and economy of the United States. The tasks stem from typical tasks that general users of the Website come to the site to accomplish. Two tasks we categorized as easy, and three tasks we categorized as hard. See Appendix A for a list of the task questions. The participant and the test administrator sat in separate rooms during the sessions and communicated via microphones and speakers. Although participants in the RTA condition completed all five tasks, due to time constraints, they only spoke retrospectively about the last task. The other four tasks were completed in silence, without a retrospective think aloud.

RESULTS

Usability Metrics: We examined the mean accuracy (across all tasks) and the mean efficiency for correctly completed tasks only. For satisfaction, we examined responses to a modified version of the Questionnaire for User Interaction Satisfaction (QUIS) that included 11 questions (Chin, Diehl, & Norman, 1988).

First we examined accuracy. A one-way ANOVA comparing average accuracy among groups showed significant age-related differences between young adults (63% task accuracy) and older adults (42% task accuracy) ($p < 0.01$) and between middle-age (57% task accuracy) and older adults ($p < 0.05$), such that young and middle-age adults completed more tasks successfully than older adults, while there were no differences between young and middle-age adults. We compared accuracy by think-aloud protocol, and we found no significant difference between CTA and RTA for any of the groups. Thus, although older adults completed fewer tasks successfully than both young and middle-age adults, think-aloud protocol did not affect overall accuracy for any of the groups. See Table 2 for mean usability metrics.

Next we examined accuracy by task difficulty. We examined the first easy task and the first hard task that participants encountered and attempted to complete. For young and older adults, a one-way ANOVA showed no significant differences in accuracy by think aloud condition for either task. However, for middle-age adults, a one-way ANOVA comparing think aloud condition showed a significant difference in accuracy for the difficult task only, such that the middle-age adults in the CTA condition correctly completed significantly more tasks (75%) than their RTA counterparts (42%) ($p = 0.07$). Thus, there was a difference in think-aloud condition by age for middle-age adults for accuracy such that CTA yielded higher accuracy for the difficult task but not for the easy task, and we did not find this same pattern for young or older adults.

[3] In early 2012, a new American FactFinder (AFF) was released. The study described in this paper was conducted on the old AFF site. The old site is no longer available online and is referred to as the "legacy" version.

Next we examined efficiency. Across both think-aloud conditions, a one-way ANOVA showed a significant difference between young adults (155 seconds) and older adults (223 seconds) ($p = 0.05$), such that young adults completed tasks faster than older adults. There were no differences between middle-age (193 seconds) and young adults or between middle-age and older adults. We compared efficiency by think-aloud condition and found that overall, there was no difference between CTA and RTA for young and older adults. However, for middle-age adults, a one-way ANOVA showed a trend, such that those in the RTA condition (163 seconds) completed tasks faster than their CTA counterparts (235 seconds) ($p = 0.10$).

We also examined efficiency by task difficulty and found no differences in performance by think-aloud condition for the easy or hard tasks for young and older adults. However, for middle-age adults, a one-way ANOVA comparing think-aloud condition showed a significant difference in efficiency for the difficult task only, such that the middle-age adults in the RTA condition completed the task faster (271 seconds) than their CTA counterparts (576 seconds) ($p < 0.05$). Thus, there was a difference in think-aloud condition for middle-age adults for efficiency such that RTA yielded faster completion time for the difficult task but not for the easy task, and we did not find this same pattern for young or older adults.

For subjective satisfaction score, across think-aloud conditions and across all satisfaction items, a one-way ANOVA comparing age group showed no significant age-related differences. Because each statement measures satisfaction with a specific aspect of the interface, we examined each statement separately. A one-way ANOVA comparing groups showed a significant difference between middle-age and older adults on two (out of 11) aspects of the site and a significant difference between young and older adults on one item, such that older adults rated their satisfaction on the two items significantly lower than middle age and younger adults. The two aspects were "Organization of information on the site (confusing – clear)" ($p = 0.10$, middle-age vs. older adults) and "Overall experience of finding information (difficult – easy)" ($p < 0.05$, middle-age vs. older adults; $p = 0.05$, young vs. older adults). However, when we examined satisfaction by think-aloud condition, we found no differences for any of the groups. Thus, although older adults reported lower satisfaction for two items, think-aloud protocol did not affect satisfaction for any of the groups. See Table 2 for accuracy, efficiency and satisfaction ratings by age group and think-aloud condition.

	Young adults		Middle-age adults		Older adults	
	CTA	RTA	CTA	RTA	CTA	RTA
Accuracy	75%	62%	65%	50%	44%	41%
Efficiency	155 s	153 s	235 s	163 s	215 s	226 s
Satisfaction*	5.14	4.46	5.39	4.91	5.14	4.42

Table 2. Accuracy, Efficiency and Satisfaction Ratings by Age Group and Think-Aloud Condition
* Scale 1 to 9: 1 = terrible, 9 = wonderful.

Verbal Quality: During testing, we noticed that participants in the CTA condition seemed to speak more about what they were presently doing, and participants in the RTA condition stayed "on task" less often but provided more insight about their experience. We also noticed that older adults appeared to speak less often and needed more prompts by the test administrator. However, these were merely impressions; to test whether there were differences in verbal comments by age group and by think-aloud condition, we first had a colleague, who was unaware of the objectives of the study transcribe verbatim the CTA and the RTA sessions. Then, using a random number generator (Random Number, 1998) we randomly selected five of the verbatim transcriptions by age group and by think-aloud condition for task 5 and used the Linguistic Inquiry and Word Count (LIWC) software (1997) to analyze the text.

The LIWC software is a text analysis program that calculates the degree to which participants verbalize, among other things, positive or negative emotions during the session (Pennebaker & Francis, 1999). LIWC quantitatively codes words based on an internal dictionary. For example, words identified as carrying positive emotions include "yes," "like," and "good," and words identified as carrying negative emotions include "no," "dislike," and "bad." For these findings we do not have a measure of variation associated with each percentage, therefore we can't reliably say whether these differences are significant or not. Future work will

take into account the variation and test for statistical significance; for now, we highlight what we noticed in the results.

First we examined the number of present and past tense utterances to ascertain that the LIWC tool would work effectively with the verbalizations. We found, as we expected, that across all groups, participants in the CTA condition spoke a higher percentage of present tense words than participants in the RTA condition. Within the CTA condition, young adults used about 5% less present tense words than the other two age groups. As well, also as expected, participants in the RTA condition spoke a higher percentage of past tense words than participants in the CTA condition. Within the RTA condition, there was a similar pattern of young adults using less present tense words than the other two age groups. See Table 3.

Next we examined words that we thought would be useful in identifying good usability: positive emotions. We found that across all age groups, participants in the CTA condition spoke a greater percentage of positive emotion words (e.g., clever, hopeful, smart) and words of assent (e.g., absolutely, awesome, cool) than participants in the RTA condition. Within the CTA condition, older adults used twice as many words of assent than both young and middle-age adults.

We also examined words that would be useful in identifying poor usability: negative emotions. We found that negative emotion words (e.g., annoy, awful, bad) and negations (e.g., aren't, can't, doesn't) occurred pretty comparably across the two conditions, though there were some age group differences. Across the two think-aloud conditions, young adults in the CTA condition used slightly more negative emotion words than their RTA counterparts, however they used slightly fewer negations in the CTA condition than in the RTA condition. The use of negations is different by think-aloud condition for the middle-age and older adults, such that for both age groups, they use negations more often in the CTA condition than in the RTA condition.

Next, we examined insight words. We found that, across all age groups, participants in the RTA condition spoke a higher percentage of insight words (e.g., complex, meaning, prove) than participants in the CTA condition. This result is consistent with what the test administrators observed during testing. Similarly, words that had to do with cognitive processes (e.g., ambiguous, hesitate, solution) occurred for all age groups more often in the RTA condition than in the CTA condition. For the RTA condition, the older adults used words that had to do with cognitive processes slightly less often than the other two age groups. The percentage of words that had to do with perceptual processes (e.g., heavy, sight, noisy) occurred fairly even across think-aloud conditions and age groups.

Next, we examined words that had to do with affective processing because this we believed could lend insight into a user's attitude or mood as they interacted with the site. The percentage of words that had to do with affective processes (e.g., caring, clever, insult) occurred slightly more often in all age groups in the CTA condition than in the RTA condition. Within the think-aloud conditions, the middle-age adults used affective language slightly more often than the other two age groups.

Finally, we examined words that had to do with work because we felt there might be differences with respect to the focus of the participant on the work element of finding an answer to their task question in the CTA and the RTA condition. The percentage of words that had to do with work (e.g., employ, exam, wage) occurred more often in the CTA condition than in the RTA condition (Table 3). This appears to reflect what the test administrators observed during the sessions: that the participants in the CTA condition seemed to stay "on-task" more consistently than their RTA counterparts.

	Young adults		Middle-age Adults		Older adults	
	CTA	RTA	CTA	RTA	CTA	RTA
Past tense	1.6	9.2	2.3	7.2	2.3	11.3
Present tense	9.9	5.8	15.1	9.6	14.6	7.4
Negations	1.4	2.5	3.1	1.8	2.7	1.9
Negative emotions	1.2	0.7	1.1	1.1	0.7	0.6
Positive emotion	2.2	1.4	2.3	1.6	2.1	1
Assent	1	0.6	1.5	0.6	3.1	0.6
Insight	3.3	5.3	3	5.4	2.3	5
Affective processes	3.3	2	3.4	2.5	2.8	1.6
Cognitive processes	19.3	22.6	16.2	23.6	17.1	19.4
Perceptual processes	4.9	3.8	4.3	3.5	4.4	4.4
Work	5.8	2.1	4.9	2.3	4.6	2.9

Table 3. Percentage of Word Type by Age and Think-aloud Condition

DISCUSSION

Our study leads to five primary conclusions. First, young and middle-age adults completed more tasks successfully (higher accuracy) than older adults. Second, young adults completed tasks faster than older adults. Third, age did not affect satisfaction. Fourth, middle-age adults who completed tasks in silence (i.e., the RTA condition) completed the most difficult task faster and with higher accuracy than middle-age adults who thought aloud while completing tasks. Finally, thinking aloud concurrently led to a higher percentage of present tense, positive, and affective utterances, whereas thinking aloud retrospectively led to a higher percentage of past tense, insight and cognitive utterances.

Our efficiency finding with middle-age adults is consistent with the literature that indicates cognitive demands influence the speed with which adults process information (Bashore, Ridderinkhof, & Molen, 1997) and that dual task interferes with performance (Hartley, 1992). We hypothesize that we did not find the identical pattern of results with the older adults because the older adults were performing at a floor level (performing very slowly), and conversely young adults were performing at a ceiling level (performing very quickly), whereas middle-age adults, who have a range of speed, demonstrated an effect of the think-aloud condition. Future research should include wider ranges of age to test the assertion that the age-related effect is incremental and occurs slowly through a lifetime.

For middle-age adults, on the hard task only, thinking aloud concurrently led to higher accuracy and slower task completion times, while thinking aloud retrospectively led to lower accuracy and faster completion times. This speed accuracy trade-off is often reported in the cognitive aging literature (e.g., Rabbitt, 1979; Salthouse, 1979), though it has not been described in prior think-aloud studies. Further work is needed on age-related differences with speed and accuracy trade-offs in CTA and RTA usability studies.

In an earlier study (Olmsted-Hawala et al., 2010) conducted in the same lab, the authors found no differences in accuracy, efficiency, or satisfaction when comparing a silent control to a speech-communication think-aloud protocol. However, in the 2010 study, the authors did not examine age-related differences. In the present study, we found that when comparing CTA to a silent condition (i.e., the RTA condition), middle-age adults, on

the difficult task only, performed differently in terms of accuracy and efficiency. This suggests that age influences usability metrics when using a think-aloud protocol in usability testing. Our findings for middle-age adults are in contrast for accuracy, and they are in line for efficiency with Van Den Haak et al. (2003) who found (without considering age) that CTA degraded accuracy and efficiency. The differences in performance by age and think-aloud protocol highlight the importance of including demographic characteristics, such as age, when reporting on usability testing.

Over the years, researchers have identified a connection with the words people use to predict their mental and physical health (Gottschalk & Gleser, 1969; Stiles, 1992). For example, a higher use of positive words predicted a person would be healthy, while the use of more negative emotional words predicted a person would less healthy. So too we believed that the use of language would vary depending on whether the language was spoken at the moment of working on the Web site, or later in a retrospective review of the task. As the LIWC output demonstrated, for some categories of analysis there were no differences, while for other categories, such as positive emotions and words of insight, there was a difference. Depending on the goals of the usability study, a researcher might opt to use CTA when interested in getting a real sense of a user's experience with the interface. Conversely, if a researcher is interested in gaining a user's insight into what the issues are with a user interface, he/she might opt to do an RTA study.

One of the challenges when proctoring the RTA session with video playback is to ensure that the participant speaks about what they were thinking about as they were doing what is displayed in the video replay. The test administrators in this study noticed that in general, the RTA participants appeared at times to be speaking about something that occurred later in the task or of something that was apparently off the topic of the task itself. When we examined the LIWC results, the retrospective condition, regardless of age group, demonstrated that the percentage of words that had to do with cognitive processes and insight occurred more often in the RTA condition. This finding is in line with previous studies (Bowers & Snyder, 1990; Ohnemus & Biers, 1993; Page & Rahimi, 1995) that have shown that RTA participants tend to give explanatory or design-related comments. In contrast, words that had to do with participants' emotions, such as their emotional reaction to the user interface of the Web site, occurred more often in the CTA condition.

We cannot determine whether CTA or RTA is "right," but we can say that each yields different results. Overall, it is good practice, when the test administrator communicates with the participant, to use the same communication with *all* participants or the data may be invalid (at least for middle-age adults). Researchers must decide which is more important—the usability metrics or the insight we receive from participants who think aloud while they complete tasks during usability studies.

Suggested applications for professional technical communicators doing usability testing:

- Recruit participants of varying age groups for usability testing; conduct usability tests with older adults.
- Consider the pros and cons of both methods when creating the protocol and deciding on a think-aloud mode.
- Be aware of the cognitive demands and the effects that thinking aloud has on adults as they age.
- Consider using CTA for a richer verbal commentary, particularly with respect to participants' emotional reaction to the site.
- Consider using RTA when interested in verbal feedback related to participants' insight on what they felt the issues were that they were having with the application.

APPENDIX A

Task level: Easy

1. You want to learn more about Maryland, and specifically about how many people live there. How many people live in Maryland?

2. You are interested in knowing about countries the US purchases products from. What percent of U.S. imports currently come from China?

Task level: Difficult

3. You are doing a report on schooling in the U.S. What percent of the population in Florida, California and Texas completed college in 2008?

4. You are doing research for a project on poor people in the U.S. and want to know which three states had the most poor people in 2008. What are the three states?

5. You are working on a project that involves work environments in the US, and you are interested in the history of coal mining. How many coal-mining companies were in the US in 2007?

REFERENCES

Bashore, T., Ridderinkhof, K., and Molen, M. "The Decline of Cognitive Processing Speed in Old Age." Current Directions in Psychological Science, 6.6 (1997): 163-169.

Boren, T., and Ramey, J. "Thinking aloud: Reconciling theory and practice." IEEE Transactions on Professional Communication. 43.3 (2000): 261-278.

Bowers, V., and Snyder, H. "Concurrent versus retrospective verbal protocol for comparing window usability." (Proceedings of the Human Factors Society 34th Annual Meeting, 1990): 1270-1274.

Capra, M. "Contemporaneous versus retrospective user-reported incidents in usability evaluation." (Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, 2002): 1973-1977.

Chin, J., Diehl, V., and Norman, K. Development of an instrument measuring user satisfaction of the human-computer interface. (Proceedings of CHI 88 ACM Press 1988): 213-218.

Ericsson, K. and Simon, H. Protocol analysis: verbal reports as data. (Cambridge, MA: MIT Press), 1984.

Gottschalk, L., and Gleser, G. The measurement of psychological states through the content analysis of verbal behavior. (Berkeley, CA: University of California Press), 1969.

Hartley, A. "Attention." In F.I.M. Craik and T.A. Salthouse (Eds.), The handbook of aging and cognition (Hillsdale, NJ: Erlbaum), 1992. pp. 3-50.

Krahmer, E., and Ummelen, N. "Thinking about thinking aloud: A comparison of two verbal protocols for usability testing." IEEE Transactions on Professional Communication. 47.2 (2004): 105-117.

Linguistic Inquiry and Word Count (LIWC). Text analysis software. (2007). http://www.liwc.net/.

Madden, M. "Older adults and social media: Social networking use among those ages 50 and older nearly doubled over the past year." Pew Internet & American Life Project: Pew Research Center. (2010). http://pewinternet.org/Reports/2010/Older-Adults-and-Social-Media.aspx.

Ohnemus, K., and Biers, D. "Retrospective versus Concurrent Thinking-Out-Loud in Usability Testing." (Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting, 1993): 1127-1131.

Olmsted-Hawala, E., Murphy, E., Hawala, S. and Ashenfelter, K. "Think-Aloud Protocols: A Comparison of Three Think-Aloud Protocols for use in Testing Data Dissemination Web Sites for Usability." (Proceedings of CHI 2010, ACM Conference on Human Factors in Computing Systems. ACM Press 2010): 2381-2390.

Page, C., and Rahimi, M. (1995). Concurrent and retrospective verbal protocols in usability testing: Is there value added in collecting both? Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting, 223-227.

Pennebaker, J. and Francis, M. "Linguistic Inquiry and Word Count (LIWC): A text analysis program." (Mahwah, NJ: Erlbaum Publishers), 1999.

Rabbitt, P. "How old and young subjects monitor and control responses for accuracy and speed." British Journal of Psychology, 70 (1979): 305–311.

Random Number Generator. (1998). http://www.random.org/.

Salthouse, T. "Adult age and the speed–accuracy trade-off." Ergonomics, 22.7, (July 1979): 811-821.

Stiles,W. Describing talk: A taxonomy of verbal response modes. (Newbury Park, CA: Sage): 1992.

US Census Bureau. Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2000 to July 1, 2009 (NST-EST2009-01). Population Division, US Census Bureau. (December 2009). http://www.census.gov/popest/data/historical/2000s/vintage_2009/index.html.

Van Den Haak, M., De Jong, M., and Schellens, P. "Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue." Behaviour & Information Technology, 22.5 (2003): 339-351.

Van Den Haak, M., De Jong, M., and Schellens, P. "Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: A methodological comparison." Interacting with Computers, 16.6 (2004): 1153-1170.

AUTHOR CONTACT INFORMATION

Erica Olmsted-Hawala
User Experience Researcher
U.S. Census Bureau
Washington DC, 20233
301.763.4893

Jennifer C. Romano Bergstrom
Senior Research Associate
Fors Marsh Group
1010 N Glebe Rd, Suite 510
Arlington, VA 22201
571.858.3795

AUTHOR BIOGRAPHIES

Erica Olmsted-Hawala has an M.A. in technical communication. She has been working at the Human Factors and Usability Research Lab in the directorate of Research and Methodology at the U.S. Census Bureau for over 11 years. At the usability lab she leads studies on the Census Bureau's Web dissemination site and on electronic data collection instruments. Her research interests include think-aloud protocols in usability testing, strategies for communicating usability results, interface displays for Web sites, and the usability of handheld devices to collect survey data.

Dr. Jennifer Romano Bergstrom has over 12 years of experience planning, conducting and managing user-centered research projects. At Fors Marsh Group, she is responsible for leading user experience (UX) research. Jen specializes in experimental design, quantitative analysis, and usability for older users. She also teaches training courses in usability, accessibility, information architecture, search engine optimization and writing for the Web. Prior to joining Fors Marsh Group, Jen completed a post doc at the US Census Bureau where she conducted numerous usability studies, many of which included examining age-related differences in Internet performance and improving the usability of Web sites and Web-based surveys for older adults. Prior to working at the Census Bureau, Jen studied cognitive aging and lifestyle factors, such as bilingualism and piano-playing that promote healthy cognition in old age.