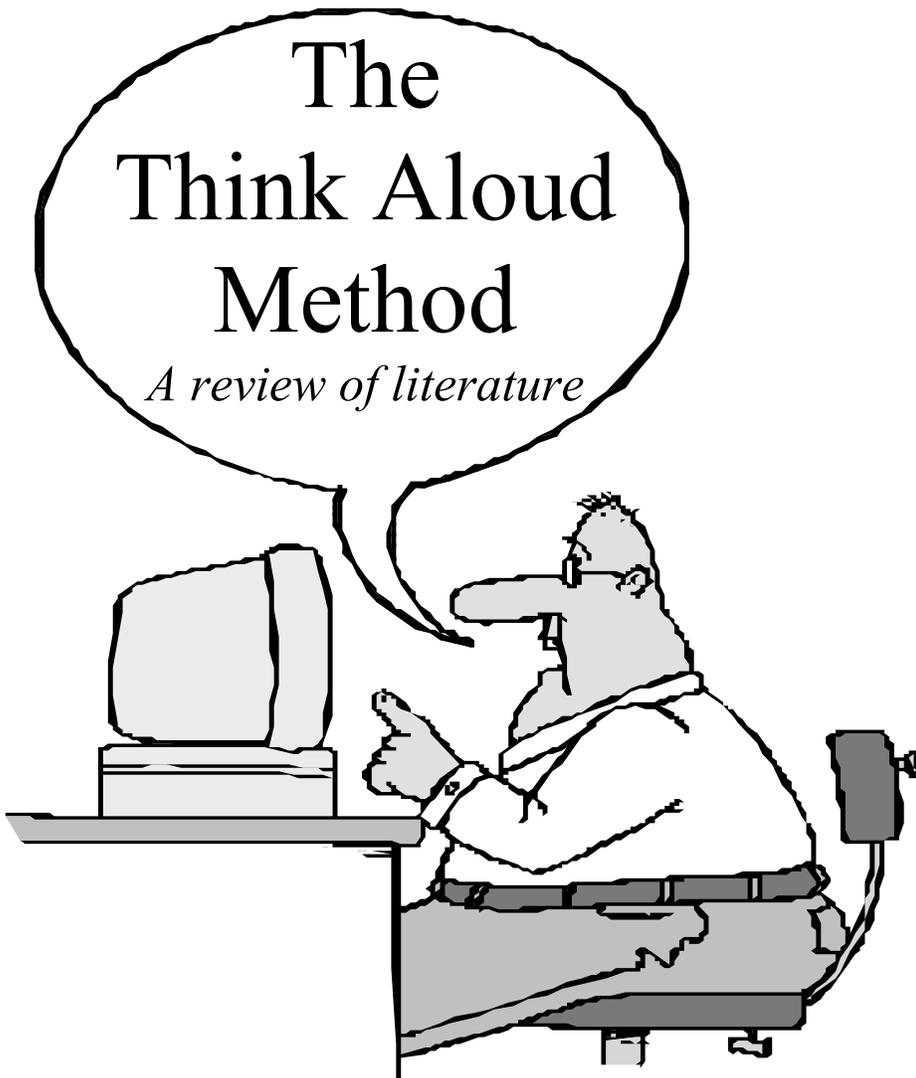


The Think Aloud Method

A review of literature



Definition of Think Aloud

“The Think Aloud Method consists of asking people to think aloud while solving a problem and analyzing the resulting verbal protocols.” (Someren 1994).

Introduction

Before we plunge into the review of literature, I'd like to do a little “Think Aloud” thought experiment to set the scene:

We're going to tap into the silent internal dialog that often comes with thinking. Here's the thought experiment.

Engage your mind's dialog as you begin this challenge: Find a blue thumbtack. Silently tell yourself why you choose each step it would take to find such an item.

My example goes like this. "OK, I'm checking the desk because there's always a few things around the bottom of the drawer. Nope, not there, damn. I could ask a neighbor, but I'm going to go to the store because I need some anyway to hang photos."

Notice I'm focusing on not only what I do, but also why I made those decisions, which is part of what I was thinking. Vocalize this process and you've got the basic idea of Think Aloud. Now imagine vocalizing what you're thinking while you attempt to complete a task on a Web site. This is usability testing for Web sites using Think Aloud.

Our mind is not like a brilliantly lit and perfectly ordered room; it is much more like an encumbered garret inhabited by moths born and grown up in half lights: our thoughts; the moment we open the door to see them better the drab little butterflies vanish. (Dimmet 1928, p 1)

Given this scenario the Think Aloud method is like a butterfly catching net.

While this method may seem simple, and, like a hammer, used more than studied this was not always so.

History

The foundation of Think Aloud is what psychologists in the 1920s and 1930s referred to as the “introspection method.”

“Introspection is based on the idea that one can observe events that take place in consciousness more or less as one can observe events in the outside world.” (Someren, 1994 p.29).

In 1920s several researchers — each working more or less independently — asked subjects to “think” or “talk” aloud when facing a given problem. J.D. Watson, an early behaviorist documented the first analysis of an experiment using thinking aloud, or “Think Aloud,” as a research tool. His findings were the beginning of the Think Aloud method breaking away from the introspection method. He summarized his findings with the following revelation:

The present writer has often felt that a good deal more can be learned about the psychology of thinking by making subjects think aloud about definite problems, than by trusting to the unscientific method of introspection. (Watson quote in Ericsson & Simon, p.59)

But the history of Think Aloud is not without controversy. E. Titchener and Willam James, and other popular psychologists of the time took the idea of introspection to extremes. They asserted that trained psychologists could observe their own thought processes by using the Introspection Method on themselves to gather data — just as physicists may throw a rock in the air to study the Law of Gravity.

One of the main problems with the Introspection Method was that it was, simply, the mind of a psychologist studying the mind of a psychologist. For this

reason, it is inherently subjective and biased because the process of introspection does not separate from the process of thinking. If we imagine an early psychologist sitting in a chair introspecting silently we can immediately see the difficulty of trying to use this method as a scientific tool. As soon as valuable data was gathered and shared with other scientists they would find it impossible to reproduce the results.

These problems of the process not being separate from the thinking and of reproducibility soon became evident. This led to widespread doubt that the Introspection Method could be used as a reliable research tool. Such trepidation continues to this day. As Someren, (1994) explains in the book *The think aloud method. A practical guide to modeling cognitive process.* “The history of the Introspection Method in psychology make psychologists suspicious of methods that resemble introspection.” (p.30). Such suspicion still exists, but due to computers it is steadily diminishing for certain type of research.

The increased use of computers in the late 1960s renewed interest in the study of cognitive processes. Computers offered a new, efficient means of processing verbal data. This new ability naturally caused computer scientists to re-examine the value of the Think Aloud Method as a way to gather verbal data about cognitive processes.

The definitive work on this subject, *Protocol Analysis: Verbal Reports as Data.* was published in 1984, just as computer science began building expert systems and relying more and more on Think Aloud techniques. The only book in the San Diego State library with “Think Aloud” in its title was published ten years later, in 1994. This book, *The Think Aloud Method: a Practical Guide for Modeling Cognitive Processes.* explains that, in the mid-eighties, computer scientists quickly discovered that the Think Aloud Method was one of the best ways to uncover critical thought processes of subject matter experts (SMEs).

They used data from Think Aloud testing, and their new computers, to build models of the experts' knowledge. These models were then used to build computers that attempted to simulate the way the SMEs made decisions. The computer scientists recognized that "...it became possible to obtain knowledge of which the expert was not aware and the free verbal format avoids distortions and misrepresentations." (p. 32)

Today, Think Aloud plays an important role in educational research, as a tool for studying both teacher and student thought processes (Wittrock 1986). Think Aloud is also a popular method in the scientific communities of psychology and computer science. It is a staple method in usability laboratories across the world. In fact, according to some researchers, Think Aloud is often described as the most widely used evaluation method in the computer industry (Jacob 1998).

Strengths & Weaknesses of the Think Aloud Method

Smilowitz (1994) exemplifies the strengths and weakness of the Think Aloud Method in a study titled, “Are we Overlooking some Usability Testing Methods: A Comparison of Lab, Beta and Forum Tests.” This study involved a piece of productivity software and a group of employees within a large corporation. To gather feedback, the researchers used 1) Think Aloud in a lab setting, 2) take-home questionnaires for beta testing and 3) a discussion forum on an intranet-based bulletin board system (BBS). Results showed that the lab tests were significantly better at uncovering usability problems than the forum tests. The beta and lab testing were equally effective, except in one crucial area--in which the lab prevailed. By using the Think Aloud Method in the lab, the researchers were able to uncover many more high severity problems (Smilowitz 1994).

Several other studies illustrate the effectiveness of the Think Aloud Method as well. One study, called “Preventing Usability Disasters,” proposed a system for preventing user interface disasters (UIDs). Researchers successfully used the Think Aloud as the primary tool for uncovering UIDs potential (Molich 1994).

A third paper by (Cennamo 1995) described the application of the Think Aloud Method to find out how learner's process videotaped content. In this study, the subjects briefly watched the videotape. The researchers then stopped the tape and asked the subjects to verbalize their thoughts. The Think Aloud Method was successful at providing information on the cognitive processing of video taped information.

An anecdotal example of the value of Think Aloud can be found in Lewis & Rieman's, Task-Centered User Interface Design: A Practical Introduction (1993).

One of Clayton's favorite examples of a thinking-aloud datum is from a test of an administrative workstation for law offices. This was a carefully designed, entirely menu-driven system intended for use by people without previous computing experience. The word "parameter" was used extensively in the documentation and in system messages to refer to a quantity that could take on various user-assigned values. Test users read this word as "perimeter", a telling sign that the designers had stepped outside the vocabulary that was meaningful to their users." (Lewis 1993 see URL).

This literature paints a very rosy picture of the Think Aloud Method, especially for problem solving (Someren 1994). But Think Aloud has limitations when used for other purposes.

Someren (1994) summed up one key limitation to the Think Aloud method:

"...if the information is nonverbal and complicated then verbalization will not only cost time, but also space in working memory because it becomes a cognitive process by itself. This will cause the report of the original process to be incomplete and it can sometimes even disrupt this process." (p. 33).

In other words, the Cennamo, (1995) study that used Think Aloud to analyze how users processed video tape could not have subjects think aloud while they watched the tape because watching video takes up too much working memory.

Working Memory and the Roll of the Observer

Patrick Jordan (1998) writes that the Think Aloud method is asking participants to perform two tasks. The first being the task they are performing for the researcher, and the second being the task of talking about their thoughts. Jordan suggests that it is quite possible that the tasks can clash. This can manifest itself as difficulties in the first task caused by having part of working memory being used taken up by verbalize their thoughts.

The essential concept here is that verbal reporting is the verbalization of information your mind has stored in working memory (Someren 1994). Researchers have long been aware of this and have invented protocols that tax working memory as little as possible. One common protocol is to have users practice thinking aloud on a mock task before they begin the real task (Cennamo, 1995).

Another protocol is to have the observer prompt the subjects in ways that do not ask "Why" questions. Such questions encourage subjects to think even harder about an action, and often end up prompting answers that are inaccurate, or invented just to answer the question.

Examples of effective prompting methods are "keep talking" and "what are you thinking?", instead of "Why did you do that?"

The final protocol used to avoid taxing working memory is one of good old avoidance. Simply put, don't use the Think Aloud method to study tasks that involved a lot of cognitive load. An example would be solving a math problem vs. finding the price of a

book on a Web site. Solving a math problem involves constant prolonged thought that would be better suited for retrospective focused interviews using analysis of video tape showing the scratch paper, and facial expressions, of the subject solving the math problem. Finding the price of a book on a Web site involves a tiny bit of working memory to keep the "find price" in working memory. It is this kind of task that the Think Aloud Method is best at.

Brock Allen

Comment: This conclusion and some of the preceding thinking on WM effects and think aloud protocols deserve some expansion because WM effects are so critical to usability in general.

Three methods

Below are examples of three different Think Aloud methods. They start simple and increase in complexity. The first method, "guerilla Think Aloud", is the simplest, The next "most bang for the buck" is more complicated, but still not complex scientific based protocol. The last "for Profs. and pros" is a true data crunching protocol.

These examples are oriented towards the CEDAR project and thus focus on the literature as it relates to Web site and software testing.

Method #1: Low Level, AKA, "guerilla Think Aloud"

Jacob Nielsen, (1994) the current guru of Web site usability (Richtel 1998) is a big believer in amateur usability testing as an efficient way to gather large amounts of valuable data. Nielsen (1994) states that:

It is important to realize that it is possible to achieve major improvements in usability even if one does not utilize all the most advanced techniques and even if one has a fairly primitive usability lab (or sets up a temporary lab in a conference room as described by Szczur). The single-most important decision in usability engineering is simply to do it! (p. 7)

As we will see in the next method using 3 to 5 subjects has been shown to be sufficient for gathering very valuable data (Nielsen, 1994).

This is especially true for iterative design of Web sites and software, during which developers can't stop the design process for hours, much less days, and wait for a usability lab to do a through study.

Given this, I propose a very simple "guerilla" form of usability testing. Here's a basic example:

1. Post your project on the Web.
2. Run around the office asking folks to visit your site do a task while talking aloud about what they are thinking as they attempt to complete the task.
3. Take notes
4. Go back and build another version taking into consideration what you found.

It's not pretty, but it will flush out major problems in a short amount of time, using very little resources.

Method #2: Mid-Level, AKA, "most bang for the buck"

Nielsen's (1994) "Heuristic Evaluation" is a good example of a mid-level method for Think Aloud testing.

Heuristic Evaluation part of Nielson's "discount usability engineering methods for quick, cheap, and easy evaluation of a user interface design."

This systematic technique is for iterative design processes and uses a small set of subjects who judge interfaces by a set of usability heuristics. This technique uses the Think Aloud method with videotape, audio recordings, or written notes by isolated subjects.

A quote from Nielsen, (1994) explains the immense value of such a mid-level method.

Much user testing is simply aimed at generating qualitative insights that are communicated through lists of usability problems and highlights videos showing striking cases of user frustration. Such insights are sufficient for most practical usability engineering applications where the goal is the improvement of a user interface through iterative design. Often, there is no real reason to expend resources of gathering statistically significant data on a user interface that is known to contain major usability problems and has to be changed anyway. By using a probabilistic model of the finding of usability problems and an economic model of project management, Nielsen and Landauer (1993) found that one often gets the optimal cost-benefit ratio by using between three and five test users for each confirming the general observation that heuristic evaluation seems to work best with three to five evaluators. In the example, a heuristic evaluation with four evaluators would cost \$6,400 and would find usability problems worth \$395,000. (???)

Another example of a mid-level method in action is Szczur, (1994) who used the Think Aloud Method to successfully implement a usability lab with limited resources at NASA. Data showed that, even on a tight budget in empty conference rooms with video

camcorders recording events, Think Aloud usability testing flushed out many problems--ninety percent of which were fixed before the software was released.

Method #3: High Level AKA, "for pros and pros"

A good example of a high level method is a study published by Swan (1998) which used the Think Aloud method to study how people make sense of the Web.

Fifteen adults were recorded using computers to research on the Web for half an hour. Using a video mixer, both the subjects' face and the computer screen action were recorded. This is a common "picture in picture" technique wherein a smaller picture of the subject's face is displayed in one corner of the screen, with the rest of the screen being filled with a recording of the user's screen activity and most movements.

Recordings of the subject's voice while thinking aloud were gathered on the videotape as well. In addition the researchers took notes for later data triangulation.

After the subject's completed their task, the video recordings were transcribed and coded by the researchers, then re-coded using consensus and inter-rater reliability. Categories were established and the data was analyzed for patterns of behavior. The paper then presented a working model of how people read the Web, finding mainly that there was a lot more scanning than they anticipated.

This example serves to illustrate what kind of data you can gather when well funded and with the right equipment and schedule. Corporate usability labs and academic institutions usually have the resources to complete such high level analysis testing.

For situations where time is not a factor, either after the major release of a Web site, or in the above situation where a something static is being studied, this type of high level methodology gathers the highest quality data.

Having looked at three very different methods, the next page will summarize the strengths and weaknesses of the Think Aloud method.

Summary Table

The key points of the Think Aloud method:

When to use	To get verbalized data on cogitative processes. To find problems Finding the “Why” of problems
When not to use	Tasks that involve heavy cognitive loads For long tests For opinions
Pros	Gathers data on mental models, cognitive processes Obtains prescriptive data for iterative design Gathering a lot of data from a few individuals
Cons	Unnatural for some users and sometimes kids Data can be hard to analyze Users can rationalize and distort data
Prompts	Minimal, users will create fictional answers Good: “Keep talking” Bad: “Why did you do that?”
Data display	Observer notes Audio tape and notes Transcribed transcripts coded and reduced. Video examples with picture in picture of screen and face

Conclusion

We have explored the Think Aloud method, from its birth as psychology's introspection, to its development as a scientifically respectable method. We have seen Think Aloud mature to its current popularity as a usability staple, and for modeling thought process. One part of the Think Aloud story we haven't covered is its future.

Thinking about the future of the Think Aloud method is ironically sparked by a quote from the past:

All we can say is: 1. That most of our mental operations are inseparable from images, or produced by images. 2. That those images closely correspond to wishes or repulsions, to things we want or do not want, so that this want or not wanting seems to be the ultimate motive power in our psychology, probably in connection with elementary conditions in our being. 3. That inevitably, people will reveal in their thoughts and speeches, in their outlook on life and in the lives themselves, the quality of the images filling their minds. Investigation and estimation of these images, together with investigation and estimation of our likes and dislikes, will tell us what we are worth morally more accurately than even our actions, for they are the roots of action. (Dimnet, 1928, p. 10)

Using Dimnet's (1928) ideas, imagine the Think Aloud Method in 100, or even a 1000 years. Imagine being able to *see* the "quality of images" in someone else's mind. Imagine data in the form of "mind movies".

Imagine when Think Aloud becomes just “Think”.

Bibliography

- Cennamo, S. K., & Chung, M., & Leuck, V., Mount, V., Turner-Vorbeck, T. (1995). International Journal of Instructional Media 22(4): 305-320.
- Dimnet, E. (1928). The art of thinking. New York, Simon and Shuster Inc.
- Jacob, R. J. K. (1998). Computer Science 171 Human-Computer Interaction, Dept. of Electrical Engineering and Computer Science, Tufts University. 1998.
- Lewis, C., & Rieman, J. (1993). Task-centered user interface design: A practical introduction. Boulder, Colorado: University of Colorado, Boulder. 1998.
- Molich, R. (1994). "Preventing user interface disasters." Behaviour & Information Technology 13: 154-159.
- Richtel, M. (1998). Making Web Sites More 'Usable' Is Former Sun Engineer's Goal, The New York Times, Cybertimes. 1998.
- Smilowitz, D. E., & Darnell, J. M., & Benson, E. A. (1994). "Are we overlooking some usability testing methods? A comparison of lab, beta, and forum tests." Behavior & Information Technology 13: 183-190.
- Someren, W. M., & Barnard, F. Y., & Sandberg, A.C. J. (1994). The think aloud method. A practical guide to modeling cognitive process. London: Harcourt Brace & Company.
- Wittrock, M. C. (1986). "The think aloud procedure: a research technique for gaining insight into the student perspective." Physical Educator 52: 93-98.

END OF FILE