

Journal of • Virtual Worlds Research

jvwresearch.org ISSN: 1941-8477

**Volume 3, Number 1
October 2010
The Researcher's Toolbox**



Volume 3, Number 1

The Researcher's Toolbox

November 2010

Editor-in-Chief

Jeremiah Spence

Guest Editors

Tom Boellstorff, University of
California, Irvine, USA

Celia Pearce, Georgia Tech
University, USA

Dmitri Williams, University of
Southern California, USA

Thomas Malaby, University of
Wisconsin, Milwaukee, USA

Elizabeth Dean, RTI, USA

Tracy Tuten, East Carolina
University, USA

Technical Staff

John Brengle
Sil Emerson



The Journal of Virtual Worlds Research is owned and published by the Virtual Worlds Institute, Inc. – Austin, Texas, USA. The JVWR is an academic journal. As such, it is dedicated to the open exchange of information. For this reason, JVWR is freely available to individuals and institutions. Copies of this journal or articles in this journal may be distributed for research or educational purposes only free of charge and without permission. However, the JVWR does not grant permission for use of any content in advertisements or advertising supplements or in any manner that would imply an endorsement of any product or service. All uses beyond research or educational purposes require the written permission of the JVWR. Authors who publish in the Journal of Virtual Worlds Research will release their articles under the Creative Commons Attribution No Derivative Works 3.0 United States (cc-by-nd) license. The Journal of Virtual Worlds Research is funded by its sponsors and contributions from readers. If this material is useful.

Journal of • Virtual Worlds Research

jvwresearch.org ISSN: 1941-8477

**Volume 3, Number 1
The Researcher's Toolbox
November 2010**

What are users thinking in a virtual world lesson? Using stimulated recall interviews to report student cognition, and its triggers.

Lyn Henderson

James Cook University, Australia

Michael Henderson

Monash University, Australia

Scott Grant

Monash University, Australia

Hui Huang

Monash University, Australia

Abstract:

Stimulated recall is an empirically rigorous introspection data collection tool that allows the interviewer to elicit, identify and explore participants' thinking. In this study it was used to identify the types of thinking skills and strategies employed by first year university students engaged in a Chinese language and culture lesson in Second Life. A valuable affordance of this technique is the ability to account for stimuli from both the virtual and physical environments, thus strengthening the researchers' claims about the relationship between thinking and instructional design. This was accomplished through the use of screen capture software to record both the avatar's on-screen activity in Second Life as well as the face of the participant (via the web camera). This data was then used during the interview, within strict methodological protocols, to stimulate participants' recall of their thinking at the time of carrying out the activity. The value of stimulated recall over other introspection tools, within the context of this study, is discussed. In addition, methodological concerns, especially those relating to reliability and validity of data, are outlined in this article; we also use data from the study to explicate strategies to minimize those concerns.

Keywords: Stimulated recall, Introspective process tracing methodology, Second Life, Virtual World, Thinking skills, Think aloud, language acquisition, education, tertiary

What are users thinking in a virtual world lesson? Using stimulated recall interviews to report student cognition, and its triggers.

This article details and exemplifies the utilization of an empirically rigorous, qualitative methodology called stimulated recall. The aim of the study was to identify the types of thinking skills and strategies used by first year university students engaged in a Chinese language and culture lesson in Second Life (SL). The study also intended to identify the triggers or stimuli, both in-world and in-class, which resulted in those thinking skills and strategies. In achieving these aims a stimulated recall approach was selected because it has been validated as a reliable introspection tool and consequently allows researchers to make claims about participant in-situ activity; for instance, what participants were thinking when completing a particular action or response in SL.

The study employed the stimulated recall method of using video to record participants during the lesson and then replaying it during each individual's retrospective interview. During the interview the video was paused by the student when recalling a specific thought and/or by the researcher when he considered a pause might elicit mediating processes. A major strength of the stimulated recall interview is using the video to explore the visual cues, such as frowning or taking notes, that have been found as major indicators of mediating processes (e.g., Gass & Mackey, 2000; Henderson & Tallman, 2006).

While the stimulated recall interview holds many advantages over other introspection tools (think aloud, questionnaire, and interview), no method is without limitations. The key methodological concerns—especially those relating to reliability and validity of data—are outlined in this article along with strategies to minimize those concerns. This article also provides an unbroken trail from theoretical and methodological choice and continuing through to analysis and discussion of findings. The implications will be of interest to researchers trying to understand participant activity—in this article, cognition—while engaged in a virtual world as well as to discretely account for simultaneous real life and in-world stimuli triggers.

The aim of this paper is to provide readers with a useful introduction to using stimulated recall within the context of research in virtual worlds. It uses the context and data from a study of tertiary level students enrolled in a Chinese language and culture subject to illustrate the discussion about the research design and process. While some of the findings are presented here, this paper primarily concerns itself with a clear description of how researchers can enable their own research utilising a stimulated recall approach.

Research Context

Universities and other educational institutions continue to invest heavily in acquiring virtual worlds. However, rigorous research exploring the value of these environments in terms of cognition is scant. In response, this qualitative empirical research gathered evidence of student cognitive skills and strategies utilised during one of a series of 90 minute SL lessons in a first year university core Chinese language and culture subject. Apart from the use of SL, the subject consisted of the traditional weekly lectures and tutorial, independent study based on textbooks, and their associated language DVD. Of the 153 students enrolled in the subject, 11 students volunteered to be video recorded and interviewed.

The lesson under scrutiny in this study was specifically designed to reinforce and extend the language and culture content previously covered in the lectures and textbook. In this lesson the students had a number of learning objectives aligned with those of the subject which was contextualised within a collaborative learning activity—identifying and ordering appropriate food in Mandarin in a Chinese restaurant setting (see Figure 1). In a dynamic semi-spontaneous scenario, communication was centered around, but not limited to, practicing key vocabulary and phrases related to ordering food learned from the textbook and to expand on the textbook through new phrases introduced by the tutor or robots (automated avatars). Successful completion of the SL lesson could only be achieved through reading, writing and negotiating choices in Chinese text. The interaction between avatars was limited to text because one of the lesson objectives was to strengthen student use of Pinyin to input Chinese characters on a computer as well as to be able to read Chinese characters.



Figure 1. Chinese restaurant in Second Life

Theory and Methodology

Stimulated recall is a tried and tested data collection tool that allows the interviewer to elicit, identify and explore participants' thinking (Gass & Mackey, 2000; Henderson & Tallman, 2006). For instance, this approach has been used to study:

- teacher librarians' mental models (Henderson & Tallman, 2006)
- first year medical students' engagement with interactive multimedia (Pausawasdi, 2001)
- the relationship between teachers' beliefs and classroom practices (Meade & McMeniman, 1992)
- physician reasoning (Barrows, Norman, Neufeld & Feightner, 1982)
- sports coaches' decision making (Gilbert, Trudel & Haughian, 1999)
- second language acquisition (Gass & Mackey, 2000).

While stimulated recall is versatile, it also has a number of limitations. Consequently, in order to strengthen validity and reduce concerns over limitations, such as reliability, it is important for researchers to recognise that stimulated recall is inherently grounded in information processing theory, the mediating process paradigm and introspection process tracing methodology. Figure 1 illustrates the relationship between the theory, paradigm and methodology that are discussed in greater detail below.

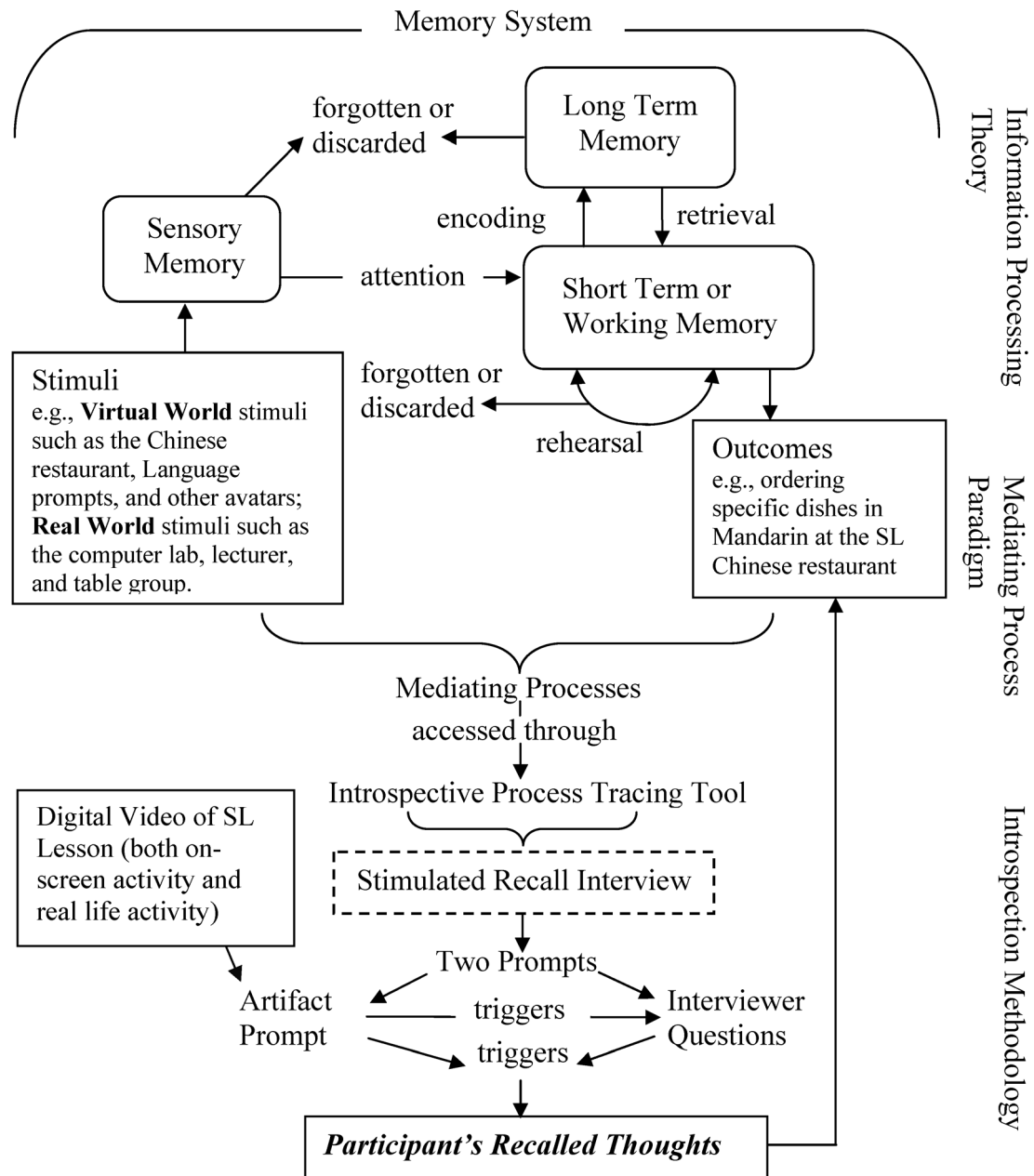


Figure 2. Diagrammatic Model of Theoretical and Methodological Framework for Stimulated Recall (adapted from Henderson & Tallman, 2006; Pauswasdi, 2001).

In brief, the *information processing theory* delineates how learning and remembering occur by examining the way we take in, discard, act on, store, and retrieve instructional input (e.g., from the instructionally designed Chinese restaurant environment in SL) to produce an outcome (e.g., correctly using Mandarin to order appropriate food) (Anderson, 1990; Atkinson & Shiffron, 1968; Craik & Lockhart, 1972; Miller, 1956). As depicted in the

simplified diagram (Figure 2), the stimuli are sent to sensory memory where it is forgotten, discarded, or given attention and forwarded to the short term working memory for processing. Next, after being processed, it is discarded or sent to long term memory to be categorized, stored and retrieved back into working memory. Finally, the student would deliver their cognitions as learning outcomes.

The *mediating processes paradigm* allows identification of the thinking skills, strategies, and processes of participants when carrying out a task in a virtual world. These mediating processes mediate, that is, come between stimuli (e.g., lesson task) and outcomes (e.g., selecting appropriate dishes) and are “the fine-grained elements of cognition through which, and by which, learning outcomes are realized” (Henderson, Putt, Ainge, & Coombs, 1997, p. 103). The mediating process paradigm is in contrast to the prevalent process-product paradigm in which thinking processes (e.g., recall) are assumed to have occurred because an outcome has been observed (e.g., ordering the correct meal in Mandarin). Consequently, the process-product paradigm is flawed as it cannot capture the choices, decisions and judgments made by students during a learning activity nor what actually triggered these cognitive processes (Marland, Patching, & Putt, 1992; Shulman, 1986).

Introspective process tracing methodology is utilized to access mediating processes based on the tested assumption that a participant can access these at some level and verbalize their thoughts (Ericsson & Simon, 1999; Gass & Mackey, 2000; Henderson & Tallman, 2006; Lyons, 1986; Shavelson, Webb, & Burstein, 1986; Vermersch, 1999). Mediating processes can be accessed through verbal protocol analysis including: (a) introspection think aloud, talk aloud or, as it is sometimes referred to, self-revelation, (b) self-observation (retrospective interview) and (c) the retrospective (as it is occurring after the event), stimulated recall introspective-prompted interview (in short, stimulated recall interview).

(a) Think-aloud Introspection Method

The “think aloud” method requires the participant to voice his thoughts as a running commentary when performing a task in SL (or other situation). However, thinking aloud is not a straightforward task as it does not come easily for the participant (Branch, 2000, Nielsen & Yssing, 2003, 2004). There are three major problems reported in the literature.

One: think aloud is alleged to be the most “invasive method” (Karasti, 2000) as it interferes with the participant’s normal thought routines, processes and actions. If we had utilized “think aloud” as the data collection method in our SL study, then the participants’ perceptions and actions ... [would] be transferred to talk, and even if the speech is immediate and run concurrently with the thoughts – users’ attention has to shift focus from ... interacting with a multi-modal interface: colours, layout, sound, graphics, animations ... [that is, from] visual perception ... to understanding to verbalization (Nielsen & Yssing, 2003, p.69). Thus, think aloud generates an artificial situation that adds cognitive strain in working memory (Branch, 2000; Karsenty, 2001).

Two: Afflerbach and Johnston (1984), Ericsson and Simon (1984) and Langer (1993) pointed out decades ago that participants are unable to verbalize as fast as they think or act, unless they are slowed down by a particularized problem. Hence, think aloud is inappropriate for research on synchronous collaborative activities such as group work in Second Life.

Three: think aloud is seen to require participant training (Gass & Mackey, 2000). For instance, Freeman (2003) described a study in which participants had to undergo “intensive training to ensure reliability of results ... providing at least 10,000” separate verbalized think alouds before being judged reliable (Freeman, 2003, p. 11)! Although this demonstrates a fanatical adherence to clinical protocols, participant training remains an issue of validity and reliability in concurrent think aloud research (Bielaczyc, Piroll, Brown & Brow, 1995; Branch, 2000; Gass & Mackey, 2000; Hertzum, Hansen, & Andersen, 2009).

These three weaknesses supported the decision to discount using the concurrent think aloud method. The study aimed to identify the thinking skills and strategies of students engaged in the lesson that required in-world interaction, including collaborative decision making. An intrusive approach that slows down the interaction would seriously influence the data being collected. In addition, since the study participants were university students with limited time, there would have been little or no opportunity for training to strengthen reliability.

(b) Retrospective Self-Observation Method

The retrospective self-observation method—usually an interview or written/online questionnaire based on the researcher’s interests—requires participants to report what they had been doing, thinking, strategizing, and/or feeling during a past activity, such as a SL session. It relies heavily on the participant’s memory and, without a visual (e.g., video of, or actual picture or diary used), aural (audiotape) or written (e.g., notes taken during a lecture) artifact reminder, it is open to charges of non-reliability and non-validity (Ericsson & Simon, 1993; Meade & McMeniman, 1994). This is confirmed by current research. For instance, Clarke, Fieberg and Gerdtham (2008) reported that, although retrospective self-observation can deliver large amounts of information, “the longer the period before recall, the greater the likelihood of error” (p.1275). One example is the study of 144 undergraduate students who increasingly made inferential errors about events and causes as the time increased from the original event (Hannigan & Reinitz, 2001). Similarly, Ottesen, Foss and Gronhaug (2004) found substantial perceptual errors made by their participants while Andersen and Mikkelsen’s (2008) research indicated considerable under-reporting—by 82 participants—in a retrospective questionnaire of occupational injuries over the previous month.

Of particular relevance to this study is the research by Harris and Wear (1993) in which 15 undergraduate education students were asked to recall when they deviated from their planned lessons and when their students deviated from the lesson topic. Despite the short time between lesson and the retrospective self-observation these pre-service teachers averaged only 42 percent accuracy for their own deviations and 35 percent for their students’ deviations. This was considered to be a serious concern for the current study because, unlike Harris and Wear’s studies that simply required participants to recall their behaviour, it required participants to recall their specific thinking processes while completing activities in-world.

One of the memory dangers with retrospective self-observation without an artifact stimulus trigger is that participants may infer plausible stories without consciously realizing that they are doing so. One reason is because we are “essentially sense-making beings and tend to create explanations” (Gass & Mackey, 2000, p.6), such as filling in omissions that were not present because they are typical of that situation (Hannigan & Reinitz, 2001; Waldmann, Holyoak, & Fratianne, 1995). This weakness has been well documented, especially the way in which details from one event or context intrude upon, and are incorporated into, memories

of a completely different event and, consequently, result in erroneous data (Allen & Lindsay, 1998; Hannigan & Reinitz, 2003; Lindsay, Allen, Chan, & Dahl, 2004).

(c) Stimulated Recall Introspection Interviews

Stimulated recall is also referred to as retrospective stimulated recall (as it occurs after the event to be introspected), “retrospective think aloud” (e.g., Guan, Lee, Cuddihy, & Ramey, 2006) and “retrospective user-reported” interview (e.g., Card, Pirolli, Van Der Wege, Morrison, Reeder, Schraedley, & Boshart, 2001). The goal of stimulated recall, and the other introspection tools, is to record the participant’s reported thoughts. It does not require them to become knowledgeable about their own subjective experience of reliving and reporting the initial experience.

Reliability and Validity Issues.

No methodology is without critics and no methodological tool is without limitations. It would be as foolish to claim infallibility for introspective reports (Lyons, 1986) as it would for other perceptual reporting. Stimulated recall has been criticized most notably on the issue of reliability (see Ericsson & Simon, 1993; Lyons, 1986; Nisbett & Wilson, 1977; Smagorinsky, 1994). In response, Henderson and Tallman (2006) proposed six issues that need consideration in order to strengthen reliability and validity: (1) timing of the stimulated recall interview, (2) capturing the data, (3) conduct of the interview, (4) interviewer training, (5) interviewee and interviewer fatigue, and (6) coding and categorizing the data.

(1) Timing of the stimulated recall interview

Recall accuracy is a reliability issue. Bloom’s (1954) stimulated recall study is particularly relevant. Prompting participant recall with aids within two days of the task, recall was 95 percent accurate but declined to 65 percent two weeks later. Garner (1988) also demonstrated significantly fewer recalls 48 hours later than on the day of the event. As mentioned previously, there is greater likelihood of plausible, schematic and/or causal-inferential gap-filling errors, the longer the timeframe between the event and the recall. Consequently in our study, ten of the eleven stimulated recall interviews were conducted within 15 minutes of the video-recorded SL lesson. The eleventh occurred 70 minutes after the lesson.

(2) Capturing the data

This study followed the well established method of video recording each participant during the SL computer lab lesson and then replaying the video during the participant’s stimulated recall interview. Eliciting participant thoughts through a video prompt has the advantage of staying close to the actual events in the sequence they occurred. Perceived as a memory retrieval cue with high associative strength, it is more likely to trigger accurate memory recall, thus avoiding guessing or reporting thoughts that they think they might have had [or what they think the interviewer would like to hear]. In fact, the process of recalling one’s thoughts appears to strengthen the overall memory of the immediate past event (Ericsson, 2002, p.984).

Guan, Lee, Cuddihy, & Ramey (2006), after comparing their 24 undergraduate participants’ stimulated recall verbalizations with their actual eye movements during the task, affirmed the “validity and reliability ... of stimulated RTA [retrospective think aloud] as it provides a valid account of what people attended to in completing tasks, it has a low risk of introducing fabrications, and its validity isn’t affected by task complexity” (p.1253).

A significant enhancement in this study was the use of innovative screen capture software (Screen Flow) to record both the on-screen SL activity as well as the face of the participant (via the inbuilt web camera). Unlike previous use of videotape technology, this allowed full resolution capture of the screen, crystal clear pause during the interview, full-facial expression capture, synchronization of face and screen video capture, as well as audio capture from both the SL-world and participants' in-class learning site. Additionally, Screen Flow did not require rendering time and allowed instant visual scrubbing (fast forward and rewind) along the video. Consequently, the participant or interviewer could move quickly forwards, backwards, jump to places of interest or return to points in the timeline as the need arose.

Additionally, Screen Flow allowed for a particularly non-invasive data collection process as opposed to participant head or lapel microphones, video cameras with tripods, and roving camera operator. It provides the distinct advantage of unobtrusively capturing rich data of what students are doing in SL synchronously with their verbals and non-verbals in-class. It also captured other stimuli, such as the lecturer giving instructions to the class in the computer laboratory.

(3) Conduct of the stimulated recall interview

The validity and reliability of the data being collected is highly dependent on the way in which the stimulated recall interview is conducted. Three issues need consideration: (a) promoting interviewee confidence, (b) instructions to the interviewee and (c) interviewer prompts.

(a) Promoting interviewee confidence: Establish rapport, mutual trust, and respect with the interviewee by being supportive, a good listener, interested in, and non-judgemental of, what the interviewee is saying. Becoming aware that they are capable of telling the interviewer about their thoughts and feelings allows interviewee realization that they are the authority on their cognitive processes during the activity (e.g., Henderson & Tallman, 2006; Marland, et al., 1992).

(b) Instructions to the interviewee: These are simple, minimal, and implement the previous stipulation. At the beginning of each stimulated recall interview, there are two general statements.

The first statement situates what is required within of the purpose of the interview: “What we want you to do is tell us what you were thinking during the lesson. [Interviewer pauses.] If you cannot remember what you were thinking or if you were thinking about something other than the lesson, either is okay. Just tell us.” Specific instructions about what to recall are not permissible as this contaminates the validity of the data. For instance, Ethell and McMeniman (2000) undermined the reliability of their collected data by requiring multiple tasks of the interviewee who was asked to walk the researcher through the lesson, stopping the video at any point where he would like to elaborate on what he was thinking, how he was feeling, or what thinking, beliefs, knowledge, or theories influenced any in-class decisions and observable teacher practices (p.93). As Ericsson and Simon (1993) cautioned, asking participants to vocalize their thoughts, seek out specific targets (e.g., theories), and explain them, increases cognitive load in short term memory and consequently impedes memory and recall.

The second statement delineates who is to pause the video. For example, “As you are watching the replay, when you remember what you **were** thinking, please pause the video and

tell me those thoughts. I will also pause the video and prompt you to tell me what you **were** thinking and feeling.” This instruction reduces researcher bias of determining when the student reported (Gass & Mackey, 2000; Henderson & Tallman, 2006; Marland, et al., 1992).

(c) Interviewer stimulated recall prompts: Gass and Mackey (2000) contend that researchers’ questions and responses are “the most serious of difficulties” (p. 89) as they have the potential to compromise the data. Exposure of the participant’s thoughts through adherence to tested protocols helps maximize reliability and credibility of recall. These protocols involve attention to non-directive questioning, interviewer versus participant initiation of responses, the timing of questions, and checking to ascertain recall versus the plausible story.

The nature of the interviewer’s prompt is an important validity issue in participant recall and researcher bias. Prompts are used to focus respondents so that they can access their memory and then vocalize what it was that they were thinking during the activity. There is a high likelihood of obtaining validity and accuracy of recall if the interviewer’s prompts (a) are non-directive, “What were you thinking then?”; (b) also include a specific context that is being viewed, “When you heard the pronunciation of the food, what were you thinking?”; and (c) seek confirmation: “Did you think that just now or when you were in the SL lesson?” If an interviewee’s response is “I can’t remember” or “No thoughts!” then asking, “So what were you *doing* there [interviewer points to the screen where the video has been paused]” is usually an effective memory aid that provides valuable thoughts to the follow-up question, “Okay, can you remember what you were thinking when you did that?” If prompts are more directive, such as those used in Sim’s (2004) study – “What is the basis for your using this technique/directing the lesson in this way” (p. 353) – then the prompts are keying the participant to report those thoughts that align with the research focus. There is a high likelihood of the participant reporting and creating plausible stories as explanation rather than recalling what they actually do remember thinking.

Notwithstanding their emphasis on non-directedness, Gass and Mackey (2000) commented that “unstructured situations do not always result in useful data” (p. 55). In our research, this has not been the case. Our prompts followed the strict protocols adapted from Marland, et al. (1992) and authenticated in other research (Burton, 2006; Henderson & Tallman, 2006; Patching, Putt & Henderson, 1996; Pausawasdi, 2001). What this research and our experience suggest is that researchers can take confidence in open non-directedness in terms of instructions and prompts rather than trade off reliability and validity against the fear of not obtaining information about their specific research focus.

A solution to this trade-off is to conduct a modified self-observation interview (Henderson & Tallman, 2006), that immediately follows the stimulated recall interview. Questions, usually asked in a survey or interview, are now permissible. Examples are: “Can you tell me something about your relationship with your avatar?”; “What are your thoughts about learning with SL?” Notice that these deliberately remain non-directive, unlike “What strategies did you use in your SL lesson, today?” which was also included to confirm and augment those identified from the stimulated recall data.

(4) Interviewer Training

In contrast to the participant training in the think aloud introspective interview, interviewers need to be trained to maximise reliability and accuracy of recall. Some guidelines are proposed by various researchers (e.g., Ericsson & Simon, 1999; Henderson & Tallman, 2006; Marland, et al., 1992) as well as in this study. Three critical strategies are advocated. Two

have been covered above: promoting interviewee confidence and conduct of the stimulated recall interview. The third examines interviewer training.

The literature generally does not identify how the researchers should be trained. However, based on the authors' experiences in this and other research projects, it is recommended that interviewer training occurs at least twice, and that one of those sessions is scheduled just before the actual research is conducted as a timely refresher. To promote optimum interviewer success, training includes a video replay of both the neophyte and skilled stimulated recall researchers carrying out a short task (e.g., 5-8 minutes is adequate). First, the inductee is interviewed by the researcher. Then the roles are reversed. This allows a "walking in my moccasins" experience in both roles. In this research project there were 3 interviewers (one experienced in stimulated recall). After each practice the interviewers compared notes, talked through issues such as establishing rapport, appropriate prompts, and interview fatigue.

(5) Interviewee and interviewer fatigue

Fatigue in both interviewee and interviewer, but especially interviewer, are infrequently discussed in stimulated recall or other research literature, despite having a potentially dramatic influence on the reliability and validity of the data. The process of stimulated recall can be particularly fatiguing since the one-on-one interview typically lasts, at minimum, twice as long as the activity under study (Henderson & Tallman, 2006). The interviewee and interviewer are required to watch the video replay as well as to pause the video to allow participant verbalization of their thinking. In addition, back-to-back interviews may cause increasing fatigue across the interviews.

As interviewers become fatigued their adherence to protocols and observation skills can deteriorate, not to mention their willingness to follow up potentially significant actions or utterances from the participants. Similarly, participants who are fatigued are more likely to make recall errors and overlook potentially valuable incidents.

The risk of fatigue needs to be weighed against how much of the video needs to be played in order to provide sufficient data to achieve the research aims. Options are: to play all of the video; only particular segments; or to use a fast-play option wherein the play-back is set at a slightly faster speed than normal. The researchers in this study chose the fast-play option. This allowed the interview to cover more of the lesson while still allowing the interviewee and interviewer ample time to slow down, pause, or rewind the video at points of interest.

(6) Coding and categorizing the data

An accurate transcription of the interview is important since some utterances can convey significant implications in terms of categorizing thinking skills. The interviews are audio-recorded, usually professionally transcribed, and the transcription checked by the researchers. The video recording allows confirmation.

Coding reliability is critical. In the coding process, there is constant referral to a coding schema based on the research aim. Our study followed a well established list of thinking processes, including their definitions (see Henderson, 1996; Marland, et.al., 1992; Table 1). When there is a difference, consensus among the researchers occurs through justification of the rationale for that categorization. If there is no consensus, then that mediating process is excluded as usable data. In this study, guided by the lead researcher, all researchers coded the same interview transcript and subsequently debated coding differences. Coding reliability

was further strengthened by every transcript being coded a second time by a different researcher; differences were debated. This means that every identified thinking skill reported in this research was approved by at least two researchers.

Table 1: Cognitive Skills and Strategies Utilized by Participants with Total Number Used and Percentage of Total

Type	Definition of Cognitive Skill *	Interviewee Example	Thinking Skills #	% #
Affect	reports feelings aroused by SL and by real-world stimuli	“I was feeling sort of excited!” “[When he said, ‘You have to write in Chinese’], I was really scared.”	145	19.7
Analysing	reduces, breaks down whole (e.g., problem, task, issue) into parts	“Is this going to come back and is it going to come up with my name and say what I put down and will I be assessed as to what I’d done.”	14	1.9
Anticipating	wonders about the possibility of an event, relevance of material, or content; predicts	“I thought: ‘What if I click on something and it does something wrong?’ ” “I thought, ‘I wonder what kind of dishes is there?’”	31	4.5
Applying	considers the use of an idea, tactic in a different context.	“I’m going to order this dish in Mandarin when I go to my favourite Chinese restaurant.”	1	0.1
Categorising	sorts items, ideas, skills into different groups	“I was thinking, that dish is some kind of meat because roe is meat.”	18	2.5
Comparing	identifies similarities, differences between two statements, concepts, ideas, models, situations, points-of-view, etc.	“They know it and I don’t.”	34	4.6
Confirming	judges that the ideas in SL support own tactics, practices; verifies actions, thoughts, and significance of SL	“I already have it – the non-spicy dish!” “Oh, yeah, I’m in the right Second Life place.”	24	3.3
Deliberating	engage in “thinking” about a topic, segment, etc., (type of thinking not disclosed):	“I’m just wondering if I need to talk with the waitress.”	6	0.8
Diagnosing	identifies strengths and/or weaknesses in idea, strategies, points-of-view	“Huh, that didn’t work but I still got something.”	24	3.3
Evaluating	worthwhileness of SL, activities, own strategies, graphics, and issues of	“It’s hard. It’s too hard for me to understand properly.” “I thought, ‘She’s sitting on the	107	14.6

	morality and ethical behavior in SL	food! Yeah, well, it's still inappropriate'!"		
Generating	formulates one's own questions, examples, ideas, or problems; interpolating; beyond what was in SL & class	"I was thinking: 'If I ever go to China, if I ever go and order something in Chinese, even if I don't like this meal, this is the one I'm going to order'."	32	4.4
Imaging	creates a mental image to gain a fuller understanding	"I'm a very picture person. I remember, 'this one [picture of a dish] was here on the food counter and this one was over here ... and because I have a memory picture, I draw connections'." (see Fig.1)	1	0.1
Justifying	rationalizing, explaining, and providing reasons for their actions and thoughts	"I was thinking: 'Where is this stupid thing? I have so many things in my inventory, how am I supposed to find one thing?'"	63	8.6
Linking	associates or brings together two or more ideas, topics, experiences, tasks	"[I heard the voices] and then I was wondering, 'Which teachers of the Chinese faculty they are' [sic]."	15	2.0
Metacognising	thinks about, reflects on, evaluates their certainty of understanding; directs own thinking, ways to troubleshoot lack of comprehension	"I thought: 'Don't expect to learn anything straight away. You know, like, a little bit, but don't expect to know the dishes at the end or remember what they were made of. Come and do it again before the exams.'" (I.6)	73	9.9
Recalling	brings back into working memory an idea, opinion, fact stored in long-term memory	"At the start he [the lecturer] said, 'You should ask other people'." "What were the different names of food [in the text book]?"	14	1.9
Reflecting	general indication of consideration over past action and response	"I was thinking, 'I didn't realise I was typing something wrong! [Then I did realise and] I thought, 'Oh, no! I've done it four times'."	2	0.3
Selecting	identifies key material, gist material, or that which is relevant to assessment.	"I thought, 'I know those dishes don't have the win sign, so don't order them'."	1	0.1
Strategy Planning	plans ways for processing and carrying out the lesson tasks, manoeuvring in SL, and	"If I don't finish I can do it, like, double check at home." (I3) "Try to find some food that	121	16.5

	subsequent strategies	you know in Mandarin, like Mee-fun. It's only rice, so go there and order some rice."		
Translating	using their own words to interpret, explicate, and/or clarify what is happening	"I was thinking, 'She's not taking it too seriously; that's irritating. It's a difficult task and we really have to coordinate properly and we're at different stages'."	7	1.0
Thinking Skills TOTAL			733	100

* definitions adapted from Henderson (1996); Marland, et.al. (1992).

the numerical tally and percentage of the total number are discussed below (see Analysis).

There are two issues that require close attention in order to maximize validity and reliability of the coding. These are: types of thoughts accessed and leading questions by the interviewer

Types of thoughts accessed. Both the video and interviewer prompts trigger the participant's thoughts. As depicted in Figure 3, there are two types of thoughts reported during stimulated recall interviews. These are "recall thoughts" and "hindsight reports".

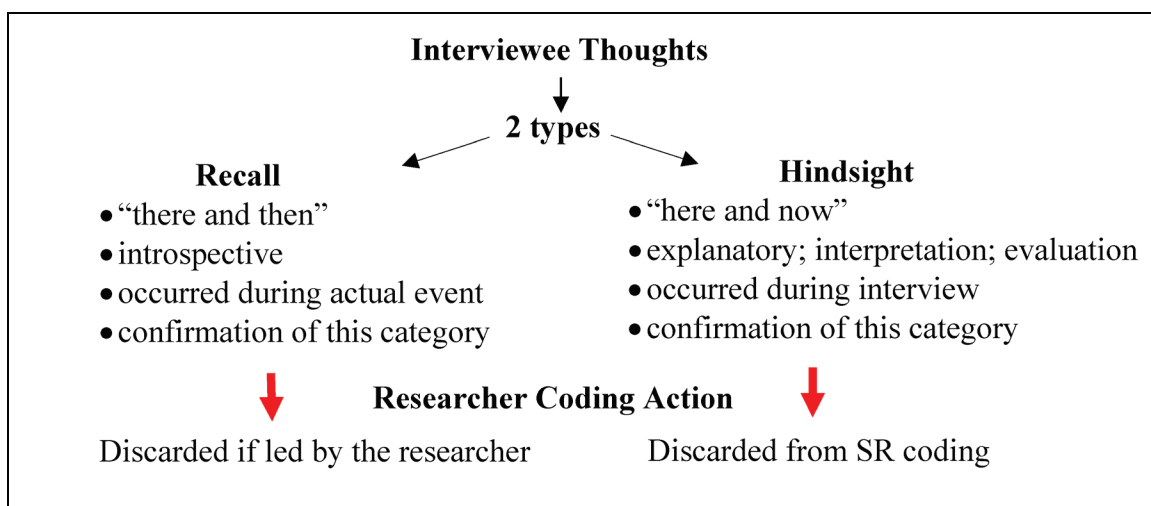


Figure 3: Types of interviewee thoughts and researcher action

Stimulated recall thoughts occur during the performance of the task and are the "back *there and then* during the SL lesson" thoughts. For instance, "Can you tell me what you were thinking when you were trying to find a way around it?", elicits "I was thinking, I'm never going to find this, and what am I going to do if everyone else is speaking in Chinese characters and I'm not?" Thus they are introspective thoughts and therefore coded and tallied. In contrast, hindsight thoughts only occur during the interview, and can be labelled "*here and now*" thoughts and are sometimes explanatory, as is seen in Figure 4 (see participant statement labelled "A"). Such hindsight thoughts may be interesting but they cannot be coded as stimulated recall.

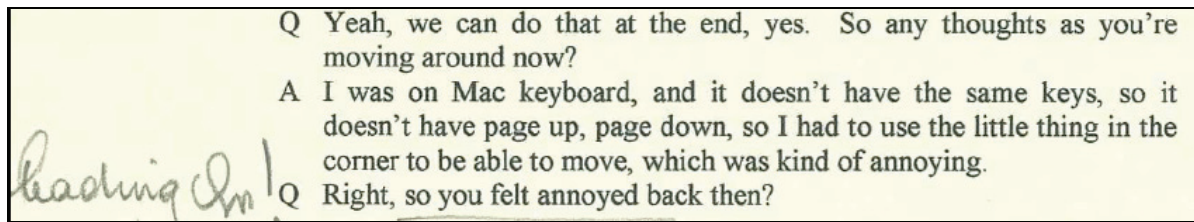


Figure 4: Excerpt from an interviewee's transcript demonstrating a leading question.

Leading the interviewee. This excerpt from a transcript (Figure 4) demonstrates, in the first "Q" (meaning Interviewer Question) a non-leading prompt, "So any thoughts as you're moving around now?" that only elicits explanatory "here and now" thoughts (Figure 3). The interviewer then asks a leading question—"Right, so you felt annoyed back then?" (Figure 4)—and consequently the interviewee's subsequent thought of feeling annoyance had to be excluded from the "affective" tally (see Table 1).

In the following excerpt (Figure 5), the typed section is the verbatim transcript from an audio recorded stimulated recall interview. The left hand margin is used to record the thinking skills and strategies utilized by the student during the SL lesson. The underlined text delineates the thinking skill and/or the SL and real world triggers. The right hand margin identifies the trigger—those from the SL Chinese restaurant and those from the real world of the SL lesson in the computer lab.

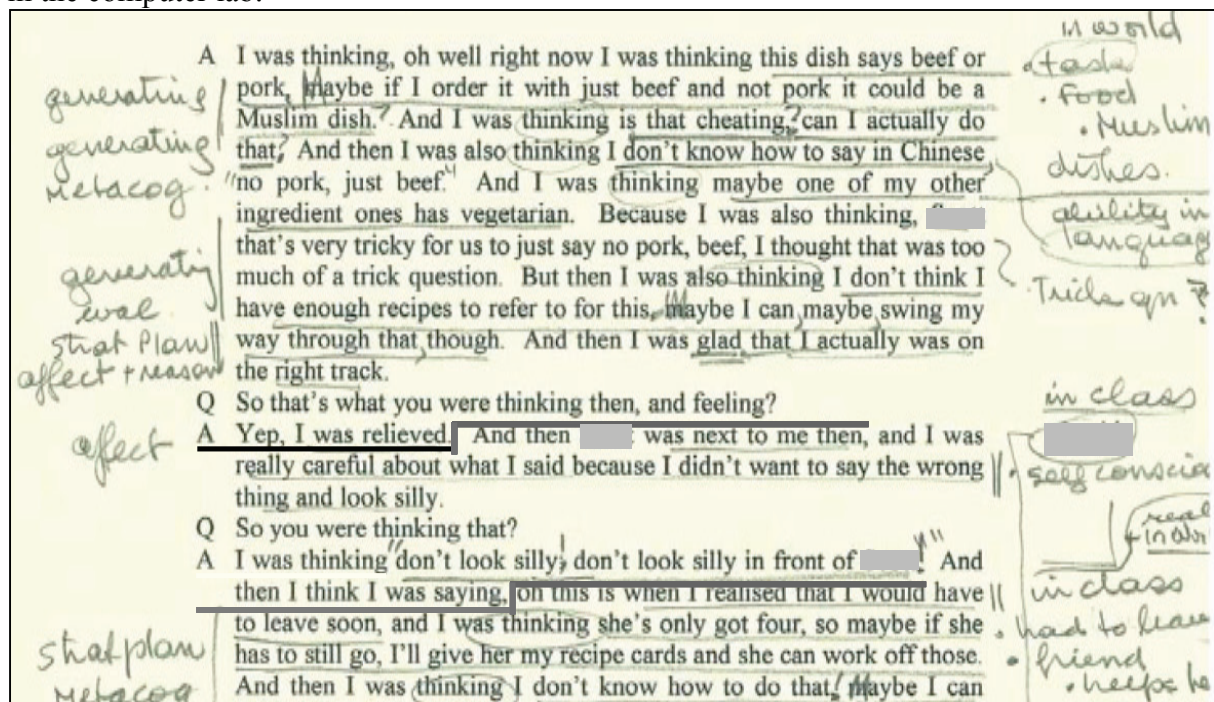


Figure 5: Delineation of cognitive skills and SL and real world triggers

Analysis

The coding process resulted in the identification of 733 instances of thinking skills or strategies used by the 11 participants (see Table 1). A more detailed table of each participants' thought types and their associated triggers (not presented here due to limited space) reveals considerable variability between the participants. For instance, Interviewee 7

reporting 178 thoughts and Interviewee 6, a mere 23, that had occurred while they were working through the SL lesson.

Although each student engaged in the same instructional activity in SL they varied in the number and types of thinking skills employed. Such a variation across the 11 participants clearly begged to be investigated. One such line of inquiry for the researchers was to see if the instances of one type of thought process was related to a particular trigger and/or occurred in conjunction with another type of thought process. For instance, while a substantial number of complex thinking skills (e.g., metacognition and strategy planning) were indicated by the participants, the most common category of thinking was *Affect*, defined as: feelings aroused by SL and real-world stimuli (e.g., “I’m happy that I now have three dishes”). The second most common thinking skill was *Strategy Planning*, for example, devising tactics to accomplish the lesson objectives and for passing the exam. As a result the researchers were prompted to consider if the frequency of *Affective* thinking was a result of instructional stimuli (e.g., an effective learning task in which success elicits an affective state), or is related in some way to the more complex Strategy Planning, or is perhaps explained by some other theory. The data provided by stimulated recall allows for this kind of exploration.

The value of stimulated recall becomes particularly evident when the types of thinking skills and strategies are linked to their triggers (the event that made them think, such as an automated waitress greeting them at the door of the restaurant) and then compared over time with a specific participant as well as across the participants. This enables the researchers to explore such things as the influence of types of in-world instructional stimuli, in-world collaborative activity, real life lecturer instructions, audio and text media, and pace.

For instance, researchers interested in instructional design of SL lessons could match the trigger with the mediating process (cognitive skill) providing them with possible teaching and learning design strategies that cognitively engage students to utilize more complex thinking skills. For example, generative thinking (see Table 1) was triggered in all of the participants at the time of choosing and ordering dishes from the Chinese restaurant menu for people with particular dietary requirements, such as a vegetarian or a diabetic. Another example of the ways in which stimulated recall can inform pedagogy is how it highlighted the effect of the lecturer moving around the class to work with students. The researchers were able to analyse if the lecturer’s movement and interaction triggered thinking across the gamut of students. In this case it was revealing to note that some of the students reported feeling self-conscious when he stopped near them in the real world of the classroom (e.g., Figure 5 includes the example of a student who said “[the lecturer] was next to me then, and I was really careful about what I said because I didn’t want to say the wrong think and look silly”). However, no student indicated similar thinking being triggered when the lecturer’s avatar stood next to the students’ avatar.

A more detailed analysis of this research data will be conducted in another forum. The intent of this paper has been to demonstrate some of the ways in which data provided by stimulated recall can support analysis. Although the number of participants in stimulated recall research is usually small due to the intense amount of work, the data is rich and it is possible to use statistical analysis if there are sufficiently detailed coded examples of both the type of mediating process and its triggers. Additionally, one transcript affords various analyses, such as identifying and explicating: the SL and real-world triggers of the participants’ cognitive skills and strategies; identity of self and avatar when learning in a SL world; the matches and mismatches between the lecturer’s thoughts and actions and those of the students during the

SL session; ascertaining if a deliberate lesson design elicits certain cognitive processes and collaboration; charting whether the SL lesson outcomes are affected by the novelty value of one SL lesson compared with two or more SL lessons.

Conclusion

To reiterate, stimulated recall is an empirically rigorous introspection data collection tool that allows the interviewer to elicit, identify, and explore participants' thinking. In this study it was used to identify the types of thinking skills and strategies employed by first year university students while engaged in a lesson in Second Life. Stimulated recall enabled the researchers to make claims about the kind of thinking triggered by the virtual world environment in general and the instructional design in particular. In addition, a valuable affordance of this technique is the ability to account for stimuli from not just the virtual, but also from the physical environment, thus strengthening the researchers' claims about the relationship between participants' cognitive processes and instructional design.

This article described how stimulated recall could be used by other researchers. The key methodological concerns, especially those relating to reliability and validity of data, have been outlined along with strategies to minimize those concerns. Of particular note is that stimulated recall is embedded within an information processing-mediating processes-introspection theoretical framework and requires adherence to strict methodological practices to maximize reliability and validity of the data and analysis. However, it is less intrusive on thought processes than the concurrent verbal reporting of think-aloud methods and is more reliable than other introspection reporting tools.

References

- Afferbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Reading Behavior*, 16(4), 307-322.
- Allen, B.P., & Lindsay, D.S. (1998). Amalgamations of memories: Intrusions of information from one event into reports of another. *Applied Cognitive Psychology*, 12, 277-285.
- Anderson, J.R. (1990). Cognitive psychology and its implications. (3rd ed.). New York: Freeman.
- Anderson, L.P., & Mikkelsen, K.L. (2008). Recall of occupational injuries: A comparison of questionnaire and diary data. *Safety Science*, 46, 255-260.
- Atkinson, R., & Shiffron, R. (1968). Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation. Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.
- Barrows, H., Norman, G., Neufeld, V., & Feightner, J. (1982). The clinical reasoning of randomly selected physicians in general medical practice. *Clinical Investigative Medicine*, 5(1), 49-55.
- Bielaczyc, K., Pirolli, P.L., & Brown, A.L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, 13(2), 221-252
- Bloom, B.S. (1954). The thought process of students in discussion. In S.J. French (Ed.), *Accent on teaching: Experiments in general education* (pp.23-46). New York: Harper & Row.
- Branch, J.L. (2000). Investigating the information-seeking processes of adolescents: The value of using think-alouds and think afters. *Library & Information Science Research*, 22(4), 371-392.
- Burton, C. (2006). M-Trilogy: *An investigation of mediating processes, metacognition and mismatches in an ICT education subject*. B.Ed.(hons) thesis. Townsville: James Cook University.
- Card, S., Pirolli, P., Van der Wege, M., Morrison, J., Reeder, R., Schraedley, P., & Boshart, J. (2001). Information scent as a driver of web behavior graphs: Results of a protocol. Analysis method for web usability. *Proceedings of SIGCHI'01*, March 31-April 4, Computer Human Interaction. ACM Press pp. 498–504.
- Clarke, P.M., Fiebig, D.G., & Gerdtham, U.G. (2008). Optimal recall length in survey design. *Journal of Health Economics*, 27(5), 1275-1284.
- Craik, F. & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.

- Ericsson, K.A. (2002). Towards a procedure for eliciting verbal expression of non-verbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. *Applied Cognitive Psychology*, 16(8), 981–987
- Ericsson, K., & Simon, H. (1984; 1993; 1999). Protocol analysis: Verbal reports as Data. (1st ed. 1984; 2nd ed. 1993; 3rd ed. 1999). Cambridge, MA: MIT Press.
- Ethell, R., & McMeniman, M. (2000). Unlocking the knowledge in action of an expert practitioner. *Journal of Teacher Education*, 51(2), 87-101.
- Freeman, A. (2003). Consciousness: A guide to the debates. Santa Barbara, CA.: ABC CLIO Inc.
- Garner, R. (1988). Verbal-report data on cognitive and metacognitive strategies. In C. Weinstein, E. Goetz, and P. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction and evaluation* (pp. 63-76). New York: Academic Press.
- Gass, S.M., & Mackey, A. (2000). Stimulated Recall Methodology in Second Language Research. Mahwah, New Jersey: Lawrence Erlbaum.
- Gilbert, W., Trudel, P., & Haughian, L.P. (1999). Interactive decision making factors considered by coaches of youth ice hockey during games. *Journal of Teaching in Physical Education*, 18(3), 290-311.
- Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1253-1262).
- Hannigan, S.L., & Reinitz, M.T. (2003). Migration of objects and inferences across episodes. *Memory & Cognition*, 31(3), 434-444.
- Hannigan, S.L., & Reinitz, M.T. (2001). A demonstration and comparison of two types of interference-based memory errors. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27(4), 931-940.
- Harris, J.C., & Wear, S.B. (1993). *Creating the reflective practitioner: Sources and causes of student teacher deviation from planned lessons*. EDRS Document ED365 715.
- Henderson, L., & Tallman, J. (2006). *Stimulated recall and mental models: Tools for teaching and learning computer information literacy*. Lanham, MD: The Scarecrow Press.
- Henderson, L., Putt, I., Ainge, D., & Coombs, G. (1997). Comparison of students' thinking processes when studying with WWW, IMM and text-based materials. In F. Verdejo & G. Davies (Eds.), *Virtual campus: Trends for higher education and training* (pp. 124-162). London: Chapman and Hall.

- Henderson, M. (1996). *Multimedia interactivity: An investigation into learners' mediating processes during click-drag activities*. B.Ed. Hons: James Cook University, Townsville.
- Hertzum, M., Hansen, K.D., Andersen, Hans H. (2009). Scrutinising usability evaluation: Does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, 28(2), 165-181.
- Karasti, H. (2000). Using video to join analysis of work practice and system design: A study of an experimental teleradiology system and its redesign. *IRIS 2000 Conference*. Retrieved January 1, 2005, from <http://iris.informatikgu.se/conference/iris20/31.htm>.
- Karsenty, L. (2001). Adapting verbal protocol methods to investigate speech systems use. *Applied Ergonomics*, 32(1), pp. 15-22.
- Langer, J. (1993). Approaches toward meaning in low- and high-rated readers. *Report Series 2.2 Albany, NY: National Center on Literature Teaching and Learning*. Retrieved December 20, 2009, from <http://cela.albany.edu/meaning/index.html>
- Lindsay, D.S., Allen, B.P., Chan, J.C.K., & Dahl, L.C. (2004). Eyewitness suggestibility and source similarity: Intrusions of details from one event into memory reports of another event. *Journal of Memory and Language*, 50, 96–111.
- Lyons, W. (1986). *The disappearance of introspection*. Cambridge, MA: MIT Press.
- Marland, P. W., Patching, W.G., & Putt, I. (1992). *Learning from text: Glimpses inside the minds of distance learners*. Townsville: James Cook University.
- Meade, P., & McMeniman, M. (1992). Stimulated recall: An effective methodology for examining successful teaching in science. *Australian Educational Researcher*, 19(3), 1-18.
- Meade, P., & McMeniman, M. (1994). Teacher knowledge in action: The theoretical base of effective teaching. Paper presented at *Australian Association for Research in Education Conference (AARE)*. Retrieved January 1, 2005, from <https://www.aare.edu.au/94pap/meadp94404.txt>
- Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.
- Nielsen, J. & Yssing, C. (2003). "Getting beyond the disruptive effect of Think Aloud". In M. Hertzum & S. Heilesen (eds.), *Proceedings of the third Danish human-computer interaction research symposium* (pp. 67-71) Datalogiske Skrifter, Roskilde University.
- Nielsen, J., & Yssing, C. (2004). *What kind of information does an HCI expert want - on concurrent usability testing?* Working Party, HCI Research Group, NORDICHI. Tampere: NORDICHI. Retrieved January 4, 2010 from <http://openarchive.cbs.dk/bitstream/handle/10398/6465/15-2004.pdf?sequence=1>

- Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Ottesen, G.G., Foss, L., Gronhaug, K. (2004). Exploring the accuracy of SME managers' network perceptions. *European Journal of Marketing*, 38(5), 593-607.
- Parsons, J. M., Graham, N. & Honess, T. (1983). A teacher's implicit model of how children learn. *British Educational Research Journal*, 9(1), 91-101.
- Pausawasdi, N. (2001). *Students' engagement and disengagement when learning with IMM in mass lectures*. PhD thesis, James Cook University.
- Shavelson, R., Webb, N., & Burstein, L. (1986). Measurement in teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp.50-91). New York: MacMillan.
- Shulman, L. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3-36). New York: MacMillan.
- Sim, C. (2004). The personal as pedagogical practice. *Teachers and teaching: Theory and practice*, 10(4), 351-364.
- Smagorinsky, P. (Ed.). (1994). *Speaking about writing: Reflections on research methodology*. Thousand Oaks, CA: Sage Publications.
- Vermersch, P. (1999). Introspection as practice. *Journal of Consciousness Studies*, 6(2-3), 17-42.
- Waldmann, M., Holyoak, K., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General*, 124, 181-206.