Preparing to teach in a social virtual world can raise many questions for instructors. One of these questions is deceptively simple: “When delivering instruction, should I talk, type, or do a little of both?” It is difficult to answer this question with confidence because there is little empirical evidence about the effects of different communication modalities on learning in virtual worlds, and on the specific aspects of virtual worlds that enhance or detract from learning. Virtual worlds raise issues of immersion (how lifelike a virtual environment is), presence (the learner’s sense of “being there”) and co-presence (the learner’s sense of being in the virtual world with others) (Heeter, 1992; Schroeder, 2006).

The question of which communication modality to use in virtual world instruction is influenced by the generally accepted belief that the more lifelike (immersive) a virtual environment is, the more successful it is assumed to be. In instructional design, dual-coding theory (DCT) has shown that lessons containing concrete information, vivid images, and sound are easier to comprehend and better remembered than abstract lessons (Paivio, 1975; Clark & Paivio, 1991). Similarly, because the human voice is natural and deeply familiar, it is a cue that may increase the likelihood of retaining information (Sallnas, 2004). In addition, speaking rather than typing during a lesson containing concrete, how-to information with vivid images may create a more immersive experience and thus contribute to improved learning outcomes (Moreno & Mayer, 2002; Mousavi, Low, & Sweller, 1995). Because social virtual worlds are actively used for distance education and other forms of training, educators have an interest in improving the understanding of immersion, presence, co-presence, and any relationship between these and learning.
This study, though limited in scope and conducted with a small sample, explored the relationship between communication modality, presence and co-presence, and learning in a social virtual world commonly used for education. To this end, I measured learning (short-term recall and retention), cognitive load, and perceptions of presence and co-presence from participants assigned to one of three communication modalities: voice only, text only, and voice and text together.

Key Terms
Working memory and cognitive load are two concepts key to my study. Working (short-term) memory is a three-part system of limited capacity (Sweller, van Merrienboer, & Paas, 1998). Because working memory is key to learning and yet so limited in its capacity, learners are automatically placed in a difficult position as they attempt to gather new knowledge. Long-term memory, in contrast, is permanent memory.

Cognitive load is defined as the amount of mental energy required to process a given amount of information. As the amount of information increases, so does the cognitive load on our mental resources (Feinberg, Murphy, & Duda, 2003). Cognitive load theory holds that learning will be inhibited when the amount of information and instruction exceed the capacity and limitations of our mental resources.

Theoretical Background and Research Questions
The study addressed four hypotheses, presented in Wilkes, 2009 with brief theoretical and empirical material from a more extensive literature review.

*Does Communication Modality Affect Learning?*
I expected retention scores in this study to be higher for voice participants (Group V) and voice-and-text participants (Group VT) than for text-only participants (Group T). I based this assumption, somewhat roughly, on Alan Baddeley’s theory of working memory—different components of short-term memory retain certain kinds of information better over others, such as speech sounds as compared to visual information (1992, 2002). Baddeley’s theory holds that the coding processes underlying working memory are neutral with respect to input modality, including the verbal-non-verbal dichotomy, which implies that we should expect no difference in retention based on the communication modality used in the study activity. Baddeley’s theory, however, also emphasizes
the critical role of attention, in that anything that limits attention capacity will impair performance (learning).

Penney’s hypothesis departs from Baddeley’s in regard to modality. She proposed a model of the structure of short-term memory called the separate streams hypothesis, describing that the processing of auditorily and visually presented verbal items is carried out separately. Other researchers’ findings support the idea that more memory capacity is available when dual modalities are used (Clark & Paivio, 1991; Mayer & Moreno, 1998; Penney, 1989; Sweller, 1994).

Moreno and Mayer (2002) found that receiving information as on-screen text rather than narration significantly hindered learning, groups presented with verbal information as speech recalled significantly more information than those presented with verbal information in the form of text, and presentation of information as on-screen text hindered learning in all groups.

**Does Communication Modality Affect Cognitive Load?**

I expected cognitive load to be lower for voice participants (Group V) and higher for voice-and-text participants (Group VT) and text-only participants (Group T), because of subjective reports of mental strain and effort.

The modality principle (Moreno & Mayer 2002) states that when presenting a multimedia explanation (visual and verbal information), words should be presented auditorily rather than visually. The rationale for this is that, by using the auditory channel to process the words, effective working memory capacity is expanded since students are not forced to split their limited visual working memories between the on-screen text and the pictorial information. Pictures are processed through the visual information channel, while spoken words are processed through the auditory channel. Processing of words is off-loaded onto the auditory channel, which is otherwise underused (Moreno, 2006).

The work of Mousavi et al. (1995) provided an important empirical foundation for that of Moreno and Mayer. These researchers found,

1. The presentation of information in mixed auditory and visual mode, rather than a single mode, had beneficial effects on learning, presumably expanding effective working memory capacity by not overloading a single working memory channel.

2. A significant modality effect between two visual-visual and two visual-auditory groups. The two visual-auditory groups required less time to solve two repeat problems.
3. Auditory solution statements to be superior over written solution statements, regardless of the way in which problem information was presented.

In summary, research on modality, working memory, and cognitive load is complex, but some researchers suggest that more items will be recalled in a memory test if some are presented in a visual modality and some in an auditory modality, rather than all in a single modality (Mousavi et al., 1995; Penney, 1989). As Penney (1989) and Baddeley (2002) point out, however, if the auditory component is too long or highly complex, it will create an excessive strain on attention and working memory, increase cognitive load, and ultimately decrease recall and retention.

Does Voice Communication Increase Perceptions of Presence in a Virtual World?

Presence (one's sense of “being there” in a virtual world) and co-presence (one's sense of “being there with others”) are considered two major factors that differentiate virtual worlds from other online and multimedia applications. According to Mikropoulos and Strouboulis “Presence is the main attribute of virtual reality (VR) that differentiates it from other information technologies, giving learners an active role,” which instructors hope will increase recall and retention of material (2004, 582–591). “Active learning” takes place when participants themselves carry out an action or series of activities (as in the how-to style instruction conducted as part of this study), rather than watch someone else perform them. Slater, Sadagic, Usoh, and Schroeder (2000) point out that if humans are required to perform tasks within virtual environments, then surely it is beneficial for them to feel present in that environment.

Some researchers (Sallnas, 2004) have indicated that modality (specifically voice) can impact perceptions of presence. Others (Moreno & Mayer, 2002) found that modality did not impact perceptions of presence, while the level of immersion made possible with the media environment (a desktop vs. head-mounted display) did affect presence. Based on prior findings by Sallnas (2004) that voice contributed to co-presence, and that co-presence was a factor of presence, I thought that Group V would experience greater perceptions of presence.

Do Perceptions of Presence Affect Learning?

Some of the more compelling reasons for developing highly immersive environments that create stronger perceptions of presence are based on theories of
Communication Modality, Learning, and Second Life

limited working memory; making interactions with technology more immersive involves adding more lifelike and natural “cues” to the environment (visual, audible, contextual, textual, and so on) that might combine to trigger recollection later, easing and enhancing the recollection of course material. More immersive environments will supposedly attract students’ limited attention to learning content rather than to the interface (Moreno & Mayer, 2002).

In regard to virtual environments specifically, some researchers have claimed that presence is causally related to task performance and learning (Schuemie, van der Straaten, Krijn, & van der Mast, 2001) and many of the factors that appear to affect presence, such as vividness, are known to enhance learning and performance (Witmer & Singer, 1998). Biocca, Harms, and Burgoon (2003) note that communication and Human–Computer Interaction (HCI) researchers are typically interested in social presence because it may mediate the effects of other variables of central concern to the researcher, including interface features, learning, and memory. There is, however, no conclusive empirical evidence to support these claims. Whether or not presence contributes to better task performance remains controversial based on reported findings (Schuemie et al., 2001).

In research conducted by Moreno and Mayer (2002), groups presented with a higher level of immersion did not differ in the number of items they recalled from groups presented with a lower level of immersion. Immersion (media environment) did not affect responses to cognitive assessment questions. Participants in more immersive conditions reported higher levels of presence, but groups did not differ in their learning outcomes.

Study Methodology and Procedures

This study took place between January and July 2009 and was conducted entirely online and anonymously.

In the interest of full disclosure, I managed the implementation of voice chat in Second Life and was a full-time employee at Linden Lab, creators of Second Life, from February 2007 to August 2009. Though I was obviously close to the voice project as a result of this, it is worth noting that no internal or known external voice research with Second Life residents had been designed or conducted prior to this study, and that no comprehensive voice user data had been collected or disseminated. That said, I believe I had no prior knowledge that might have skewed or otherwise positioned me to expect or look for certain types of findings over others. In addition, research has indicated that researchers of games should play the games they are studying. If they do not, they cannot...
know what questions to ask, decipher the local language, understand the game mechanics, or have any sense of the social context of play (Williams, 2004). Finally, Linden Lab provided no financial or other support of this work, which was completed entirely in my own personal time. Study sessions, for example, were conducted exclusively during evenings and weekends.

**Recruitment and Screening for Study Participation**

First, study participants were recruited offline (through flyers posted at libraries, for example) and online (through various websites and email lists). All recruiting materials and study communication informed participants that they would receive $10 USD (or equivalent currency) as compensation following study completion. Prospective study participants were referred to a study information web page and required to submit consent forms. After submitting the consent form, potential subjects were directed to and completed a screening questionnaire.

Both new and existing Second Life users, known as Residents, were recruited for study participation. In order to account for the factor of Second Life experience, the screening questionnaire included three questions about experience, which asked if the participant had ever logged in to Second Life before, how long ago their account had been created, and how many total hours they had spent in the virtual world. These questions were taken into account during the intercorrelations phase of data analysis.

In order to qualify for study participation, prospective participants had to meet three conditions:

1. They had to agree to accept the test condition (communication modality) they were assigned to, and specifically to use spoken voice communication if assigned to it.

2. They had to meet the minimum system requirements (asked across two questions about Internet connection speeds and RAM). Otherwise, Second Life may not have been able to run on their computers, and it was easier to disqualify these participants at the outset.

3. They had to have a high degree of availability (i.e., be available at several dates and times listed in the screening questionnaire).

Next, qualified subjects were notified by email as to whether or not they had been selected for study participation and, later, randomly assigned to study conditions. Table 2.1 presents the total number of participants recruited and
the number of participants who did not qualify or who were eliminated for other reasons.

**Table 2.1  Reasons for Elimination of Prospective Study Participants and Final Sample Size (N = 60)**

<table>
<thead>
<tr>
<th>Reason for elimination</th>
<th>Number of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total recruited</td>
<td>149</td>
</tr>
<tr>
<td>Did not agree to accept assigned condition</td>
<td>24</td>
</tr>
<tr>
<td>Did not meet minimum system requirements</td>
<td>6</td>
</tr>
<tr>
<td>Did not attend scheduled study session (no-show)</td>
<td>46</td>
</tr>
<tr>
<td>Attempted to but could not attend or complete study session (technical difficulties)</td>
<td>10</td>
</tr>
<tr>
<td>Total study participants</td>
<td>63</td>
</tr>
<tr>
<td>Removed during analysis phase due to missing data or as outliers</td>
<td>3</td>
</tr>
<tr>
<td><strong>Final study sample (N)</strong></td>
<td><strong>60</strong></td>
</tr>
</tbody>
</table>

* The sample size of 60 participants was the bare minimum required to achieve the desired power for statistical analysis (Cohen, 1998).

**Sample Demographics**

Of the 60 participants in the final sample, 37 (62%) were female and 23 (38%) were male. The ages of participants ranged from 18 to over 80 years of age: 23 (38%) were under age 35, 25 (42%) were under age 55, and 12 (20%) were age 55 or more. Most of the participants (51 or 85%) identified as White/Caucasian; of the remaining nine participants, two (3.2%) were African American, one (1.6%) was Hispanic, four (8.1%) were Asian, one (1.6%) was Native American, and one (1.6%) was a Pacific Islander.

Three aspects of prior Second Life use were considered: whether or not a participant had ever logged into Second Life before, the age of their account, and the total time spent in the virtual world. Forty-one (68%) participants had logged into Second Life at least once before the Introduction to Building in Second Life class, while nineteen (32%) had not. In regard to account age, 34 (57%) participants had created their Second Life account less than six months ago, while 26 (43%) had their Second Life account for more than six months. In response to the question: “How much time (total hours) have you spent in Second Life?” 28
Stephany F. Wilkes reported that they had spent less than 2 hours total in Second Life, 9 (15%) had spent 2 to 20 hours total in Second Life, and 23 (38%) had spent 20 or more total hours in Second Life.

Introversion and extroversion, video game use, availability for study participation, and primary interest in Second Life (Table 2.2) were also recorded in the screening questionnaire. Typing proficiency was not self-reported, an omission that may have later influenced text chat outcomes in Groups T and VT. All participants were, however, asked to describe their computer experience through questions about their usual computer input device (mouse, keyboard, and so on): if they already used voice chat in Second Life, if they owned a headset with a microphone for speaking, what kind of computer they would use to participate in the study, the speed at which they usually connected to the Internet, and the amount of computer memory their machine had. Finally, at the beginning of each activity session for Groups V and VT, the instructor described how to use Second Life volume controls, how to mute one’s own microphone and that of other participants, and how to use Talk controls.

Table 2.2 Frequencies of Primary Interest in Second Life \( (N = 60) \)

<table>
<thead>
<tr>
<th>Topic of interest</th>
<th>Not at all</th>
<th>Somewhat</th>
<th>Fairly</th>
<th>Extremely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socializing</td>
<td>9 (15%)</td>
<td>20 (33%)</td>
<td>20 (33%)</td>
<td>7 (12%)</td>
</tr>
<tr>
<td>Building/content creation</td>
<td>13 (22%)</td>
<td>11 (18%)</td>
<td>15 (25%)</td>
<td>16 (27%)</td>
</tr>
<tr>
<td>Virtual business</td>
<td>26 (43%)</td>
<td>13 (22%)</td>
<td>11 (18%)</td>
<td>6 (10%)</td>
</tr>
<tr>
<td>Providing education</td>
<td>22 (37%)</td>
<td>13 (22%)</td>
<td>9 (15%)</td>
<td>10 (17%)</td>
</tr>
<tr>
<td>Receiving education</td>
<td>11 (18%)</td>
<td>21 (35%)</td>
<td>11 (18%)</td>
<td>13 (22%)</td>
</tr>
</tbody>
</table>

Note: Responses on a 4-point, Likert-style scale.

Participant Assignment to Condition

For the study, I used a stratified random sample controlled for age and gender. The results of a Chi-squared test (Table 2.3) indicated that assignment to condition was random.

Each participant and study session had one of three randomly assigned communication conditions: either the Voice (V), Text (T), or Voice and Text (VT) communication modality. Everyone attending a particular study session was automatically “in” their test condition simply by using the communication modality provided and stated at the beginning of class. Group V had 18 participants while Groups T and VT had 21 participants each.
Table 2.3  Chi-square Scores for Assignment to Test Condition Based on Age, Gender, and Race

<table>
<thead>
<tr>
<th>Participant attribute</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.34</td>
<td>4</td>
<td>.86</td>
</tr>
<tr>
<td>Gender</td>
<td>1.48</td>
<td>2</td>
<td>.48</td>
</tr>
<tr>
<td>Race</td>
<td>0.49</td>
<td>2</td>
<td>.78</td>
</tr>
</tbody>
</table>

Study Activity

Next, using a typical computer setup (i.e., either a desktop and mouse, or laptop and mouse) participants attended one learning activity, the *Introduction to Building in Second Life* class, for approximately 45 minutes. During the class, participants completed building exercises with an online instructor, a female avatar with female voice. Both avatar appearance and instructor were kept constant throughout the study.

Following the *Introduction to Building* class, subjects completed an online questionnaire with three sections on perceived mental strain (cognitive load), presence, and retention, comprised of questions about material covered during the *Introduction to Building* class. First, participants completed two 7-point Likert-scale format questions that asked them to report perceived mental strain. Reports of mental strain were followed by a Presence Questionnaire (PQ) (Witmer & Singer, 1998; Gerhard, Moore, & Hobbs, 2001), a 7-point Likert-scale format with 28 questions rating their experience in different areas of the 3D environment. After completing the online PQ, participants completed a learning (retention) assessment, a test of how well they remembered material from the learning activity. Only short-term retention was measured. Participants were compensated immediately after the online questionnaire was submitted.

Study Instrumentation and Reliability

All of the study instruments described in this section were part of a single, online post-activity questionnaire. Three dependent variables were measured: presence, retention, and cognitive load. Two of the instruments used in this study were the creation of the author (the cognitive load and retention instruments) while one instrument, the Presence Questionnaire (PQ), was created and is copyrighted by Witmer and Singer (1998) with permission to make copies of part or all of their work for personal, classroom, and commercial use.
Witmer and Singer’s PQ is frequently employed in virtual reality research and cited in the literature on immersion and presence in virtual environments. I employed the PQ (version 3.0) in this study for many sound reasons: because Moreno (2006) used it in her work on modality and presence, and I aimed to expand on her work; it is complete in that it measures multiple aspects of presence; its validity and statistical reliability is well established and no serious shortcomings of the instrument have been detected in more than a decade; it measures presence data collected from participants immediately following exposure to a virtual environment, which also took place in this study; and because the items in the PQ applied neatly to Second Life without requiring alteration. Witmer and Singer have described their tests of reliability on the PQ at length in the literature (Witmer, Jerome, & Singer, 2005).

Though Second Life is not the same (or even the same kind of) virtual environment in which the PQ was originally applied, the relevance of PQ questions to Second Life inspired application to a social virtual world, just as Moreno (2006) applied the PQ to an environment different from those studied by Witmer and Singer (1998) and Witmer et al. (2005).

The 28-item Presence Questionnaire was measured on a 7-point Likert-type scale ranging from “No/None” to “High/Extremely” on presence variables. Example items included: “How well could you move or manipulate objects in the virtual environment?” and “Were you able to anticipate what would happen next in response to the actions that you performed?”

Reliability of All Presence Items
The PQ was supplemented with 11 post-experiment questions from Gerhard et al. (2001). These questions are more attitudinal in nature and measure immersion, communication, involvement, and awareness. I added these because they were considered highly salient based on the design of the study, and because they directly ask participants about social presence (co-presence) while the PQ does not. The 7-point Likert-style scale was applied to these questions in order to keep response format consistent. Example items included, “How responsive were the avatars of other participants to verbal communication that you initiated?” and “How compelling was your sense of other participants being present?”

The total presence score was a sum of all responses to all presence questions, beginning with the PQ items followed by Gerhard et al.’s items (2001). Gerhard et al.’s presence sub-scales were not included or evaluated as part of this study. Because the addition of Gerhard’s questions could have compromised the
established reliability of Witmer and Singer’s PQ, reliability was tested for all presence items and was strong with Cronbach’s alpha .95.

**Reliability of Cognitive Load Measures**

The cognitive load measures were brief and typical of cognitive load evaluations, in which study participants are asked to report their perceived levels of mental effort and strain. The cognitive load assessment was comprised of two 7-point Likert-type scale options that ranged from “Not at all” to “Extremely.” They were, “How difficult was the Introduction to Building class?” and “How much mental strain did you experience during the Introduction to Building class?”

**Reliability of Retention Measures**

The retention instrument was originally comprised of nine, and later eight primarily multiple-choice style questions designed to measure different types of learning outcomes: declarative knowledge, concept learning, and procedural knowledge (Smith & Ragan, 2005). Example items included, “Which of the following shapes are ‘prims’ in Second Life? Please check all that apply” and “Below is the list of steps required to construct a table in Second Life. The steps are listed in the incorrect order. Please re-order the steps shown below to successfully build the table shown here (a table with an oval top and four cylindrical legs). Select the #1 button for the first step you should take, with #8 marking the last step in the procedure.” The total retention score was the sum of all correct answers.

To establish internal scale reliability for each of the three scales presented above, reliability was calculated using Cronbach’s alpha for the dependent variables (retention, presence, and cognitive load). All of the items that constituted a particular measure were entered and a Cronbach’s alpha was run for that measure, resulting in the internal consistency of that particular scale. The procedure and results of scale reliability for all items on the test instrument are presented in Table 2.4.

**Table 2.4** Chi-square Scores for Assignment to Test Condition Based on Age, Gender, and Race

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Scale alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention</td>
<td>.51</td>
</tr>
<tr>
<td>Cognitive load</td>
<td>.94</td>
</tr>
<tr>
<td>Presence</td>
<td>.95</td>
</tr>
</tbody>
</table>
The original retention alpha was 0.495, too low to be considered reliable. Factor loadings indicated that one question reduced scale reliability when included. It was “Look at the image below. Three lines are shown: a red line, a green line, and a blue line. Each line represents one angle, or axis. Choose the colour of the line that represents the Z axis.” This question was not retained, bringing Cronbach’s alpha to 0.51, which still did not meet the standard of 0.7 desired reliability. Upon further examination, I noted that reliability was underestimated: The retention scale was not a continuous measure, but dichotomous, and another measure of reliability would be more appropriate. I also note this limitation in the Discussion section.

Results

Did Demographic Variables Influence Outcomes?

First, to determine if demographic variables might affect observed outcomes, I ran descriptive statistics and Pearson’s two-tail correlations for all variables (independent, dependent and demographic; see Table 2.5). As I noted previously, I used a stratified random sample controlled for age and gender, and a Chi-square test confirmed that the assignment to condition was indeed random and not significant for age ($p = .855$), gender ($p = .477$), or race ($p = .782$). Other demographic variables (time and experience) were not included in the stratified random control or Chi-square test, so I conducted analyses to investigate whether the independent variable and/or demographic variables (time, experience, age, race, or gender) impacted cognitive load, retention, or presence.

The results of the Pearson’s correlations (Table 2.5) were negative and significant for retention and presence ($r[54] = -0.36, p = .01$), for cognitive load and retention ($r[54] = -0.26, p = .05$), for experience and presence ($r[54] = -0.31, p = .02$), for experience and retention ($r[54] = -0.39, p = .001$), and for time and cognitive load ($r[54] = -0.36, p = .01$). The results of the Pearson’s correlations were positive and significant for time and presence ($r[54] = 0.41, p = .001$) and for time and experience ($r[54] = 0.50, p = .001$).

Next, I calculated means and standard deviations for the independent variable of communication modality (Group V, Group T, and Group VT) on each of the three dependent measures (presence, retention, and cognitive load). These are shown in Table 2.6.

Then, I completed a one-way analysis of variance (ANOVA) for each dependent variable (cognitive load, retention, and presence). The results of the ANOVA between groups were significant for cognitive load ($F[2, 54] = 4.58, p = .01$) and
retention \( (F[2, 54] = 3.53, p = .04) \) but were not significant for presence \( (F[2, 54] = .65, p = .53) \).

**Table 2.5** Correlation Matrix for Independent, Dependent, and Demographic Variables \( (N = 60) \)

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>V/T/VT</td>
<td>2.05</td>
<td>0.81</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence</td>
<td>185.47</td>
<td>27.14</td>
<td>-0.124</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention</td>
<td>.72</td>
<td>0.17</td>
<td>-0.25</td>
<td>-0.36**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive load</td>
<td>2.14</td>
<td>1.22</td>
<td>-0.22</td>
<td>-0.24</td>
<td>-0.26*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>3.80</td>
<td>1.54</td>
<td>0.28**</td>
<td>-0.31*</td>
<td>-0.39**</td>
<td>0.09</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1.92</td>
<td>0.93</td>
<td>-0.15</td>
<td>0.41**</td>
<td>0.48**</td>
<td>-0.36**</td>
<td>0.50**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>3.25</td>
<td>1.54</td>
<td>0.09</td>
<td>-0.11</td>
<td>-0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>1.62</td>
<td>0.49</td>
<td>0.13</td>
<td>0.12</td>
<td>-0.11</td>
<td>0.04</td>
<td>-0.17</td>
<td>0.15</td>
<td>-0.03</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>1.43</td>
<td>1.18</td>
<td>0.08</td>
<td>-0.02</td>
<td>0.11</td>
<td>-0.08</td>
<td>-0.18</td>
<td>-0.04</td>
<td>-0.23</td>
<td>0.12</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*p < .05 (2-tailed).
**p < .01 level (2-tailed).

**Table 2.6** Means and Standard Deviations for Cognitive Load, Retention, and Presence by Communication Modality

<table>
<thead>
<tr>
<th>Condition</th>
<th>V</th>
<th>T</th>
<th>VT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Cognitive load</td>
<td>18</td>
<td>5.56</td>
<td>2.57</td>
</tr>
<tr>
<td>Retention</td>
<td>18</td>
<td>5.19</td>
<td>0.9</td>
</tr>
<tr>
<td>Presence</td>
<td>18</td>
<td>188.07</td>
<td>27.33</td>
</tr>
</tbody>
</table>

I hypothesized that retention scores would be higher for voice participants (Group V) and voice-and-text participants (Group VT) than for text-only (Group T) participants. The results of a one-way ANOVA to test this hypothesis were significant \( (F[2, 54] = 3.53, p = .04) \). Retention scores were highest for Group T \( (M = 5.38, SD = 1.14) \), lower for Group V \( (M = 5.19, SD = 0.89) \), and lowest for Group VT \( (M = 4.49, SD = 1.32) \).
Cognitive load was not lower for voice participants, as expected; in fact, cognitive load was lowest for text participants. The results of a one-way ANOVA to test this hypothesis were significant ($F[2, 54] = 4.58, p = .01$). Cognitive load was lowest for Group T ($M = 3.33, SD = 2.08$), higher for Group VT ($M = 4.14, SD = 2.26$), and significantly higher for Group V ($M = 5.56, SD = 2.57$).

I expected voice use to contribute to greater perceptions of presence but it did not. The results of a one-way ANOVA to test this hypothesis were not significant ($F[2, 54] = .65, p = .53$). Group T had the highest presence scores ($M = 188.70, SD = 24.14$), while Group V had just slightly lower presence scores ($M = 188.07, SD = 27.33$), and Group VT the lowest ($M = 180.00, SD = 30.11$).

Finally, I expected that perceptions of presence would not correlate with retention, but results indicated the opposite. The results of a Pearson’s two-tail correlation to test this hypothesis were positive and significant ($r(60) = .36, p = .01$). Perceptions of presence did correlate with short-term retention.

To determine if the means were significantly different, I ran a post-hoc analysis (Tukey) for all dependent variables (cognitive load, retention, and presence) by condition (Group V, Group T, and Group VT). Post-hoc analysis showed mean differences were significant ($p < .05$) between groups V and VT on retention ($p = .03$), between groups T and VT on retention ($p = .04$), and between groups V and T on cognitive load ($p = .01$). See Table 2.7.

**Table 2.7 Post-hoc Differences in Means of Cognitive Load, Retention, and Presence by Condition**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean difference</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive load</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V and T</td>
<td>1.11*</td>
<td>.37</td>
<td>.01</td>
</tr>
<tr>
<td>V and VT</td>
<td>.11</td>
<td>.37</td>
<td>.15</td>
</tr>
<tr>
<td>T and V</td>
<td>-1.11*</td>
<td>.37</td>
<td>.01</td>
</tr>
<tr>
<td>T and VT</td>
<td>-.40</td>
<td>.35</td>
<td>.50</td>
</tr>
<tr>
<td>VT and V</td>
<td>-.71</td>
<td>.37</td>
<td>.15</td>
</tr>
<tr>
<td>VT and T</td>
<td>.40</td>
<td>.35</td>
<td>.50</td>
</tr>
<tr>
<td><strong>Retention</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V and T</td>
<td>-.03</td>
<td>.05</td>
<td>.86</td>
</tr>
<tr>
<td>Condition</td>
<td>Mean difference</td>
<td>SE</td>
<td>p</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------</td>
<td>-----</td>
<td>--------</td>
</tr>
<tr>
<td>V and VT</td>
<td>0.10</td>
<td>0.05</td>
<td>.15</td>
</tr>
<tr>
<td>T and V</td>
<td>0.03</td>
<td>0.05</td>
<td>.86</td>
</tr>
<tr>
<td>T and VT</td>
<td>0.13*</td>
<td>0.05</td>
<td>.04</td>
</tr>
<tr>
<td>VT and V</td>
<td>-0.10</td>
<td>0.05</td>
<td>.15</td>
</tr>
<tr>
<td>VT and T</td>
<td>-0.13*</td>
<td>0.05</td>
<td>.04</td>
</tr>
</tbody>
</table>

### Presence

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean difference</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>V and T</td>
<td>0.63</td>
<td>8.77</td>
<td>1.00</td>
</tr>
<tr>
<td>V and VT</td>
<td>8.06</td>
<td>8.77</td>
<td>.63</td>
</tr>
<tr>
<td>T and V</td>
<td>0.63</td>
<td>8.77</td>
<td>1.00</td>
</tr>
<tr>
<td>T and VT</td>
<td>8.7</td>
<td>8.43</td>
<td>.56</td>
</tr>
<tr>
<td>VT and V</td>
<td>-8.06</td>
<td>8.77</td>
<td>.63</td>
</tr>
<tr>
<td>VT and T</td>
<td>-8.7</td>
<td>8.43</td>
<td>.56</td>
</tr>
</tbody>
</table>

* *p < .05.

### Time and Experience

Finally, I conducted exploratory analysis for time and experience, as the correlation matrix (Table 2.5) showed a relationship between these two extraneous variables, the independent variable (communication modality), and the three dependent variables (cognitive load, retention, and presence). While not hypothesized, strong correlations (both positive and negative) indicated that time and experience should be treated as second independent variables in future analyses. I ran six 2-way ANOVAs with the independent variables of communication modality and either time or experience, and one of the dependent variables (cognitive load, retention, or presence).

Cognitive load was lowest for Group T ($M = 1.67$) regardless of experience. There was no significant difference between Group T participants who had created their Second Life accounts less than six months ago ($M = 1.82$) and more than six months ago ($M = 1.50$).

Cognitive load was higher for Group V ($M = 2.77$) than for Group T ($M = 1.67$) and Group VT ($M = 2.07$), without a significant difference between Group T and
Group VT. There was a significant between-groups effect for communication modality and time spent in Second Life ($p = .23$). The text condition (Group T) and time spent in Second Life (> 20 hours) correlated with a greater decrease in cognitive load.

**Discussion**

*Experience and Time are Significant*

In this study I sought to determine whether or not communication modality (assignment to Group V, T, or VT) had an impact on cognitive load, retention, and presence. The most significant finding is that experience and time had significant effects on measures of cognitive load, retention, and presence. Significant between-subjects effects were found for cognitive load and time ($p = .23$), for retention and time ($p = .21$), and for retention and experience ($p = .03$).

These findings are not surprising. According to Eyring, Johnson, and Francis (1993), task familiarity (an individual’s possession of declarative knowledge and procedures relevant to performance of a given task) can be gained through prior experience with the task, or through experience with tasks similar in terms of declarative knowledge and procedures. Eyring et al. (1993) cite Ackerman (1989), who suggested that previous experience may allow an individual to pass more rapidly through early stages of skill acquisition and reduce cognitive demands. It is possible, based on my findings about the impact of time and experience, that more experienced Second Life users had learned how to better use the virtual world’s interface (which was necessarily a strong component of the learning activity) and may have also known other content included in the learning activity script.

Retention scores were approximately the same for participants in the voice condition group (Group V), regardless of the time (either < 20 hours or > 20 hours) they had spent in Second Life. The text condition group (Group T) correlated with increased retention, but only for those participants who had spent more time in Second Life (i.e., > 20 hours, $M = .81$). Based on these results, it is impossible to determine whether time spent in Second Life, the communication modality, or both led to improvements in learning. For retention, between-groups effects were significant for communication modality and time spent in Second Life ($p = .21$).

Though retention increased slightly for Group T, it did not increase significantly for both experience levels. Retention scores were highest for Group VT participants who also had more than six months’ experience ($M = .88$). The VT
condition had the opposite effect on less experienced participants, for whom retention dropped significantly \((M = .57)\). For retention, the between-subjects effects for communication modality and experience were significant \((p = .03)\).

For presence, between-subjects effects were significant for time \((p = .01)\) but not for communication modality and time \((M = .975)\). The text condition (Group T) correlated with increases in perceptions of presence for both experience levels in Second Life. For presence, between-groups effects were significant for experience \((p = .014)\) but not for communication modality and experience in Second Life \((p = .938)\).

**Cognitive Load was Significantly Higher for Voice Conditions**

Cognitive load was lowest for Group T \((M = 3.33, SD = 2.08)\), higher for Group VT \((M = 4.14, SD = 2.26)\), and significantly higher for Group V \((M = 5.56, SD = 2.57)\). This is the opposite of the result I expected. As I reviewed earlier, using a visual modality to present both pictorial (in this case, the 3D objects in the *Introduction to Building* class) and verbal information (text-based instruction) can create an overload situation for the learner. More memory capacity is thought to be available when dual modalities are used (Penney, 1989).

It would be careless not to explore the most simple explanations first; the availability of chat log text transcripts for Groups T and VT. It is possible that cognitive load may have appeared higher for Group V because the text modality provides transcripts, and a text record would be very helpful in aiding recall and retention. Group V participants may have had to work harder to remember everything without a record to refer to. A chat window of text may also have reduced cognitive load not just after class, but during as well; Group T participants would have quickly realized that, if they missed a step, they could scroll back and catch up later by scrolling forward. This ability could have reduced perceptions of mental strain or a feeling of falling behind during instruction. Unfortunately, the nature of remote, online research makes it impossible for the instructor to see or control whether or not participants in Groups T and VT retained full or partial text transcripts of the material.

Another simple explanation is participant familiarity with the communication tools themselves and extraneous cognitive load. While text-based chat and email communication is common online, voice communication may be less so. At the time I conducted the study, Second Life had its own voice communication system; familiarity with generally available voice chat applications like Skype would not necessarily have applied or extended to the virtual world. The fact that participants had to learn how to use voice chat in Second Life (even
with instructor guidance) may have increased extraneous cognitive load during instruction.

Another possible interpretation is that auditory load may be intrinsically more demanding (Mousavi et al., 1995), or that voice communication in combination with social virtual worlds creates higher auditory load than might be expected for other multimedia learning contexts. The results seen here, of Group V experiencing the highest cognitive load, support those of Mousavi et al.’s (1995) experiment in which two groups, a visual-visual mode and an auditory-auditory mode, were studied to see if visual processes were inherently more demanding. Mousavi et al.’s (1995) results suggest that the auditory-auditory mode may have been more intrinsically demanding of cognitive load; a significant difference favoured the visual-visual group (which was most like Group T in this study).

Finally, in Group V, complementary rather than different simultaneous information was presented. According to Penney’s (1989) separate streams hypothesis, different content coming in through two streams is expected to free up working memory. In Penney’s (1989) work, lists of different items presented simultaneously to different sensory modalities improved short-term memory retention. The instructor’s script for the Introduction to Building class was comprised of complementary information to the actions taking place in Second Life. The nature of the study activity, then, may have played a role in contributing to higher load for Group V.

Voice Use did not Contribute Significantly to Greater Perceptions of Presence

Group T had the highest presence scores ($M = 188.70$, $SD = 24.14$) while Group V had just slightly lower presence scores ($M = 188.07$, $SD = 27.33$) and Group VT the lowest ($M = 180.00$, $SD = 30.11$). Some research (Sallnas, 2004) indicated that modality (specifically voice) could impact perceptions of presence. Other research (Moreno & Mayer, 2002) found that modality did not impact perceptions of presence, while the level of immersion made possible with the media environment (a desktop vs. head-mounted display) did affect presence. The results of this study support the findings of Moreno and Mayer (2002) and do not extend those of Sallnas (2004).

One possible explanation for these results (though such data was not collected as part of this study) is that the combination of real-world voices with inworld avatars may have lessened the “fantasy impression” of avatars, reminding participants more of the real world than the virtual one. To address this, future research should evaluate perceptions of presence with real voices.
and voice fonts, which are filters that make human voices sound different. Comparisons between two voice groups (Real Voice [RV] and Voice Font [VF]), studied along with Text Only, Real Voice and Text (concurrently) and Voice Font and Text (concurrently) groups could provide more substantive data on voice types and perceptions of presence in virtual worlds.

Finally, no or low levels of conversation may have contributed to a decrease in participants’ sense of co-presence. A sense of being with others includes things like primitive responses to social cues, such as laughter, and these cues are thought to ultimately aid recall. Participants were not required to engage in any conversation, whether text or voice-based, while attending the Introduction to Building class. While data on non-instruction conversation was not recorded as part of this study, anecdotal recollection brings to mind only two livelier classes that evidenced these types of primitive responses. It is possible that little overall communication may have made the voice modality less salient to the study experience than it would have been in other contexts, such as a virtual support group for people recently diagnosed with a disease, in which human connection through speaking with others is of primary interest.

Related to this is the fact that participants were randomly assigned to their conditions and were unlikely to have prior acquaintance with one another; they simply may not have felt comfortable striking up a conversation with unfamiliar avatars. In addition, the Introduction to Building class was a one-time-only experience for study participants, perhaps giving little reason to engage in conversation since users could be quite confident of not encountering the person (or avatar) again. For these reasons, these results may not hold up in a context where students attend a class over a longer period of time, get to know one another better, attend a class both in real life and Second Life, and/or possibly engage in more voice communication. Conflicts in prior research on voice, immersion, and presence still exist and are worthy of further study.

**Perceptions of Presence did Correlate with Retention**

Perceptions of presence did correlate with retention. The results of a Pearson’s two-tail correlation to test this hypothesis were positive and significant: \( r = .36, p = .01 \). These findings should be viewed with caution, however, because the data I collected as part of this study do not really tell us why. Unfortunately, this study does not shed much light on the specific aspects of presence (whether social cues, attributes of the virtual environment itself, and so on) that most influenced retention; how presence may have influenced attention, which Baddeley and others have shown is so important in retention, or generally how
the mechanisms between perceptions of presence, retention, and cognition function. As Miller (2011) points out, the field of memory research is complex and recommendations to instructors should not be overly simplified.

The retention measures I used in this study could have been, and should become, stronger. First, I only measured short-term retention. Results may differ if longer-term retention is measured. In addition, the retention instrument was brief in terms of the number of questions it contained and focused on Second Life-specific how-to content, which is unusual, not common in instruction, and difficult to judge for robustness. This study should be replicated with non-Second Life specific content (ideally multiple types of subject matter) to see if the results for retention and presence hold up across different types of subject matter. Varying subject matter, of course, introduces the possibility for high experience or existing knowledge on the part of participants, so those factors should be examined and, if necessary, controlled for.

Though limited, then, this finding does add some empirical evidence to the idea that greater perceptions of presence do contribute positively to the type of learning (active) studied here. The relationship between and cognitive mechanisms behind presence and retention is a rich area for future study by educational and cognitive psychologists.

Future Directions and Conclusion

It is important to remember that these study results are based on a very small sample (N = 60). Replicating this study with a larger sample may strengthen or change the results, especially if a larger sample size is used with striated random assignment based on experience and time (suggested) instead of race, gender, and age (used here).

As Miller (2011) states, the memory theory literature suggests powerful linkages between attention and memory, and that what has been traditionally known as short-term memory may, in fact, be attentional focus. Miller also points out that information that is “vivid, emotionally arousing and meaningful” tends to win out in memory recollection and retention (2011, pp 11–122). Future work on virtual environments should investigate the role of and measure attention and whether the ability of a virtual environment to provide vivid, emotionally arousing information influences learning in both the short term and long term.

This study dealt only with short-term memory and retention, in that the post-activity assessment always took place immediately after the Introduction
to Building class. No participants, for example, attended the class and then returned to the post-activity assessment hours, days, or weeks later. Retention should be measured over a longer period to see if differences are found over time and improved long-term memory storage and retention is, most likely, the goal of most educators.

The prior spatial, Second Life interface, and other abilities of participants were not evaluated or taken into account but should be for similar studies. In addition, “real world” subjects that are commonly taught should be used in instruction rather than the Second Life-specific material taught here (useful as an equalizing factor across a small sample but uncommon and not highly relevant).

These results should not be generalized across virtual worlds; I confined this study to surveying, interviewing, and observing Residents within Second Life only. A study of additional social virtual worlds and the comparison of communication habits between users of each is an interesting direction for further research, especially to see how challenging it is to generalize findings across populations (and test how world-specific some findings may be).

Acknowledgements

The author gratefully acknowledges her adviser, Dr. Susan Feinberg and the study participants who made this work possible.

REFERENCES


doi:10.15215/aupress/9781771991339.01


