Running head: Effects of Socially Relevant Representations on Multimedia Learning

--ROUGH DRAFT--DO NOT CITE!--

The Effects of Socially Relevant Representations on Multimedia Learning.

Joel D. Galbraith

Christopher M. Hoadley

Penn State University

--ROUGH DRAFT--DO NOT CITE!--

Abstract

Much research suggests that students learn better when deeply engaged in learning tasks. One possible engagement strategy is to use socially relevant representations to make instruction more personable. This study looks at the effects of two social factors (authorial voice and human speech) on both learning and perceptions of social presence in multimedia instruction. While no learning benefits were apparent involving either lower or higher processing, significant differences did appear with respect to social presence measures. Implications of these mixed results are discussed with regard to their potential relationship to learner motivation and satisfaction. The robustness of the newly developed Solitary Learner's Inventory of Social Presence (SLISP) is also discussed.

Introduction:

Consider the following instructions given to faculty developing distance education courseware at a large western university.

> Choose one person you know who could be a prospective student in the course. You may wish to jot down this student's name at the top of the course outline so you can keep him or her in mind as you write. Whenever you work on the course, visualize this person. As you write, carry on a dialogue, a sort of tutorial, in your mind with this person. Using this technique will often help you create a course that has a warm and friendly tone.

Why are such instructions given to course developers? And what effect does a "warm and friendly tone" have on learners and their comprehension? In the literature surrounding the design of multimedia courseware and textbooks, such personable and socially relevant representations (SRR) are believed to positively influence learning (Hoadley & Kirby 2004; Mayer, Fennell, Farmer & Campbell, 2004; Mayer, Sobko & Mautone, 2003; Paxton, 2000; Beck, McKeown & Worthy, 1995). Mayer and colleagues write specifically of a "personalization effect" achieved through "using words in a conversational style rather than a formal style" (Mayer, Fennell, Farmer & Campbell, 2004 p. 389). Others further describe "personalized messages", whereby directly addressing learners using words like "I" and "you", is believed to elicit more active participation of students (Moreno & Mayer, 2000, 2004).

Both of these related strategies are held, by these authors, to be more meaningful when seen as socially-relevant ways of representing information, rather than as a personalization issue per se. Hoadley and Kirby (2004) define SRR's as "any representation (presumably in software, but also in other media) which contributes information that is not part of the traditional domain area content, but rather is used for (or derived from) social interaction." (p. 264). In this view, induced conversational style and personalized messages in multimedia learning provide the learner opportunities to perceive a richer and more dialogic social encounter with an "other"(often the course author, the textbook author, a teacher, a student peer, or a computer agent etc.), than if the materials are written in a more formal, third person, expository style. The term "personalization", is also potentially problematic in that it can connote an adaptive or intelligent system that customizes instruction based on a learner's profile or tracked behavior. This form of personalization was not a part of the present study, nor of any of those reviewed for this paper. Thus, in this study, we call what others label "personalization" an example of socially relevant representations (SRR) (Hoadley, 1999; Hoadley and Kirby, 2004), and more specifically, in the present study, the two socially socially relevant representations employed are authorial voice and human speech.

Authorial voice refers to the relative prominence of the author in the texts or other learning material. Such texts can have a "visible" author that is highly evident in the text, who often "speaks" in first person and directly engages the reader. Indicators of authorial voice might include more dialectic rhetorical strategies (such as rhetorical questions), the use of first-person or a particular conversational style, or other engagement devices.

Materials presented with high authorial voice are believed to promote the development of social behaviors and relationships between a reader and the author, or student and instructor (Nolen, Johnson-Crowley & Wineburg, 1994). As noted by Eco, (Eco, U. (1994). #746 . *Zwischen Autor und Text. Interpretation und Überinterpretation.*, München: Hanser) all texts have some degree of authorial voice. The rhetorical style, however, employed in most multimedia instruction, college-level textbooks, online courses, and certainly academic journals, often exhibits very limited authorial voice. This empty and anonymous "textbookese" may leave students wondering "who writes this stuff, and why should I care?" Little is done to engender any mutuality between author and solitary, individual learner. The learner's failure to perceive any social relationship or sense the author's voice or presence may hinder engaging that student in a direct or indirect dialogue on the material to be learned.

Does limited authorial voice matter for learning? Do perceptions of social presence lead to more active learning and deeper comprehension? Is there any truth to the old maxim that "the student won't *care* how much you *know*, until s/he *knows* how much you *care*"? And can such empathy be communicated through SRRs?

Purpose:

Existing research does not appear to adequately address the impact of certain SRRs such as authorial voice and human speech on learner comprehension and perceptions of social presence in multimedia instruction—particularly in independent learning/training environments. To address this, the present study explores the hypothesis that the addition

of strong social cues induced by authorial voice and human speech will significantly increase student comprehension of a multimedia lesson and student perceptions of social presence over control conditions. By better understanding the effects of these SRRs in multimedia instruction, designers will be able to create more effective instruction and instructional environments.

Throughout this article, two distinct types of "voice" manipulations are employed and should not be confused. The terms "voice" or "authorial voice" will be used to represent a literary device or style of writing where the author's presence is readily evident in the prose. The author uses a more conversational tone and may speak in first person with phrases like "I think…" and "Have you ever wondered…". This use of "voice" must be distinguished from the second type of "voice" manipulation described in this text—referred to as "human speech" or "audio", which describes conditions in which audible human voice is used to convey messages and encouragement.

Review of Literature

Socially Relevant Representations (SRR) and Social Presence(SP)

SRRs are defined as representations or cues which contribute information that is not part of the traditional domain area content, but rather is used for (or derived from) social interaction (Hoadley & Kirby, 2004). Authorial voice is an SRR identified by Hoadley & Kirby (2004) that is believed to influence both a readers' comprehension and affective responses (Paxton, 2002; Nolen, 1995; Beck, McKeown & Worthy, 1995; Nolen, Johnson-Crowley & Wineburg, 1994). Instruction displaying clear authorial voice is

instruction that generally features a strong narrative voice in the text, clueing learners into what is the author is presenting as fact and perhaps more importantly, what should be considered the author's observations and opinion (Gibb 2002). Adding voice often involves employing a personal, conversational tone and heightening connections between the author and the reader (Beck, McKeown & Worthy, 1995). The visible author will often reference him or herself in the text using personal pronouns and phrases such as "I think", in my opinion", and "I believe" (Paxton, 2002), or directly address the reader/learner with phrases like "have you ever considered", "Now you try" and so forth.

Social Presence has many definitions, but can be generally defined as a sensory experience or perception indicating the presence of, and an interpersonal relationship with, another intelligent entity. (Biocca, 1997; Short, Williams, & Christie, 1976; Walther, 1992; Walther & Burgoon, 1992) Thus, when a learner is directly addressed by an author using such conversational dialogic phrases as presented in the previous paragraph, it can be expected to influence their perceptions of Social Presence. When such phrases are also communicated through audible human speech, the effects should be even greater.

Two ways to interpret instruction

Particularly relevant is research from the fields of communications and educational psychology that address the degree to which learner interactions with computers are interpreted as either social activity or information delivery. Where computers employ social cues, particularly those paralleling human to human interactions, the conventions

of natural human conversational exchange are triggered (Reeves & Nass, 1996) thus learners will tend to interact socially with such instructional content (Mayer, Sobko & Mautone, 2003).

Other researchers have gone further to describe the experiences of individuals engaged in computer-based instruction. Mayer, Sobko & Mautone (2003) describe at some length their related theory of Social Agency. They hypothesize that learners generally interpret multimedia learning episodes "as either a case of information delivery or a case of social communication" (p. 420). Indeed, learning software interfaces and environments help establish how the learner interprets the experience (Hoadley & Enyedy, 1999). Importantly, the way that learner interprets the episode—as social communication versus information delivery—influences the type of schemas that are activated in the learner, the type of cognitive processing that occurs during learning, and ultimately the quality of the learning outcome.

Mayer and his colleagues continue to describe a scenario where the learner receives a multimedia message with strong social cues (such as a human voice speaking in a conversational tone). Here, the learner is more likely to interpret the episode as a case of social conversation and, according to Reeves & Nass (1996), the conventions of human conversational exchange would be triggered in the learner.

So what mechanisms are at play here? Mayer et al (2003), address issues beyond the simple appeal or satisfaction of SRRs (although they call it a "personalization effect").

They suggest that social cues in a multimedia lesson elicit a learning scenario consistent with constructivist orientations where:

> The learner engages in a sense-making process including selecting relevant information, organizing it into a coherent representation, integrating it with other knowledge, and encoding it in memory. The result of sense-making processing is the construction of a meaningful learning outcome, which supports good performance on transfer tests. (p.420)

In contrast, consider their scenario for the information-delivery interpretation. This is likely the more traditional interpretation where an individual learner interacts with a multimedia lesson. Here, the weak social cues (such as text only, written in 3rd person), elicits a scenario where:

> human-to-human conversational rules are not activated, so the learner uses cognitive processing aimed solely at acquiring information rather than trying to understand it…paying attention to key ideas and trying to store them in memory. This processing leads to rote learning outcomes that lead to poor performance on transfer tests. (p.420)

Research in this area is not new but remains surprisingly inconclusive. A number of years ago, Salomon (1981) made similar conclusions. He found that when events were perceived by learners as communicational rather than informational, more mental effort was invested in them, and they invited more interaction. Both

Mikhail Bakhtin (1981) and Roland Barthes (1977) made foundational contributions to this topic. Baktin's d*ialogism*, and Barthe's *readerly* and *writerly* texts both describe similar notions of the author-reader relationship and the construction of knowledge/meaning.

Authorial Voice

"Textbookese" is a term used by Paxton to describe what may elicit the information-delivery interpretation. It describes a collection of objective facts composed by anonymous, authoritative author(s) (Paxton, 2002). Much multimedia instruction, including the control treatment of this study is written largely devoid of any social cues--using the rhetorical style of "textbookese". The omniscient, non-visible author(s?) gives little or no clue as to his or her personal point of view, nor makes any attempt to connect with the reader. It takes on a third-person voice that discourages questioning by the reader. Instead of being drawn in and becoming a participant of the construction of knowledge by the author, the non-critical reader defaults to a more instructive mode, adopting the information-delivery interpretation, and becoming a passive recipient of language (Olson, 1989; Mayer et al, 2003).

In support of Mayer et al's social communication scenario, Beck, McKeown & Worthy, (1995) and Schraw & Brunning, (1996), found that by personalizing the text, and imbuing it with voice, young students viewed the author as communicating directly with them, which increased their engagement and comprehension. In a similar vein, numerous university students reported strong affinity for, and a sense of interpersonal relationship

with, an instructor of a prerecorded-video course whom they had never met. (Galbraith & Spencer, 2002). The research cited show how individuals can interpret reading a textbook or interacting with a multimedia lesson as a communicative and social experience when appropriately cued—even though these modes of instruction would hardly be thought of as particularly social.

Human Speech

The sound of a human voice is a strong socially relevant and communicative cue, but is not always considered advantageous. For example, Walter (1999) argued that social cues in general were not all that relevant to learning. He found, that when it came down to just efficiency of information exchange, the vocal channel (human speech) offered little more than a carrier of language. He further stated that when partners were engaged in a task, vocal cues and phatics took a backseat to the content being conveyed verbally. Phatics are the utterances used to share feelings and establish a mood of sociability, but do not convey information or ideas. Similarly, Galegher and Kraut (1994) found that although text-based CMC groups were less satisfied with their communication than audio-enabled task groups, there were no significant differences in the quality of their task-related outputs that these conditions produced. Indeed many distance learning findings in this area indicate that social presence will more likely influence a student's motivation and desire to enroll in future online courses, than it will directly affect student learning outcomes. In other words, it appears that some hold that the instructional materials and a learner's desire to complete the task may be more pertinent to learning than socially relevant factors like authorial voice and human speech. The findings just cited, however,

come from data involving largely synchronous, two-way communication studies.

But, in a particularly challenging and lengthy multimedia lesson, can a communicative experience cued by highly evident authorial voice, trigger the deeper processing found by Mayer and colleagues with their shorter length and less complex treatments? Paxton (2002) might respond affirmatively. He cites "a small and as-yet-tentative body of research" that suggests that when a sense of authorship is heightened, "students tend to read more critically, more flexibly, and with a view to negotiating meaning for themselves. (p. 200) If this is true, significant results should also be realized in the current study, particularly so given the challenging instructional content of this study. More fundamentally, will the socially relevant representations of this study, authorial voice and human speech, be strong enough cues to really elicit social communicative behavior—even between a student and computer-based multimedia lesson? Might the learner really be drawn into a mental "conversation" with the author? Will the learning effects be readily observable throughout the range of educational objectives?

Comment [JDG1]: Could officially state RQs here

Methods:

The participants were 182 undergraduate college students from a variety of classes (Astronomy, English literature, Business, Spanish, Engineering). Approximately 48 participants were randomly assigned to each one of four conditions. Each participant viewed the multimedia lesson and completed the associated tests using a computer in one 60 minute session in a campus computer lab. (See Table and Figure 1) Participants in the two conditions with audio (human speech), were supplied with headphones, and asked to

put them on before beginning the materials. Most participants completed all online materials and quizzes well within the allotted time (60 minutes). Only a handful needed, and were given, a few extra minutes to complete the study.

Table 1

Condition Descriptions

Condition 1	No Authorial Voice, No Human Speech
Condition 2	No Authorial Voice, With Human Speech
Condition 3	With Authorial Voice, No Human Speech
Condition 4	With Authorial Voice, With Human Speech

Figure 1

Conditions Matrix

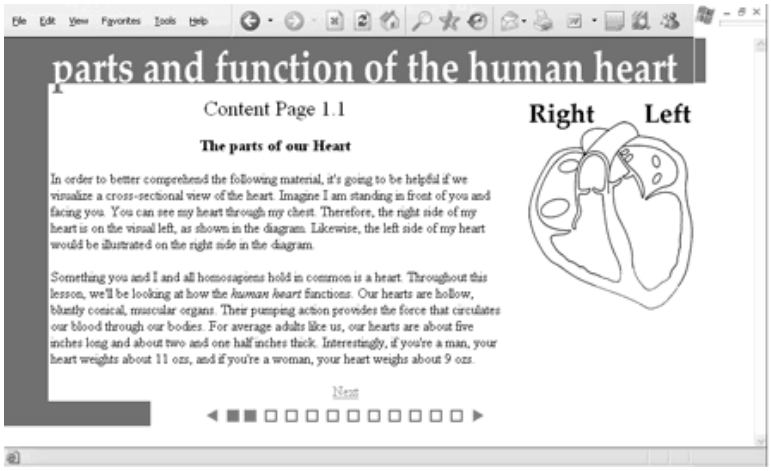
		Authorial Voice *(Instructional text in 1st person, with conversational tone)*	
		No	***Yes***
Human Speech (adult female voice for instructions with instructional text)	**No**	**Condition 1** *No Authorial Voice* No Human Speech	**Condition 3** *Yes Authorial Voice* No Human Speech
	Yes	**Condition 2** *No Authorial Voice* Yes Human Speech	**Condition 4** *Yes Authorial Voice* Yes Human Speech

Treatments:

The Dwyer Heart Content is designed to teach and then assess students' knowledge of the basic functioning of the human heart (Dwyer, 1978). It has been used in hundreds of empirical studies and has been thoroughly validated. The instructional materials and assessments take approximately 60 minutes to complete. In this study, all instructional content was presented in 20, non-scrolling computer screens (frames). Five additional screens were used to explain study procedures, provide progress feedback and quiz instructions. Each of the 20 instructional screens was split into two sections with textual instruction on the left two thirds of the screen and a static image the heart on the right third of the screen illustrating relevant features or concepts. Navigational controls were placed at the bottom of the screen. All materials were coded in HTML, and presented in the Microsoft Internet Explorer web browser. Human speech was recorded in Microsoft Windows Media Audio format, and presented in "autoplay" mode using a controllable embedded player within the HTML page. (see Figure 2)

Figure 2

Sample browser screen from condition 3 with Authorial Voice



To avoid potential information processing confounds related to presentation modality in this study, human speech was used in only two of the four conditions (Conditions 2 & 4), and occurred only in non-instructional screens which gave only general procedures for the study. Furthermore, where speech (adult female voice) was used, almost all of the screen text was removed except for select phrases aimed at reinforcing important research study procedures. Thus, audio was used simply to heighten social or relational cues between author and learner, and not to present any lesson materials. Learners were also not tested on any information provided on screens that had both human speech and text, thereby controlling for any problems related to dual processing (Mayer & Moreno, 1998) and limited capacity (Lang, 1995).

The SRRs employed to make the materials more personable were authorial voice and human speech as described previously. In many instances, phrases in the non-SRR conditions like "*the* heart is…" were changed to "*our* hearts are…" or "*your* heart is…" in the conditions displaying voice. Likewise, on study procedure screens, phrases like "Part 1 is now complete" and "when ready, click 'next' to begin the quiz", were changed to "Ok, we've now completed Part 1 and you're half way done" and "when you feel you're ready to continue, click 'next'". Audible messages were presented in like manner in the speech conditions. During instrument review processes, it was determined that simply substituting possessive pronouns like "your", "my" or "our" for the word "the" (i.e. "your heart" vs. "the heart"), would not sufficiently cue a social response from the learner.

Thus, in this study, SRR-ed conditions had closer to 13-38 more words per individual screen than their non-SRRed counterparts. Changes ranged from simple word (pronoun) substitutions to more significant additions and verbage to convey sociability. This level of change contrasts with the 13 pronoun substitutions made in a recent related study (Mayer et al., 2004). Care was also taken to add elements of authorial voice and conversational tone without "improving" the instruction. It was important to avoid adding anaphoric and cataphoric references (references to previously covered concepts—reminding; nor foreshadowing references--hinting of concepts to come in future screens), as these would in effect act as rehearsal or cueing strategies for learners. Such language devices and teaching techniques are arguably part of natural, conversational instruction, making it particularly challenging to create a natural conversational tone without unduly advantaging the treatment text and audio. Figure 3 provides side-by-side example text from both control and treatment conditions.

In addition, care was taken not to disproportionately affect the prepositional density (and therefore readability) of the treatment. Adding words to make treatments sound conversational was inevitable, but too much conversational language could have made the treatment text more readable and understandable than the control, thereby confounding any observed effects.

Figure 3

Sample Treatment Text (3rd person, formal style, 123 words) Conditions 1,2	**Sample Treatment Text (1st person, with authorial voice--conversational style, 139 words) Conditions 3, 4**
Content Page 1.2 The heart lies toward the front of the body and is in a slanting position between the lungs, immediately below the breastbone. The wide end points toward the right shoulder. The small end of the heart points downward to the front of the chest and toward the left. The lower portion of the heart is called the apex and is the part that can be felt beating. The human heart is two pumps combined in a single organ which circulates blood to all parts of the body. The heart is divided longitudinally into two halves by the septum. The two halves may be compared to a duplex house. Each house is independent of the other, separated by a common wall, which is the septum.	Content Page 1.2 The heart lies toward the front of our human body, in a slanting position between our lungs, and immediately below our breastbone. The wide end points toward our right shoulder. The small end of the heart points downward to the front of the chest and toward our left. The lower portion of our heart is called the apex. It is the part that we can feel beating. Put you right hand on the upper left of your chest; do you feel it? The heart is really like two pumps combined in a single organ which circulates blood to all parts of our body. It is divided longitudinally into two halves by the septum. We can compare its two halves to a duplex house. Each house is independent of the other separated by a shared wall which is our septum.
Sample Instructions Text (3rd person, formal style, 88 words) Conditions 1,2	**Sample Instructions Text (1st person, with authorial voice--conversational style, 126 words) Conditions 3, 4**
End of Unit 1 Unit 1 of the heart lesson is now complete. To review any material before taking the first quiz, click one of the boxes below to go back. If satisfied, click "Next" to start the first twenty-item multiple choice quiz. **Remember to enter the assigned ID number from the small slip of colored paper.** After completing the quiz, directions will be provided to move to Unit 2 of the heart lesson. If any questions arise, raise your hand and a research assistant will respond shortly.	End of Unit 1 Congratulations, you have now completed Unit 1 of our discussion on the human heart. If you want to quickly review anything before this first quiz, just click one of the boxes below to go back. If you are satisfied that you understand what we've been talking about, simply click "Next" to start your first 20-item multiple choice quiz. **Remember to enter the assigned ID number from off that small slip of colored paper that the researchers gave you.** When you've completed the quiz, you will be directed to move on to the last unit, and we'll pick up our heart discussion again. As always, if you have any questions raise your hand and a research assistant will come around to help you.

Instruments:

The Dwyer Heart Content assessments were designed to assess a learner's achievement on varied educational objectives from less complex identification activities to more complex measures of comprehension. The assessments have been systematically tested for reliability and validity. Assessing varying levels of learning was believed to be important and beneficial as most literature reviewed failed either to measure and/or report what type of learning was affected with different modalities or perceptions of presence.

Comment [JDG2]: Add citations

The 60 multiple-choice items used in this investigation were developed by Dwyer (1978) and consist of typical verbal stem and verbal response options. Dwyer's 20-item drawing test was not used in the present study. A brief description and historical reliability data of the tests follow.

Identification Test (IT)

The identification test (α =.82) was designed to evaluate the participant's ability to identify parts or positions of an object. The participants were required to identify parts of the heart numbered in a drawing by answering 20 multiple-choice questions. The objective of this test was to measure the student's ability to recall facts from the heart content, involving generally lower-level cognitive processing.

Terminology Test (TT)

This test (α =.82) was designed to measure knowledge of specific facts, terms and definitions pertinent to the heart content. This 20-item multiple choice test was used to evaluate the participant's ability to learn concepts.

Comprehension Test (ST)

This test consists of 20 multiple-choice items (α=.80). It was designed to evaluate the participant's knowledge of the functions of the heart as they occur during operational phases of the heart. The comprehension test was designed to measure participant's understanding of relationships between concepts and facts, and assessed a participant's grasp of related rules and principles involving generally higher level processing. Foe example, given the location of certain parts of the heart at a particular moment of its functioning, the student was asked to determine the position of other specified parts or positions of other specified parts of the heart at the same time.

While the tests remained verbally unchanged from their print originals, they were ported to an online HTML format for this study, where a server captured participant responses and stored them in a database.

Solitary Learner's Inventory of Social Presence (SLISP)

An original 25-item instrument was devised because no existing social presence survey adequately addressed the solitary learner scenario presented by this study—that is, none were suitable for use with a solitary learner interacting in a self-paced manner with a computer-based multimedia lesson written by an author both temporally and spatially distant from them. Data were collected using a 7-point Likert scale with the following response options: 1) Strongly Disagree, 2) Disagree, 3) Somewhat Disagree, 4) Neither Agree nor Disagree (neutral), 5) Somewhat Agree, 6) Agree, 7) Strongly Agree.

A host of questions were developed and drawn from other related questionnaires

addressing the "Presence" construct in other contexts. Kumar and Benbassat (2002) developed a "para-social presence" (PSP) instrument that most closely approaches the notion of social presence forwarded in this study. Their survey assessed the level of social presence felt by an Internet patron to a commercial website. Other social presence surveys assume an interlocutor, either human or virtual, whose communicative behavior is contingent or responsive to the actions of the learner. Simply put, they measure the social presence felt by participants communication with another human or agent—in either an online class, or while emailing, or in a chatting or instant messaging session. Some surveys also dealt with presence in terms of sharing a virtual space with others in an online game or a videoconference—all these scenarios and their associated presence measures were ill-suited models for this study.

In designing any instruments to measure social presence, one needs to be cautious in how one frames statements of relational or social judgments. It is possible that by overtly equating social-like interactions with a self-paced multimedia lesson with interactions between fellow human beings, users might be cued into rejecting such an "absurd" attribution according to Social Response Theory (Kumar & Benbasat, 2002; Reeves & Nass, 1996). The social presence survey, SLISP, developed for this study took this into account and remains similar to, but qualitatively distinct from other social presence measures.

Candidate items were paired down and refined to 35 questions and then validated with numerous peers and experts. Questions were also assigned to the four subscales of

Immediacy, Empathy, Positivity and Involvement identified by Kumar & Benbasat (2002). Finally, the instrument was modified and revalidated to ensure item relevancy to the subscale constructs. The resultant 25-item Solitary Learner's Inventory of Social Presence (SLISP), performed reliably (Cronbach's alpha= .94). Six miscellaneous non-social presence items were added to the final instrument for exploratory purposes such as " .

Data Analysis:

A 2x2 factorial design was employed to examine the effects of authorial voice and the addition of human speech on both achievement and perceptions of social presence. (see Figure 1) All statistical analyses were conducted using SPSS.

According to the descriptive analysis, the mean student achievement scores are all very closely clustered around 36 points out of 60 possible indicating no statistically significant difference between groups. Condition 1, with neither voice nor speech, had the highest mean score (M=37.26, SD=10.92), followed by condition 4, with both voice and speech, (M=36.89, SD=13,45). Condition 3 with voice but no speech, came next (M=36.65, SD=10,96) followed lastly by condition 2 with no voice, but with speech (M=35.67, SD=10,90). Looking at any of the three individual learning measures, but specifically at the higher-order comprehension measure, which was hypothesized to be most impacted by SRRs, still yielded no significant differences between conditions (M=11.62, SD=4.60). **Furthermore, dropping low achieving students from the data analysis likewise produced no significant difference across any of the tests. This was tested**

due to the hierarchical nature of the heart content materials. If students do poorly on tests 1 and 2, there is an expectation that they will also perform poorly on test 3 as, test 3 relies on a solid understanding of knowledge assessed in test 1 and 2.

SLISP scores measuring students' perceptions of social presence did indicate significant differences between conditions. Conditions 3 (M=106.57, SD 22.16) and 4 (M=106.82, SD 22.67) with authorial voice both scored noticeably higher than conditions 1 (M=83.76, SD 21.51) and 2 (M=89.82, SD 22.52) without authorial voice (See Table 2).

Table 2

Means and Standard Deviations Comparing Achievement on Heart Content Scores and Social Presence Scores by Condition

		Achievement		Social Presence	
Condition	*n*	*M*	*SD*	*M*	*SD*
1-- No Authorial Voice, No Human Speech	46	37.26	10.92	93.43	22.16
2-- No Authorial Voice, With Human Speech	45	35.67	10.90	93.18	22.67
3-- With Authorial Voice, No Human Speech	46	36.65	10.96	116.24	21.51
4-- With Authorial Voice, With Human Speech	45	36.89	13.45	110.18	22.52
Total	182	36.62	11.52	103.27	24.29

After determining that the data met the appropriate assumptions, the data were analyzed using one-way analyses of variance (ANOVA) for each of the dependent variables. Analysis of authorial voice (see table 3) produced a significant effect on perceptions of social presence (F=36.70, p=.000), but not on achievement (F=.030, p=.863). Human speech (see table 4), on the other hand, appeared to have no significant impact on achievement (F=.157, p=.692) or perceptions of social presence (F=.769, p=382).

Table 3

One-way Analysis of Variance Comparing the Effects of Authorial Voice on Achievement and Social Presence Measures.

Source	*SS*	*df*	*MS*	*F*	*p*
Achievement					
Between Groups	4.005	1	4.01	.030	.863
Within Groups	24016.835	180	133.43		
Total	24020.841	181			
Social Presence					
Between Groups	18080.198	1	18080.20	36.699	.000
Within Groups	88678.066	180	492.65		
Total	106758.264	181			

Table 4

One-way Analysis of Variance Comparing the Effects of Human Speech on Achievement and Social Presence Measures.

Source	*SS*	*df*	*MS*	*F*	*p)*
Achievement					
Between Groups	20.959	1	20.959	.157	.692
Within Groups	23999.882	180	133.333		
Total	24020.841	181			
Social Presence					
Between Groups	454.054	1	454.050	.769	.382
Within Groups	106304.210	180	590.58		
Total	106758.264	181			

Table 5

H3 Analysis of Variance for text-only condition with authorial voice, and the human speech condition with no authorial voice.

Source	*SS*	*df*	*MS*	*F*	*p (sig.)*
Social Presence					
Between Groups	18917.434	3	6305.81	12.778	.000
Within Groups	87840.829	178	493.49		
Total	106758.264	181			

Voice conditions exhibited higher means regardless of whether or not speech was present in the condition. Accordingly, table 4 shows that human speech did not have a significant effect on social presence (F=.769, p=.382), standing perhaps in the shadow of

the more dominant effect of authorial voice.

Student interest in the material and perceived lesson difficulty were both examined for exploratory purposes against achievement and social presence and exhibited significant (<.001) moderate to high positive correlations (see Table 6). Perhaps not surprisingly, feeling like they did well on the test was the highly correlated with achievement (r=.775) and moderately correlated with social presence perceptions (r=.404). Feeling high levels of social presence was also moderately correlated with interest in the lesson materials (r=.507), as was feeling like they learned a lot from the lesson (r=.498).

Table 6

Correlations (2-tailed) of Miscellaneous Items (n=182)

	Pearson Correlation		
Survey Item	Social Presence	Achievement	Sig.
The lesson material was interesting to me	.507	.561	.000
I think I did well on the tests	.404	.775	.000
I learned a lot from this lesson	.498	.464	.000
The lesson was easy	.305	.578	.000

Findings:

The primary hypothesis in this study was partially supported. Both secondary hypotheses were, however, not supported. An analysis of the 25-item Solitary Learner's Inventory of Social Presence Inventory (SLISP) also returned promising findings with a .94 reliability score (see table 7).

$H_0 1$, that no differences in achievement and perceptions of social presence will be realized by means of authorial voice manipulations, is partially supported. Indeed, no significant differences were observed in achievement across any of the four conditions. Significant differences were however observed between groups in student perceptions of social presence. Mean ratings of social presence were significantly higher in both conditions manipulated to include authorial voice either through text alone or with text and audio. That is, where efforts were made to elicit a heightened social response in the learner through the addition of authorial voice elements, the results were successful.

$H_0 2$, that no differences in achievement and perceptions of social presence will be realized by means of human speech manipulations, was supported. Here again with regard to achievement, no significant differences existed between groups. Perhaps surprisingly, the sound of human speech alone, sans authorial voice--that is a presenter speaking in 3rd person, making no attempt to directly address the reader, nor sharing any phatic expressions—also made no significant difference in student perceptions of social presence. In this case, we retain the null.

$H_0 3$, that no differences in perceptions of social presence will be observed between the text-only condition with authorial voice, and the audio condition with no evident authorial voice, is rejected. That is, differences did in fact exist between these groups attributable to the authorial presence evident in the text. Table 7

Effect sizes of SRRs on Achievement & Social Presence

	Hypotheses (itemized)	Conclusion	Effect size
$H_0$1-a	No significant difference in *achievement* will be found due to level of authorial voice in the multimedia lesson.	Supported	0.03
$H_0$1-b	No significant difference in *achievement* will be found due to human speech in the multimedia lesson.	Supported	-0.06
$H_0$2-a	No significant difference in *perceptions of social presence* will be found due to level of authorial voice in the multimedia lesson.	NOT supported	.89
$H_0$2-b	No significant difference in *perceptions of social presence* will be found due to human speech in the multimedia lesson.	Supported	-0.14
$H_0$3	No significant difference in *perceptions of social presence* will be found between text-only condition *with* authorial voice, and the human speech condition with *no* authorial voice.	Supported	1.07

Sig=.000

Discussion:

Consistent with some of the findings of Mayer and his colleagues, the personalization effect brought about by authorial voice and human speech, did not have an affect on achievement at the shallower processing level of simple retention and factual recall. On the other hand, contrary to some of the literature reviewed including that of Mayer and colleagues' findings on deeper processing and transfer, those same Socially Relevant

Representations or personalization tactics, had no effect on achievement associated with the deeper cognitive processing levels of comprehension. While transfer and comprehension are both indicative of deeper processing, it might be arguable that they are sufficiently different to make comparison of these measures inappropriate.

To some extent, university student populations have likely developed coping strategies, "school smarts" and abilities to learn at a minimal level from any materials or instructor regardless of the presentation style or modality employed. Indeed, learning was arguably lacking across the board, with mean scores ranging from 59% to 62% on all tests. The effects of at least one SRR, however, did significantly help create the "warm and friendly tone" introduced at the outset of this article, promoting higher perceptions of social presence overall.

This may point to significant motivation effects. Students may have found the treatment too long or the content too boring to be particularly motivated. The short experimental conditions of Mayer and colleagues' studies may not have triggered sufficient negative affect to influence student motivation, yielding therefore the positive findings they report. Thus, additional research in this area should look more closely at the effect of interest and motivation in the task.

The study design did not permit independently measuring the impact of human speech with authorial voice on social presence, as authorial voice was added simultaneously to the text and audio. Its effects however can be seen as negligible as no significant

differences existed between conditions 3 and 4. It is however possible that expository text could have been perceived as more social if human speech with authorial voice was added. This combination was not tested, but it's implications would be interesting and of practical value, since adding a shell of "personality" or "personality bookends" to existing instructional materials is certainly more efficient than editing or rewriting entire texts or courses to include elements of authorial voice.

Comment [BLG3]: Name the conditions, rather than us ea number.

Future Research

More questions invariably arise such as, under what conditions do any potential findings apply and not apply? Are such social cues more important and relevant for independent study and distance learning courses than they are for more traditional face to face courses? With what types of learners? Future research should assess student desire for social connectedness or attunement to socially relevant representations like authorial voice. For some students, such personalization attempts may cue undesired or unwelcome social responses and inhibit optimal learning. Future research might also simultaneously look at student factors such as reading ability, learning style, field dependence and locus of control. Additionally, there was an anecdotal (but unmeasured) sense that students in authorial voice conditions completed the task in less time than others despite their minimally increased word count. In many training or learning situations, any significant time efficiencies realized could certainly be considered valuable even if average scores remained constant.

Conclusion

This study suggests that the effects of SRRs such as authorial voice, human speech and personalization may not be as obvious or positive on learning as some would advocate.

Differences in findings may stem from the complexity and length of learning tasks that study participants were subjected to. The effects of SRRs on learner perceptions of social presence were more obvious, but this did not translate into higher achievement. This effect should not be too readily dismissed as its impact on student motivation and persistence in learning from multimedia materials may be significant.

References:

Power Analysis for ANOVA Designs

Sample-size Table

The power parameters you specified were:

- a = '4' (levels of factor for power)
- b = '1' (levels of factor(s) crossed with A)
- delta = '0.25 0.5 0.75 1.0 1.25' (effect size(s))
- alpha = '0.05' (significance level)

```
Power analysis for ANOVA designs
     4x 1   layout Ha: T1=GM-Delta/2, T2=T3=...=T(k-1)=GM,
Tk=GM+Delta/2   tested at Alpha= 0.050

                   DELTA (in units of sigma=Std. Dev.)

      N  0.250  0.500  0.750  1.000  1.250
----------------------------------------

      2  0.051  0.056  0.065  0.077  0.094
      3  0.053  0.064  0.084  0.112  0.152
      4  0.055  0.072  0.103  0.150  0.215
      5  0.057  0.080  0.123  0.190  0.281
      6  0.059  0.089  0.145  0.231  0.347
      7  0.061  0.097  0.166  0.273  0.412
      8  0.063  0.106  0.189  0.315  0.475
      9  0.065  0.115  0.212  0.357  0.534
     10  0.067  0.124  0.235  0.399  0.590
     12  0.071  0.143  0.282  0.479  0.688
     14  0.075  0.162  0.329  0.554  0.767
     16  0.079  0.182  0.376  0.622  0.830
     18  0.083  0.202  0.422  0.683  0.878
     20  0.087  0.222  0.467  0.736  0.913
     25  0.098  0.274  0.572  0.840  0.965
     30  0.108  0.327  0.663  0.907  0.987
     35  0.120  0.379  0.740  0.947  0.995
     40  0.131  0.430  0.802  0.971  0.998
     50  0.155  0.527  0.891  0.992  0.999
```

Sample-size Table
The power parameters you specified were:

- a = '2' (levels of factor for power)
- b = '2' (levels of factor(s) crossed with A)
- delta = '0.25 0.5 0.75 1.0 1.25' (effect size(s))
- alpha = '0.05' (significance level)

```
Power analysis for ANOVA designs
     2x 2   layout Ha: T1=GM-Delta/2, T2=T3=...=T(k-1)=GM,
Tk=GM+Delta/2   tested at Alpha= 0.050

                   DELTA (in units of sigma=Std. Dev.)

      N  0.250  0.500  0.750  1.000  1.250
----------------------------------------

      2  0.058  0.085  0.131  0.195  0.275
      3  0.067  0.119  0.209  0.332  0.477
      4  0.074  0.151  0.281  0.452  0.632
      5  0.082  0.183  0.350  0.556  0.746
      6  0.089  0.214  0.416  0.644  0.829
      7  0.097  0.245  0.478  0.718  0.887
      8  0.104  0.276  0.535  0.779  0.926
      9  0.112  0.307  0.587  0.828  0.953
     10  0.120  0.337  0.636  0.867  0.970
     12  0.135  0.395  0.719  0.923  0.988
     14  0.150  0.450  0.786  0.956  0.995
     16  0.166  0.503  0.839  0.975  0.998
     18  0.181  0.552  0.880  0.986  0.999
     20  0.197  0.597  0.911  0.992  0.999
     25  0.235  0.696  0.960  0.998  0.999
     30  0.274  0.775  0.982  0.999  0.999
     35  0.311  0.835  0.992  0.999  0.999
     40  0.349  0.881  0.997  0.999  0.999
     50  0.420  0.940  0.999  0.999  1.000
```

The sample size values given are those for each of the 2 levels of the factor called 'Factor A'.With 2 combinations of other factors at each level of Factor A, divide the sample size by 2 to determine the sample size per treatment cell.