8.

How Music Affects Learning in a 3D Gaming Environment

An Experiment

Ryan L. Sittler (California University of Pennsylvania)

Abstract

This study examined the effects of music, in a 3D gaming environment, on educational achievement as measured by a series of criterion-referenced tests (identification, terminology, and comprehension). A sample of 81 undergraduate students, from a medium-sized public institution in western Pennsylvania, were recruited to take part in the experiment and then randomly assigned to one of three groups that each played a short educational game module about the human heart. Each group played an identical module save for the fact that one contained no music, one module included music by Mozart, and another allowed participants to self-select the music they would like to receive. They then each took a series of identical tests to ascertain performance. Results are discussed herein.

Introduction

The study of games as educational tools rarely examines the role, if any, that audio plays in the learning process. Specifically, music in game environments is even more rarely studied and little empirical evidence exists as to whether or not it affects learning outcomes (Zehnder & Lipscomb, 2006; Fassbender et al, 2012). Some researchers, however, have put forth ideas about how music may influence a learner while playing an educational game. In some instances these concepts are based on broader educational media research.

Gredler (2002) suggested that music may act as a distractor—this would interfere with information processing and negatively affect learning performance. Clark and Mayer (2008; 2012) support this notion by suggesting that extraneous auditory information reduces performance and therefore should be limited or eliminated. This may be due to overloading the system that deals with auditory information in working memory—this is referred to as the phonological loop. Working from the theoretical position of information processing—that humans have limited mental resources (Miller, 1956; Atkinson & Shiffrin, 1968; Lang, 2000)—it does make sense that music could interfere with learning. Other authors suggest that the research is not clear in this area and background music could positively affect a learners' performance and improve learning outcomes (Bishop, Amankwatia, & Cates, 2008; Brown & Green, 2011).

Despite the dearth of empirical research on this topic specifically, there is evidence that music can be harnessed for positive learning effects. The colloquially named Mozart Effect, for example, showed that subjects listening to 10 minutes of Mozart's *Sonata for Two Pianos in D Major* exhibited short term improvement in performance of abstract and spatial reasoning tests (Rauscher, Shaw, & Ky, 1993). Other studies have since extended the research in this area and have determined that while music alone may not provide a desired outcome, specific types of music may work better than others in improving performance—particularly if the listener enjoys the musical selection (Nantais & Schellenberg, 1999; Ilie & Thompson, 2011). In short, certain songs in certain educational applications (such as a virtual game environment) can improve learning outcomes (Richards et al, 2008; Fassbender et al, 2012). Further, listening to music that a person enjoys—self-selecting per their own tastes—may lead to increased improvement.

This paper describes an experiment in which undergraduate students were taught about the human heart, in a 3D gaming environment, and then tested on different educational objectives that included facts, concepts, procedures, and principles (Merrill, 1983; Ragan & Smith, 2004). It utilized, with permission, a modified version of an existing instructional script and criterion-referenced tests based on the work of Dwyer & Lamberski (1977).

The overarching research question this study addresses is: what are the effects of audio, within a 3D gaming environment, on the achievement of different educational objectives? Based on theoretical knowledge of the subject, and information contained in the review of literature, this then leads to two hypotheses:

1. Participants that listen to Mozart's *Sonata for Two Pianos in D Major* during their treatment will perform better on criterion-referenced tests than those that listen to silence.

2. Participants that listen to self-selected music during their treatment will perform better on criterion-referenced tests than those that listen to Mozart's *Sonata for Two Pianos in D Major*.

Method

This study utilized a 1 X 3 factorial post-test only control experimental design (Cresswell, 2009). A pre-test was not used because the author did not want to provide the students with any leading information about the questions in the post-test. The independent variable was music at three levels: none, Mozart, and Self-Selected. Participants (N = 81) were recruited from undergraduate classes in the Department of Communications Media at Indiana University of Pennsylvania via an in-class presentation by the author. This included 43 males (53.1%) and 38 females (46.9%) aged 18 or older. 32 (39.5%) self-identified as gamers whereas 49 (60.5%) did not. Three participants (3.7%), all female, self-identified that they had more than three previous college level classes that taught information about the human heart—this question was asked to identify those with prior knowledge on the subject. Involvement was voluntary and food was offered as compensation for their time. Each subject was randomly placed into one of three separate groups: Control (n = 27), Mozart (n = 27), and Self-Selected (n = 27).

The members of each group worked in a computer lab that was equipped with a series of identical computers, software, and Encore AE-06 disposable stereo headphones. They each answered a series of demographic questions and then played a functionally identical short educational module, in a 3D gaming environment, which taught them facts, concepts, procedures, and principles of the human heart and its functions (see Figure 1).

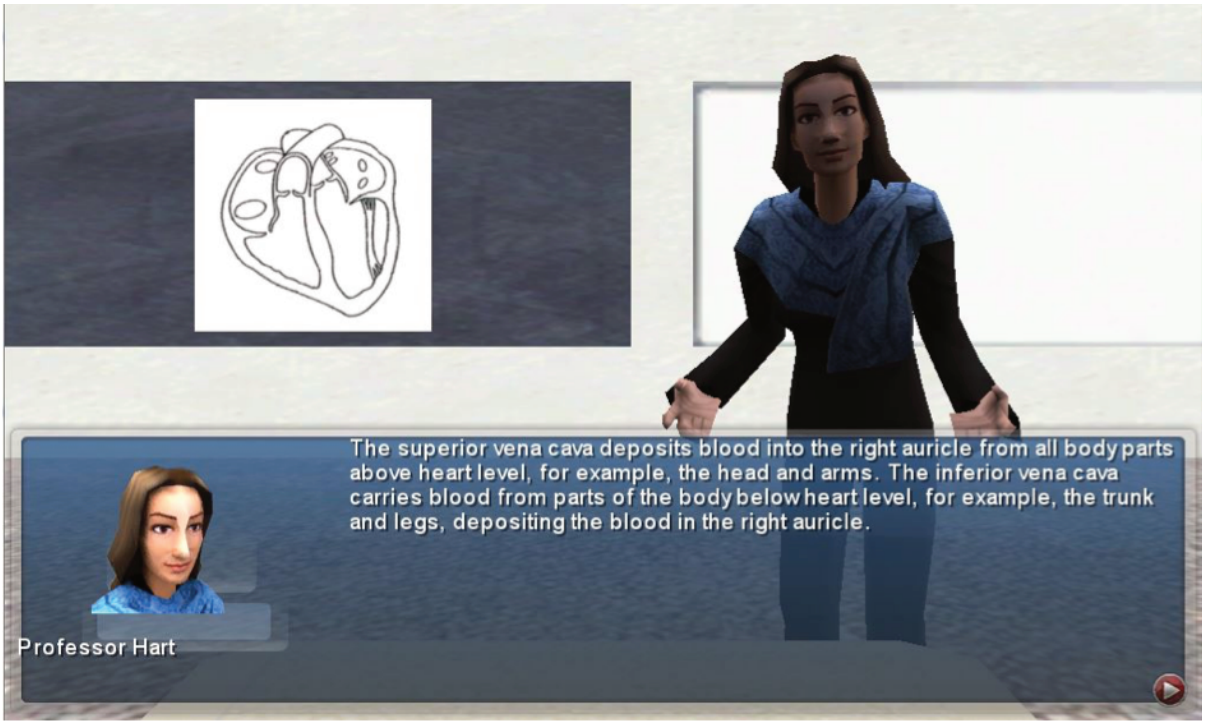

Figure 1. Screenshot of the educational module with Professor Hart Non-Player Character (NPC).

The module opened with a brief introduction, from an NPC named Professor Hart, explaining that participants needed to wear headphones and adjust volume as needed, within a prescribed range, regardless of whether or not they were receiving music in their treatment. This was required to account for the sensory experience of wearing headphones as it could be a confounding variable—subjects were to be kept under similar conditions as much as possible. They were then familiarized with the interface and how to control it by pointing and clicking with the mouse. Some brief narrative action was included to help draw participants into the environment; they got to see some other NPC students in the virtual classroom, were instructed to take a seat, and some humor was used in the text to lighten the mood. The instructional script, which was broken into three sections (the parts of the heart, the circulation of blood through the heart, and the cycle of blood pressure in the heart), was delivered via the Professor Hart NPC and included facts, concepts, procedures, and principles of the human heart and its functions.

Participants had the option to repeat each section of the instructional script until they felt comfortable with the content. They then each took an identical battery of three criterion-referenced tests—identification, terminology, and comprehension. Each test contained 20 multiple-choice questions. The identification and terminology tests each included five potential answers for each question. The comprehension test included four potential answers for each question. The instructional script and tests were based on the work of Dwyer & Lamberski (1977).

Participants in the Control group wore headphones but received no audio treatment during their experience working through the module. Conversely, participants in the Mozart group listened to Mozart's *Sonata for Two Pianos in D Major* and those in the Self-Selected group picked the music of their choice from Spotify. Subjects were not given a time limit but generally completed the module in 30 minutes or less. The study took place over three days as the lab was not big enough to include all participants at one time. Data were collected via the Qualtrics platform and analyzed using IBM SPSS version 22. Eight questions were eliminated *post hoc*—resulting in a 52 question test set—due to a Cronbach alpha coefficient of α = .64. Items with negative inter-item correlation were dropped from each test. The resulting Cronbach alpha coefficient for the composite of tests was acceptable at α = .71 (Pallant, 2013).

Results

Taking the three test sets as a composite score, the subjects in both the Control (*M* = 14.59, *SD* = 5.37) and Mozart (*M* = 15.74, *SD* = 6.65) groups performed better than those subjects in Self-Selected (*M* = 11.67, *SD* = 4.04). Mozart participants were the most successful overall. This was true of both the aggregate scores and the scores on the three individual identification, terminology, and comprehension tests (see Table 1).

		Identification Test[a]		Terminology Test[b]		Comprehension Test[b]		Composite Score[c]	
Group	*n*	x̄	SD	x̄	SD	x̄	SD	x̄	SD
Control	27	5.70	2.83	4.44	2.76	4.44	2.45	14.59	5.37
Mozart	27	5.59	3.30	5.30	3.16	4.85	2.93	15.74	6.65
Self-Selected	27	3.85	2.27	3.64	1.66	4.19	1.84	11.67	4.04
Total	81	5.05	2.92	4.46	2.67	4.49	2.43	14.00	5.66

[a] 18 Items
[b] 17 Items
[c] 52 Items

Table 1. Subjects' scores by test and aggregate.

A one-way between groups analysis of variance (ANOVA) was used to further examine the data. There was a statistically significant difference at the **p* < .05 level on scores for the three groups: *F* (2, 78) = 4.00, *p* = .02. The effect size, calculated using eta squared, was .09. However, the difference in mean scores between each group was small (see Table 2).

					95% Confidence Interval for *M*			
Group	*n*	x̄[a]	SD	*SE*	Lower Bound	Upper Bound	Minimum	Maximum
Control	27	14.59	5.37	1.03398	12.4672	16.7180	6.00	25.00
Mozart	27	15.75	6.65	1.28057	13.1085	18.3730	7.00	39.00
Self-Selected	27	11.67	4.04	.77717	10.0692	13.2642	6.00	23.00
Total	81	14.00	5.66	.62903	12.7482	15.2518	6.00	39.00

[a] 52 Items

Table 2. Scores in aggregate.

Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the Control (*M* = 14.59, *SD* = 5.37) was not significantly different from either the Mozart (*M* = 15.75, *SD* = 6.65) or Self-Selected (*M* = 11.67, *SD* = 4.04) groups. However, Mozart and Self-Selected did differ significantly at the **p* < .05 level (see Table 3).

	Group (I)	Group (J)	*M* Difference (I-J)	*SE*	*p*	95% Confidence Interval Lower Bound	Upper Bound
Tukey HSD	Control	Mozart	-1.14815	1.48615	.721	-4.6990	2.4027
		Self-Selected	2.92593	1.48615	.127	-.6249	6.4767
	Mozart	Control	1.14815	1.48615	.721	-2.4027	4.6990
		Self-Selected	4.07407*	1.48615	.020	.5233	7.6249
	Self-Selected	Control	-2.92593	1.48615	.127	-6.4767	.6249
		Mozart	-4.07407*	1.48615	.020	-7.6249	-5.233

* The mean difference is significant at the .05 level.

Table 3. Post-hoc comparisons.

Though not the main focus of this study, prior knowledge was assessed via a question in the demographic survey that each participant completed. In it, they were asked to self-identify if they had at least three prior college level courses that taught information about the human heart. Three female subjects indicated that they had. No males met this criterion. Taking the three test sets as a composite score, two females outperformed the average of others in their gender group (see Table 4). Likewise, though not a focus of this study, aggregate scores showed that females (M = 15.21, SD = 6.71) outperformed males (M = 12.93, SD = 4.34).

Subject Group	Major	Identification Test[a] Score	$\bar{x}^d$	Terminology Test[b] Score	$\bar{x}^d$	Comprehension Test[b] Score	$\bar{x}^d$	Composite Score[c] Score	$\bar{x}^d$
Control	Nutrition	8.00	6.08	10.00	6.50	6.00	5.00	24.00	17.58
Mozart	Dance / Psychology	3.00	6.08	9.00	6.50	5.00	5.00	17.00	17.58
Self-Selected	Journalism	8.00	4.15	6.00	5.32	4.00	4.49	18.00	15.21

[a] 18 Items
[b] 17 Items
[c] 52 Items
[d] Mean of female subjects score on this test. Within own treatment groups only.

Table 4. Scores by participants self-identifying as having had more than three prior college classes that taught about the human heart.

Discussion

The overarching research question this study addresses is: what are the effects of audio, within a 3D gaming environment, on the achievement of different educational objectives? Specifically, two hypotheses were to be tested:

1. Participants that listen to Mozart's *Sonata for Two Pianos in D Major* during their treatment will perform better on criterion-referenced tests than those that listen to silence.
2. Participants that listen to self-selected music during their treatment will perform better on

criterion-referenced tests than those that listen to Mozart's *Sonata for Two Pianos in D Major*.

As evidenced by the data presented in this paper, there was not a statistically significant difference in participants' achievement on criterion-referenced tests between the Control group (which received no music) and the Self-Selected group which listened to Mozart. Hypotheses 1 is not supported. Further, there was a statistically significant difference in participants' achievement on criterion-referenced tests when comparing participants in the Mozart group and those in the Self-Selected group. However, the hypotheses predicted that Self-Selected would be more successful and that is not the case. Therefore, hypothesis 2 is not supported. Overall, the subjects that listened to Mozart had the highest mean score $\bar{x} = 15.74$ out of 52 questions (30.3%).

Limitations & Future Research

Overall, subjects performed poorly on each of the tests. This leads the author to believe that there was a fundamental problem with the design of the instructional module since it was the delivery—and not the assessments—that largely deviated from the original work on which this study was based. The highest performing group listened to Mozart and scored just 30.3% and the lowest performing group, Self-Selected, scored 22.4% ($\bar{x} = 11.67$ out of 52 questions). These scores are generally better than random guessing but still not impressive. A different pedagogical approach may have been more successful.

Prior knowledge was only assessed from the perspective of whether or not participants had extensive previous coursework on the human heart. A pre-test would have also been able to help better understand prior knowledge on the subject and is another avenue worth exploring in future research. Further, it is possible that the instructional approach used in this study would work better as a review activity as opposed to a way of introducing content. This is slightly hinted at in the data collected, but the sample size is too small to make any strong statements without further research being conducted.

Past research has suggested that as long as participants are listening to music that they enjoy, they should perform better on certain types of mental tests and exercises. The evidence here stands in opposition to that concept. One possible explanation for this is that subjects had to expend extra mental resources while choosing and selecting their music from Spotify. This is typically called decision fatigue. This could have caused more cognitive load for participants which ultimately interfered with their information processing abilities (Kalyuga, 2009). Additionally, as participants in the Self-Selected group were free to listen to the music of their choice, they may have inadvertently made a selection that impeded their ability to perform. This is due to the fact that they likely did not have a concept of what music may or may not help them in studying (e.g., choosing music with too much singing versus instrumentation, tempo, music they are familiar with versus something new). Indeed, some participants commented after the experiment that they did not enjoy the music they self-selected. When pressed as to why, they stated that the Spotify playlist was not as good as they anticipated. All of these problems may be avoided in the future by having subjects select music ahead of time, and ensuring that it is ready to play when they arrive for the experiment.

There were other confounds in this study that made it difficult to strike a balance between a pure lab setting and something that participants might experience in the real world. Noise-cancelling headphones, for example, may have mitigated the opportunity for background noise to interfere with information processing. Similarly, white noise could have been used—instead of silence—to establish baseline scores. All of these possibilities can be addressed in future research.

Acknowledgments

The author offers sincere thanks and appreciation to Dr. F.M. Dwyer for granting me permission to adapt his research for use in this context.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory (Vol. 2, pp. 89-195). New York, NY: Academic Press.

Bishop, M. J., Amankwatia, T. B., & Cates, W. M. (2008). Sound's use in instructional software to enhance learning: A theory-to-practice content analysis. Educational Technology Research and Development, 56(4), 467-486. doi:10.1007/s11423-006-9032-3

Brown, A., & Green, T. D. (2011). The essentials of instructional design: Connecting fundamental principles with process and practice (2nd ed.). New York, NY: Pearson.

Clark, R. C., & Mayer, R. E. (2008). E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning (2nd ed.). San Francisco, CA: Pfeiffer.

Clark, R. C., & Mayer, R. E. (2012). Using rich media wisely. In R. A. Reiser & J. V. Dempsey (Eds.), Trends and issues in instructional design and technology (Third edition. pp. 309-320). New York, NY: Pearson.

Cresswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches (3rd ed.). Los Angeles, CA: Sage.

Dwyer, F. M., & Lamberski, R. (1977). The human heart: Parts of the heart, circulation of blood and cycle of blood pressure. State College, PA: Learning Services.

Fassbender, E., Richards, D., Bilgin, A., Thompson, W. F., & Heiden, W. (2012). VirSchool: The effect of background music and immersive display systems on memory for facts learned in an educational virtual environment. Computers & Education, 58(1), 490-500. doi:10.1016/j.compedu.2011.09.002

Gredler, M. E. (2002). Games and simulations and their relationships to learning. In D. H. Jonassen (Ed.), Handbook of research for educational communications and technology (Second edition. pp. 571-581). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ilie, G., & Thompson, W. F. (2011). Experiential and cognitive changes following seven minutes exposure to music and speech. Music Perception: An Interdisciplinary Journal, 28(3), 247-264.

Kalyuga, S. (2009). *Cognitive load factors in instructional design for advanced learners*. New York, NY: Nova Science Publishers.

Lang, A. (2000). The limited capacity model of mediated message processing. Journal of Communication, 50(1), 46-70. doi:10.1111/j.1460-2466.2000.tb02833.x

Merrill, M. D. (1983). Component display theory. In C. Reigeluth (Ed.), Instructional-design theories and models (pp. 279–333). Mahwah, NJ: Lawrence Erlbaum Associates.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81-97. doi:10.1037/h0043158

Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference. Psychological Science, 10(4), 370-373.

Pallant, J. (2013). SPSS survival manual: A step by step guide to data analysis using IBM SPSS (5th ed.). New York, NY: McGraw-Hill.

Ragan, T. J. & Smith, P. L. (2004). Conditions theory and models for designing instruction. In D. Jonassen (Ed.), Handbook of Research on Educational Communications and Technology (pp. 623-649). Mahwah, NJ: Lawrence Erlbaum Associates.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. Nature, 365(6447), 611. doi:10.1038/365611a0

Richards, D., Fassbender, E., Bilgin, A., & Thompson, W. F. (2008). An investigation of the role of background music in IVWs for learning. ALT-J, Research in Learning Technology, 16(3), 231-244.

Zehnder, S. M., & Lipscomb, S. D. (2006). The role of music in video games. In P. Verderer & J. Bryant (Eds.). Playing video games: Motives, responses, and consequences (pp.241-258). Mahwah, NJ: Lawrence Erlbaum Associates.