Effects of task difficulty and music expertise in virtual reality: an observation of cognitive load and task performance in Beat Saber

Kyla Ellahiyoun

(s3791337)

Word count: 4998

Due date: October 16th 2023

Extended due date: November 13th 2023

Abstract

Music psychology has extensively studied the effect of music on cognitive functions, notably executive functioning and listening behaviours. Few studies, however, have examined the effect of musical training on perceived cognitive load. This experimental study investigated whether musical training could explain outcomes of cognitive load and task performance as task difficulty increased using Beat Saber. It was hypothesised that: (1) increasing task difficulty levels would predict differences in cognitive load; (2) musical training groups would rate differently on cognitive load as task difficulty increased; (3) and musical training groups would predict differences in task performance as task difficulty increased. Thirty-two participants between 18 to 35 years old were screened for their ability to safely engage in VR research and completed a series of questionnaires in addition to three main task difficulty levels using the MetaQuest Pro headset. Analyses of variance (ANOVA) revealed significant effects of task difficulty on cognitive load and task performance, with nonsignificant interactions of musical training on either outcome. These findings suggest that that participants with lower cognitive load and higher performance were more experienced with VR or digital games. Furthermore, this may indicate that musical training alone is unable to account for differences in cognitive load and task performance across all game difficulty levels. This research is the first not to display an interactional effect of musical training on cognitive load and task performance within a VR exergame. Future research should consider alternative methods of grouping musical expertise and the additional predictability of flow and self-efficacy.

Word count: [250]

Introduction

Music psychology has long been investigating the effect of music on cognitive functions. This has resulted in the realisation that music has positive outcomes on executive functioning (Chee et al., 2022), and that listening behaviour and training could protect against cognitive decline in older populations (Román-Caballero et al., 2018). Recently, this interest in cognitive outcomes is extended to observing cognitive load (CL) during music-oriented tasks. CL accounts for limitations of working memory and considers the delivery mode of novel information should follow the scientific principle of parsimony (Sweller et al., 1998), and be presented with the simplest explanation. This is important when presenting new information or tasks, as comparing the expertise between individuals or groups of people can be explained through theories of optimal performance (Yerkes & Dodson, 1908; Csikszentmihalyi, 1975). In relation to CL and expertise, there is evidence of a negative relationship between a higher level of skill and acquired mental workload during challenging tasks (Haith & Krakauer, 2018).

CL is prominently researched with regards to instructional design (Sweller et al., 1998), where it is established that the inherent difficulty of a task and the method it is delivered could significantly affect the accumulation of CL (Sweller et al., 1998). To address this effect, some researchers have opted for commercial games that are recognisable to the public (Huang, 2020). A notable example being Beat Saber (Beat Games, 2018), a popular exergame that combines upbeat music and physical movement (Mueller & Isbister, 2014). There have been few attempts to modify games such as Beat Saber or create new games for research participants to engage with (Lemmens & von Münchhausen, 2023), however, they are susceptible to limited generalisation as they can lack in elements which make games inherently enjoyable (Cutting et al., 2023).

Music Sophistication & Training

An abundance of research has defined musical abilities by the duration of formal MT, ignoring other factors that do not overtly contribute to the skillset (Levitin, 2012). A widely adopted measure of musical ability in the general population is the Goldsmiths Music Sophistication Index (Gold-MSI; Müllensiefen et al., 2014). It observes music sophistication as a psychometric construct, encompassing independent dimensions comprising skills and achievements, musical perception and music-making, amount of practice, emotional and functional usage of music, and creativity (Müllensiefen et al., 2014).

Despite accounting for engagement and perceived musical abilities, the Gold-MSI's MT subscale is used as a point of comparison between research participants. For example, studies have chosen certain items from the MT subscale to categorise low, intermediate, and high music sophistication groups (Matziorinis et al., 2023), or only considered MT to inform their definition of expertise (Chaddock-Heyman et al., 2021) as a degree of objectivity is expected. To this end, the inconsistent categorisation of musical expertise permits leniency to be exercised within musicality research, affecting direct comparison between studies.

Cognitive Load Theory

Cognitive load theory (CLT) was conceived because of competing learning and memory frameworks (Atkinson & Shiffrin, 1998; Baddeley & Hitch, 1974), demonstrating its adaptability for instructional design (Sweller et al., 1998). Furthermore, by accounting for limitations of working memory, the delivery mode of novel information should follow the scientific principle of parsimony for the simplest explanations (Sweller et al., 1998). This notion is supported by widely accepted assumptions within CL research, such as element interactivity (Sweller, 1994; Sweller 2010) and schema acquisition (Chi et al., 1982). Firstly, CL is considered a multidimensional construct which describes the mental workload that performing a given task imposes on a learner's cognition (Paas & Van Merriënboer, 1994; Sweller et al., 1998). Continuing from this, elemental interactivity refers to the number of elements being processed simultaneously in working memory for schema construction and their interactions to occur (Sweller, 1994). Essentially, it describes the inherent difficulty of a task. Furthermore, schema acquisition relies on long-term memory to guide behaviour from a repository of stored information (Kalyuga & Singh, 2016). By combining these two assumptions, researchers exploring the relationship and interactions between CL and task difficulty are better able to express differences in learning behaviour and skill acquisition. In all these contexts, CL is informed by expertise.

Expertise In Context

Expertise is characterised as a measurable degree of skill, whereby a necessary level of skill or knowledge is objectively referred to as competency (Greenfield et al., 1994). In acknowledgement of this, Kirschner and Williams (2013) argued that expertise should reframe the expert versus novice dichotomy by positioning players as an expert in their own understanding. It competes with the notion of invalidating the experiences of beginners by focusing on the agency of 'experts' (Kirschner & Williams, 2013). Following their example, Toft-Nielsen (2016) argued that to acquire expertise in games, the amount of knowledge to become an expert should be considered. Accordingly, research measuring expertise in VR exergames should appraise the contextual parameters used to define expertise. As the present study seeks to understand expertise within cognitive engagement, the discussion changes course to consolidate motivational theories into practice (Moreno & Mayer, 2007; Huang et al., 2021). A popular approach is by making VR exergames adaptive. Examples include measuring heart rate, stress, or cognitive effort, which are integrated into different game elements to modify difficulty (Mandryk et al., 2006). Few studies investigating hand-eye coordination in a Beat Saber context propose the opposite effect of exergame training, potentially leading to faster music instrument mastery (Rutkowski et al., 2021). However, this effect is outside the scope of this study, as our experiment aims to combine different research areas in the context of virtual reality.

A Case for Virtual Reality

Virtual reality offers pedagogical means for users to learn, engage, and enhance their knowledge and skills within an immersive environment. These digital simulations provide a psychological experience dependent on interplays between technology and perceptual processes (Steuer, 1992; Takac, 2023). It contributes to a reduction in confounding variables that are typical of laboratory and clinical settings. As attributable to music, VR is a primary component of entertainment and video games that seek to provide an engaging experience to the extent of distracting a user from external stimuli (Schmidt et al., 2018). Extending from this, previous studies have found that situating exergames within a VR context is not only a promising intervention for encouraging physical exercise (Huang, 2020) but also provides a more immersive experience and feeling of agency than traditional exergames (Campo-Prieto et al., 2021).

A significant concern about implementing exergames in VR is attributable to its physically intensive nature, which could lead to discomfort or severe after-effects from prolonged use, such as cybersickness. It is defined as symptoms of eye strain, headache, sweating, disorientation, vertigo, nausea, or vomiting and can occur strictly from visual perception alone (LaViola, 2020). A standardised measure for cybersickness in VR research is the Simulator Sickness Questionnaire (SSQ; Kennedy et al., 1993), initially developed to assess simulator sickness in aviators.

Research Aim and Hypotheses

Although aspects of music sophistication and CL have been jointly researched, there are limited studies that attempt to explore this interaction in a VR rhythm-based exergame. As the interaction between music-related activities and executive function (EF) activation is well-established in music psychology literature, and working memory is a component of EF, this could posit CL as a subjective measure of EF. Furthermore, game difficulty in Beat Saber is referred to as task difficulty to align with terminology used in existing literature. Thus, the present study aims to investigate the interactions of music sophistication, CL, and task performance in a VR rhythm game as task difficulty increases from easy, to normal, to hard.

Hypothesis 1: There will be a difference in perceived CL as task difficulty increases.

Hypothesis 2: There will be a difference in CL between MT groups as task difficulty increases.

Hypothesis 3: There will be a difference in task performance between MT groups clustered by the Gold-MSI MT subscale, as task difficulty increases.

Method

Participants

Recruitment of participants was performed via word of mouth, social media websites (Facebook, Reddit, and Instagram) and advertising posters at the Royal Melbourne Institute of Technology (RMIT) Melbourne City Campus. Participants were screened relative to equipment guidelines and participant characteristics in VR studies (Slater & Sanchez-Vives, 2016; LaViola, 2020; Meta, 2022). The eligibility criteria were as follows: (1) being 18 to 35 years old; (2) being a native level proficient English speaker; (3) not being currently pregnant; (4) having no vision difficulties that cannot be corrected by glasses or contact lenses; (5) having not recently undergone any medical procedure (including cosmetic); (6) not having a pacemaker; (7) not suffering from a heart or serious medical condition; (8)

having no physical disabilities that impair balance, motor action, or standing with assistance; and (9) having no history of significant motion sickness, active nausea, and vomiting or epilepsy.

Out of the 74 people who met the screening criteria, 32 participants actively completed the present study. The sample consisted of 11 females (34.4%), 18 males (56.3%) and 3 non-binary (9.4%) people aged 18 to 35 years ($M = 23.63$, $SD = 4.58$). They were given the opportunity to enter a draw for one of two $100 JB Hi-Fi gift cards at the completion of the study.

Materials

Demographic measures.

The pre-experimental survey includes questions pertaining to participants previous experience with digital games, VR and Beat Saber, on a 7-point Likert scale. The questions were as follows: what best describes your experience with (1) digital games; (2) virtual reality; (3) and Beat Saber.

Gold-MSI.

The Gold-MSI (Müllensiefen et al., 2014) is a 38-item self-report measure of music sophistication comprising five subscales (active engagement, perceptual abilities, musical training, singing abilities and emotions) that contribute to a general music sophistication factor. Out of those, 31-items are scored on a Likert scale ranging from 1 (*completely disagree*) to 7 (*completely agree*) and 7-items are scored from categorical responses (e.g., "I have had formal training in music theory for 0/.5/1/2/3/4-6/7-10/11+ years."; see Appendix B). The primary Gold-MSI, derived from 18-items, demonstrated good internal reliability $\alpha = .88$, with its subscales exhibiting acceptable to good internal reliability between $\alpha = .71$ and $\alpha = .85$ (Chee et al., 2022), where a higher score signifies higher music sophistication.

Responses were scored using the Gold-MSI scoring template (https://shiny.gold-msi.org/gmsiscorer/) for further analyses. Due to time constraints in this study, the Gold-MSI listening tasks were excluded and only the derived general and MT scores were analysed.

NASA-Task Load Index.

The NASA-TLX (Hart & Staveland, 1988) is a subjective, multidimensional questionnaire that assesses perceived cognitive workload to determine the effectiveness of a task or system based on an individual's performance. It is a prominent measure of CL in VR research (Orru & Longo, 2019) and is comprised of the following six items: mental demand, physical demand, temporal demand, performance, effort, and frustration (see Appendix A). All items are scored on a 7-point Likert scale from 1 (*very low*) to 7 (*very high*), except performance which is scored from "perfect" to "failure". Research has provided evidence of good test-retest reliability (r = .71 to .81; Devos et al., 2020) and good internal consistency α = .86 (Braarud, 2021).

Simulator Sickness Questionnaire.

Research in VR warrants the monitoring for simulator sickness or cybersickness symptoms as a necessary duty of care. The SSQ (Kennedy et al., 1993) is a 16-item questionnaire with three subscales (nausea, oculomotor, and disorientation) scored on a 4-point Likert scale from 0 to 3 (0 = *none;* 1 = *slight;* 2 = *moderate;* 3 = *severe*) and demonstrates a high reliability of α = .91 (Xu et al., 2020). The SSQ was administered after each VR play session for the researcher to determine the participants ability to continue with the study (see Appendix C).

Apparatus.

Figure 1

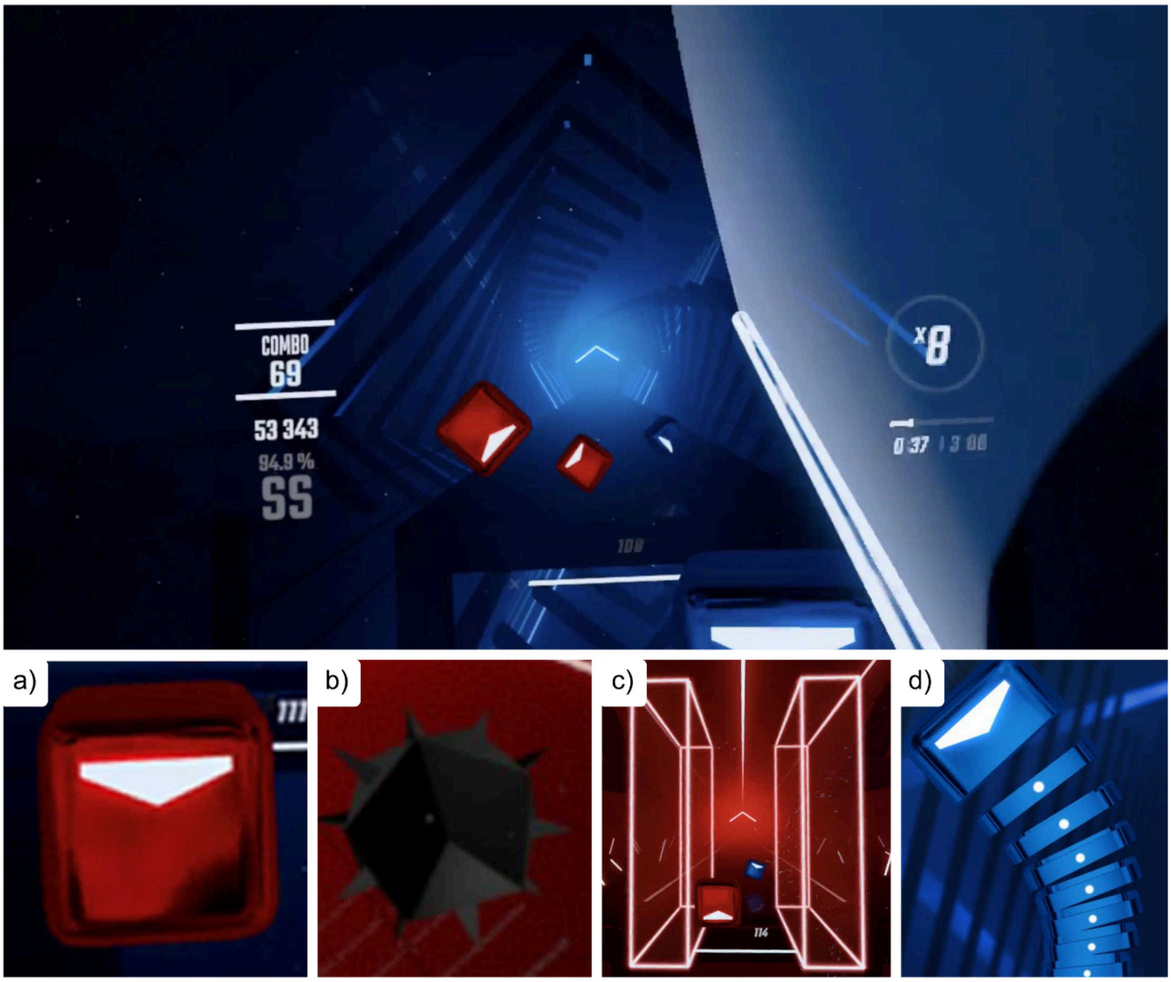
Superimposed Photo of the Experimental Layout

Note. The screen on the left displays the casting of Beat Saber from the MetaQuest Pro, which is being recorded on OBS. The laptop is used for all surveys, and the participant stands in an open space unobstructed by objects during gameplay.

The MetaQuest Pro (Reality Labs, Washington, USA) is a VR head-mounted-display with its own central processing unit. As shown in Figure 1, the surveys and recording of Beat Saber (Beat Games, 2018) version 1.31.0 gameplay were completed on separate computers to reduce distractions to participants between task and survey completion, and recordings were completed using Open Broadcaster Software (OBS version 29.1.3; Hugh, 2017) saved in MP4 format.

Figure 2

Screenshots from Beat Saber Gameplay and Main Interactable Objects



Note. a) blocks which represent notes of a song, b) mines which players should avoid, c) walls (an obstacle) that players' head should avoid, however, is interactable with their Sabers, and d) chain notes which represent sounds that are continuous.

Task performance is measured using the game Beat Saber (Beat Games, 2018). It requires players to react to visual cues of approaching coloured blocks by performing slicing movements to the rhythm or beat of a piece of music. Each block represents a note and needs to be cut in the direction (downwards, upwards, diagonally) indicated by an arrow on each block. Figure 2 demonstrates the virtual environment and objects players can interact with. The scoring system depends on the swinging angle movement when cutting a note and how precise a cut is through a note’s centre. Hitting a note perfectly has no impact on the overall score.

Previous studies did not have congruent song choices to reference from. However, from referencing Reddit forums (Fun_Committee6311, 2021; MemerGeist, 2022) and rigorous playthroughs by researchers, the songs selected needed to be distinguishable in difficulty (see Table 1). Similar song duration for the three experimental conditions (easy, normal, and hard) was to control for differences in exhaustion beyond task difficulty. For consistent gameplay, experience the following changes were made to Beat Saber settings:

- Switched on: no fail, automatic player height, fixed note (block) colours (red = right, blue = left), and the environment was set to The First.
- Switched off: arc haptics, arc visuals, environment effects and environment flickering.

The ranking system is percentage of accuracy based; however, this is only accessible during gameplay without modifying the game.

Table 1

Beat Saber Song Details for Experimental Conditions Tutorial, Easy, Normal and Hard

	Experimental Conditions [a]			
	Tutorial	Easy	Normal	Hard
Song Choice	Balearic Pumping	Rum n Bass	POP/STARS	Natural
Song Artist	Jaroslav Beck	Boom Kitty	K/DA	Imagine Dragons
Duration (minutes)	02:14	03:09	03:09	03:06
Beats Per Minute	111	132	170	100
No. of Notes [b]	170	155	341	533 (526) [c]
Notes per second	~1.26	~0.82	~1.80	~2.83
No. of Walls	9	23	41	0
No. of Mines	6	2	2	0

Note. [a] Experimental conditions align with the difficulty levels of the songs played, the only exception being the Tutorial as participants played a song on Easy difficulty level to warm up. [b] 'Notes' refers to the coloured blocks players interact with in the game.

[c] Although there are 526 notes, the maximum combo participants could achieve is 553 due to chain notes and is a more accurate representation for 'number of notes' in this song.

Procedure

This study was approved by the Human Research Ethics Committee (ethics number: 26489) at RMIT University. Eligible participants provided written informed consent upon arrival to the lab. They completed a survey battery prior to experiencing the VR conditions in the following order (1) demographic data about their previous experience with digital games, VR, and rhythm games, and (2) the Gold-MSI. Participants were provided a tutorial on how to use the MetaQuest Pro and how to play Beat Saber, which consisted of a tutorial accompanied by an Easy level song to reduce novelty effects for consecutive conditions through gameplay exposure. For each condition, participants were instructed to press 'play' after a count of 5 when the researcher said 'Go'. At the completion of each song, participants reported their scores verbally before removing the headset and being presented with the simulator sickness questionnaire and NASA-TLX. Participants completed a post-experiment survey at the completion of the hard difficulty level surveys, which asked participants to rank their preference of difficulty order by level of enjoyment and reasoning for their choice.

Design

The present study employed a within-subjects experimental design to evaluate changes to dependent variables across the main independent variable of task difficulty.

Data Analysis

The data were analysed using IBM SPSS 28. Demographic data were analysed descriptively and with independent sample t-tests. Hypothesis 1 was analysed using a one-way repeated measure of variance (ANOVA) with Bonferroni correction, and hypotheses 2 and 3 were analysed via a two-way mixed ANOVA with Bonferroni correction and post-hoc tests when appropriate.

Results

Demographics

Frequencies of responses to pre-experimental survey questions can be found on Table 3, which described players' experience with digital games, virtual reality, and Beat Saber on a 7-point Likert scale (see Table 2). On average, 78% participants played digital games once a week or more (M = 6.22, SD = 1.04), 63% have played virtual reality at least once in the last year or 5 years (M = 3.69, SD = 1.45), and 91% have played Beat Saber at least once in the last 5 years or less (M = 2.91, SD = 1.59). An independent samples t-test found no statistically significant difference between male (M = 72.78, SD = 4.48) and female (M = 69.18, SD = 4.02) participants for general musical sophistication scores, p = .588. Participants mean SSQ scores was within the expected range across all conditions: easy (*M* = 8.32, *SD* = 2.51), normal (*M* = 12.06, *SD* = 3.24) and hard (*M* = 16.29, *SD* = 4.57). Two participants scored above the standard threshold and were monitored to determine their fitness to continue the study.

Data Cleaning

Prior to running analyses, the data was inspected for outliers. Participant two's data was excluded from analyses due to reporting CL scores beyond the third IQR quartile for the easy difficulty condition, and higher CL scores across all difficulty conditions in comparison to other participants. This indicated that the participant might not have understood the task as instructed. Data for the remaining 31 participants was analysed. Assumptions of normality for CL, MT and task performance scores were tested via Shapiro-Wilk as the sample size was small ($n < 50$; Elliott & Woodward, 2007). The data was normally distributed with no outliers as assessed by boxplots and the Shapiro-Wilk test ($p > .05$) for hypothesis 1 and 2 and was violated with one outlier for hypothesis 3.

Table 2

Frequency Statistics of Participants Experience with Digital Games, Virtual Reality, and Beat Saber

Measure	Score	*f*	Relative *f*	c*f*	Percentile
Digital Games Experience	7	17	0.53%	32	100
	6	8	0.25%	15	46.875
	5	5	0.16%	7	21.875
	4	1	0.03%	2	6.25
	3	1	0.03%	1	3.125
	2	0	0.00%	0	0
	1	0	0.00%	0	0
VR Experience	7	1	3.13%	32	100
	6	3	9.38%	31	96.875
	5	3	9.38%	28	87.5
	4	11	34.38%	25	78.125
	3	9	28.13%	14	43.75
	2	2	6.25%	5	15.625
	1	3	9.38%	3	9.375
Beat Saber Experience	7	0	0.00%	32	100
	6	2	6.25%	32	100
	5	1	3.13%	30	93.75
	4	12	37.50%	29	90.625
	3	5	15.63%	17	53.125
	2	1	3.13%	12	37.5
	1	11	34.38%	11	34.375

Note. 1 = Never played. 2 = I have not played it in the last 5 years. 3 = I have at least played once in the last 5 years. 4 = At minimum, I play once a year. 5 = At a minimum, I play once a month. 6 = At a minimum, I play once a week. 7 = I play daily.

Qualitative Results

Following the main experiment, participants ranked the difficulty levels by which they found the most enjoyable accompanied by reasoning. Of interest was their highest ranked difficulty, which was hard (56%, n = 18), normal (31%, n = 10), and easy (13%, n = 4). Feedback from the hard condition was clustered thematically: the need to concentrate (*"the harder it was, the more I had to actually focus"*), high stimulation (*"it was a fun and* *novel challenge"*) and flow (*"appropriate difficulty for my skill level"*; *"...helped me to get into the zone"*). The normal difficulty was described as *"the perfect balance for my current expertise in the game"*, additional to it being the *"right amount of difficulty for the game"*. Participants who enjoyed the easy difficulty the most emphasised on performance, stating it was *"easier to slash the cubes, resulting in [a] higher score and better ranking"*.

NASA-TLX & Task Difficulty

By conducting a one-way repeated analysis of variance (ANOVA), we found a significant increase in perceived CL as task difficulty (easy, normal, hard) increased over time (see Table 3). Mauchly's test suggested that the data violated the assumption of sphericity, $\chi^2(2) = 6.16$, $p = .046$, therefore a Greenhouse-Geisser correction was applied ($\varepsilon = 0.883$). The changes to task difficulty showed that perceived CL varied significantly across task difficulty levels ($F(1.679, 50.36) = 112.79$, $p < .001$, partial $\eta^2 = .79$) with CL increasing from easy level ($M = 97.85$, $SD = 69.89$) to normal level ($M = 219.89$, $SD = 116.62$) to hard level ($M = 320.43$, $SD = 137.01$). Post hoc analysis with a Bonferroni adjustment revealed that CL significantly increased from easy to normal difficulty ($M = 122.04$, 95% CI [89.46, 154.62], $p < .001$), and from normal to hard difficulty ($M = 100.54$, 95% CI [66.56, 134.51], $p = .001$), and easy to hard difficulty ($M = 222.58$, 95% CI [177.48, 267.68], $p = .001$).

Table 3

Means, Standard Deviations, and One-Way Analysis of Variance in Cognitive Load and Task Difficulty Conditions Easy, Normal and Hard

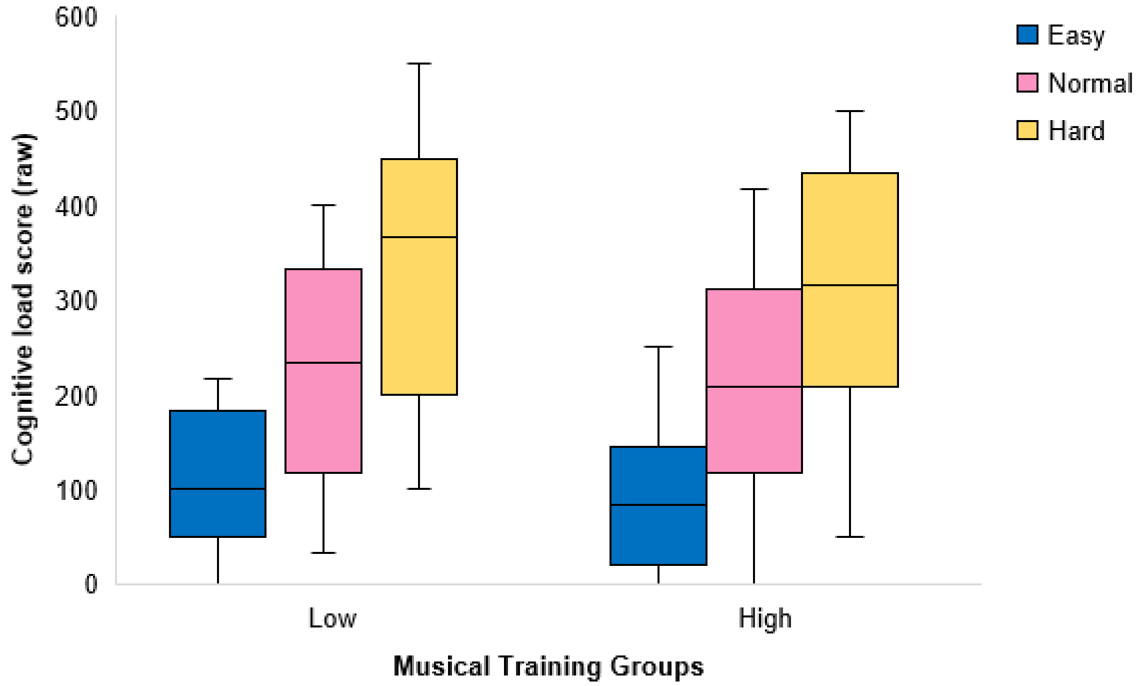
Measure	Easy		Normal		Hard		*F*(1.679, 50.362)	η²
	M	*SD*	*M*	*SD*	*M*	*SD*		
Cognitive Load	97.85	69.89	219.89	116.62	320.43	137.01	112.79***	.790

***$p < .001$.

Musical Training & Cognitive Load

Figure 4

Changes in Cognitive Load between Musical Training Groups as Task Difficulty Increases



A two-way mixed ANOVA was conducted with music training groups as the between-subject factor, task difficulty (easy, normal, hard) as the within-subject factor, and

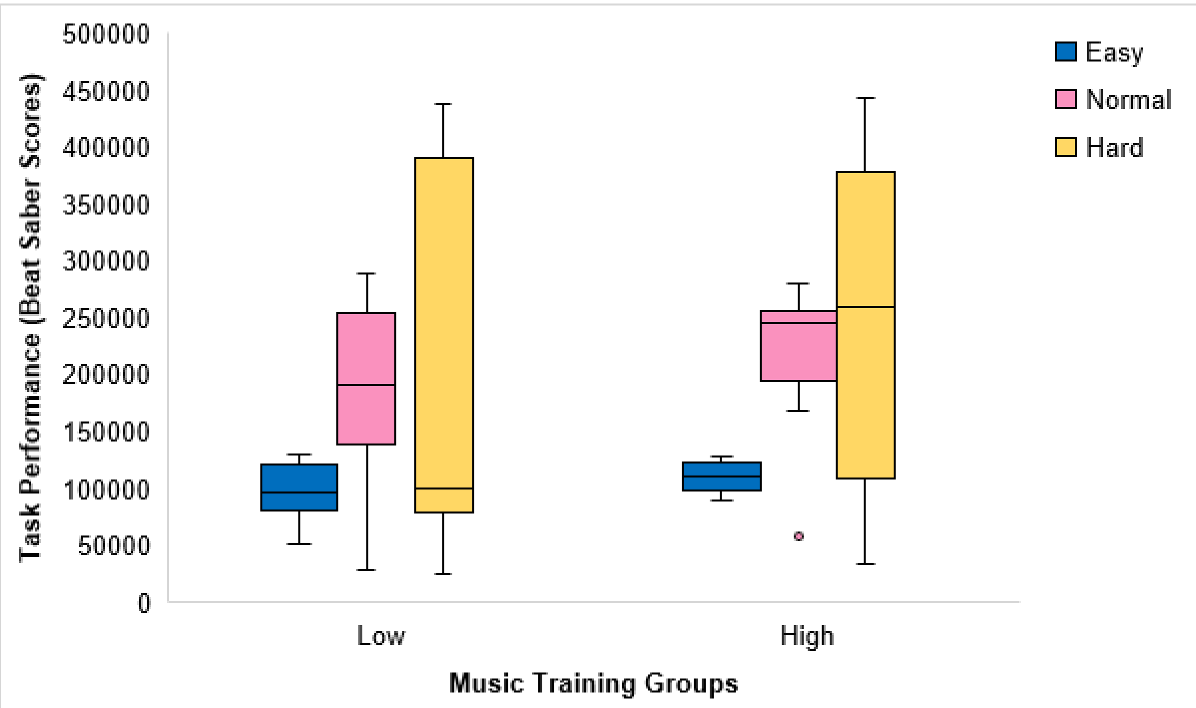
CL as the dependent variable. Homogeneity of variance ($p > .05$) and covariances ($p = .232$) were met, as assessed by Levene's and Box's M test, respectively. Mauchly's test of sphericity was met for the two-way interaction, $\chi^2(2) = .810$, $p = .053$.

The ANOVA was able to detect a significant effect of task difficulty ($F(2, 58) = 109.59$, $p < .001$, partial $\eta^2 = .79$) in CL for low and high MT as expected per the previous ANOVA for hypothesis 1. The main effect of MT showed no significant difference in CL between low and high MT groups $F(1,29) = .41$, $p = .527$, partial $\eta^2 = .014$. Furthermore, the ANOVA failed to detect a significant interaction between music training and task difficulty $F(2, 58) = .12$, $p = .90$, partial $\eta^2 = .004$. A visual representation of changes in CL between low and high music groups is presented in Figure 4.

Musical Training & Task Performance

Figure 5

Changes in Task Performance Between Musical Training Groups as Task Difficulty Increases



A two-way mixed ANOVA was conducted with MT groups as the between-subject factor, task difficulty (easy, normal, hard) as the within-subject factor, and task performance (Beat Saber scores) as the dependent variable. Participant 30 was a significant outlier, however retained, as running the ANOVA with and without this datapoint showed no difference in significance for the following assumptions and interaction effects. Normality was violated as assessed by Shapiro-Wilk's test ($p > .05$), and Mauchly's test indicated that sphericity was violated for the two-way interaction $\chi^2(2) = 20.59$, $p = .001$, therefore a Greenhouse-Geisser correction was applied ($\varepsilon = .700$). There was heterogeneity of variances, as assessed by Levene's test of variance ($p > .05$), and homogeneity of covariances as assessed by Box's M test ($p = .086$).

The main effect of task difficulty showed a significant difference in task performance for all difficulty levels, $F(1, 29) = 20.24$, $p < .001$, partial $\eta^2 = .411$.There was no significant interaction between task difficulty and MT on task performance, $F(1.32, 38.14) = .35$, $p = .618$, partial $\eta2 = .012$. A visual representation of changes in task performance between low and high music groups is presented in Figure 5.

Discussion

The present study aimed to explore how music expertise affects CL and task performance as task difficulty increases using Beat Saber. The study followed the prediction that higher music training elicits lower subjective cognitive workload and higher task performance scores in terms of easy, normal, and hard difficulty conditions. Hypothesis 1 predicted that a difference in perceived CL will occur as task difficulty changes, which was supported by significant results. The main effect of increasing task difficulty on CL reflected participants felt further challenged as difficulty increased, moreover, indicated the song choices for each difficulty level were distinctly different. The remaining results from this study did not support hypotheses 2 and 3. Hypothesis 2 and 3 predicted that MT groups could explain for the difference in CL and task performance as task difficulty increased. However, the interactions were nonsignificant. That being said, the main effects of task difficulty on CL and task performance were significant as informed by analyses ran for hypothesis 1. These findings seem to indicate that MT alone cannot account for differences in CL and task performance as the difficulty of a task increases. It should be noted that this is the first study to not display an interactional effect of music training on CL and task performance within the context of a rhythm-based VR exergame.

Task Difficulty & Cognitive Load

The repeated-measures ANOVA showed that CL was significantly different between task difficulty levels. This finding suggests the requirements of each difficulty level were different enough from each other, potentially providing a difference in challenge. Within the Beat Saber context, this reported change could be explained by an increase in the frequency of elements in higher difficulty levels, as the gameplay elements seen in Figure 2 can be learned independently (Sweller et al., 1998). This corroborates the assumption of element interactivity in CL literature (Sweller, 2010). The intrinsic load accrued from a task relies on its clarity, and our attempt to reduce extraneous load through a tutorial provided an opportunity for schema construction or consolidation before the three main conditions. Furthermore, element interactivity always related to an individual's level of knowledge or expertise (Chen et al., 2015). Qualitative responses of participants' most enjoyed difficulty provide insight into this nuance, as those who enjoyed the easy difficulty most emphasised the speed of the task and ease of interaction with incoming notes. Conversely, those who ranked the hard difficulty first highlighted that the need to focus on elements simultaneously was enjoyable and challenging. Henceforth, we can infer that participants who ranked harder difficulties first in enjoyment are likely more experienced.

Although CL was a defining point of difference as task difficulty changed, it should be noted that our results could be affected by the novelty of VR. Considering our participant sample consisted of minimal previous VR experience, and despite instructions to report CL based on the task alone, CL results can be conflated by way of exposure to a virtual environment (Anglin et al., 2017; Frederiksen et al., 2020).

Music Training & Cognitive Load

MT groups were not a significant predictor of CL as task difficulty increased. A point of difference to hypothesis 1 results was the main effect of task difficulty on CL between low and high MT groups without violating ANOVA assumptions, perhaps due to the robustness of mixed ANOVAs. The nonsignificant interactions between low and high music groups for CL might imply that irrespective of the grouping variable, the difficulty of the tasks was evaluated similarly across all participants (see Figure 4).

Music Training & Task Performance

MT groups did not demonstrate a significant interaction with task performance, contradicting hypothesis 3. Previous research suggests that individuals with higher music training, as opposed to those with lower music training, perform better on tasks that involve visual-spatial judgement (Strong & Mast, 2019). However, this study was unable to reproduce and directly measure this effect. In a general sense, Beat Saber requires players to make accurate visual-spatial judgements to interact with approaching objects within a virtual environment (Tian et al., 2023). Perhaps regardless of MT as difficulty increases, other factors are splitting the participants besides low and high MT groups (see Figure 5). Possibly due to unfamiliarity with VR or the latency sensitivity of Beat Saber (Li et al., 2020). An alternative speculation involves the potential skill transfer from games that involve hand-eye coordination, as 78% of our sample reported playing digital games at least once a week or more.

Another finding that stands out is that although the higher MT group exhibited scores closer to each other, the low MT group had a wider spread of performance scores across all difficulty conditions notwithstanding lower medians (Figure 5). This addresses a critical pitfall of the musicians versus non-musicians dichotomy within the literature, which Talamini et al. (2017) underscore to be from expertise acquired via informal training, as opposed to formal and the ignorance of other traits that contribute towards musical abilities. A possible explanation for this observation in our results is the grouping method for low and high MT. The single outlier raises attention to this notion, as their reported years of MT qualified their inclusion in the high group, while their overall music sophistication score was below the median of n = 74. This discrepancy is not unexpected, as there is no congruent metric for grouping participants using the Gold-MSI (Baker et al., 2020; Dahary et al., 2020; Matziorinis et al., 2023). To this end, although our findings broadly contribute to the inconclusive evidence between music sophistication and executive functions (Bowmer et al., 2018), it emphasises the need to examine musical expertise from a holistic set of experiences beyond years of formal training.

Limitations & Future Directions

The pilot nature of this study imposes inherent limitations for real-world applications beyond providing insight into interactions between CL and task difficulty. Notably, the small sample size for music research, the method of categorising groups, and the lack of additional measures considering the limited time available. Firstly, the sample may be biased because of recruiting on social media and one physical location, resulting in a younger sample than in previous studies. However, this variates from the literature, which is known for observing CL in young children or older adult populations within educational research (Yeung et al., 1998; Berends & van Lieshout, 2009; Chaddock-Heyman et al., 2021). Secondly, MT groups were categorised by a singular item, as opposed to multiple or an entire subscale, which is a deviation from previous research (Dahary et al., 2020; Matziorinis et al., 2023). In supplement to this, sociodemographic factors such as occupation and income were not assessed despite eliciting a high impact on music sophistication collectively in the original Gold-MSI study (Müllensiefen et al., 2014). Furthermore, future research would benefit from including objective listening tasks, as participants may underestimate or overestimate their responses on the Gold-MSI survey battery, potentially skewing results. It is recommended for future work using the Gold-MSI to account for sociodemographic and objective listening task differences to better compare participant characteristics between studies.

Although not assessed, additional insight could be derived from observing flow (Csikszentmihalyi, 1975), given that qualitative statements from participants infer different levels of challenge being appropriate for their current skillset. Furthermore, in observing subjective CL within a VR rhythm game task, there remains uncertainty in how much performance influenced participants' reporting. Future research is encouraged to explore the relationship between knowledge of performance and subsequent subjective CL, in addition to examining the effects of each NASA-TLX item and accounting for self-efficacy regarding task performance.

Conclusion

In summary, this thesis attempted to answer the assumption that musical expertise could explain the accumulation of CL, and task performance in the VR rhythm game Beat Saber as task difficulty increased. This investigation found that changes in task difficulty led to an increase in CL and task performance and that MT alone is not a strong predictor for either of those outcomes. Although there were significant effects within groups for CL and task performance, the relevance of MT was less than expected and should be interpreted with caution due to a potentially biased sample and the chosen method of categorising MT groups. From a theoretical viewpoint, this study adds to CL literature by meeting assumptions of element interactivity via a simple and engaging commercial game. Moreover, the findings align with prior VR literature, suggesting that participants with more digital games, VR, and Beat Saber experience tend to exhibit better task performance. Future research would benefit from categorising musical expertise groups through more robust methods to account for other factors that influence musical ability. Additionally, future research should measure flow and self-efficacy as qualitative responses inform their relevance. Thus, the present study offers the initial basis for how task difficulty explains some variability in CL and task performance and that interactions for music expertise may require different classifications to better account for differences in perceived cognitive workload and performance in a VR rhythm game.

References

[Fun_Committee6311]. (2021, December 27). *Beat Saber music packs from best to worst.* [Online forum post]. Reddit.https://www.reddit.com/r/beatsaber/comments/rpb8i7/beat_saber_music_packs_from_best_to_worst/

[MemerGeist]. (2022, March 20). *I honestly find the Beat Saber OSTs a lot better than some people think.* [Online forum post]. Reddit.https://www.reddit.com/r/beatsaber/comments/thwl8c/i_honestly_find_the_beat_saber_osts_a_lot_better/

Anglin, J. M., Sugiyama, T., & Liew, S.-L. (2017). Visuomotor adaptation in head-mounted virtual reality versus conventional training. *Scientific Reports*, *7*(1), 45469–45469. https://doi.org/10.1038/srep45469

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, 89-195. Academic Press. https://doi.org/10.1016/s0079-7421(08)60422-3

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation*. *8*, 47-89. New York: Academic press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baker, D. J., Ventura, J., Calamia, M., Shanahan, D., & Elliott, E. M. (2020). Examining musical sophistication: A replication and theoretical commentary on the Goldsmiths Musical Sophistication Index. *Musicae Scientiae*, *24*(4), 411–429. https://doi.org/10.1177/1029864918811879

Beat Games. 2018. Beat Saber game. https://beatsaber.com/

Bégel, V., Di Loreto, I., Seilles, A., & Dalla Bella, S. (2017). Music Games: Potential Application and Considerations for Rhythmic Training. *Frontiers in Human Neuroscience*, *11*, 273–273. https://doi.org/10.3389/fnhum.2017.00273

Berends, I. E., & van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: Effects of increased cognitive load. *Learning and Instruction*, *19*(4), 345–353. https://doi.org/10.1016/j.learninstruc.2008.06.012

Bogost, I. (2007). *Persuasive games: the expressive power of videogames*. MIT Press. https://doi.org/10.7551/mitpress/5334.001.0001

Bowmer, A., Mason, K., Knight, J., & Welch, G. (2018). Investigating the Impact of a Musical Intervention on Preschool Children's Executive Function. *Frontiers in Psychology*, *9*, 2389–2389. https://doi.org/10.3389/fpsyg.2018.02389

Braarud, P. Ø. (2021). Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human–machine work. *International Journal of Industrial Ergonomics, 86*, 103233. https://doi.org/10.1016/j.ergon.2021.103233

Campo-Prieto, P., Cancela, J. M., & Rodríguez-Fuentes, G. (2021). Immersive virtual reality as physical therapy in older adults: present or future (systematic review). *Virtual Reality : The Journal of the Virtual Reality Society*, *25*(3), 801–817. https://doi.org/10.1007/s10055-020-00495-x

Chaddock-Heyman, L., Loui, P., Weng, T. B., Weisshappel, R., McAuley, E., & Kramer, A. F. (2021). Musical Training and Brain Volume in Older Adults. *Brain Sciences*, *11*(1), 50-. https://doi.org/10.3390/brainsci11010050

Chee, Z. J., Leung, Y., & de Vries, M. (2022). Musical sophistication and its relationship with executive functions, autistic traits, and quality of life. *Psychomusicology: Music, Mind, and Brain, 32*(3–4), 87–97. https://doi.org/10.1037/pmu0000294

Chen, O., Kalyuga, S., & Sweller, J. (2015). The Worked Example Effect, the Generation Effect, and Element Interactivity. *Journal of Educational Psychology*, *107*(3), 689–704. https://doi.org/10.1037/edu0000018

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, 7-75. Hillsdale, NJ: Erlbaum.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety* (1st ed.). Jossey-Bass Publishers.

Cutting, J., Deterding, S., Demediuk, S., & Sephton, N. (2023). Difficulty-skill balance does not affect engagement and enjoyment: a pre-registered study using artificial intelligence-controlled difficulty. *Royal Society Open Science*, *10*(2), 220274–220274. https://doi.org/10.1098/rsos.220274

Dahary, H., Palma Fernandes, T., & Quintin, E.-M. (2020). The relationship between musical training and musical empathizing and systemizing traits. *Musicae Scientiae*, *24*(1), 113–129. https://doi.org/10.1177/1029864918779636

Devos, H., Gustafson, K., Ahmadnezhad, P., Liao, K., Mahnken, J. D., Brooks, W. M., & Burns, J. M. (2020). Psychometric Properties of NASA-TLX and Index of Cognitive Activity as Measures of Cognitive Workload in Older Adults. *Brain Sciences*, *10*(12), 994-. https://doi.org/10.3390/brainsci10120994

Elliott, A., & Woodward, W. (2007). Statistical analysis quick reference guidebook: With SPSS examples. *Sage*. https://doi.org/10.4135/9781412985949

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., & Andersen, S. A. W. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surgical Endoscopy*, *34*(3), 1244–1252. https://doi.org/10.1007/s00464-019-06887-8

Gao, Z., Pope, Z. C., Lee, J. E., & Quan, M. (2019). Effects of active video games on children's psychosocial beliefs and school day energy expenditure. *Journal of Clinical Medicine*, *8*(9), 1268–. https://doi.org/10.3390/jcm8091268

Haith, A. M., & Krakauer, J. W. (2018). The multiple effects of practice: skill, habit and reduced cognitive load. *Current Opinion in Behavioral Sciences*, *20*, 196–201. https://doi.org/10.1016/j.cobeha.2018.01.015

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human Mental Workload*, 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9

Huang, K.-T. (2020). Exergaming Executive Functions: An Immersive Virtual Reality-Based Cognitive Training for Adults Aged 50 and Older. *Cyberpsychology, Behavior and Social Networking*, *23*(3), 143–149. https://doi.org/10.1089/cyber.2019.0269

Huang, W., Roscoe, R. D., Johnson‐Glenberg, M. C., & Craig, S. D. (2021). Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *Journal of Computer Assisted Learning*, *37*(3), 745–758. https://doi.org/10.1111/jcal.12520

Hugh, B. (2017). the OBS Project Contributors. *Open Broadcasting Software*. Retrieved from https://www.obsproject.org/

Kalyuga, S., & Singh, A.-M. (2016). Rethinking the Boundaries of Cognitive Load Theory in Complex Learning. *Educational Psychology Review, 28*(4), 831–852. https://doi.org/10.1007/s10648-015-9352-0

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology, 3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *SIGCHI Bulletin*, *32*(1), 47–56. https://doi.org/10.1145/333329.333344

Lemmens, J. S., & von Münchhausen, C. F. (2023). Let the beat flow: How game difficulty in virtual reality affects flow. *Acta Psychologica*, *232*, 103812–103812. https://doi.org/10.1016/j.actpsy.2022.103812

Levitin, D. J. (2012). What does it mean to be musical? *Neuron, 73*(4), 633–637. https://doi.org/10.1016/j.neuron.2012.01.017

Li, Y.-C., Hsu, C.-H., Lin, Y.-C., & Hsu, C.-H. (2020). Performance Measurements on a Cloud VR Gaming Platform. *Proceedings of the 1st Workshop on Quality of Experience (QoE) in Visual Multimedia Applications,* 37-45. https://doi.org/10.1145/3423328.3423497

Matziorinis, A. M., Gaser, C., & Koelsch, S. (2023). Is musical engagement enough to keep the brain young? *Brain Structure and Function*, *228*(2), 577–588. https://doi.org/10.1007/s00429-022-02602-x

Meta. (2022). *Safety & Warranty Guide* (Meta Quest Pro). Retrieved from https://scontent.fmel7-1.fna.fbcdn.net/v/t39.8562-6/387101845_1376554836597789_

6607604066392692464_n.pdf?_nc_cat=109&ccb=1-7&_nc_sid=b8d81d&_nc_ohc=mdB77UAyAlEAX_cW89x&_nc_ht=scontent.fmel7-1.fna&oh=00_AfAY4LQxKreCAQr1at-iMtBZtAPoF_PWh38LdBuGhkD_5w&oe=652E1330

Moreno, R., Mayer, R. Interactive Multimodal Learning Environments. *Educ Psychol Rev,* 19, 309–326 (2007). https://doi.org/10.1007/s10648-007-9047-2

Mueller, F., & Isbister, K. (2014). Movement-based game guidelines. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2191–2200. https://doi.org/10.1145/2556288.2557163

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PLoS ONE, 9(*2), e89642. https://doi.org/10.1371/journal.pone.0089642

Orru, G., & Longo, L. (2019). The Evolution of Cognitive Load Theory and the Measurement of Its Intrinsic, Extraneous and Germane Loads: A Review. In *Human Mental Workload: Models and Applications*, 23–48. https://doi.org/10.1007/978-3-030-14273-5_3

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional Control of Cognitive Load in the Training of Complex Cognitive Tasks. *Educational Psychology Review, 6*(4), 351–371. https://doi.org/10.1007/BF02213420

Reason, J. T., & Brand, J. J. (1975). Motion sickness. Academic press.

Rutkowski, S., Adamczyk, M., Pastuła, A., Gos, E., Luque-Moreno, C., & Rutkowska, A. (2021). Training using a commercial immersive virtual reality system on hand–eye coordination and reaction time in young musicians: A pilot study. *International*

Journal of Environmental Research and Public Health, *18*(3), 1–9. https://doi.org/10.3390/ijerph18031297

Schmidt, S., Ehrenbrink, P., Weiss, B., Voigt-Antons, J.-N., Kojic, T., Johnston, A., & Moller, S. (2018). Impact of Virtual Environments on Motivation and Engagement During Exergames. *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 1–6. https://doi.org/10.1109/QoMEX.2018.8463389

Strong, J. V., & Mast, B. T. (2019). The cognitive functioning of older adult instrumental musicians and non-musicians. *Aging, Neuropsychology, and Cognition*, *26*(3), 367–386. https://doi.org/10.1080/13825585.2018.1448356

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4(*4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review, 22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Takac, M., Collett, J., Conduit, R., & De Foe, A. (2023). A cognitive model for emotional regulation in virtual reality exposure. *Virtual Reality: The Journal of the Virtual Reality Society*, *27*(1), 159–172. https://doi.org/10.1007/s10055-021-00531-4

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PloS One*, *12*(10), e0186773–e0186773. https://doi.org/10.1371/journal.pone.0186773

Tian, M., Cai, Y., & Zhang, J. (2023). The Impact of Virtual Reality-Based Products on Mild Cognitive Impairment Senior Subjects: An Experimental Study Using Multiple Sources of Data. *Applied Sciences*, *13*(4), 2372. https://doi.org/10.3390/app13042372

Xu, W., Liang, H.-N., Zhang, Z., & Baghaei, N. (2020). Studying the effect of display type and viewing perspective on user experience in virtual reality exergames. *Games for Health Journal, 9(*6), 405–414. https://doi.org/10.1089/g4h.2019.0102

Yerkes, R. M., Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*. *18*(5): 459–482. https://doi.org/10.1002/cne.920180503

Yeung, A. S., Jin, P., & Sweller, J. (1998). Cognitive Load and Learner Expertise: Split-Attention and Redundancy Effects in Reading with Explanatory Notes. *Contemporary Educational Psychology*, *23*(1), 1–21. https://doi.org/10.1006/ceps.1997.0951

Appendix A

Table 1

NASA Task Load Index Questionnaire (Short form) and Subscales

Subscales	Questions
Mental Demand	How mentally demanding was the task?
Physical Demand	How physically demanding was the task?
Temporal Demand	How hurried or rushed was the pace of the task?
Performance	How successful were you in accomplishing what you were asked to do?
Effort	How hard did you have to work to accomplish your level of performance?
Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?

Note. Participants were instructed mark a circle from 1 (v*ery low)* to 7 (*very high)* for each question.

Appendix B

Table 1

Goldsmith Music Sophistication Index Questionnaire

Subscale	Item
Active Engagement	I spend a lot of my free time doing music-related activities.
	I enjoy writing about music, for example on blogs and forums.
	I'm intrigued by musical styles I'm not familiar with and want to find out more.
	I often read or search the internet for things related to music.
	I don't spend much of my disposable income on music.
	Music is kind of an addiction for me - I couldn't live without it.
	I keep track of new of music that I come across (e.g. new artists or recordings).
	I have attended ___ live music events as an audience member in the past twelve months. **
	I listen attentively to music for ___ per day. **
Perceptual Ability	I am able to judge whether someone is a good singer or not.
	I usually know when I'm hearing a song for the first time.
	I find it difficult to spot mistakes in a performance of a song even if I know the tune.
	I can compare and discuss differences between two performances or versions of the same piece of music.
	I have trouble recognizing a familiar song when played in a different way or by a different performer.
	I can tell when people sing or play out of time with the beat.
	I can tell when people sing or play out of tune.
	When I sing, I have no idea whether I'm in tune or not.
	I don’t like singing in public because I’m afraid that I would sing wrong notes.
Musical Training	I have never been complimented for my talents as a musical performer.
	When I hear a piece of music I can usually identify its genre.
	I engaged in regular, daily practice of a musical instrument (including voice) for ___ years. **
	At the peak of my interest, I practiced ___ hours per day on my primary instrument. **
	I have had formal training in music theory for ___ years. **
	I have had ___ years of formal training on a musical instrument (including voice) during my lifetime. **

	I can play ___ musical instruments. **
Singing Ability	If somebody starts singing a song I don't know, I can usually join in. I can sing or play music from memory. I am able to hit the right notes when I sing along with a recording. I am not able to sing in harmony when somebody is singing a familiar tune. I don't like singing in public because I'm afraid that I would sing wrong notes. After hearing a new song two or three times, I can usually sing it by myself. I only need to hear a new tune once and I can sing it back hours later.
Emotions	I sometimes choose music that can trigger shivers down my spine. Pieces of music rarely evoke emotions for me. I often pick certain music to motivate or excite me. I am able to identify what is special about a given musical piece. I am able to talk about the emotions that a piece of music evokes for me. Music can evoke my memories of past people and places.

Note. *Contributes to the converted General Sophistication score.

** All items excluding those marked with double asterisks are scored from 1 to 7. 1 = completely disagree. 2 = strongly disagree. 3 = disagree. 4 = neither agree nor disagree. 5 = agree. 6 = strongly agree. 7 = and completely agree.

(Müllensiefen et al., 2014)

Appendix C

Table 1

SSQ Questionnaire and Subscale Weighting Computations

	Weights		
Symptom	Nausea (N)	Oculomotor (O)	Disorientation (D)
General discomfort	1	1	0
Fatigue	0	1	0
Headache	0	1	0
Eyestrain	0	1	0
Difficulty focusing	0	1	1
Increased salivation	1	0	0
Sweating	1	0	0
Nausea	1	0	1
Difficulty concentrating	1	1	0
Fullness of head	0	0	1
Blurred vision	0	1	1
Dizzy (eyes open)	0	0	1
Dizzy (eyes closed)	0	0	1
Vertigo	0	0	1
Stomach awareness	1	0	0
Burping	1	0	0
Total	[1]	[2]	[3]

Note. The total scores are calculated according to the following equations:

N = [1] x 9.54.

O = [2] x 7.58.

D = [3] x 13.92.

Total Score = ([1] + [2] + [3]) x 3.74.

(Kennedy et al., 1993).