UNIVERSITÉ DE GENÈVE

FACULTÉ DE PSYCHOLOGIE ET DES SCIENCES DE L'ÉDUCATION

THE INFLUENCE OF BACKGROUND MUSIC ON LEARNING FROM TEXT

MÉMOIRE REALISÉ EN VUE DE L'OBTENTION DE LA MAITRISE UNIVERSITAIRE EN SCIENCES ET TECHNOLOGIES DE L'APPRENTISSAGE ET DE LA FORMATION

PAR

Edouard Adam

Directeur du mémoire

Mireille Bétrancourt

Jury

Mireille Bétrancourt, TECFA, FPSE, Université de Genève

Huib Tabbers, Université Erasmus de Rotterdam

Julien Venni, TECFA, FPSE, Université de Genève

Genève, Août 2019
Université de Genève

Université de Genève
Faculté de Psychologie et des Sciences de l'éducation

ABSTRACT

This thesis aims to provide the reader with an objective look at the past scientific research done about the influence of background music on cognition. It also describes the findings of an experimental approach focusing more specifically on the case of background music's effect on learning from text. 48 participants from the University of Erasmus Rotterdam had to answer comprehension questions after learning texts varying in reading difficulty while being exposed to background music or silence. The experiment aims to prove that background music can have different effects (positive or negative) on learning, based on the reading difficulty of the learned material.
Arousal levels and cognitive load measures were reported as well. Statistical analysis failed to show significance from the effects of background music, reading difficulty, arousal levels and cognitive load on comprehension scores. Reasons for these results are discussed, along with the directions to take for future research.

UNIVERSITÉ
DE GENÈVE

FACULTÉ DE PSYCHOLOGIE
ET DES SCIENCES DE L'ÉDUCATION

Déclaration sur l'honneur

Je déclare que les conditions de réalisation de ce travail de mémoire respectent la charte d'éthique et de déontologie de l'Université de Genève. Je suis bien l'auteur-e de ce texte et atteste que toute affirmation qu'il contient et qui n'est pas le fruit de ma réflexion personnelle est attribuée à sa source ; tout passage recopié d'une autre source est en outre placé entre guillemets.

Genève, le 18 Août 2019

Edouard Adam

Signature :

Table of Contents

Introduction and context

Music has been an important part of human life since the dawn of times. The oldest known musical melody dates back to 1400 B.C. While traditionally music required performers, either through singing or via rhythmical or melodic instruments, nowadays, our ability to record music has allowed us to listen to music without the need for musical performers. Moreover, music streaming through the use of the internet now enables us to play any type of music at any given moment. The world-wide known website *Youtube* has now dedicated channels available at all times for music listening. Some of these music channels are for instance dedicated to helping to study and can average up to 15 thousand users live simultaneously.

Additionally, according to a recent scientific report, the prevalence of most leisure activities, such as watching TV, movies or reading books, has been overtaken by music listening (Rentfrow & Gosling, 2003). In addition to that, thanks to the invention of the radio, television, computers, smartphones and other connected devices, it has become easier than ever to be listening to music in the background while focusing on something else (working, studying, cooking, cleaning, etc.). Because music has become such an overwhelming presence in our everyday life, we cannot ignore its potential impact on human behavior. Not surprisingly, the field of scientific research has been investigating how listening to music can affect humans across varying contexts.

For example, listening to music has been found to alter mood and modify consumer behavior (e.g., Alpert & Alpert, 1990; Bruner, G., 1990 for reviews, cited by Zentner, Grandjean & Scherer, 2009). Music has been also shown to influence the speed of eating and drinking depending on the speed of the music being listened to (McElrea & Standing, 1992). On a different level, music has also shown beneficial effects in treating emotional disorders, by inducing positive moods (Gold, Voracek, & Wigram, 2004, cited by Thompson, Schellenber & Letnic, 2012).

While many individual studies have shown music to have a wide range of effects across various contexts, scientific reviews studying the effects of background music, more specifically on task performance, cognition and learning have offered mixed evidence (Dalton & Behm, 2007; h & Harms, 2018; Schellenberg, 2005). Depending on the situation, listening to background music seems to either help performance, worsen it, or have no effect at all.

These scientific reviews point out to a lack of experimental consistency across studies which makes any findings hard to generalize. Music is sometimes being listened to previously to the task, sometimes during the task. Sometimes the choice of music is left to the subjects, other times it is forced upon them. The loudness of the music chosen can vary, so as the length of the musical excerpt, the style of the music, and the presence of lyrics. All of these variables that differ from study to study makes it challenging to yield a clear-cut conclusion about the influence of background music on task performance.

The aim of this thesis is to present and discuss the main theories regarding how background music affects cognition and more specifically learning and text comprehension. A second goal is to create an experimental design in which all of the variables presented before are either controlled or at least mentioned, in order to provide an answer about the impact of background music in a specific learning situation, which is learning from text. We also aim to provide researchers with a clear insight for future replications.

1. Background music and task performance

The Mozart effect

One of the first studies to catch enough empirical attention in the field of background music and its influence on general task performance was a study conducted by Rauscher, Shaw, & Ky (1993) about a potential "Mozart effect". In this study, the authors found that participants performed better on spatial-temporal tasks right after listening to a Mozart sonata for about 10 minutes, in comparison to participants who sat in silence or listened to relaxation instructions before the test.

Note that this study paradigm differs from multiple studies that will be conducted later on as listening to music took place before the test, instead of listening and learning taking place simultaneously.

Listening to classical music before passing the test was thought to prime spatial abilities because of the direct effect the sonata would potentially have on specific brain areas. These areas were thought to be typically used for spatial-temporal reasoning and would be activated by this specific music. The pattern of activation would in term facilitate short-term performance on spatial-temporal tasks. The main explanation for the positive results was thought to be provoked by priming effects.

The Mozart effect, a priming effect?

Priming refers to the activation of particular representations or associations in memory right before carrying out an action or a task. For example, a person who sees the word *blue* will be slightly faster to recognize the word *ocean* in comparison to someone who saw an unrelated word. This associative priming takes place because the concepts of blue and ocean are closely associated with each other in long-term memory, leading to facilitation in retrieval.

Priming can be either positive or negative depending on the effect it has on the task (Stadler & Hogan, 1996). It can be unconscious (Draine & Greenwald, 1998), can result in changes in behavior (Payne, Brown-Iannuzzi, & Loersch, 2016) or changes in cognition (McKoon & Ratcliff, 1979).

According to Husain, Thompson, & Schellenberg, (2002), a priming effect cannot apply in the case of the Mozart effect. Indeed, the authors explain that priming should only occur on two occasions. Either, when the same stimulus is being exposed at least two times (repetition priming) or when the subsequent stimulus has a relation to the first one (associative or semantic priming).

In the case of the Mozart effect, the Mozart sonata, being an auditory stimulus, has no relation whatsoever to the spatial-temporal task, which is a visual stimulus. The authors demonstrate that the priming effects are very unlikely because the nature of the two stimuli are completely unrelated: "*If the Mozart effect were indeed an example of priming, it would be a surprising* *instance of cross-modal priming between an auditory event and a visual task in which the priming stimulus (music) is seemingly unrelated to subsequent stimuli (tests of spatial reasoning presented visually)*." (Husain, Thompson, & Schellenberg, 2002)

Indeed, a study by Klimesch, Schimke, & Schwaiger (1994), cited by Husain, Thompson, & Schellenberg, 2002, has shown priming effects tend to disappear when the prime and the target have few features in common, which is the case with the Mozart effect. Therefore, the priming hypothesis behind the effect is seemingly unlikely.

Moreover, a significant number of researchers have been unable to replicate the findings from Rauscher and colleagues (e.g. Carstens, Huskins, & Hounshell, 2011); (Steele, Bass, & Crook, 1999); (Stough, Kerkin, Bates, & Mangan, 1994), as cited by Gonzalez & Aiello, (2019). .

However, Rauscher, Shaw, & Ky (1993), by being among the first to explore the effects of music on cognition, opened up a new door which lead researchers to investigate more deeply the potential impact of listening to music on task performance, through emotion, cognition and behavior.

Background music's outcomes on task performance, a modern approach

Background music on general behavior

Behne (1999) is one of the first researchers to have conducted a meta-analysis about the influence of background music on general behavior.[1] Behne analyzed 153 studies that examined the impact of background music on 'non-musical behavior'. These behaviors involved for instance achievement on test scores and in school, understanding of documentaries, behavior in the workplace and public places, and even reaction to driving a car or playing sport while listening to music.

In his analysis, Behne found out that about one-third of all studies showed non-significant results and another third pointed out to 'inconsistent results'. Moreover, he found that the proportion of non-significant results had increased over time across studies. Behne concluded that background music had a negligible impact on everyday behavior and that this very small impact had decreased over time through habituation effects of humans being more and more exposed to music.
However, a deeper look into Behne's analysis, reveals some biases. For instance, it does not differentiate between the positive and the negative effects from music listening and only looks at p-values while ignoring effect sizes.

According to Field (2009), a renowned statistician, a significant p-value does not tell us about the importance of an effect. The only way to measure the size of an effect is to have a standardized measure of the relationship's strength between the variables studied, which is defined as the effect size. Moreover, having a standardized measure of power makes it possible

[1] A meta-analysis is a statistical analysis that combines the results of multiple scientific studies. A crucial benefit from this approach is obtaining a higher statistical power than with a sole study.

to compare studies that have assessed different variables or have used different scales of measurement. Reporting only the p-value is not enough to conclude anything about the power of a studied effect.

Background music on behavior, cognition and emotion

Years later, Kämpfe et al. (2011) acknowledged these problems and aimed to conduct a meta-analysis including reports of effect sizes and distinguishing between positive and negative effects. They first analyzed background music's influence on general behavior, similarly to Benhe's analysis.
Their global analysis on 97 studies indicated no general effect of background music. However, because the studies were not comparable due to their differences in experimental designs, the authors chose to separate the various studies into four categories in order to analyze them separately: "Mundane behavior"; "Cognition" (judgment); "Cognition" (achievement) and "Emotion".

Their analysis suggests that background music impacts motoric behavior (e.g. playing sports) and emotional reactions positively. However, music seems to impact cognitive behavior negatively such as in tasks involving memory processes.

Amongst other explanations, the authors put attentional limitations as one of the main reasons for the negative impact of background music on cognition. According to them, the attention that should be put into performing the cognitive task is being pulled away by the process of listening to the music, which in terms impairs cognitive performance. The attention distraction hypothesis takes place especially in tasks that require conscious efforts such as a reading task (Treisman, 2006, as cited by Kämpfe et al. (2011). However, the authors also announce that a greater number of studies which examine specific cognitive effects in more detail would be necessary to conclude any general tendencies, whether positive or negative.

Background music on learning

The effect of background music on learning is still debated, with researchers adopting two opposite positions. Music can either be seen as hindering learning, or improving it. A number of studies also conclude that background music has no overall effect on learning. The important studies and theories for each side will be briefly presented here.

Music hinders learning, research and theories

Music can be viewed as a potential distractor. In reading tests for example, studies have found either a negative effect of listening to music (Henderson, Crews, & Barlow, 1945) or no effect at all (Freeburne & Fleischer, 1952), as cited by Landay & Harms (2018) . Additionally, music is capable of distracting or deterring performance of certain tasks (e.g., Cassidy & MacDonald, 2007); (Furnham & Bradley, 2005); (Furnham & Strbac, 2002); (Reynolds, McClelland, & Furnham, 2014); cited by Landay & Harms (2018).
For instance, Furnham & Strbac (2002) discovered that performance during a reading comprehension task was significantly worse with music in comparison to silence.

The negative effects from background music found in multiple studies and meta-analysis has lead researchers to develop theoretical models to puzzle out these results. We chose to present two models that seem particularly relevant because they involve crucial factors in learning situations, such as attentional resources and working memory.

One of these theoretical models is called the **irrelevant sound effect** (ISE; e.g., Jones, Alford, Bridges, Tremblay, & Macken, 1999) and states that auditory stimuli, including music, impairs task performance if two conditions are simultaneously present.
First, the learner must rely on encoding and retrieving information in a particular order, as it is for instance common in mental arithmetic, free recall, and serial recall tasks (Beaman & Jones, 1997); (Perham, Banbury, & Jones, 2007), as cited by Gonzalez & Aiello, (2019). Second, the auditory stimuli must be made of a number of different sounds in the stimuli (a high degree of acoustic variation)(D. Jones et al., 1999).

If these two conditions are met, auditory distractions such as music interfere with task-related items that are being held in working memory (Perham, Marsh, Clarkson, Lawrence, & Sörqvist, 2016); (cited by Gonzalez & Aiello, 2019). The ISE is robust, and can occur regardless of the modality of the task (auditory or visual; Campbell, Beaman, & Berry, 2002) or the volume of the auditory stimuli (Ellermeier & Hellbrück, 1998) (cited by Gonzalez & Aiello, 2019).

Although the ISE is robust, the rigorous conditions under which it takes place cannot be met by any experimental conditions. For this reason, the ISE should not be regarded as a general model to explain the negative effects of background music on learning because it is difficult to spread the results found under a very precise experimental design to a real-world learning situation.

Another theoretical model which condition might be met with more ease, takes root in the field of multimedia learning. Under a learning situation, **the seductive details effect** describes how an information that is irrelevant to the instructional goal, but still alluring, can "seduce" our attention, leading to negative effects on learning (Harp & Mayer, 1997).
For example, according to the seductive details effect, it would be useless and even detrimental to present a decorative picture of a bee while presenting a material about reproduction in bees because that information is irrelevant to the instructional goal in the sense that it does not add any useful information about the subject. This is only exact under the assumption that the learner knows what a bee is.

The seductive detail can be anything from an image, an illustration, written texts to spoken text, background noise and even background music. In a study by Moreno & Mayer (2000), participants heard thunderstorm noise while learning about how lightning forms. They were then tested on retention and transfer tests about the learned material. The results showed that adding entertaining but irrelevant auditory material to a multimedia instructional message was detrimental to learning, similarly to what the seductive details effect entails.

This result was obtained across two different dependent measures (i.e., retention and transfer) within two distinct multimedia learning messages (i.e., an explanation of lightning and an explanation of hydraulic braking systems). In the same study, it was also found that subjects listening to background music during learning performed worse in the transfer and retention tests when compared to those who did not listen to music.

The explanation made by the authors for these results was that auditory stimuli related to the learning material seemed to have overloaded auditory working memory, thus affecting learning negatively. But an overload from the working memory is not the only consequence from adding seductive but irrelevant details to a learning material. Other explanations include distracting attention, disrupting coherence and hindering schema interference (e.g.,Harp & Mayer, 1998); (Lehman, Schraw, McCrudden, & Hartley, 2007). In conclusion, Mayer & Harp (1998) assure that seductive but irrelevant details will always result in decrements in learning, in whichever form they are (visual or auditory).

Music improves learning, research and theories

While research has shown music to have negative effects on learning, a number of studies concluded that listening to music could as well improve performance during a variety of tests, including reading comprehension, arithmetic, and memory in both children and adults (Doyle & Furnham, 2012); (Hallam S., Price J., & Katsarou G., 2002), as cited by Landay & Harms (2018).

Additionally, studies have found that students who heard music played during class lectures earned higher scores on an exam than students who did not (Schlichting Jr. & Brown, 1970) (cited by Landay & Harms, 2018). Other studies explored how undergraduate students who listened to classical music during a pre-recorded lecture had significantly higher scores on a following multiple-choice exam (Dosseville, Laborde, & Scelles, 2012) (cited by Landay & Harms, 2018).

Moreover, studies have even found that moderate volumes of background music facilitate performance in activities that involve high levels of concentration and attention (Corhan & Gounard, 1976); (Fontaine & Schwalm, 1979); (Davies, Lang, & Schackleton, 1973), as cited by Dalton & Behm, (2007). But how can one explain this wide range of positive effects from listening to music?

Many researches have suggested that the Mozart effect presented before cannot be retained as a general explanation for the positive effects of music on cognition or learning, due to its experimental biases and overall poor replicability (Husain et al., 2002); (Steele et al., 1999). However, a majority of researchers agree to say that music indeed facilitates such performance, because the auditory stimulus from the music is considered stimulating, in that it increases motivation, arousal, perception of energy and mood (Atkinson et al., 2004; Davies et al. 1973; Matthews et al., 1998).

Indeed, rather than listening to Mozart directly affecting cognitive performance, such as what the Mozart effect proposes, music listening in general may be a method of inducing arousal which in turn provides temporary improvement in learning outcomes. This particular hypothesis was primarily developed by Husain et al. (2002), in answer to the Mozart effect.

According to the mood-arousal hypothesis, the positive effects from listening to music may be best explained as an artifact of arousal or mood (e.g., Cassity, Henley, & Markley, 2007); (M. H. Jones, West, & Estell, 2006);(Thompson, Schellenberg, & Husain, 2001).

The arousal-mood hypothesis attracted scientific attention over the last years and is even today seen as providing one of the best framework for explaining the favorable effects of background music on cognition (Nguyen & Grahn, 2017). It seems therefore essential to present this theory, along with various studies assessing the link between arousal, mood and learning performance.

We will start by explaining how the theory was born, followed by a more specific presentation of more modern studies confirming - not only an effect of music on arousal and mood - but also the effects of arousal and mood on learning performance.

The Arousal-Mood hypothesis

After having elaborated a critique to the Mozart's effect, Husain et al., (2002) proposed an alternative theory. According to their theoretical view, listening to background music would not have any direct influence on spatial-temporal reasoning as proposed by Rauscher et al., (1993) but would affect performances indirectly through the mediation of arousal and mood. The authors even speculate further by postulating that this mediation should not only influence spatial abilities, but also any type of cognitive performance, therefore including learning.

History of arousal and performance

The link between arousal and performance has been studied since a long time. More than a century ago, Yerkes & Dodson, (1908) published an article on the "Relation of Strength of Stimulus to Rapidity of Habit-formation", in which they discovered that rats could be motivated to complete a maze with slight electrical shocks. However, contrary to their expectation, they found out that the performance level of rats decreased when the shocks were the strongest.
At this time, the authors did not perform any statistical analysis and the term arousal was not part of the scientific jargon yet. But they started speculating on why the performance level only increased up to a certain point. They assumed the strength of the stimuli to account for better performances only up to a certain point, at which performance would diminish.

About 50 years later, Donald Hebb, a Canadian neuropsychologist, proposed in an article about drive and the central nervous system that arousal in relation to performance followed an inverted U-curve, in a similar way to Yerkes and Dodson's hypothesis (Hebb, 1955). Hebb was the first one to introduce the term arousal which is nowadays common jargon among researchers, although the term is wildly misused. For this reason, we find it essential to give a definition of the term. Arousal in our case will be defined according to Sloboda & Juslin's definition (2001), which is the degree of physiological activation or the intensity of an emotional response.

In a similar way to Yerkes and Dodson theory, Hebb suggested that increased arousal could help improve performance, but only up to a certain point. Once arousal becomes too high, performance diminishes. Figure 1 shows the relationship between arousal and performance according to the hebbian law.

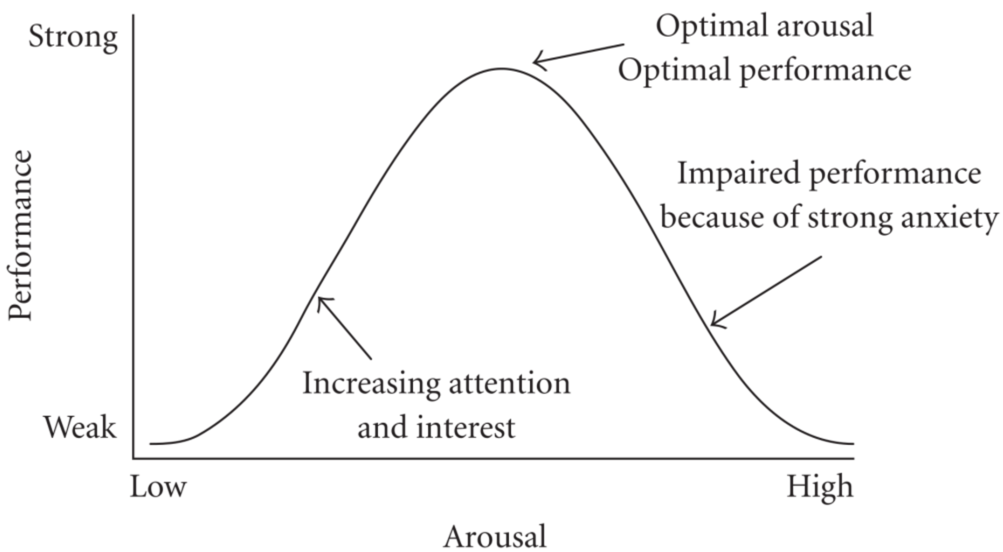

Figure 1: Hebbian version of the Yerkes-Dodson law.

As years went by, with the help of new recording tools, researchers started being able to report more accurate measures of arousal. Some of these measures include physiological responses to music and recorded brain activity.

Effect of music on arousal, through physiological responses

In 1996, Bartlett analyzed the effect of music and sound stimuli on physiological responses. In the majority of the studies he investigated (61%), an effect of music on different physiological parameters was found. For music previously characterized as arousing, there was an increase in heart rate and muscle tension. Whereas, for calming music, a decrease in heart rate, muscle tension, as well as an increase in skin temperature and skin conductance was found.

In the same year, Witvliet and Vrana (1996; Witvliet et al. 1998) also investigated the effect of musical excerpts on physiological measures such as skin conductance, heart rate, as well as changes in zygomatic, corrugator, and orbicularis oris muscles in the face. Differently to Bartlett's analysis, the music excerpts had been this time pre-selected with respect to their scores on the valence and arousal dimensions. Their results showed clear arousal effects with an increase in levels of skin conductance and heart rate, as well as an increase in tension of the orbicularis oculi muscle, during listening to excerpts judged as high on arousal levels. Witvliet and Vrana were among the first to show arousal effects on physiological responses for music that was characterized as such.

More recently, it was found that specific brain activation patterns are associated to music listening. A study by Schmidt and Trainor (2001) examined whether pattern of regional EEG (Electro encephalograms) activity could differentiate emotions induced by musical excerpts. The excerpts were varying in affective valence (i.e., positive vs. negative) and intensity (i.e., intense vs. calm). Based on the valence of the musical excerpts, the pattern of asymmetrical frontal EEG differed. Indeed, participants showed greater relative left frontal EEG activity to joy and happy musical excerpts and greater relative right frontal EEG activity to fear and sad musical excerpts. Not only research has been shown to affect arousal levels through music listening, but also mood.

Effect of music on mood, through mood manipulation

According to Sloboda & Juslin (2001), mood is a long-lasting psychological state not elicited by concrete events at a particular point in time but rather experienced as a diffuse affective state. A mood is low in intensity but relatively long in duration, often without apparent cause and thus persisting across encounter of multiple events. As mentioned by Husain et al. (2002) in the arousal-mood hypothesis, mood is similar to valence, which can be either positive or negative. Examples of positive moods can include happiness, excitement, or calmness whereas negative moods include sadness, anger, nervousness, or boredom.

Studies of mood manipulation have increased in recent years, placing music among the most widely used techniques for mood induction (Juslin and Sloboda, 2010) For instance, in an experiment by Fried and Berkowitz (1979), various groups of subjects heard musical excerpts that were either soothing, stimulating, or aversive in nature, while a fourth group was not exposed to any music. Subjects who heard the soothing music were most likely to volunteer for another study immediately afterwards. The authors believe it is evidence for music to impact mood positively and even affect behavior, leading participants to an altruistic comportment. Mood ratings pointed out that the soothing and stimulating musical excerpts elicited varying positive moods whereas aversive music tended to arouse negative feelings.

Although music is known to influence mood, it is still under debate whether people perceive the expressed state within the music (cognitivist view) or whether music can actually induce moods in listeners (emotivist view). Moreover, a potential problem with mood induction is that a similar mood induction condition can result in various emotional reactions from different participants. Thus, some authors believe that eliciting a mood trough music might vary too much from individuals to individuals to be able to draw any general tendency (Västfjäll, 2001, p. 182).

So far, we have examined how music influences arousal and mood, but the arousal-mood hypothesis not only pre-suppose an existing relationship between arousal, mood and music, but also between arousal, mood and cognitive performances. We will briefly explore this relationship by presenting studies about mood and arousal influencing learning.

Mood and arousal, influences on learning

In a recent study by Nadler, Rabi, & Minda (2010), positive, neutral, and negative moods were induced in subjects. Subjects in the positive-mood condition performed better than subjects in the neutral or negative mood conditions in classifying stimuli. More generally, positive mood has been associated with better learning outcomes (Isen, 2002) while negative mood or boredom hinders learning (O'Hanlon, 1981; Pekrun, 2006).

Additionally, Kleinsmith & Kaplan (1963) found arousal to influence long term memory. In their study, paired associates learned under high arousal exhibited low immediate recall but high permanent memory, meaning that arousal influenced long term memory, but not short-term memory. This study was nonetheless not the only one to find effects of arousal on memory. In another experiment by Bradley, Greenwald, Petry, & Lang (1992), arousal was found to be the only dimension (amongst others like pleasantness) to have a stable effect on memory performance: pictures rated as highly arousing were remembered better than low-arousal stimuli.

However, arousal is not only known to directly influence memory, but was also found to amplify the effects of competition between mental representations, influencing memory consolidation of representations that are active concurrently (Ponzio & Mather, 2014).

These different studies all show in some way or the other that arousal and mood influence either learning or memory, which are closely related, as to learn is first to remember. As cited by Lehman & Seufert, (2017), there is broad evidence of the impact of arousal on learning (e.g., Kleinsmith and Kaplan, 1963; Eysenck, 1976; Heuer and Reisberg, 2014).

Background music and task performance, a summary

We started by presenting a prominent study introducing the positive effect from listening to music: the Mozart effect. Although the study showed experimental biases and consequent studies challenged its findings, it generated a movement amongst researchers who felt the need to seek the potential power of music on cognition, whether positive or negative.

We then analyzed these studies, showing how background music can influence general behavior, cognition, emotion and learning, both negatively and positively. The main theories that revolved around these opposite sides were presented, while putting a particular attention on the seductive details effect and the arousal-mood hypothesis. The first theory explains how music can deteriorate performance under specific conditions, while the latter explains best how music affects performance positively.
After presenting the mood-arousal hypothesis, we presented studies that evaluated the link between music and arousal, music and mood, and finally about how mood and arousal influenced learning.

Now that we have examined how background music influences behavior, cognition, emotion, mood and arousal more specifically, we still need to clarify what is meant by learning. Learning possesses such a wide meaning that we need to specify in the first place we are interested in how background music influences learning in the case of learning from text more specifically. The next chapter will therefore be dedicated to what is learning, starting from multimedia learning to learning from text, thus presenting the main models in the respective fields.

One could consider that the learning theories from multimedia learning do not apply to our case because we are interested in how learning from text takes place, and that is correct to a certain extent. However, if one considers the presence of background music while learning, along with the fact that in our case learning will take place on a computer, then one can consider this situation as a form of multimedia learning. Nevertheless, the field of multimedia learning has had such important breakthrough towards learning that, even if one would not accept this earlier assumption, presenting researches from the field of multimedia learning is fundamental, if not essential.

2. Learning from multimedia

Learning from multimedia material is getting more frequent nowadays, partly because of the ever-growing use of multimedia documents, thanks to computers and smartphones. In the early 80s, a number of studies found that pictures seemed to facilitate comprehension and memorization of the information from a text. In response to these results, theoretical models were developed to come up with explanations to why texts benefited from adding relevant pictures. The Cognitive Theory of Multimedia Learning (CTML; Mayer, 2002) was born. Today, it might be considered as one of the most insightful theories about multimedia learning. The theory aims to give a research-based insight of how students learn, in order to provide a more effective use of technology in learning.

The cognitive theory of multimedia learning, inspiration

Mayer's work was greatly inspired by Atkinson and Shiffrin's work (1986) which proposes that three different types of memory storage work side by side: the sensory memory, the short-term memory and the long-term memory.

- Sensory memory stores a great number of basic auditory and visual stimuli from our perceptual system for a very short time. It is used to filter important information. It can as well be defined as the cognitive structure that allows us to perceive new information. (Sweller, 2005).

- Short term memory is used for actively processing information in order to build a mental construct while doing a task. The storage capacity is limited to 7+-2 separate elements according to Miller (1956). The exact storage capacity, which is also called memory span, changes from individual to individual.

- Long-term memory stores our knowledge base for long term. The storage capacity is hypothetically unlimited, but, although information from the long-term memory cannot be erased, it can become irrecoverable, and thus forgotten. The more elaborated a mental model about an information is, the easier an item will be remembered.

Mayer's CTML also took inspiration by three largely know theories from the field of cognitive sciences: the limited capacity from working memory (Baddeley, 1986, 1998), the dual coding theory (Paivio, 1986), and the active processing theory (Mayer, 1999, 2001).

Baddeley's model and definition of working memory (1986, 1998) overlaps with the short-term memory such as mentioned by Atkinson and Shiffrin (1986). According to Baddeley, working memory helps processing information. It is made out of three main mental structures: the phonological loop, the visuo-spatial scratchpad and the central executive.

- The phonological loop treats auditory information. A visual stimulus can also be processed by the phonological loop depending on its nature (i.e. text).

- The visuo-spatial scratchpad stocks and process visual information.

- The central executive controls and coordinates these two structures.

Pavio's double coding theory (1986), similarly to Baddeley's model of working memory, states that information is treated by separated structures depending on the nature of the stimulus - visual or auditory. Although the two structures are independent, the information can travel from one to another depending on the need to visualize a word or put a word on an image for example.

Lastly, Mayer's active processing principle (1999, 2001) states that learners benefit from actively processing the information. For example, focusing our attention on important elements, or organizing mental elements with our previous knowledge can benefit learning, because it leads to us creating a better mental model. As mentioned before, the more elaborated a mental model is, the easier the information contained in it will be remembered.

Based on these three theories, Mayer (2001) distinguishes three steps for processing information: learning is an active process of selecting, organizing and integrating information based upon prior knowledge. The next chapter describes how learning takes place in the light of these three mechanisms.

Selecting, organizing and integrating information

First, the bits of information such as the images and the written letters from the material get logged in the learner's sensory memory. Pictures and printed texts are only briefly held as visual and auditory images. Indeed, sensory memory only briefly maintains an exact copy of what was presented.

After that quick initial moment, the learner's working memory selects information from the sensory memory in order to maintain and manipulate it. At this state, information becomes accessible to consciousness. With the help of the working memory, the learner can select which images and words to remember and work with, and which ones not to.

With two distinct channels, the learner is able to work with more than one information at a time based on the type of stimulus (auditory or visual). Indeed, different types of stimuli are processed in a distinctive manner. Each of these sets of information (auditory and visual) are processed and organized into coherent schemas that allows the learner to understand and remember the information. The fact that one can process both visual and auditory stimuli as the same time is one of the main reasons why Mayer declares we learn more deeply from texts with visual stimuli than from text alone, as we can handle more information at the same time.

Finally, the learner integrates the visual model and the auditory model altogether with their prior knowledge and experiences from their long-term memory to form one mental model. Once all the material has been associated in a functional way, the newly acquired knowledge can travel from the working memory to the long-term memory. From this point, one can say that the information has been learned.

Figure 2 summarizes visually where the three different processes take place, as well as how a learner process new information.

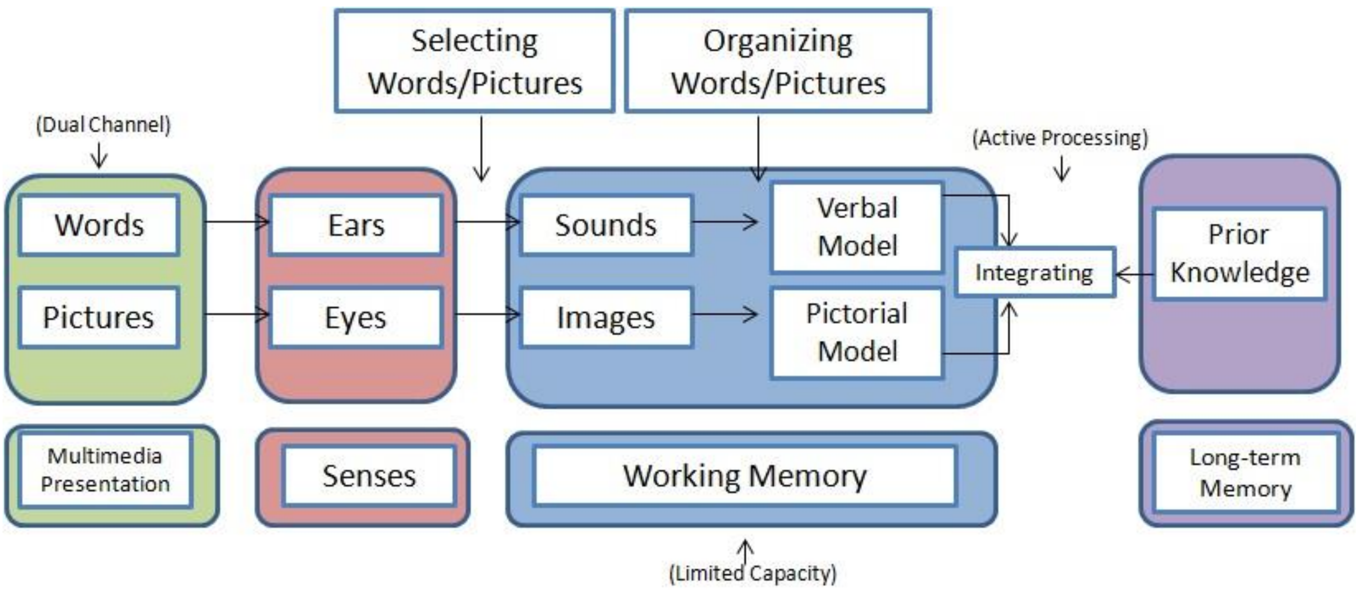

Figure 2: A visual representation of the cognitive theory of multimedia learning, Mayer (2001)

We can see how visual and auditory stimuli follow a different path, even though they can be processed simultaneously, thanks to dual channels.

Although Mayer's CTML presents a very robust model of learning. the theory might not have elaborated enough on how exactly the integration of information takes place. Fortunately, another model coming from the field of text comprehension provides a more detailed explanation of the integration process: Kintsch's model of construction and integration (1988).

Because of the precise insight it gives on the integration process, along with the fact that our experiment will be focused on learning from text, we will present this model in which learning from text is explained into more details.

3. Learning from text

Construction-Integration model from Kintsch

Although Kintsch and Van Dijk's model of construction and integration (C-I) about text comprehension was born in 1983, it is still very relevant today because of how good of an insight it provides on the cognitive processes involved in comprehending a text. The C-I model describes the comprehension of any text in two distinct phases. First, a text base is built, made out of the interaction between the linguistic input from the text and the learner's knowledge base. Then, in a second phase, the text base is integrated by the reader into a coherent whole, the situation model.

The text-base model and the situation model

The text base model consists of a network of concepts and relations directly and only derived from the text. The main content is taken from the syntactic and semantic structure of the text, which is then formed in terms of propositions by the reader's mind. A proposition represents a complete idea from the text. For instance, the sentence "I gave my toothbrush to my sister" turns into the propositional representation: gave (toothbrush, I, sister). The following example consists of the predicate 'gave' followed by three arguments including an agent, (I), an object (toothbrush) and a goal (sister). But the text-base mental model by its own is not enough to completely understand a text.

The reader must add nodes and establish links between nodes from knowledge about the world and their experience, in order to produce a coherent structure. The information provided by the text is constructed with the integration of prior knowledge. This mental representation is called the situation model (Kintsch, 1988). In more simple terms, the situation model refers to the global understanding of the text and corresponds to a deeper level of understanding than the text-base model.

Figure 3 below propose a graphical representation of the process of construction and integration from a text.

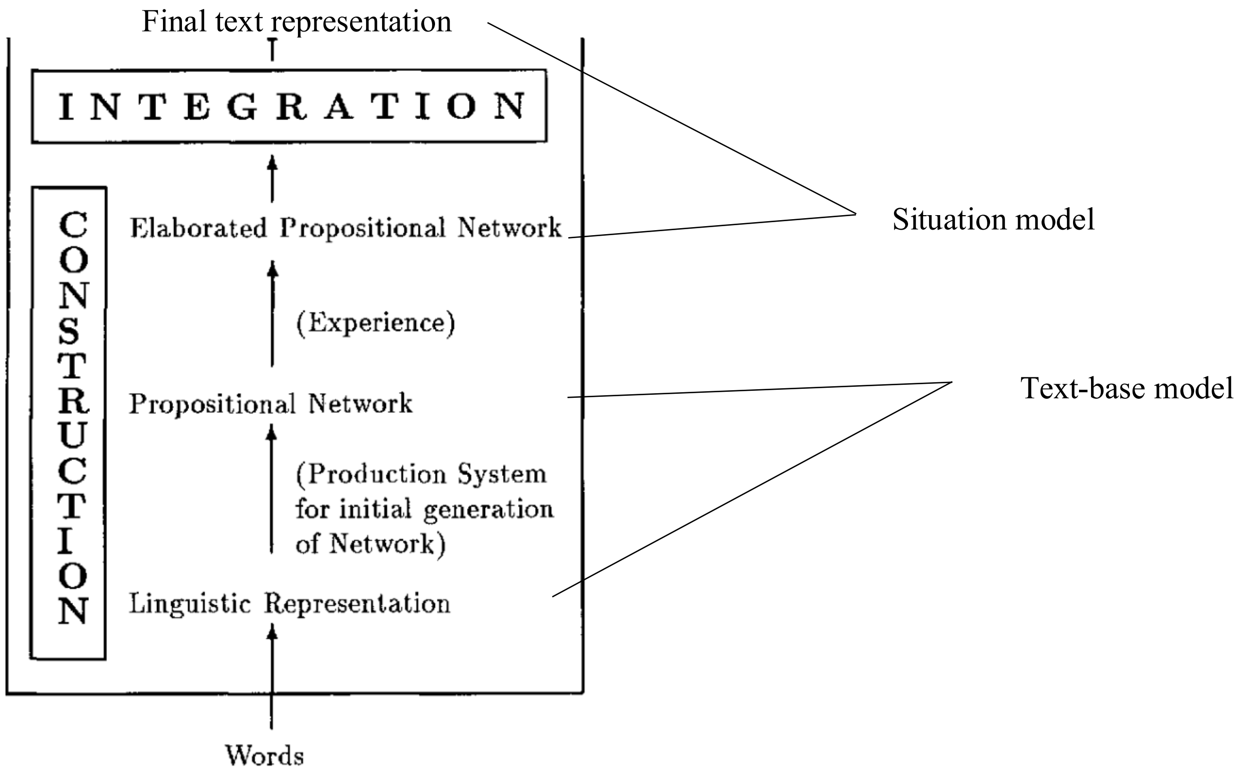

Figure 3: Visual representation of the C-I model (retrieved from Wharton & Kintsch, 1991), with text-base and situation models added

The words are turned by the reader into a linguistic representation, then into a propositional network, which form the text-base model. From that is added past experience, to form an elaborated propositional network, which will eventually be integrated into the final representation (situation model). The construction-integration model differentiates between memorizing a text (text-base understanding) and understanding the concepts of a text (situational understanding). Remembering a text implies that one reproduces the content of the text in some form, without necessarily understanding it. Learning therefore implies the ability to use the information provided by the text and combine it with previous knowledge. The new information from the text can thus be associated and integrated. A typical example of learning from a text would be a subject answering questions requiring the need for inferences (derive by reasoning) without being able to access the text while answering.

Main factors involved in text comprehension

In the field of text comprehension, crucial factors involved in text comprehension are prior knowledge and the text general structure (Bransford & Johnson, 1972). Prior knowledge can be defined as the amount of knowledge a learner has accumulated about a specific domain prior to learning about that domain (adapted from O'Donnell, Sharp, Wade, & O'Donnell (2014)). Prior knowledge is assumed to be necessary for the reader to fill contextual gaps within the text and to develop a global understanding of the text.

When referring to the text structure, an essential characteristic is coherence. Coherence can be defined as the extent to which the relationship between ideas in a text are explicit. Coherence is influenced by the number of conceptual gaps (information not explicitly stated) from the text and left for the reader to infer. Therefore, text comprehension depends on a great level on the ability to fill in missing information that is not explicitly stated in the text, which in other words is the ability to generate inferences.

Consider for example the following text: *The student quickly threw his thesis in the fire because he could not work anymore. The ashes floated up the chimney*. Here the reader must draw the inference that the thesis burned in the fire. If the reader did not know that fire burns paper into ashes, they would not be able to fully comprehend the text (adapted from Singer & Ferreira, 1983).

Most of the information extracted from text is not understood or remembered from the direct content of the text, but more so from abstractions, inferences, and the use of prior knowledge in order to help maintain a coherent representation of ideas from the text and derive an understanding from the text's content. It is thus not surprising that McNamara, Kintsch, Songer, & Kintsch, (1996) found that when prior knowledge about the learned subject is low, texts that are more coherent are easier to understand.

The ability to generate inferences and to perceive coherence from a text, along with our prior knowledge are all dimensions processed at some point by our working memory. However, our working memory has limited capacity to handle information, whether textual or auditory. The amount of information that can be processed at the same time by our working memory is referred as **cognitive load**. If too many information is presented, our ability to read and understand a text is altered. The next chapter is dedicated to presenting the cognitive load theory (Sweller, 1994) which is a crucial instructional theory aiming to provide advices in order to facilitate learning by preventing cognitive overload.

The cognitive load theory

The cognitive load theory from Sweller gives us a more precise insight about the role of working memory and cognitive load when learning from instructional material. Processing too much information at the same time can lead working memory to saturate and worsen learning. According to Sweller, because short-term memory is limited, learning experiences should always be designed to reduce cognitive load in order to promote better acquisition of the material.

Different types of cognitive load

Processing of information, reading, selecting, and integrating takes cognitive resources. According to Sweller (1999), there are three sources of cognitive demands: intrinsic load, extraneous load, and germane load.

Intrinsic cognitive load consists of the cognitive load provoked by direct exposition to the instructional material itself. It is caused by the natural complexity of the information or the amount of element interactivity involved in the studied material. It is also referred as the interaction between the nature of the material being learned and the expertise of the learner.

Extraneous cognitive load, on the other hand, is the source of cognitive load caused by factors that are not central to the material itself but the presentation of it. A high amount of extraneous load does not serve and even impairs successful learning.

For example, the process of switching our attention from a spot to another on a document, in order to acquire information is not optimal and can thus be defined as a source of extraneous load. To reduce the extraneous load provoked by attention switching, Yaghoub Mousavi, Low, Sweller, & Mousavi, (1995) for instance suggest that rather than having labels alongside a diagram, labels should be placed directly on the diagram. The rationale behind this tactic is that putting the labels directly on the diagram reduces the student's need to switch their attention between the text and the visual image thus decreasing extraneous load and improving learning.

Finally, **germane cognitive load** is dedicated to the processing, construction and automation of schemas. Learning in general takes cognitive resources, and this use is called germane load. This type of cognitive load is "effective" and directly associated to success in learning. For this reason, it is crucial to design instructional material that increase germane load, but limits the two precedent sources of load as much as possible (intrinsic load and extraneous load).

Other crucial components that are crucial when designing instructional material comes from the field of readability, which aims to assess how easily a reader can understand a text.

Readability and reading difficulty, a quick review

Readability can be defined as the ease with which a reader can understand a text. The primary aim of the classic readability studies was to create handy strategies to match reading materials with the capacities of students and grown-ups. These endeavors focused on making effective readability formulas which educators and librarians could use.

Lively & Pressey (1923), as cited by DuBay (2004), were concerned about the practical implication of choosing science textbooks for middle school. The books were so overlaid with specialized words that instructors invested too much energy teaching vocabulary to students They argued that developing an approach to quantify and lessen the "vocabulary load" of reading material would be useful. Their article featured the first readability formula. It measured the number of various words in each 1,000 words and the number of words that were not on the Thorndike vocabulary list of 10,000 words. Their method produced a correlation coefficient of .80 when tested on 700 books, meaning that a great number of words from the books were not on the Thorndike vocabulary list.

Other researchers started developing readability formulas to provide an objective and replicable measure of text reading difficulty. For instance, one of the most famous readability formula, which is called the Flesch reading-ease test, was developed by Kincaid (1975). The formula calculates the mean length from the sentences of a chosen text, as well as the mean number of syllables per words, in order to assess the overall reading difficulty from any chosen text.

While this was a fairly good measure to detect elaborated sentences and words, it was biased. Indeed, because sentences that are very long do not necessary imply that they are more challenging to understand and many words that contain multiple syllables are still easily comprehensible (i.e. crocodile). For this reason, although readability formulas became very diverse, we believe they lack enough parameters to be able to provide a replicable and unbiased prediction of a text's reading difficulty.

However, more recently, and thanks to the growing use of artificial intelligence, a more modern approach was taken by Graesser, McNamara, & Kulikowich (2011) to evaluate text's reading difficulty. They designed an algorithm, *Coh-Metrix*, which based its prediction on over 40 characteristics, ranging from text characteristics to cognitive characteristics.

Perhaps the central argument of the article presenting the algorithm according to the authors themselves, is that the algorithm possesses advantages over earlier unidimensional metrics of text difficult and can identify problems and solutions at a more precise degree, within a multilevel framework of text comprehension.

Learning from text, a small summary

In the previous chapter, we started by analyzing how learning from multimedia takes place, in the light of the CTML by Mayer. This theory covers a wide range of cognitive processes implicated in learning, but lacks deeper explanation for the integration of information.

From the field of text comprehension, the Construction-Integration model from Kintsch (1998) covers the integration of information in more detail. After presenting the C-I model, explaining how the reader constructs a situational model from the text-base model, we analyzed the main factors involved in text comprehension.

We then presented the cognitive load theory from Sweller (1994; 2005), which aims to provide ways to limit unnecessary cognitive load from our working memory. Finally, we presented the concept of readability and reading difficulty, from the earliest scientific attempts to calculate it, to a more modern approach that enables a multidimensional understanding of reading difficulty.

We will now present our theoretical frame, upon which we will base our experiment about the influence of background music on learning from text.

4. Theoretical frame and research question

So far, we have introduced three different theories regarding how music could affect cognitive performance: the Mozart effect, the ISE, the seductive details effect and the arousal-mood hypothesis. As proposed before, the Mozart effect, due to the overall poor replicability and the multiple issues from the theoretical perspective, will be discarded. The ISE will also not be a theory relevant to our study because of its specific requirements in design that do not match with our experimental goals. This leaves us with two opposing theoretical views about the potential influence of background music on learning from text: the arousal-mood hypothesis and the seductive details effect.

Earlier on we have stated that a seductive detail could be anything ranging from a picture to written texts or even background noise or music. If background music can be considered as a seductive detail, a safe assumption would be that background music would have the same impact as any other type of seductive detail, leading to worsen learning. The main reasons behind this harmful effect would be of overloading working memory and distracting attention.

So far, to our knowledge, only one research has looked into the effects of background music on learning, with both the arousal-mood hypothesis and the seductive details effect in the same experimental design. Lehmann & Seufert (2017) investigated the role of working memory and its interaction with background music when considered in the light of the Mozart effect, the arousal-mood hypothesis and the seductive detail effect.

The authors speculate that subjects who listen to background music would have to suffer from attention distraction, because of their uncontrolled need to divide their attention between learning and listening to the music being played in the background. In this scenario, cognitive resources (which we could consider as extraneous load) are used to process background music, in addition to the learning task, leading to disruption in comprehension.

In their study, Lehmann & Seufert (2017) assumed that learners with low working memory capacity would be overwhelmed by processing both the learning material and the background music. While learners with sufficiently high working memory capacity could benefit from the potential positive effect of listening to background music which would compensate for the additional cognitive burden.

As hypothesized, their analysis revealed a significant interaction between background music and working memory capacity, meaning that the higher the learners' working memory capacity, the better they learned with background music. Moreover, the presence or absence of background music influenced comprehension outcomes with no background music leading to better comprehension of the material presented. This result gives support to the hypothesis that background music can indeed act as a seductive detail and cause harmful effects on learning, more precisely on comprehension in this case.

While the interaction between background music and working-memory capacity was significant, and the presence of background music influenced comprehension negatively, they did not find background music to influence significantly arousal levels or mood, therefore not being able to replicate the arousal-mood hypothesis.

In the study, they used an instrumental version of a song with lyrics for the background music, which was thought as a mistake in the discussion, as subjects could potentially sing the possibly-known lyrics in their head, interfering with working memory while reading the material. Lehmann & Seufert (2017) failed to prove that listening to background music had an influence on arousal levels, leading to better learning performances.

We aim to design an experiment in which the arousal-mood hypothesis has a better chance of being elicited, while intending to replicate the effects of the seductive details effect at the same time.

Although, these two antagonistic views are opposed in the effect they have on learning performance and are for this reason usually not being studied simultaneously, we believe that the overall effect of background music on comprehension might be a balance between the costs and benefits associated to these conflicting effects. This hypothesis is in line with what Kämpfe et al., (2011) discussed in their meta-analysis about the influence of background music. The authors suggest that listening to background music interferes with cognitive related task such as reading and memory, but seems to have positive effects on emotions possibly due to the arousal-mood hypothesis.

For this reason, we believe it is crucial to design an experiment where these opposing forces are manipulated to put more weight on one or the other effect, depending on the condition. Indeed, the potential *cost* of background music listening for comprehension is that it *places demands on attention* and *overburdens working memory*. The potential *benefit* of background music listening for comprehension is that it *enhances arousal levels and mood* which then affect learning positively. We aim to create an experimental condition in which both effects take place simultaneously.

In order to do that, we will use a musical excerpt known to enhance arousal levels and mood – a Mozart's sonata for two pianos in D Major, K. 448 (Husain et al., 2002; Schellenberg et al., 2007; Thompson et al., 2001). The musical excerpt is expected to be long enough to not have to repeat itselft, to exclude any bias from the repetition (i.e. boredom or irritation). The musical piece does not contain any lyrics, because lyrics are known to interfere with cognitive load. (Perham and Currie, 2014) (Shih et al., 2012; Shih, Huang, & Chiang, 2009).

A report from the mean decibels level of the musical piece will be given, as we think it is a crucial piece of information for any future replication. Surprisingly enough, we have only found very few studies in the field of background music which reported the volume of the music played to the participants' ears. Moreover, the majority of them reported volume levels as a percentage only. An information which, by itself, is not sufficient, since volume level (in percentages) is not a comparable measure when used with various devices.

What differs from previous research is that we aim to manipulate the studied material itself instead of having a measure of working memory from subjects. Although it might be more challenging than having reports of working memory capacity from participants, we are interested in manipulating the cognitive load experienced by the participants, through modifying the difficulty of the material presented. More specifically, we want to manipulate the reading difficulty (RD) of the material presented, along with the presence or absence of background music, in order to explore both the arousal-mood hypothesis and the seductive details effect in the same study.

Namely, we created two versions of each text that will be read by participants before answering comprehension questions about them: one low reading difficulty (RD) condition and one high reading difficulty (RD) condition. The modifications for RD were based on characteristics used by an algorithm presented before - *Coh-Metrix* - created by Graesser et al. (2011), which we believe to be the best existing measure for assessing text RD. The reason behind using the characteristics of the algorithm and not the algorithm in itself were the following: first, it was unfortunately inaccessible for our use. Second, even if it would have been available, the algorithm would have required a minimal number of words from the texts to be able to work properly. This number would have been unreachable with our current material.

Nevertheless, because *Coh-metrix* provides currently one of the best measure of RD, we chose to use its main characteristics, provided by the article from Graesser et al. (2011) to create manually two texts of varying RD (low and high) with the same content. The precise parameters we chose to modify are described in the method chapter.

Once again, the aim of our research is to explore the effects of background music and text RD on learning, more specifically on text comprehension. We expect the arousal-mood hypothesis and the seductive detail effect to take place simultaneously, but we think it is possible to put more weight on one or the other effect by manipulating the texts' RD. This manipulation should in turn increase cognitive load in learners, such that in the high RD texts, participants should experience a higher cognitive load than in the low RD texts. Moreover, the increase in cognitive load provoked by the manipulation of RD should lead to different learning outcomes based on the interaction between the level of RD and the presence/absence of background music.

Indeed, when the task is easy enough in terms of cognitive resources (low RD), music should enhance mood and arousal levels – facilitating learning outcomes. In contrast, when the task is more cognitively demanding (high RD), music should act as a seductive detail and be more prevalent than the effect it has on arousal, resulting in worse comprehension scores.

5. Method

Participants

Data was collected on 48 participants (90 % women //43 women, 5 men) who were recruited through the ERAS system of Erasmus University Rotterdam. Age of participants ranged from 17 to 27 years old (M=20, SD= 2.20). These students passed the experiment in order to receive credits for their bachelor in psychology. Informed consent was given when participants applied through the system. Requirements for the experiment were to not have any hearing problems and being able to read in English.
An auto-evaluation of their reading competence was given as part of a pre-test questionnaire at the beginning of the experiment. 56% of the participants evaluated their reading competency to a C2 level, 35% to a C1 level, and 9% between B1 and B2. Participants main language was either English (14.6%), Dutch (22.9%), German (27.1%) or others (35.4%).

Participants were randomly assigned to one of the experimental group of the between-subject factor: background music - present or absent. Reading difficulty was a within subject factor, meaning that each participant had to read both easy and hard texts.

Material

The four texts used in this experiment had been previously created and tested by Meteyard, Bruce, Edmundson & Oakhill (2015) in an experiment profiling text comprehension impairments in aphasia patients. Each of the four texts came with four comprehension questions, requiring both global and local inference, which are important attributes to evaluate a text comprehension, as seen before. Each of the four texts were modified to create two versions varying in RD: a high RD version and a low RD version. The main differences between the two versions is in the syntactic structure of the sentences. The modification to make the text harder to read were based on characteristics used by an algorithm -*Coh-Metrix* - created by Graesser et al. (2011) and cited in their article. These characteristics are:

- The number of low-frequency words, in proportion to the higher-frequency ones. A low-frequency word is a word that is not usual and can thus make the entire sentence more difficult to comprehend (Graesser et al., 2011)
- The frequency of passive voice, which is more difficult to process than active voice (Just & Carpenter, 1987)
- Lexical diversity: The most well-known lexical diversity index is “type-token ratio”. Type token ratio is the number of unique words in a text (i.e., types) divided by the overall number of words (i.e., tokens). A high “type : token ratio” indicates that words are not repeated many times in the text, which should generally decrease the ease and speed of text processing (Graesser, McNamara, Louwerse & Cai, 2004).
- Word polysemy: polysemy is measured as the number of senses of a word. A word with more senses will generally be more ambiguous and slow to process, particularly for less skilled (Gernsbacher & Faust, 1991) and less knowledgeable readers (McNamara &

McDaniel, 2004).
- Number of polysyllable words vs fewer syllable words. Polysyllable words are usually harder to process (Graesser et al., 2011)
- Sentence syntax, which is harder to process when the sentences are longer, with more words before the main verb of the main clause, and more logic-based words (i.e. and, or, not) (Graesser et al., 2011).
- Sentences with many words before the main verb tax the reader's working memory (Graesser, Cai, Louwerse, & Daniel, 2006).

An example of the two RD versions for one of the text is available in annex number 1

Each participant read two low RD texts, and two high RD texts. The order in which they read them was counterbalanced (report to the table below for further details).

Table 1: Distribution of participants

Background music	Present		Absent	
Order of text Difficulty (4)	easy-easy-hard-hard	hard-hard-easy-easy	easy-easy-hard-hard	hard-hard-easy-easy
Participants (48)	12	13	12	11

The music we chose to play was an instrumental piano piece – a Mozart's Sonata for two pianos in D Major, K. 448 and was chosen for reasons explained before. The music did not contain any lyrics and was played to participants on headphones at a volume of 30% inside a closed cubicle of the Erasmus Behavioral Lab. The loudness of the music was averaging 55 dB when putting the microphone near the headphones. This volume is thought to be reasonable, based on other similar experiment design.

For the measure of cognitive load, the traditional items from cognitive load of perceived effort and difficulty were measured. The measure of perceived effort from Paas (1992) was presented as followed: "In the text that just finished, I invested: "very, very low mental effort to very, very high mental effort". Participants had to answer on a 9 scales-point. The measure of perceived difficulty created by Ayres (2006) was as followed: "The text that just finished was: "very, very easy - to very, very difficult". The scale was on 9 points too.

The questions were open-ended (i.e. what was the weather like?) and participants had to type in the answer, without being able to look at the previously read text. 16 comprehension questions were asked in total. 8 of them pointed out to the low RD texts, the 8 others to the high RD texts. Participants could either score 1 point for answering correctly, 0.5 points for answering partly correctly, or 0 points for not answering correctly, leading to a maximum of 16 points in total.

Arousal was measured throughout the experiment four times to account for the effects of background music on arousal levels with the subscale of the Self-Assessment Manikin (SAM), created by (Bradley & Lang, 1994), which was previously used for measuring arousal by Lehmann & Seufert (2017). Participants needed to answer to a question regarding their arousal state by selecting one of the 5 manikins with a certain level of arousal indicated by a bigger or smaller explosion in its belly. Other measures from the SAM were excluded because they had no purpose for this experimental design.

A post-test questionnaire was given about their habits regarding the use of music and classical musical when studying. An item measuring the self-perceived help of the condition (music or silence) on task performance was given.

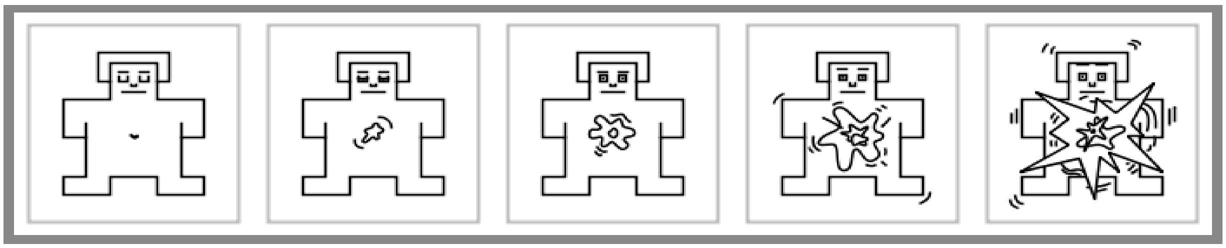

Figure 3: the 5 manikins representing arousal levels from which participants had to choose one (from low to high)

Procedure

Participants were greeted in the Erasmus behavioral lab's waiting room and guided to the experimental room, which consisted of 12 isolated small cubicles with a computer in each of them, and a control room from which the experimenter could observe and control all 12 computers. Participants were told to sit in a cubicle, close the door behind them, sit on the chair and wait for the computer's task to start.

As the experiment was about to start, participants had to read instructions telling them to put headphones on (whether they were in the silence condition or the music condition). Participants had to click on a button when they were ready and the experiment started with a small pretest-questionnaire including age, mother tongue and an auto-evaluated measure of their English reading level. Music was played as soon as they pushed the button to start the test and would not stop until the participant finished to complete the experiment. The musical piece was quite long (21 minutes), and no participants took longer than that, but the song was planned to repeat just in case.

Right after finishing this questionnaire, they had to read the first text and could take as long as they wanted to (time was measured for each text and each series of questions). Text RD was counterbalanced so that in each experimental group, half of the participants started with the easy version of the texts, while the other half started with the hard version of the texts. Note that, even though text difficulty was counterbalanced, the texts themselves were always read in the same order.
Once they were done reading the text, the perceived measures for cognitive load, including perceived effort and difficulty were given. They then had to answer 4 comprehension questions regarding the text previously read without being able to go back to it. Again, they could take as much time as needed.

Right after answering the question, the same measures of cognitive load were given, this time regarding the effect of answering the questions. The same process was repeated for the second, third and last text. The four measures of arousal were taken at four different times during the test, to make sure participants were surprised enough to not answer mindlessly. The first measure was taken right after the pre-test questionnaire, before starting to read the first text. The second and third measure were taken after answering the questions from the second and third text. The last measure of arousal was also taken after answering the questions from the fourth text, but before reporting the perceived cognitive load and intrinsic load from the questions.

After having answered the last text's comprehension questions, participants had to fill out a questionnaire about their usual use of music when studying and use of classical music when studying. There was an item regarding the perceived help of the condition (music or silence) on task performance.

Operational hypotheses

Main hypotheses

H1: Participants will score higher on comprehension scores on average in the low RD condition when compared to the high RD condition, regardless of the presence of music.

H2a: When RD is low, participants listening to background music will obtain better comprehension scores on average than those without any background music but in the same RD condition, according to the arousal-mood hypothesis.

H2b: When RD is high, participants listening to background music will obtain worse comprehension scores on average when compared to those without any background music but in the same RD condition, according to the seductive detail effect.

To account for the arousal-moody hypothesis

H3: Participants' levels of arousal will be higher in the background music condition, when compared to the silence condition, regardless of the RD.

To account for the cognitive load from the background music

H4a: Average cognitive load levels from participants reading the texts will be higher when background music is being played in comparison to silence, regardless of the RD.

H4b: Average cognitive load levels from answering the questions will be higher when background music is being played in comparison to silence, regardless of the RD.

To account for the cognitive load induced by the RD

H5a: Average cognitive load levels from reading the texts will be higher when RD is high, in comparison to low RD, regardless of the presence of background music.

H5b: Average cognitive load levels from answering the questions will be higher when RD is high, in comparison to low RD, regardless of the presence of background music.

Independent and dependent variables studied in this experiment

Independent variable 1: Presence or absence of background music (between participants)

Independent variable 2: Manipulated reading difficulty - low vs high (within participants)

Dependent variable 1: Comprehension score from the texts

Dependent variable 2: Perceived arousal levels throughout the task

Dependent variable 3: Perceived cognitive load from reading the texts

Dependent variable 4: Perceived cognitive load from answering the questions

Dependent variable 5: Time spent on reading text, answering question and on overall task

6. Results

All data from the experiment was collected using *qualtric*s, a well-known online software designed to create and run experiments and collect the data. Data from *qualtrics* was exported into an excel sheet, cleaned up, organized. The questions were recoded manually. From that point on, the data was imported to the software SPSS, which was used for all analysis (statistical and exploratory).

Comprehension scores from the texts

Four comprehension questions were asked for each of the texts to account for the degree of comprehension from the varying reading difficulty (low vs high) and from the presence or absence of the background music.

The texts were grouped two by two, based on their difficulty (low or high), leading to a total of 8 comprehension questions per condition. The score obtained per question could be either 1 if the answer was correct, 0 if incorrect or 0.5 if the question was answered partially. The maximum comprehension score obtainable was therefore 8.

A two-way mixed ANOVA to investigate the impact of background music, text RD and their interaction on comprehension scores was conducted. The main effect of text reading difficulty on comprehension scores was not statistically significant [$F(1,46) = 0.017$, $p = .896$]. Additionally, no significant main effect of background music on comprehension scores was found [$F(1,46) = 0.29$, $p = .593$]. The interaction between background music (present or absent) and the text reading difficulty (low RD vs high RD) on comprehension scores failed to show statistical significance [$F(1,46) = 0.323$, $p = .673$].

These results indicate that neither the varying RD, or the presence or absence of background music nor their interaction with one another could account for a mean comprehension score difference between participants.

Figure 4 below gives a visual representation of the results.

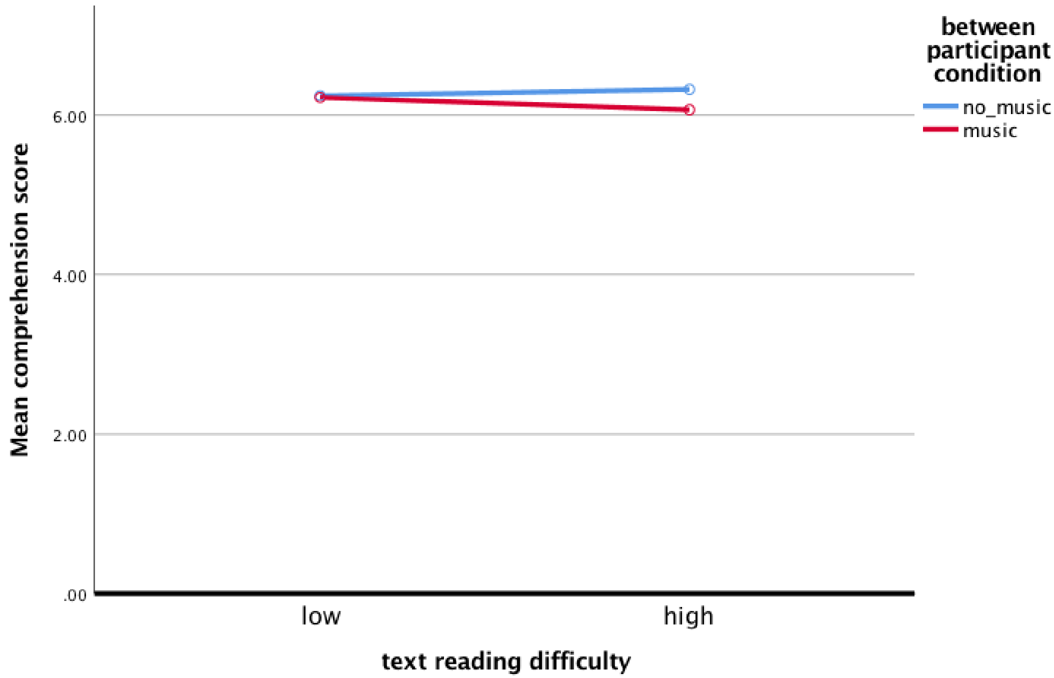

Figure 4: mean comprehension scores in the low and high reading difficulty condition, in interaction with the presence of background music

As we can observe, participants generally scored considerably high on comprehension scores (maximum 8) and had very similar scores regardless of the RD or the presence of background music.
According to our hypotheses, we would expect subjects listening to music (in red) to score higher in the low RD condition, when compared to subjects not listening to any music (in blue) in the low RD condition. This is not the case. We would also expect subjects listening to background music (in red) in the high RD condition to score lower than subjects not listening to music (in blue) but in the same condition. Although this seems to be the case, the difference is not significant. Overall, this figure is very illustrative of a non-significant interaction.

Please report to table 3 for the exact mean scores for each condition

Table 3: Mean comprehension scores for participants in the low and high RD conditions, in interaction with background music.

Reading difficulty	Background music	Mean comprehension scores (maximum of 8)
low	present	6.22 (1.16)
	absent	6.24 (1.27)
high	present	6.06 (1.46)

	absent	6.32 (1.30)

Figure 5 below presents the same result with error bars and histograms, as putting the error bars in figure 4 prevented a good display of the results.

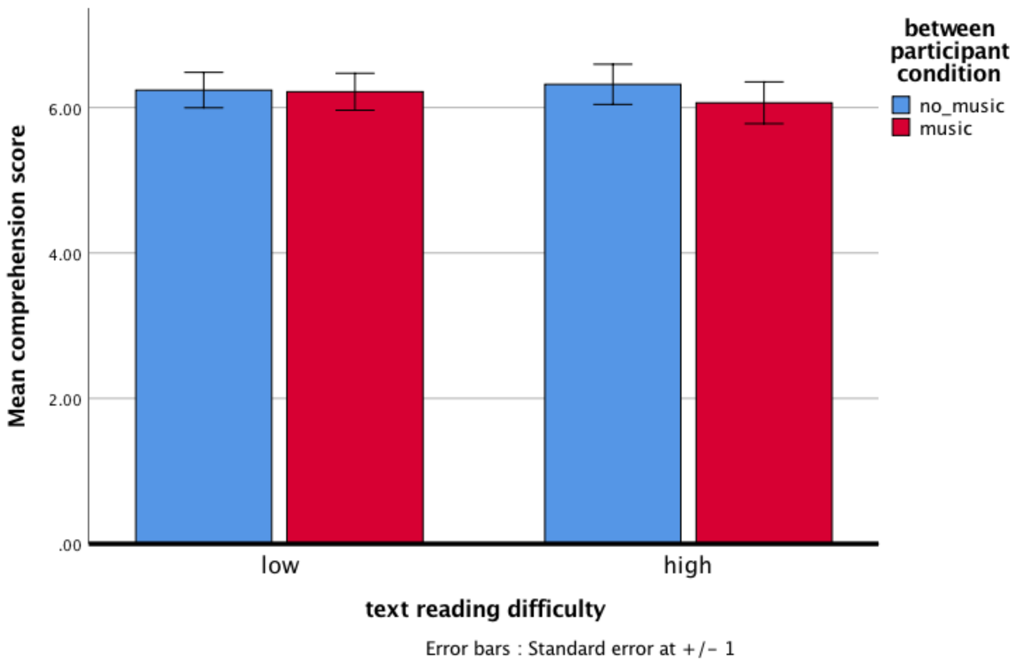

Figure 5: Mean comprehension scores in the low and high RD condition, in interaction with background music.

Manipulation of reading difficulty

Since the reading difficulty factor did not have a significant impact on comprehension scores, and the low and high RD conditions were calculated by adding the texts together two by two, we thought it would be essential to explore whether the manipulation of reading difficulty worked as expected in any of the texts separately, since it did not work when they were grouped. To verify if the reading difficulty manipulation applied on any of the texts, four paired sample t-tests were conducted to compare each of the low and high RD version of the texts.
If the manipulation of reading difficulty worked correctly, we expect the mean comprehension scores from the high RD texts to be lower (worse comprehension) than the ones from the low RD texts.

Table 2 below offers a summary of this exploration.

Table 2: Mean comprehension scores for each text in both high and low RD version, regardless of background music, with standard deviation, t-scores and p-values.

	Average comprehension score				
	low RD (max 4)	*high RD (max 4)*	*Mean per text (max 4)*	*t*	*p-val.*
Text about injury from a bull	2.80 (.77)	3.04 (.95)	2.92 (.87)	-1.00	.32
Text about house-cleaning	2.77 (.88)	2.67 (.80)	2.72 (.84)	0.43	.67
Text about football game	3.25 (.51)	3.08 (.79)	3.17 (.66)	-0.87	.39
Text about birthday party	3.65 (.43)	3.60 (.44)	3.63 (.43)	-0.33	.74

The maximum score attainable per text is 4, since there are 4 comprehension questions per text. A high score means a better comprehension of each text. We expected comprehension scores to be higher in the low RD version than in the high RD version, because, regardless of the presence of music, they should have been easier to understand. If this seems to be the case for the last three texts, the difference remains not statistically significant: as we can observe from the table, the *t-scores* are quite low and the p-value is over 0.05 (ranging from .32 to .74). Note that the last two t-scores are inverted because of the counterbalance in reading difficulty

A low t-test score and a high p-value implies this: the hypothesis stating that subjects performed generally better in the low RD condition when compared to the high RD condition, due to the RD manipulation cannot be confirmed for any of the texts (one to four). Even more surprisingly, the direction of the t-test for the first text is opposed to what we expected. Indeed, the high RD version of this text was answered with more ease (M: 3.04) than the low RD version (M=2.80).

Overall, contrary to what we expected, these results imply that the manipulation of the reading difficulty was not statistically significant. We will give various hypotheses to account for this unexpected finding in the discussion

Arousal levels and cognitive load

Arousal

Four measures of arousal were taken during the experiment to account for the effect of background music on arousal levels while music was being played throughout the task. We expected arousal levels to be higher when music was played when compared to when no music was being played. It was chosen not to average out these four measures, because we believe it is important to have arousal level measures at different times during the musical excerpt. This was done in case various parts of the musical excerpt would have elicited different levels of arousal.

Four independent t-tests were conducted to account for the possible effect of background music on arousal levels. The four independent t-tests failed to show significance (t (48) = 0.22 p = 0.83); (t (48) = -0.62 p = 0.54) ;(t (48) = -1.01 p= 0.32); (t (48) = -1.08 p = 0.29). Figure 5 below gives a visual representation of these reports.

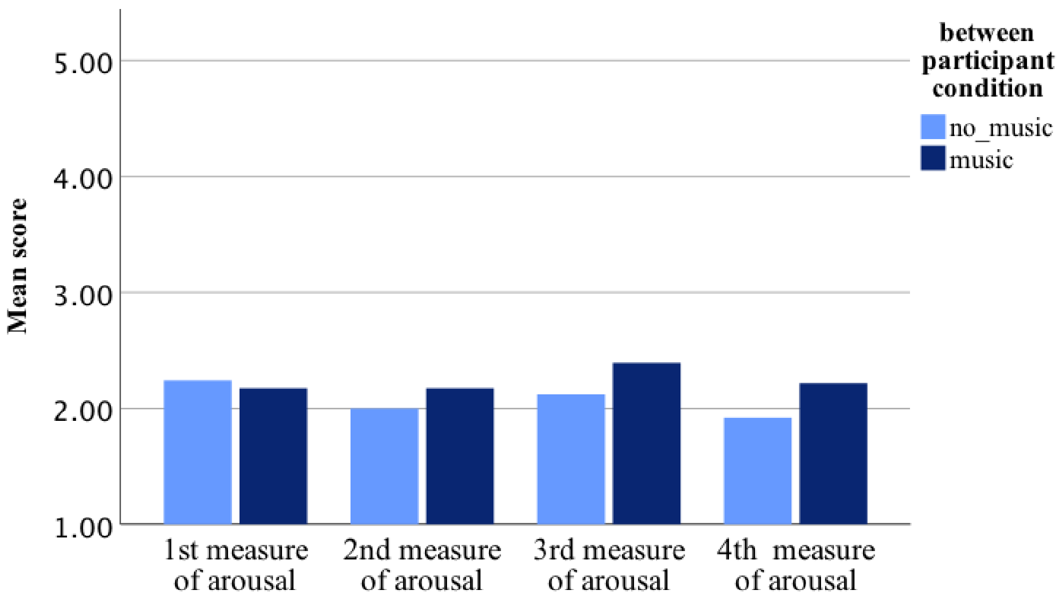

Figure 5: Average levels of arousal between participants listening to background music or silence as the learning experience last

As we can observe from the figure, fairly low arousal levels for each measure were reported, since the maximum is 5 and the average levels are around 2. We can observe a very slight increase (not statistically significant) in the arousal levels difference between with and without music as the experiment goes. Indeed, as the arousal levels diminish in the silence condition, the arousal levels increase in the music condition (at least up to the third measure). This marginal effect suggests that as time goes, differences in arousal between the silence condition and the music condition increase, as table 5 shows in more detail.

Table 5: Average arousal levels reported by participants performing with or without background music (with t-scores and p-values).

Arousal measures	*average arousal levels (max 5)*				
	without music	*with music*	*difference*	*t-score*	*sig.*
First measure	2.24 (1.01)	2.17 (1.07)	0.07	0.22	.83
Second measure	2.00 (0.87)	2.17 (1.07)	0.17	-0.62	.54
Third measure	2.12 (0.88)	2.40 (0.98)	0.28	-1.01	.32
Fourth measure	1.92 (0.91)	2.22 (1.00)	0.30	-1.08	.29

Note that, in line with this increase in difference, the p-values decreases as well, meaning that more time might have potentially lead to a significant difference between the two conditions, if the trend was followed. This interpretation is still highly speculative and the overall conclusion is that our measure of arousal could not account for any difference between participants listening to background music and those who did not. Reasons for this unexpected result will be given in the discussion section.

Cognitive load: perceived difficulty and perceived mental effort

The measure of cognitive load contained two items (effort and difficulty). 8 measures of perceived effort were taken, along with 8 measures of perceived difficulty. One after reading each text, one after answering each set of questions, leading to 16 measures in total.
The measures were grouped based on the reading difficulty of the texts. One was named, easy_cognitive_load and grouped the measures from reading the low RD texts. The other was named hard_cognitive_load, which regrouped the measures from reading the high RD texts. The same procedure was repeated for the measures taken after answering the questions.

Four two-way mixed ANOVA were conducted to investigate the impact of background music, text reading difficulty and their interaction on cognitive load from reading the texts and answering the comprehension questions. As the amount of statistical analysis was quite important for this section, we chose to report all of them in the two tables below.

Table 6: p-values for the multiple two-ways mixed ANOVAs conducted on the impact of text difficulty and background music on cognitive load measures of perceived effort and difficulty

Cognitive load		Main effect of text RD	Main effect of music	Interaction between music and text RD
Perceived effort	from reading texts	p =.053	p =.50	p =.91
	from answering questions	p =.173	p=.171	p =.35
Perceived difficulty	from reading texts	p =.076	p =.123	p =.77
	From answering questions	p =.95	p=.102	p =.826

The p-values that are close to being significant are marked in green, the others are in black.

We can observe from this table that none of the effects were significant from the two way mixed ANOVAs. However, there is a marginal effect of the text RD on the measures of perceived effort and difficulty from reading the texts, meaning that there was a marginal difference in perceived difficulty and effort between reading the low and high RD texts, regardless of the music being played. Please report to table 7 below for the directions from the differences

Table 7: Cognitive load measures (perceived effort and difficulty) from reading high and low RD texts and answering the questions, in interaction with the presence of background music

				Background music	
				absent	present
Cognitive	Perceived effort	From reading texts	low RD texts	3.70 (1.37)	3.96 (1.31)
			high RD texts	4.04 (1.35)	4.26 (1.27)
		From answering questions	low RD texts	3.98 (1.18)	4.30 (1.19)
			high RD texts	4.04 (1.25)	4.63 (1.32)

load	*Perceived difficulty*	*From reading texts*	*low RD texts*	3.02 (1.20)	3.41 (0.94)
			high RD texts	3.24 (1.14)	3.72 (1.00)
		From answering questions	*low RD texts*	3.89 (1.16)	4.39 (1.20)
			high RD texts	3.84 (1.25)	4.41 (1.29)

Marked in green and red are the two marginal differences mentioned before

We can see that the marginal difference found between the low RD texts and the High RD text is in the expected direction. Indeed, the high RD texts were marginally perceived as more difficult and requiring more effort than the low RD texts, regardless of the background music being played or not. Note that although, these effects are very close to significance (p= .053 and p= .076), they cannot be interpreted as having an effect

Overall, the two way mixed ANOVAs failed to show significance from the main effects of music and reading difficulty, as well as, the interaction between them on cognitive load measures of perceived effort and difficulty.

Time spent

Various measures of time were taken, mainly to control for any potential advantages from taking more time reading the texts or answering the questions. Another reason for measures of time was to account for any effects of background music or text reading difficulty on time taken to read the texts, answer the questions, or complete the task.

Time spent on reading the texts

A two-way mixed ANOVA to investigate the impact of background music and text reading difficulty and their interaction on time taken to read the texts was conducted. There was no significant main effect of the text RD (low or high) on time spent reading the texts [$F(1,46) = 1.715$, $p = .196$] and no significant effect from the interaction between background music (present or absent) and the text difficulty (low RD vs high RD) on time taken to read texts was found [$F(1,46) = 0.482$, $p = .491$].

However, there was a significant main effect of background music (present or absent) on time taken to read the texts. When subjects were listening to background music, they read the texts faster, regardless of the reading difficulty of the texts. [$F(1,46) = 6.718$, $p = .013$]. The power observed was of 0.782 which is considered as a medium power. Below is a figure of the results

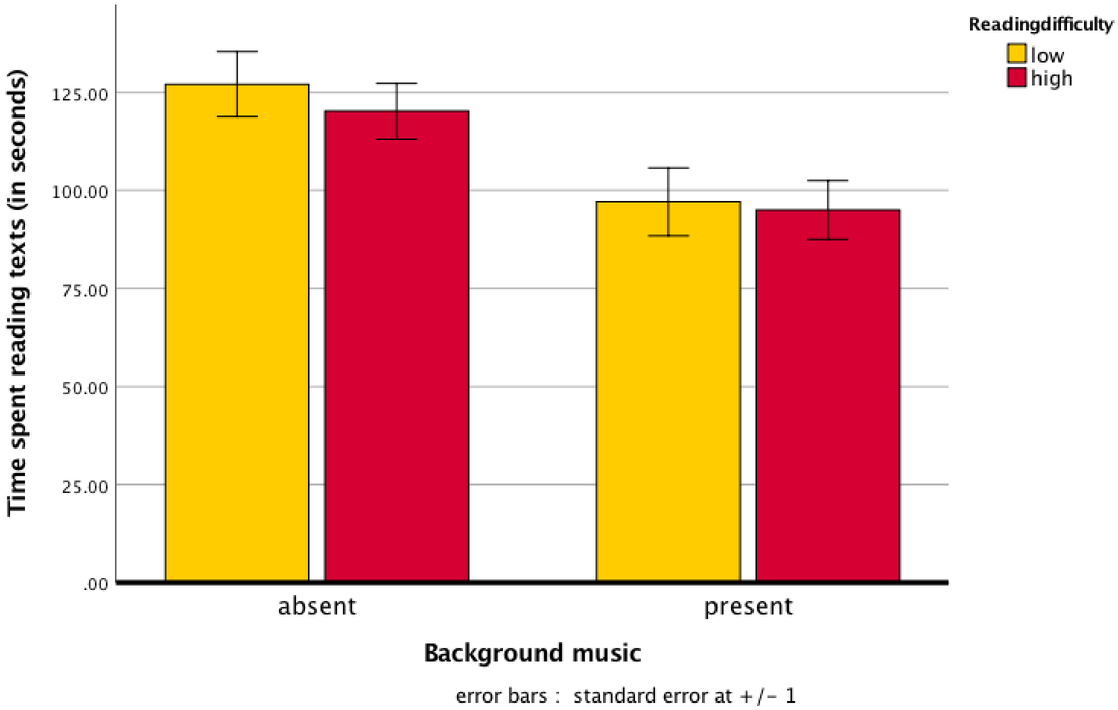

Figure 6: time difference when reading texts (in seconds) from subjects in the silence condition and in the music condition in interaction with reading difficulty.

As mentioned previously, the presence of background music, regardless of reading difficulty has a significant effect on time taken to read the texts. Participants read texts faster when listening to background music.

Time spent on answering the questions

Another two-way mixed ANOVA to investigate the impact of background music and text reading difficulty and their interaction on time taken to answer the questions was conducted.

There was no significant main effect of the text RD (low or high) on time spent answering questions [$F(1,46) = 0.818$, $p = .370$] and there was no significant effect for the interaction between background music (present or absent) and the text difficulty (low RD vs high RD) on time taken to answer the questions [$F(1,46) = 0.00$ 0.482, $p = .995$].

However, a significant main effect of background music (present or absent) on time taken to answer questions was found [$F(1,46) = 17.248$, $p < .001$]. When subjects were listening to background music, they answered the question significantly faster than subjects who were not listening to any music. The power observed was of 0.982 which is considered as a very high power.

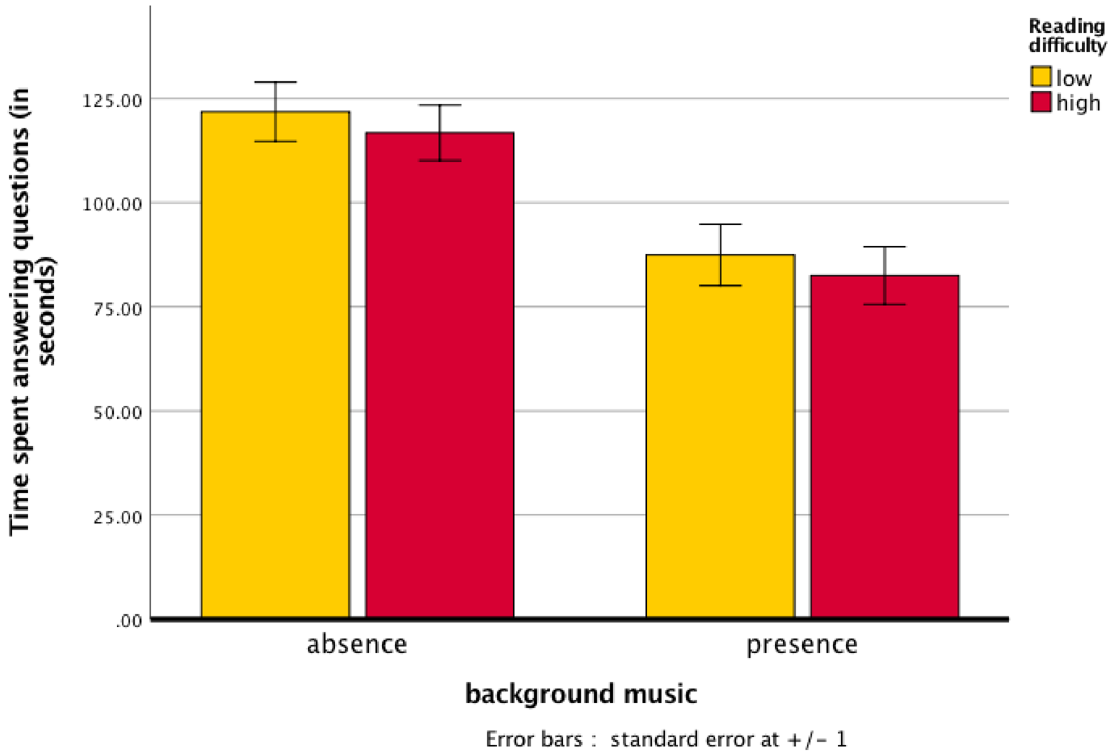

Figure 7: time difference when answering questions (in seconds) from subjects in the silence condition and in the music condition in interaction with reading difficulty.

As mentioned previously, the presence of background music, regardless of reading difficulty has a significant effect on the time taken to answer the questions, making them answer faster.

Time spent on overall task

When conducting an independent sample t-test to compare means for time taken between the no background music condition and the background music condition, the overall time spent on the task was significantly shorter for subjects who listened to background music while learning, when compared to those who learned in silence. t (46)= 3.04 p = 0.004. The results remained significant even when a subject that was considered as an outsider because of the time taken to complete the task was removed. Figure 8 shows the difference between conditions graphically.

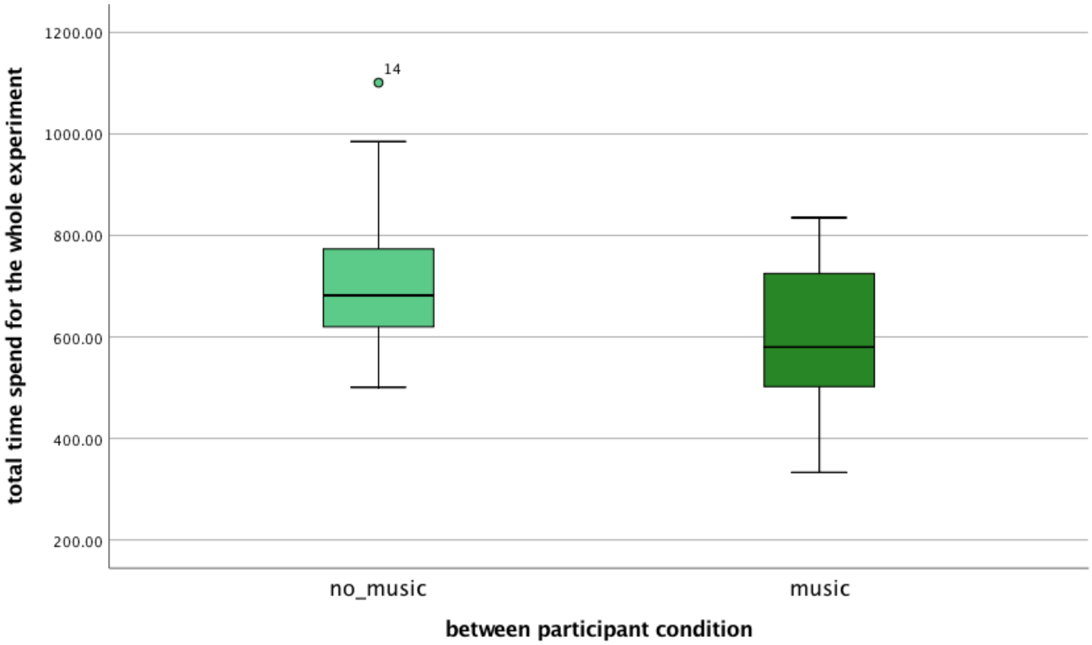

Figure 8: overall time difference for task completion in seconds from subjects in the silence condition and in the music condition.

Overall, over the three different time measurements, background music had a main effect on how fast participants performed, whether it was when reading [$F(1,46) = 6.718$, $p =.013$], answering the questions [$F(1,46) = 17.248$, $p < .001$] , or completing the task $t\ (46)= 3.04$ $p = 0.004$. This effect happened without any significant decrease in comprehension. This unexpected effect will be discussed furthermore in the following chapter.

7. Discussion

The aim of this thesis was to investigate how listening to background music affected text comprehension, in interaction with reading difficulty. More precisely, our experimental design aimed to assess both positive and negative effects from listening to background music on comprehension scores.
Indeed, when the task was easy enough in terms of cognitive resources (low reading difficulty), music was thought to enhance mood and arousal levels – facilitating learning outcomes. In contrast, when the task was more cognitively demanding (high reading difficulty), music was expected to add to the cognitive burden, resulting in worse comprehension scores.

The texts from which subjects learned from were modified to create a low and high reading difficulty version, expected to provoke various degrees of cognitive load in interaction with the presence or absence of background music.
More precisely, we expected an increase from both the arousal and the cognitive load from listening to background music in the two RD conditions. With the difference that, in the low RD, the positive effects of arousal and mood on learning would be more prevalent than the negative burden from the cognitive overload, resulting in better comprehension scores. In opposition, when the RD was high, the cognitive burden from listening to music should be stronger than the positive effect of arousal, resulting in worse comprehension scores.

Here are the exact hypotheses as they were previously stated:

Main hypotheses

- H1: Participants will score higher on comprehension scores on average in the low RD condition when compared to the high RD condition, regardless of the presence of music.

- H2a: When RD is low, participants listening to background music will obtain better comprehension scores on average than those without any background music but in the same RD condition, according to the arousal-mood hypothesis.

- H2b: When RD is high, participants listening to background music will obtain worse comprehension scores on average when compared to those without any background music but in the same RD condition, according to the seductive detail effect.

To account for the arousal-moody hypothesis

- H3: Participants' levels of arousal will be higher in the background music condition, when compared to the silence condition, regardless of the RD.

To account for the cognitive load from the background music

- H4a: Average cognitive load levels from participants reading the texts will be higher when background music is being played in comparison to silence, regardless of the RD.

- H4b: Average cognitive load levels from answering the questions will be higher when background music is being played in comparison to silence, regardless of the RD.

To account for the cognitive load induced by the RD

- H5a: Average cognitive load levels from reading the texts will be higher when RD is high, in comparison to low RD, regardless of the presence of background music.

- H5b: Average cognitive load levels from answering the questions will be higher when RD is high, in comparison to low RD, regardless of the presence of background music.

Difference in comprehension scores

A two-way mixed ANOVA to investigate the impact of background music, text reading difficulty and their interaction on comprehension scores was conducted. Contrarily to what we expected, it showed no significant main effect of text reading difficulty on comprehension scores, no significant main effect of background music on comprehension scores and no significant effect for the interaction between background music (present or absent) and the text reading difficulty (low vs high difficulty) on comprehension scores (please report to the result section to see the exact numbers).

The biggest mistake a researcher can make when finding non-significant results is to conclude that, if there is little or no evidence that the alternative hypothesis is true, then the null hypothesis must be true. Indeed: when the results are not statistically significant, it means that the data provide little or no evidence that the null hypothesis is false. However, it does not provide evidence that the null hypothesis is true.

In our specific case, the mistake would be concluding that background music, reading difficulty and their interaction have no effect on comprehension scores. This is not accurate. What would be correct is to conclude that there is no credible evidence that background music, reading difficulty and their interaction have an effect on comprehension scores, but that it's however not complete proof that they do not have any effect at all.

While keeping this in mind, our result seems in line with an earlier reading comprehension study by Freeburne & Fleischer, (1952). in which they found that background music did not seem to affect reading comprehesion.

However, our finding differs from previous literature which found that music can have positive effects on cognition (i.e., Husain et al., 2002; Thompson et al., 2001 Schmidt and Trainor, 2001) or that background music affects cognition negatively (e.g., Cassidy & MacDonald, 2007); (Furnham & Bradley, 1997); (Furnham & Strbac, 2002); (Reynolds, McClelland, & Furnham, 2014); (Lehmann & Seufert, 2017).
Our result is therefore not in line with the seductive details effect, which states that adding irrelevant details to learning (such as a background music), should lead to poorer comprehension.

It is as well not in line with the arousal-mood hypothesis, which states that listening to background music should improve learning (through enhance in mood and arousal).

Our results which did not confirm our main hypotheses led us to analyze our data in more depth. We found an explanation by taking a deeper look at the manipulation of reading difficulty, which leads us to understand something important about our chosen material. The next part is dedicated to this analysis.

Manipulation of reading difficulty

Overall, contrary to what we expected, the manipulation of the reading difficulty was not statistically significant. We will give various hypotheses to account for this unexpected finding.

Hypothesis 1: The manipulation of reading difficulty was not sufficiently robust to create a difference in reading difficulty.

This hypothesis can be discarded by taking a deeper look at the comprehension scores. Indeed, subjects scored generally very high on comprehension ($M= 3.12$ $SD= .65$) for the low RD versions and ($M= 3.10$ $SD= 0.75$) for the high RD versions; maximum score attainable: 4). Knowing that the maximum score attainable is 4, these results suggest an overall ceiling effect, regardless of the manipulation of reading difficulty, for each text.
The problem in this case has to do with the overall easiness of the material, more than the manipulation conducted. With a base material that is not challenging enough, any manipulation would be hard to show significance. This leads us to our second hypothesis.

Hypothesis 2: The learning material used in the task was not challenging enough for the subjects, therefore not providing enough variability in participants' answer to account for differences between the conditions of varying reading difficulty.

This hypothesis is highly plausible, due to the fact that our data seems to assume a ceiling effect, due to the high scores from participants (please refer to table 2 provided in results to see participants comprehension scores in more details). This changes our focus from the reading difficulty manipulation to the material itself. If this hypothesis is true, then this means the material presented was not challenging enough. Why was that the case? Could not this have been prevented by pre-testing the material's difficulty?

In our experiment, the material was indeed pre-tested in a pilot study on five subjects, in order to account for any issues. Although the pilot study was useful to some extent, the exploratory results from these five participants did not show any general tendency towards a ceiling effect. Some subjects even reported that the material was too challenging when asked what they thought. Reasons for this occurrence might have been that five participants were not enough to find a general trend, or that these five subjects differed slightly from the subjects who participated in the study. This leads us to a third hypothesis explaining why participants performed with such ease.

Hypothesis 3: Participants performing in the experiment were better than the average population. This hypothesis is backed-up by the fact that, out of all participants from our study, 91% of the participants evaluated their English reading competency between C1 and C2 levels.

Although we did specify that a requirement for participating to the study was to be able to read English, we did not expect such a proficient level of English from participants, knowing that subjects were mainly first year psychology students from the university of Erasmus Rotterdam. This proficient level of reading competency might have been one of the reason the experiment was done which such ease.

Overall, we still believe that the most probable hypothesis for the lack of significance from the main effect of text reading difficulty on comprehension scores is in fact really due to the overall easiness of the base material presented, regardless of the manipulation of difficulty. This affected the statistical significance of other variables manipulated in our experiment, such as reading difficulty, cognitive load and background music, as they all depended to a certain degree on the material. Although this was hard to avoid, recommendations for future studies that aim to manipulate text reading difficulty will be given later on.

Difference in arousal levels

Contrarily to what we expected, our findings did not show a significant difference in arousal levels from subjects who listened to background music when compared to subjects who learned in silence. This result does not support the arousal-mood hypothesis, which states that listening to music increases arousal and mood, which in turns lead to better learning performance.

This is contradictory to the results from previous studies (e.g., Nantais and Schellenberg, 1999; Sloboda and Juslin, 2001; Husain et al., 2002; Pelletier, 2004) in which music was found to influence arousal levels and mood.

However, this dissimilarity might be best explained by a variation in experimental designs. In all of the previous studies mentioned, listening to music was done previously to the task, as an induction. Indeed, subjects had to either be listening to music for about 10 minutes on average, or be sitting in silence for the same time before the task started. Usually subjects were told to pay attention to the music in a way or the other, while subjects in silence were either told to relax, not told anything, or it was not mentioned by the article.

A critique to this approach might be that the difference in arousal between the two conditions (music vs no-music) might have to do with the negative effects of sitting in silence before starting a task for one group, on top of experiencing the benefits from listening to music previously to the task for the other group. It might thus be a combination of both the negative effects of sitting in silence and the positive effects of listening to music previously to the task.

Another important issue from these studies, is that, if background music is played simultaneously to the task such as in Thompson et al. (2001) and Schmidt & Trainor, (2001), the authors do not report any measures of arousal or mood. This means that any potential effect of music on arousal and mood is only hypothetical, as there are not any measures to prove the presence of the effects.

More surprisingly, our study, along with Lehman & Seufert (2017), might be the first one to report arousal as music is played concurrently to the comprehension task. Our result is in line with Lehman et al. (2017) in which they did not find arousal to differ significantly between the background music vs silence condition, even though our task included a Mozart sonata known to enhance arousal levels and mood from the previous studies mentioned.

While acknowledging the limitations from our study, this might imply that the effect of music on arousal can only take place when music is listened previously to the task, as an induction. Indeed, studies who chose to play music concurrently to the task have either simply not taken measures of arousal into account (Thompson et al., 2001; Schmidt & Trainor 2001) or failed to show a significant difference in arousal levels (Lehman et al, 2017; Adam, 2019).

A very recent study suggests that the arousal-mood hypothesis can only take place when music is listened previously to the task (Gonzalez & Aiello, 2019). More research needs to be done to explore whether the arousal-mood hypothesis can take place even when concurrent to the task. As of now, it does not seem to be the case (Lehman et al, 2017; Adam, 2019). However, limitations from our study, including how arousal was measured, will be discussed later on.

Perceived cognitive load

We expected both the change in reading difficulty and the process of listening to music to increase cognitive load in subjects. Our statistical analysis was not able to show a significant difference on the perceived cognitive load of students who learned with music during the task, in comparison to those who learned in silence. However, we reported a marginal difference from reading difficulty on measures of perceived effort and perceived difficulty.

This suggests that even though participants did not perform better on comprehension scores, there might have been a difference in how they perceived the reading difficulty (low vs high). However, since these results only show marginal effects, more participants would be needed to account for proof that reading difficulty indeed increased participants cognitive load.

This result is therefore not in line with the seductive details effect, which states that adding irrelevant details to learning (such as a background music), should lead to increase cognitive load and lead to poorer comprehension (Harp & Mayer, 1998). Indeed, for instance, Lehman et al. (2017), found learners with higher working memory capacity to learn better with background music than participants with lower working memory. Whereas, learners with lower working memory capacity reached higher comprehension scores without background music. Their finding implies that background music can indeed affect cognitive load, because learners with higher working memory had less trouble understanding the material than the ones with low working memory.

In opposition, our finding concludes that it is very likely in our experiment that background music did not have an effect on perceived cognitive load. Speculations for this effect include a lack of difficulty from the material, implying it would leave enough cognitive resources from subjects to perform the task while listening to the music, or a high working memory from subjects, resulting in the same effect.

A combination of both of these effects is also possible, but to this day we cautiously advise the reader on their speculative nature, as they would need to be explored in other experiments, as for example, with Lehman & Seufert (2017).

Time taken to complete the task

To our surprise, we found out a significant difference of time spent on questions, text read, and overall task, based on whether subjects were in the background music or the silence condition. Indeed, subjects listening to background music were on average faster to read the texts, to answer the questions, and to finish the overall task. Moreover, we can state that, even though subjects were faster, they did not perform significantly worse on comprehension scores (neither better). This suggests that listening to music stimulated participants to perform in a faster way.

To our knowledge, this is the first report of an effect of speed of performance induced by listening to background music in a text comprehension task. Indeed, music has been shown to influence the speed of various behaviors in other settings. For instance, background music has been shown to influence the speed of eating and drinking depending on the speed of the music being listened to (McElrea & Standing, 1992). Background music has also been shown to affect the time to complete a dinner in restaurant, with faster music making people stay less when compared to slow music (Milliman, 1986).

In our study, we speculate that, similarly to McElrea et al. (1992) and Mimilan (1986), the effect of tempo from the music had an effect on the speed to which participants performed. A less likely explanation include subjective feelings that lead participants to perform faster because they felt trapped or bothered by listening to the music and wanted to complete the task as fast as possible in order to get out of the unpleasant situation.

Either way, future research should explore this effect of speed performance from listening to background music, as to our knowledge, it has never been observed in a learning situation.

Limitations and recommendations for future studies

Measure of Arousal (and Mood?)

From the title, one can guess our first limitation: we did not take mood into account. One of the reason for this choice is that we did not want to add more cognitive weight to the participants while performing the task, as we found it important to keep subjects as free as possible from any confounding variables external to the learning material. This is one of the reason we chose such a simple measure for arousal, to reduce distraction from the learning task as much as possible. Moreover, we thought a measure of arousal would be sufficient to account for the effects from the arousal-mood hypothesis. The measurement tool for arousal was the subscale of the Self-Assessment Manikin (SAM), created by Bradley & Lang (1994).

As our result did not show any significant difference between participants who performed while listening to music and those who performed in silence, we hypothesize that this subscale might not have been sensitive enough to account for changes in arousal. Indeed, it consisted of a five Likert scale. Participants needed to answer to a question regarding their arousal state by selecting one of the 5 manikins with a certain level of arousal indicated by a bigger or smaller explosion in its belly.

The results showed on average that participants only used the first three manikins, and very rarely the last two. This suggests that a wider scale would be needed to account for smaller changes in arousal. Although we chose the SAM because we needed a concise item to have four reports of arousal levels from participants during the task without disturbing learning, we recommend researchers to use a more complex, more adapted measure, as both our study and Lehman et al. (2017) did not find a significant difference in arousal levels, while using this subscale.

A different type of measure could be for example the revised version of The Profile of Mood States (POMS), in its short form (McNair, Lorr, & Droppleman, 1992). The POMS in its short form is still very heavy to assign because it contains more than 30 items, each requiring participant to rate (on a 5-point scale) their agreement that a particular adjective describes their affective state at the present time. With this sort of form, taking four measures of arousal during the experiment would not be realistic. We therefore advise researchers to find a balance between the number of measures of arousal needed and the complexity of the form given, in order to have a measure sufficiently sensitive, without overwhelming the subjects.

A possible alternative approach would be to record physiological measures of arousal instead of subjective measures. However, we warn researchers that such measures are known to be very invasive, thus possibly harming the authenticity of the learning experience and making it difficult to relate it to a real-life learning situation.

One last limitation from our experiment in relation to arousal was that there was not any measure of it before the music started, therefore not providing a baseline for arousal, except for the baseline from participants who learned without listening to music. This was rendered difficult to implement, as the button which participants pushed to start the task started the pre-test and automatically played the music.

Limitation from the musical excerpt

One critique which can be addressed to our experiment is that we imposed a style of music to participants that is not necessarily the most favourite amongst students. Indeed, a study by Kotsopoulou & Hallam, (2010) showed classical music to be the least favourite style of music played during studying, across all nationalities and age groups. The most favourite style was recorded pop music.

In our experiment, we included a questionnaire at the end of our task to evaluate our participants' habits regarding the use of music and classical music when studying. Here is what we found:

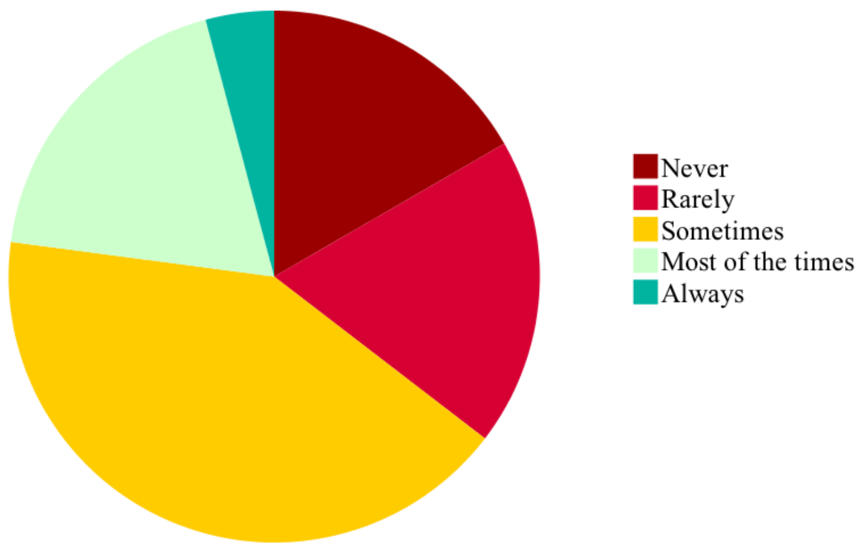

Figure 9: Usual use of music when studying (in %) :

Participants mostly answered that they sometimes listened to music while studying (41.7%), followed by rarely (18.8%), most of the times (18.8%), never (16.7%) and finally always (4.2%).

Therefore, out of the 48 participants to the study, the majority answered between sometimes and most of the times (60.5%). This allows us to know that participants from our experiment seem to use music while studying in some occasions.

The next question was about the use of classical music when studying. The 8 participants that had answered "never" in the previous question were excluded. 40 participants remained.

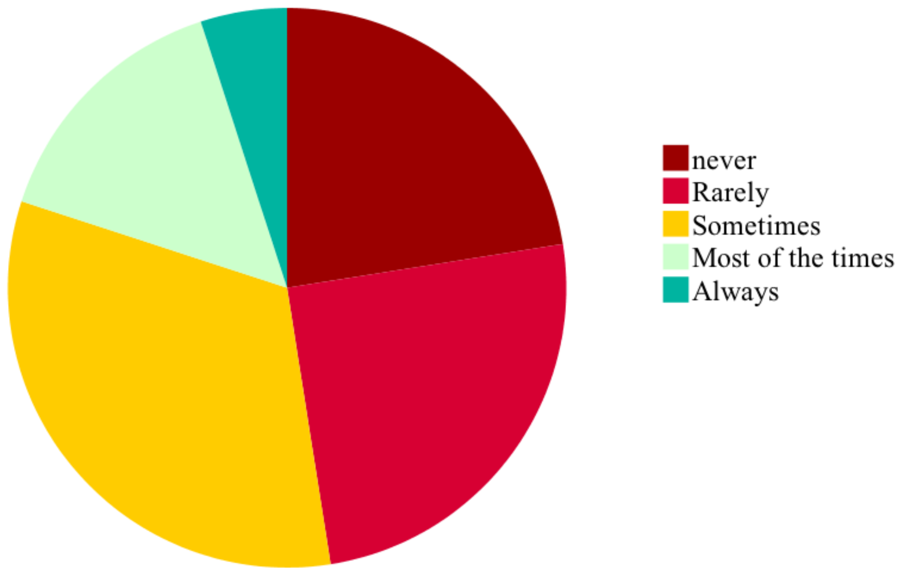

Figure 10: use of classical music when studying. (in %)

32.5 % of the participants answered that they sometimes studied with classical music, 25 % said rarely, 22.5% never, 15% most of the times, and 5% of the students said always.

Again, sometimes is the most common answer (32.5%). If we regroup the two lowest and the two highest answers of the scale, 20% answered most of the times or always, while 47.5 % answered never or rarely. In this case we can observe that participants seemed less likely to use classical music when studying. This might account for a reason to why the Mozart sonata from our experiment did not seem to induce higher arousal, as most participants were not necessarily used to studying with classical music and some of them might possibly not have liked the music.

However, it is still under debate whether preferred music enhances or hinders performance. Some research has shown that disliked music and liked music impaired reading comprehension similarly (Perham & Currie, 2015), others showed disliked music to even be potentially better for performance than liked music (Perham & Sykora, 2012).

Nonetheless, using a music that had been previously tested in multiple experimental settings and had proven to show effects on arousal and mood in a student's population was primordial to us. This would unfortunately not have been the case with pop-songs, as past research has been less consistent with the choice of songs used in experimental designs, making it difficult to find the right suit for our population. A different experimental design could be to ask participants what style of music they usually prefer to use when studying, and to base the choice accordingly to the majority. This is tricky as preferences in music varies greatly.

Future research should include a pre-test of the Mozart's Sonata for two pianos in D Major, K. 448, used in our experiment, as it seems there is a distinction in the effect it has on arousal and mood depending on whether it is played concurrently to the task or previous to the task. As of now, when played previously to the task, it has been proven to increase arousal and mood. But when play concurrently to the task, the sonata does not seem to increase arousal (Adam, 2019; Lehman & Seufert, 2017).

Learning material limitations and how to manipulate it

Earlier on we found out that the material given resulted in an overall ceiling effect. The material had been previously used and tested by two other studies in the field of text comprehension, to assess the ability to understand a text by generalizing inferences (global and local).

As mentioned earlier in the document, text comprehension depends on a great level on the ability to fill in missing information which is not explicitly stated in the text. In other words, comprehension depends a great deal on the ability to generate inferences, which is why this material was chosen. However, our results provided us with evidence that the material was not challenging enough for participants. For this reason, it seems essential to provide recommendations for future studies that aim to manipulate text reading difficulty.

In general, this might have been better prevented by pre-testing the material on more subjects (10 should be sufficient- 5 were used in our experiment). This would have hopefully allowed us to spot on the ceiling effect earlier on with more ease. Even though this is a not very cost-effective method because of how much time it requires, we encourage future researchers not only to pre-test the material before the manipulations of reading difficulty, but also after the manipulations, to verify their effect. This could be done by comparing comprehension scores from participants between the two conditions being manipulated.

A less-time consuming method would be to use a readability formula to assess the overall ease of comprehension from the two conditions and verify that the two scores differ sufficiently. Although this is cost-effective, we do not recommend to solely us a readability formula, as they usually base their results on a small number of parameters (i.e. length of words, sentences and paragraphs) which can lead to biased evaluations of difficulty. We thus also recommend to test the material directly on participants.

The most adequate procedure would be to combine the two previously mentioned methods. First, one should use a readability formula to obtain a general measure of difficulty. Then, the material should be tested on participants to assess the level of difficulty (with the use of comprehension questions and participants' opinion). Finally, one should make the desired modifications to reading difficulty and create the required numbers of conditions - in our case two: one low RD, one high RD. This should be followed by a new assessment from a readability formula for each text, along with another pre-test of the now modified material.

Regarding the reading difficulty manipulation itself, we recommend researchers to make modifications on specific characteristics, such as the ones mentioned in our article for example (Graesser, McNamara, & Kulikowich, 2011). However, we need to warn that manipulating difficulty is a hard task that requires a careful balance between manipulating enough and manipulating too much.

For instance, if one makes too many modifications between the two conditions, one takes the risk of creating two texts that differ in the content, rendering the comprehension questions easier or harder to answer, not because of the manipulation in reading difficulty, but because of difference in the amount of information one text or the other provides.

Indeed, one needs to manipulate the reading difficulty, without changing the meaning of the modified texts, because the comparison between the texts need to be on the factor of reading difficulty and not on the amount or the clarity of content from the differing texts.

Note that, we could have chosen to manipulate cognitive load by adding a cognitive burden from a stimulus external to the text, such as for instance a counting backwards task. However, our will to design a close-to–real-life-setting prevented us to use external and unnatural stimuli, as one would not be counting backwards when in a realistic learning situation. For this reason, we chosen to keep our focus on a parameter that seem more natural: the text reading difficulty.

The long and arduous process described before to control the material and the manipulation of reading difficulty would probably prevent subjects from scoring too high on comprehension scores. Understanding that this is not a process any researcher can conduct, a simpler solution, but probably less effective, could be to put all comprehension questions at the end of the material, instead of subsequently to each text.

This would in turn artificially create a more challenging material, because subjects would have to remember each of the texts to be able to answer the comprehension questions up to the first one which would have been read minutes before answering the comprehension questions at the end.

This simpler solution does not require any direct modifications from the material itself, but still demands a careful counter-balance from the order of the text's reading difficulty. Indeed, one would need to make sure that the difference in comprehension scores comes from the text reading difficulty and not because of how recent the text was read: the last texts read being remembered more easily than the first ones, leading to a difference in comprehension scores depending on how recently they were read.

8. Conclusion

The aim of this thesis was to investigate how listening to background music affected text comprehension in the light of the arousal-mood hypothesis and the seductive detail effect. More precisely, our experimental design aimed to assess both positive and negative effects from listening to background music on comprehension scores by manipulating reading difficulty.

Comprehension scores were thought to be affected by the varying reading difficulty of the task. When the task was easy enough in terms of cognitive resources (low reading difficulty), music was thought to enhance mood and arousal levels – facilitating learning outcomes. In contrast, when the task was more cognitively demanding (high reading difficulty), music was expected to add to the cognitive burden, resulting in worse comprehension scores.

In order to verify our hypotheses, an experiment on 48 subjects was conducted. Subjects had to read texts of varying reading difficulty, either in silence (23 subjects), or while being exposed to background music (25 subjects). Reports of comprehension scores, cognitive load, arousal levels, and time spent on task were taken throughout the experiment for each condition.

The main results showed that contrarily to our main hypotheses, background music, reading difficulty, and their interaction with one another did not account for the difference in comprehension scores. Regarding cognitive load, nor background music, reading difficulty or their interaction accounted for a difference in perceived effort or difficulty when reading the texts or answering the comprehension questions. Finally, while background music did not account for a difference in arousal levels, it did show an effect on the speed with which participants performed the task, with background music leading to faster reading, answering the questions, and faster task completion.

Overall, our findings are inconsistent with the view that background music influenced comprehension either negatively or positively, through the seductive detail effect or the arousal-mood hypothesis.

However, we believe that our inconsistent results came from the overall easiness of the material used in the experiment, which is why we refer the reader to our previous chapter for recommendations in choosing and designing a better material, while warning that reading difficulty is challenging to manipulate. We therefore require more researches do be done regarding the interaction between background music and reading difficulty be able to either replicate or discard our findings, as to our knowledge we are the first research that aimed to manipulate both in one experimental design.

Moreover, while acknowledging the limits of our findings, we suggest that further research needs to be done on the impact of music on arousal and mood, to pinpoint whether the arousal mood hypothesis can take place when concurrent to the task.

As for now, the effect of music on arousal and mood seem to be only significant when played previously to the task. Whereas studies who chose to play music concurrently to the task have either simply not taken measures of arousal into account (Thompson et al., 2001; Schmidt &

Trainor 2001) or failed to show a significant difference in arousal levels from listening to the background music (Lehman et al, 2017; Adam, 2019).

9. Bibliographic references

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Psychology of learning and motivation (Vol. 2, pp. 89-195). Academic Press

Atkinson, G., Wilson, D., & Eubank, M. (2004). Effects of music on work-rate distribution during a cycling time trial. *International Journal of Sports Medicine, 25*(08), 611-615.

Baddeley, A., Logie, R., Bressi, S., Sala, S. D., & Spinnler, H. (1986). Dementia and working memory. *The Quarterly Journal of Experimental Psychology Section A, 38*(4), 603-618.

Baddeley, A. (1998). Recent developments in working memory. *Current opinion in neurobiology, 8*(2), 234-238.

Bartlett, D. L. (1996). Physiological responses to music and sound stimuli. In D. A. Hodges (Ed.), *Handbook of music psychology* (pp. 343-385). San Antonio, TX: Institute for Music Research Press.

Behne, K. E. (1999). Zu einer Theorie der Wirkungslosigkeit von (Hintergrund-)Musik [On a theory of the non-impact of (background-) music]. In K-E. Behne, G. Kleinen, & H. de la Motte-Haber (Eds.), *Musikpsychologie: Jahrbuch der Deutschen Gesellschaft für Musikpsychologie* (14) (pp. 7–23). Göttingen: Hogrefe-Verlag

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering Pictures: Pleasure and Arousal in Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/0278-7393.18.2.379

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*. https://doi.org/10.1016/0005-7916(94)90063-9

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*. https://doi.org/10.1016/S0022-5371(72)80006-9

Cassidy, G., & MacDonald, R. A. R. (2007). #2 The effect of background music and background noise on the task performance of introverts and extraverts. *Psychology of Music*, *35*(3), 517–537. https://doi.org/10.1177/0305735607076444

Cassity, H., Henley, T., & Markley, R. (2007). The Mozart Effect: Musical Phenomenon or Musical Preference? A More Ecologically Valid Reconsideration. *Journal of Instructional Psychology*.

Dalton, B. H., & Behm, D. G. (2007). Effects of noise and music on human and task performance: A systematic review. *Occupational Ergonomics*, *7*, 143–152.

Davies, D. R., Lang, L., & Shackleton, V. J. (1973). The effects of music and task difficulty on performance at a visual vigilance task. *British Journal of Psychology, 64*(3), 383-389.

Dosseville, F., Laborde, S., & Scelles, N. (2012). Music during lectures: Will students learn better? *Learning and Individual Differences*. https://doi.org/10.1016/j.lindif.2011.10.004

Draine, S. C., & Greenwald, A. G. (1998). Replicable Unconscious Semantic Priming. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.127.3.286

DuBay, W. H. (2004). *The Principles of Readability*. Online Submission.

Ferguson, A. R., Carbonneau, M. R., & Chambliss, C. (1994). Effects of positive and negative music on performance of a karate drill. *Perceptual and motor skills, 78*(3_suppl), 1217-1218.

Field, A. (2009). Discovering Statistics Using SPSS. In *Discovering Statistics Using SPSS*.

https://doi.org/10.1234/12345678
Freeburne, C. M., & Fleischer, M. S. (1952). The effect of music distraction upon reading rate and comprehension. *Journal of Educational Psychology, 43*(2), 101–109.
Fried, R., & Berkowitz, L. (1979). Music Hath Charms… And Can Influence Helpfulness 1. *Journal of Applied Social Psychology, 9*(3), 199-208.
Furnham, A., & Bradley, A. (1997). Music while you work: The differential distraction of background music on the cognitive test performance of introverts and extraverts. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 11*(5), 445-455.
Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: The differential distraction of background music and noise on the cognitive test performance of introverts and extraverts. *Ergonomics*. https://doi.org/10.1080/00140130210121932
Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: a component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 245.
Gonzalez, M. F., & Aiello, J. R. (2019). More than meets the ear: Investigating how music affects cognitive task performance. *Journal of Experimental Psychology: Applied.*
Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers, 36*(2), 193-202.
Graesser, A. C., Cai, Z., Louwerse, M. M., & Daniel, F. (2006). Question Understanding Aid (QUAID) - A Web facility that tests question comprehensibility. *Public Opinion Quarterly*. https://doi.org/10.1093/poq/nfj012
Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, *40*(5), 223–234. https://doi.org/10.3102/0013189X11413260
Hallam S., Price J., & Katsarou G. (2002). The Effects of Background Music on Primary School Pupils" Task Performance. *Educational Studies*.
Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.89.1.92
Harp, S. F., & Mayer, R. E. (1998). How Seductive Details Do Their Damage: A Theory of Cognitive Interest in Science Learning. *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.90.3.414
Hebb, D. O. (1955). Drives and the CNS. *Psychological Review*.
Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of Musical Tempo and Mode on Arousal, Mood, and Spatial Abilities. *Music Perception*, *20*(2), 151–171. https://doi.org/10.1525/mp.2002.20.2.151
Isen, A. M. (2003). Positive affect, systematic cognitive processing, and behavior: Toward integration of affect, cognition, and motivation. In *Multi-level issues in organizational behavior and strategy* (pp. 55-62). Emerald Group Publishing Limited.
Jones, D., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999). Organizational Factors in Selective Attention: The Interplay of Acoustic Distinctiveness and Auditory Streaming in the Irrelevant Sound Effect. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.25.2.464
Jones, M. H., West, S. D., & Estell, D. B. (2006). The Mozart effect: Arousal, preference, and

spatial performance. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/1931-3896.s.1.26
Juslin, P. N., & Sloboda, J. A. (2001). *Music and emotion: Theory and research*. Oxford University Press.
Just, M. A., & Carpenter, P. (1987). Speedreading. In *The psychology of reading and language processing*.
Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music*, *39*(4), 424–448. https://doi.org/10.1177/0305735610376261
Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review, 95*(2), 163.
Kleinsmith, L. J., & Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of Experimental Psychology*. https://doi.org/10.1037/h0040288
Kotsopoulou, A., & Hallam, S. (2010). The perceived impact of playing music while studying: Age and cultural differences. *Educational Studies*, *36*(4), 431–440. https://doi.org/10.1080/03055690903424774
Landay, K., & Harms, P. D. (2018). Whistle while you work? A review of the effects of music in the workplace. *Human Resource Management Review*, (March 2017), 0–1. https://doi.org/10.1016/j.hrmr.2018.06.003
Lehman, S., Schraw, G., McCrudden, M. T., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology*. https://doi.org/10.1016/j.cedpsych.2006.07.002
Lehmann, J. A. M., & Seufert, T. (2017). The Influence of Background Music on Learning in the Light of Different Theoretical Perspectives and the Role of Working Memory Capacity. *Frontiers in Psychology*, *8*(October), 1–11. https://doi.org/10.3389/fpsyg.2017.01902
Matthews, G., Quinn, C. E., & Mitchell, K. J. (1998). Rock music, task-induced stress and simulated driving performance. *Transport Research Laboratory-Publications-PA*, 20-32.
Mayer, E. R. (1999). *The promise of educational psychology* (Vol. 58). Amsterdam: Noth Holland.
Mayer, E. R. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.

Mayer, E. R., & Chandler, P. (2001). When Learning is just a click away: does simple interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*(2), 390-397.
Mayer, E. R., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: when presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198.
Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85-139). Academic Press.
McElrea, H. (1992). Fast music causes fast drinking. *Perceptual and Motor Skills*. https://doi.org/10.2466/pms.75.5.362-362
McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal*

Learning and Verbal Behavior. https://doi.org/10.1016/S0022-5371(79)90255-X
McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse processes, 22*(3), 247-288.
McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction, 14*(1), 1-43.
McNair, D. M., Lorr, M., & Droppleman, L. F. (1992). Profile of mood states—revised. *San Diego, Calif.: Educational and Institutional Testing Service.*
Meteyard, L., Bruce, C., Edmundson, A., & Oakhill, J. (2015). Profiling text comprehension impairments in aphasia. *Aphasiology, 29*(1), 1-28.
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.
Milliman, R. E. (1986). The influence of background music on the behavior of restaurant patrons. *Journal of consumer research, 13*(2), 286-289.
Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology, 87*(2), 319.
Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.92.1.117
Nadler, R. T., Rabi, R., & Minda, J. P. (2010). Better mood and better performance: Learning rule-described categories is enhanced by positive mood. *Psychological Science*. https://doi.org/10.1177/0956797610387441
Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference. *Psychological Science, 10*(4), 370-373.
Nguyen, T., & Grahn, J. A. (2017). Mind your music: The effects of music-induced mood and arousal across different memory tasks. *Psychomusicology: Music, Mind, and Brain*, *27*(2), 81–94. https://doi.org/10.1037/pmu0000178
O'Donnell, E., Sharp, M., Wade, V., & O'Donnell, L. (2014). Personalised E-Learning: The Assessment of Students' Prior Knowledge in Higher Education. In *Handbook of Research on Education and Technology in a Changing Society* (pp. 744–755). IGI Global.
O'hanlon, J. F. (1981). Boredom: Practical consequences and a theory. *Acta psychologica, 49*(1), 53-82.
Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology, 84*(4), 429.
Paivio, A. (1986). *Mental representations: A dual coding approach.* London: Oxford Univ. Press.
Payne, B. K., Brown-Iannuzzi, J. L., & Loersch, C. (2016). Replicable effects of primes on human behavior. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000201
Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational psychology review, 18*(4), 315-341.
Ponzio, A., & Mather, M. (2014). Hearing something emotional influences memory for what was just seen: How arousal amplifies effects of competition in memory consolidation. *Emotion, 14*(6), 1137.
Perham, N., & Sykora, M. (2012). Disliked music can be better for performance than liked

music. *Applied Cognitive Psychology, 26*(4), 550-555.
Perham, N., & Currie, H. (2014). Does listening to preferred music improve reading comprehension performance? *Applied Cognitive Psychology*.
https://doi.org/10.1027/1618-3169/a000314
Rauscher, F. H., Shaw, G. L., & Ky, C. N. (1993). Music and spatial task performance [6]. *Nature*. https://doi.org/10.1038/365611a0
Rentfrow, P. J., & Gosling, S. D. (2003). The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.84.6.1236
Reynolds, J., McClelland, A., & Furnham, A. (2014). An investigation of cognitive test performance across conditions of silence, background noise and music as a function of neuroticism. *Anxiety, Stress, and Coping, 27*, 410–421. http://dx.doi.org/10.1080/10615806.2013.86 4388
Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *41*(1–A), 107–122. https://doi.org/10.1080/14640748908402355
Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science*. https://doi.org/10.1111/j.0963-7214.2005.00389.x
Schellenberg, E. G., Nakata, T., Hunter, P. G., & Tamoto, S. (2007). Exposure to music and cognitive performance: Tests of children and adults. *Psychology of music, 35*(1), 5-19.
Schlichting Jr., H. E., & Brown, R. V. (1970). Effect Of Background Music On Student Performance. *American Biology Teacher*. https://doi.org/10.2307/4443158
Schmidt, L. A., & Trainor, L. J. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition & Emotion, 15*(4), 487-500.
Shih, Y. N., Huang, R. H., & Chiang, H. S. (2009). Correlation between work concentration level and background music: A pilot study. *Work, 33*(3), 329-333.
Shih, Y. N., Huang, R. H., & Chiang, H. Y. (2012). Background music: Effects on attention performance. *Work, 42*(4), 573-578.
Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal Learning and Verbal Behavior*. https://doi.org/10.1016/S0022-5371(83)90282-7
Sloboda, J. A., & O'neill, S. A. (2001). Emotions in everyday listening to music. *Music and emotion: Theory and research*, 415-429.
Sloboda, J. A., & Juslin, P. N. (2010). At the interface between the inner and outer world. *Handbook of music and emotion*, 73-97.
Stadler, M. A., & Hogan, M. E. (1996). Varieties of positive and negative priming. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/BF03210745
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction, 4*(4), 295-312.
Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Aus- tralia: ACER Press.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning*, 19-30.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychological Science*. https://doi.org/10.1111/1467-9280.00345
Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, *40*(6), 700–708.

https://doi.org/10.1177/0305735611400173

Västfjäll, Daniel. Emotion induction through music: A review of the musical mood induction procedure. Musicae Scientiae 5.1_suppl (2001): 173-211.

Wharton, C., & Kintsch, W. (1991). An overview of construction-integration model: a theory of comprehension as a foundation for a new cognitive architecture. *ACM SIGART Bulletin, 2*(4), 169-173.

Witvliet, C. V., & Vrana, S. R. (1996, August). The emotional impact of instrumental music on affect ratings, facial EMG, autonomic measures, and the startle reflex: Effects of valence and arousal. In *Psychophysiology* (Vol. 33, pp. S91-S91).

Witvliet, C. V., Vrana, S. R., & Webb-Talmadge, N. (1998). In the mood: Emotion and facial expressions during and after instrumental music, and during an emotional inhibition task. *Psychophysiology Supplement, 88*.

Yaghoub Mousavi, S., Low, R., Sweller, J., & Mousavi, S. Y. S. (1995). Reducing Cognitive Load by Mixing Auditory and Visual Presentation Modes. *Journal of Educational Psychology*. https://doi.org/10.1037//0022-0663.87.2.319

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

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion, 8*(4), 494.

10. Annexes

Annex 1: excerpt of one of the texts in its two different versions of reading difficulty with the comprehension questions

Low RD version

A football fanatic from Dorset entered a Mexican bullring. He was confident that he was going to take part in a football training as a warm-up before the matador's main event. However he was pierced and tossed into the air nine times by a black bull. After twelve minutes of agony, he needed medical treatment.

“They asked for six volunteers, so I thought ‘Why not?’. The host said we would be playing a game of soccer. So the organizers brought out two small-sized goals. Then the host said, ‘watch out for the ball’ but he meant ‘bull’. We started passing the ball but, out of the corner of my eye, I saw this huge bull.”

High RD version :

When a football fanatic from Dorset entered a Mexican bullring, he was confident that he had agreed to take part in a kickabout as a warm-up to the matador's main event. However he was gored and tossed into the air nine times by a black bull. He needed medical treatment after a twelve-minute ordeal.

“They asked for six volunteers, so I thought ‘Why not?’. The compere said we would be playing a game of soccer and the organizers brought out two small-sized goals. Then he said, ‘watch out for the ball’ but he meant ‘bull’. We started having a kickabout but, out of the corner of my eye, I saw this huge bull.”

Excerpt from Comprehension questions

1. Where was the football fanatic from?

___

2. Why did the doctor treat him?

___

3. Which part of the bull injured the football fanatic?

___

4. How many times was he tossed in the air?

___

Annex 2: results of the ANOVAs for measures of cognitive load (perceived effort and difficulty) on texts and comprehension questions

1. Cognitive load perceived from reading the texts

1.1 Perceived effort

Main effect of text difficulty on perceived effort from reading text
[$F(1,46) = 3.93$, $p = .053$].

Main effect of music on perceived effort from reading text
[$F(1,46) = 0.471$, $p = .50$].

Interaction of music and text difficulty on perceived effort from reading text
[$F(1,46) = .012$, $p = .91$].

1.2 Perceived Difficulty

Main effect of text difficulty on perceived difficulty from reading text
[$F(1,46) = 3.30$, $p = .076$].

Main effect of music on perceived difficulty from reading text
[$F(1,46) = 2.47$, $p = .123$].

Interaction of music and text difficulty on intrinsic load from reading text
[$F(1,46) = .085$, $p = .77$].

2. Cognitive load perceived from answering the questions

2.1 Perceived Effort

Main effect of text difficulty on cognitive load from answering questions
[$F(1,46) = 1.91$, $p = .173$].

Main effect of music on cognitive load from answering questions
[$F(1,46) = 1.93$, $p = .171$].

Interaction of music and text difficulty on cognitive load from answering questions
[$F(1,46) = .91$, $p = .35$].

2.2 Perceived Difficulty

Main effect of text difficulty on intrinsic load from answering questions
[$F(1,46) = 0.004$, $p = .95$].

Main effect of music on intrinsic load from answering questions
[$F(1,46) = 2.78$, $p = .102$].

Interaction of music and text difficulty on intrinsic load from answering questions
[$F(1,46) = .05$, $p = .826$].

Annex 3: two-way mixed ANOVA on the impact of background music, text reading difficulty and their interaction on comprehension scores

Statistiques descriptives

	between participant condition	Moyenne	Ecart type	N
Easy_texts	no_music	6.2400	1.26754	25
	music	6.2174	1.15627	23
	Total	6.2292	1.20265	48
Hard_texts	no_music	6.3200	1.29808	25
	music	6.0652	1.46399	23
	Total	6.1979	1.37121	48

Tests multivariés[a]

Effet		Valeur	F	ddl de l'hypothèse	Erreur ddl	Signification
reading_difficulty	Trace de Pillai	.000	.017[b]	1.000	46.000	.896
	Lambda de Wilks	1.000	.017[b]	1.000	46.000	.896
	Trace de Hotelling	.000	.017[b]	1.000	46.000	.896
	Plus grande racine de Roy	.000	.017[b]	1.000	46.000	.896
reading_difficulty * music_group	Trace de Pillai	.004	.180[b]	1.000	46.000	.673
	Lambda de Wilks	.996	.180[b]	1.000	46.000	.673
	Trace de Hotelling	.004	.180[b]	1.000	46.000	.673
	Plus grande racine de Roy	.004	.180[b]	1.000	46.000	.673

Tests multivariés[a]

Effet		Eta-carré partiel	Paramètre Paramètre	Puissance observée[c]
reading_difficulty	Trace de Pillai	.000	.017	.052
	Lambda de Wilks	.000	.017	.052
	Trace de Hotelling	.000	.017	.052
	Plus grande racine de Roy	.000	.017	.052
reading_difficulty *	Trace de Pillai	.004	.180	.070

music_group	Lambda de Wilks	.004	.180	.070
	Trace de Hotelling	.004	.180	.070
	Plus grande racine de Roy	.004	.180	.070

a. Plan : Constante + music_group
Plan intrasujets : reading_difficulty
b. Statistique exacte
c. Calcul à l'aide d'alpha =

Tests de sphéricité de Mauchly[a]

Mesure: MEASURE_1

Effet intrasujets	W de Mauchly	Khi-deux approx.	ddl	Signification	Epsilon[b] Greenhouse-Geisser	Epsilon[b] Huynh-Feldt	Epsilon[b] Borne inférieure
reading_difficulty	1.000	.000	0	.	1.000	1.000	1.000

Teste l'hypothèse nulle selon laquelle la matrice de covariance des erreurs des variables dépendantes orthonormées est proportionnelle à la matrice identité.[a]
a. Plan : Constante + music_group
Plan intrasujets : reading_difficulty
b. Permet d'ajuster les degrés de liberté de la moyenne des tests de signification. Les tests corrigés sont affichés dans la table Tests des effets intrasujets.

Tests des effets intrasujets

Mesure: MEASURE_1

Source		Somme des carrés de type III	ddl	Carré moyen	F	Signification
reading_difficulty	Hypothèse de sphéricité	.031	1	.031	.017	.896
	Greenhouse-Geisser	.031	1.000	.031	.017	.896
	Huynh-Feldt	.031	1.000	.031	.017	.896
	Borne inférieure	.031	1.000	.031	.017	.896
reading_difficulty * music_group	Hypothèse de sphéricité	.323	1	.323	.180	.673
	Greenhouse-Geisser	.323	1.000	.323	.180	.673

	Huynh-Feldt	.323	1.000	.323	.180	.673
	Borne inférieure	.323	1.000	.323	.180	.673
Erreur (reading_difficulty)	Hypothèse de sphéricité	82.529	46	1.794		
	Greenhouse-Geisser	82.529	46.000	1.794		
	Huynh-Feldt	82.529	46.000	1.794		
	Borne inférieure	82.529	46.000	1.794		

Tests des effets intrasujets

Mesure: MEASURE_1

Source		Eta-carré partiel	Paramètre Paramètre	Puissance observée[a]
reading_difficulty	Hypothèse de sphéricité	.000	.017	.052
	Greenhouse-Geisser	.000	.017	.052
	Huynh-Feldt	.000	.017	.052
	Borne inférieure	.000	.017	.052
reading_difficulty * music_group	Hypothèse de sphéricité	.004	.180	.070
	Greenhouse-Geisser	.004	.180	.070
	Huynh-Feldt	.004	.180	.070
	Borne inférieure	.004	.180	.070
Erreur (reading_difficulty)	Hypothèse de sphéricité			
	Greenhouse-Geisser			
	Huynh-Feldt			
	Borne inférieure			

a. Calcul à l'aide d'alpha =

Tests des contrastes intrasujets

Mesure: MEASURE_1

Source	reading_difficulty	Somme des carrés de type III	ddl	Carré moyen	F	Signification
reading_difficulty	Linéaire	.031	1	.031	.017	.896
reading_difficulty * music_group	Linéaire	.323	1	.323	.180	.673

Erreur (reading_difficulty)	Linéaire	82.529	46	1.794		

Tests des contrastes intrasujets

Mesure: MEASURE_1

Source	reading_difficulty	Eta-carré partiel	Paramètre Paramètre	Puissance observée[a]
reading_difficulty	Linéaire	.000	.017	.052
reading_difficulty * music_group	Linéaire	.004	.180	.070
Erreur (reading_difficulty)	Linéaire			

a. Calcul à l'aide d'alpha =

Tests des effets intersujets

Mesure: MEASURE_1

Variable transformée: Moyenne

Source	Somme des carrés de type III	ddl	Carré moyen	F	Signification	Eta-carré partiel	Paramètre Paramètre
Constante	3696.503	1	3696.503	2328.138	.000	.981	2328.138
music_group	.461	1	.461	.290	.593	.006	.290
Erreur	73.037	46	1.588				

Tests des effets intersujets

Mesure: MEASURE_1

Variable transformée: Moyenne

Source	Puissance observée[a]
Constante	1.000
music_group	.082
Erreur	

a. Calcul à l'aide d'alpha =

Annex 4: T-tests for arousal between background music and silence condition

Statistiques de groupe

	between participant condition	N	Moyenne	Ecart type	Moyenne erreur standard
first measure of arousal level on participants	no_music	25	2.2400	1.01160	.20232
	music	23	2.1739	1.07247	.22363
2nd measure of arousal level on participants	no_music	25	2.0000	.86603	.17321
	music	23	2.1739	1.07247	.22363
3rd measure of arousal level on participants	no_music	25	2.1200	.88129	.17626
	music	23	2.3913	.98807	.20603
4th measure of arousal level on participants	no_music	25	1.9200	.90921	.18184
	music	23	2.2174	.99802	.20810

Test des échantillons indépendants

		Test de Levene sur l'égalité des variances		Test t pour égalité des moyennes		
		F	Sig.	t	ddl	Sig. (bilatéral)
first measure of arousal level on participants	Hypothèse de variances égales	.005	.942	.220	46	.827
	Hypothèse de variances inégales			.219	45.073	.828
2nd measure of arousal level on participants	Hypothèse de variances égales	1.239	.271	-.620	46	.538
	Hypothèse de variances inégales			-.615	42.344	.542
3rd measure of arousal level on participants	Hypothèse de variances égales	.134	.716	-1.005	46	.320
	Hypothèse de variances inégales			-1.001	44.257	.322

4th measure of arousal level on participants	Hypothèse de variances égales	.000	.994	-1.080	46	.286
	Hypothèse de variances inégales			-1.076	44.592	.288

Test des échantillons indépendants

		Test t pour égalité des moyennes			
		Différence moyenne	Différence erreur standard	Intervalle de confiance de la différence à 95 % Inférieur	Supérieur
first measure of arousal level on participants	Hypothèse de variances égales	.06609	.30082	-.53943	.67160
	Hypothèse de variances inégales	.06609	.30157	-.54127	.67344
2nd measure of arousal level on participants	Hypothèse de variances égales	-.17391	.28033	-.73819	.39037
	Hypothèse de variances inégales	-.17391	.28286	-.74461	.39678
3rd measure of arousal level on participants	Hypothèse de variances égales	-.27130	.26982	-.81443	.27182
	Hypothèse de variances inégales	-.27130	.27113	-.81765	.27504
4th measure of arousal level on participants	Hypothèse de variances égales	-.29739	.27527	-.85147	.25669
	Hypothèse de variances inégales	-.29739	.27636	-.85414	.25936

Annex 5: ANOVA to investigate the impact of background music and text reading difficulty and their interaction on time taken to read the texts

Statistiques descriptives

	between participant condition	Moyenne	Ecart type	N
text_easy_readingtime	no_music	127.1317	46.49780	25
	music	97.1338	35.26822	23
	Total	112.7577	43.76763	48
text_hard_readingtime	no_music	120.2177	36.98171	25
	music	95.0101	34.39713	23
	Total	108.1390	37.60515	48

Tests multivariés[a]

Effet		Valeur	F	ddl de l'hypothèse	Erreur ddl	Significatio n
time_taken	Trace de Pillai	.036	1.715[b]	1.000	46.000	.197
	Lambda de Wilks	.964	1.715[b]	1.000	46.000	.197
	Trace de Hotelling	.037	1.715[b]	1.000	46.000	.197
	Plus grande racine de Roy	.037	1.715[b]	1.000	46.000	.197
time_taken * music_group	Trace de Pillai	.010	.482[b]	1.000	46.000	.491
	Lambda de Wilks	.990	.482[b]	1.000	46.000	.491
	Trace de Hotelling	.010	.482[b]	1.000	46.000	.491
	Plus grande racine de Roy	.010	.482[b]	1.000	46.000	.491

Tests multivariés[a]

Effet		Eta-carré partiel	Paramètre Paramètre	Puissance observée[c]
time_taken	Trace de Pillai	.036	1.715	.250
	Lambda de Wilks	.036	1.715	.250
	Trace de Hotelling	.036	1.715	.250
	Plus grande racine de Roy	.036	1.715	.250
time_taken * music_group	Trace de Pillai	.010	.482	.104
	Lambda de Wilks	.010	.482	.104

	Trace de Hotelling	.010	.482	.104
	Plus grande racine de Roy	.010	.482	.104

a. Plan : Constante + music_group
 Plan intrasujets : time_taken
b. Statistique exacte
c. Calcul à l'aide d'alpha =

Tests de sphéricité de Mauchly[a]

Mesure: MEASURE_1

Effet intrasujets	W de Mauchly	Khi-deux approx.	ddl	Signification	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Borne inférieure
time_taken	1.000	.000	0	.	1.000	1.000	1.000

Teste l'hypothèse nulle selon laquelle la matrice de covariance des erreurs des variables dépendantes orthonormées est proportionnelle à la matrice identité.[a]
a. Plan : Constante + music_group
 Plan intrasujets : time_taken
b. Permet d'ajuster les degrés de liberté de la moyenne des tests de signification. Les tests corrigés sont affichés dans la table Tests des effets intrasujets.

Tests des effets intrasujets

Mesure: MEASURE_1

Source		Somme des carrés de type III	ddl	Carré moyen	F	Signification
time_taken	Hypothèse de sphéricité	489.229	1	489.229	1.715	.197
	Greenhouse-Geisser	489.229	1.000	489.229	1.715	.197
	Huynh-Feldt	489.229	1.000	489.229	1.715	.197
	Borne inférieure	489.229	1.000	489.229	1.715	.197
time_taken * music_group	Hypothèse de sphéricité	137.443	1	137.443	.482	.491
	Greenhouse-Geisser	137.443	1.000	137.443	.482	.491
	Huynh-Feldt	137.443	1.000	137.443	.482	.491
	Borne inférieure	137.443	1.000	137.443	.482	.491

Erreur (time_taken)	Hypothèse de sphéricité	13120.482	46	285.228		
	Greenhouse-Geisser	13120.482	46.000	285.228		
	Huynh-Feldt	13120.482	46.000	285.228		
	Borne inférieure	13120.482	46.000	285.228		

Tests des effets intrasujets

Mesure: MEASURE_1

Source		Eta-carré partiel	Paramètre Paramètre	Puissance observée[a]
time_taken	Hypothèse de sphéricité	.036	1.715	.250
	Greenhouse-Geisser	.036	1.715	.250
	Huynh-Feldt	.036	1.715	.250
	Borne inférieure	.036	1.715	.250
time_taken * music_group	Hypothèse de sphéricité	.010	.482	.104
	Greenhouse-Geisser	.010	.482	.104
	Huynh-Feldt	.010	.482	.104
	Borne inférieure	.010	.482	.104
Erreur (time_taken)	Hypothèse de sphéricité			
	Greenhouse-Geisser			
	Huynh-Feldt			
	Borne inférieure			

a. Calcul à l'aide d'alpha =

Tests des contrastes intrasujets

Mesure: MEASURE_1

Source	time_taken	Somme des carrés de type III	ddl	Carré moyen	F	Signification	Eta-carré partiel
time_taken	Linéaire	489.229	1	489.229	1.715	.197	.036
time_taken * music_group	Linéaire	137.443	1	137.443	.482	.491	.010
Erreur (time_taken)	Linéaire	13120.482	46	285.228			

Tests des contrastes intrasujets

Mesure: MEASURE_1

Source	time_taken	Paramètre Paramètre	Puissance observée[a]
time_taken	Linéaire	1.715	.250
time_taken * music_group	Linéaire	.482	.104
Erreur (time_taken)	Linéaire		

a. Calcul à l'aide d'alpha =

Tests des effets intersujets

Mesure: MEASURE_1
Variable transformée: Moyenne

Source	Somme des carrés de type III	ddl	Carré moyen	F	Signification	Eta-carré partiel	Paramètre Paramètre
Constante	1156913.769	1	1156913.769	425.791	.000	.902	425.791
music_group	18254.131	1	18254.131	6.718	.013	.127	6.718
Erreur	124986.335	46	2717.094				

Tests des effets intersujets

Mesure: MEASURE_1
Variable transformée: Moyenne

Source	Puissance observée[a]
Constante	1.000
music_group	.718
Erreur	

a. Calcul à l'aide d'alpha =

Annex 6: ANOVA to investigate the impact of background music and text reading difficulty and their interaction on time taken to answer the questions was conducted.

Statistiques descriptives

	between participant condition	Moyenne	Ecart type	N
text_easy_q_time	no_music	121.8046	42.34557	25
	music	87.4284	25.99322	23
	Total	105.3327	39.15485	48
text_hard_q_time	no_music	116.7654	40.20581	25
	music	82.4657	23.66495	23
	Total	100.3301	37.24845	48

Tests multivariés[a]

Effet		Valeur	F	ddl de l'hypothèse	Erreur ddl	Signification
time_taken	Trace de Pillai	.017	.818[b]	1.000	46.000	.370
	Lambda de Wilks	.983	.818[b]	1.000	46.000	.370
	Trace de Hotelling	.018	.818[b]	1.000	46.000	.370
	Plus grande racine de Roy	.018	.818[b]	1.000	46.000	.370
time_taken * music_group	Trace de Pillai	.000	.000[b]	1.000	46.000	.995
	Lambda de Wilks	1.000	.000[b]	1.000	46.000	.995
	Trace de Hotelling	.000	.000[b]	1.000	46.000	.995
	Plus grande racine de Roy	.000	.000[b]	1.000	46.000	.995

Tests multivariés[a]

Effet		Eta-carré partiel	Paramètre Paramètre	Puissance observée[c]
time_taken	Trace de Pillai	.017	.818	.144
	Lambda de Wilks	.017	.818	.144
	Trace de Hotelling	.017	.818	.144
	Plus grande racine de Roy	.017	.818	.144
time_taken * music_group	Trace de Pillai	.000	.000	.050

	Lambda de Wilks	.000	.000	.050
	Trace de Hotelling	.000	.000	.050
	Plus grande racine de Roy	.000	.000	.050

a. Plan : Constante + music_group
 Plan intrasujets : time_taken
b. Statistique exacte
c. Calcul à l'aide d'alpha =

Tests de sphéricité de Mauchly[a]

Mesure: MEASURE_1

Effet intrasujets	W de Mauchly	Khi-deux approx.	ddl	Signification	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Borne inférieure
time_taken	1.000	.000	0	.	1.000	1.000	1.000

Teste l'hypothèse nulle selon laquelle la matrice de covariance des erreurs des variables dépendantes orthonormées est proportionnelle à la matrice identité.[a]
a. Plan : Constante + music_group
 Plan intrasujets : time_taken
b. Permet d'ajuster les degrés de liberté de la moyenne des tests de signification. Les tests corrigés sont affichés dans la table Tests des effets intrasujets.

Tests des effets intrasujets

Mesure: MEASURE_1

Source		Somme des carrés de type III	ddl	Carré moyen	F	Signification
time_taken	Hypothèse de sphéricité	599.191	1	599.191	.818	.370
	Greenhouse-Geisser	599.191	1.000	599.191	.818	.370
	Huynh-Feldt	599.191	1.000	599.191	.818	.370
	Borne inférieure	599.191	1.000	599.191	.818	.370
time_taken * music_group	Hypothèse de sphéricité	.035	1	.035	.000	.995
	Greenhouse-Geisser	.035	1.000	.035	.000	.995

	Huynh-Feldt	.035	1.000	.035	.000	.995
	Borne inférieure	.035	1.000	.035	.000	.995
Erreur (time_taken)	Hypothèse de sphéricité	33675.420	46	732.074		
	Greenhouse-Geisser	33675.420	46.000	732.074		
	Huynh-Feldt	33675.420	46.000	732.074		
	Borne inférieure	33675.420	46.000	732.074		

Tests des effets intrasujets

Mesure: MEASURE_1

Source		Eta-carré partiel	Paramètre Paramètre	Puissance observée[a]
time_taken	Hypothèse de sphéricité	.017	.818	.144
	Greenhouse-Geisser	.017	.818	.144
	Huynh-Feldt	.017	.818	.144
	Borne inférieure	.017	.818	.144
time_taken * music_group	Hypothèse de sphéricité	.000	.000	.050
	Greenhouse-Geisser	.000	.000	.050
	Huynh-Feldt	.000	.000	.050
	Borne inférieure	.000	.000	.050
Erreur (time_taken)	Hypothèse de sphéricité			
	Greenhouse-Geisser			
	Huynh-Feldt			
	Borne inférieure			

a. Calcul à l'aide d'alpha =

Tests des contrastes intrasujets

Mesure: MEASURE_1

Source	time_taken	Somme des carrés de type III	ddl	Carré moyen	F	Signification	Eta-carré partiel
time_taken	Linéaire	599.191	1	599.191	.818	.370	.017
time_taken * music_group	Linéaire	.035	1	.035	.000	.995	.000
Erreur (time_taken)	Linéaire	33675.420	46	732.074			

Tests des contrastes intrasujets

Mesure: MEASURE_1

Source	time_taken	Paramètre Paramètre	Puissance observée[a]
time_taken	Linéaire	.818	.144
time_taken * music_group	Linéaire	.000	.050
Erreur (time_taken)	Linéaire		

a. Calcul à l'aide d'alpha =

Tests des effets intersujets

Mesure: MEASURE_1
Variable transformée: Moyenne

Source	Somme des carrés de type III	ddl	Carré moyen	F	Signification	Eta-carré partiel	Paramètre Paramètre
Constante	999319.711	1	999319.711	610.141	.000	.930	610.141
music_group	28249.190	1	28249.190	17.248	.000	.273	17.248
Erreur	75341.181	46	1637.852				

Tests des effets intersujets

Mesure: MEASURE_1
Variable transformée: Moyenne

Source	Puissance observée[a]
Constante	1.000
music_group	.982
Erreur	

a. Calcul à l'aide d'alpha =

Annex 7: T-test between background music and silence condition for total time taken to accomplish the task

Statistiques de groupe

	between participant condition	N	Moyenne	Ecart type	Moyenne erreur standard
total time spend for the whole experiment	no_music	25	740.7200	189.60497	37.92099
	music	23	592.1304	143.85798	29.99646

Test des échantillons indépendants

		Test de Levene sur l'égalité des variances		Test t pour égalité des moyennes		
		F	Sig.	t	ddl	Sig. (bilatéral)
total time spend for the whole experiment	Hypothèse de variances égales	.425	.518	3.038	46	.004
	Hypothèse de variances inégales			3.073	44.447	.004

Test des échantillons indépendants

		Test t pour égalité des moyennes			
		Différence moyenne	Différence erreur standard	Intervalle de confiance de la différence à 95 %	
				Inférieur	Supérieur
total time spend for the whole experiment	Hypothèse de variances égales	148.58957	48.90811	50.14264	247.03649
	Hypothèse de variances inégales	148.58957	48.35069	51.17282	246.00631