Modelling music selection in everyday life with applications for psychology-informed music recommender systems

Noah Felix Henry

PhD

University of York

Arts and Creative Technologies

December 2023

Abstract

Music is a highly functional and utilitarian resource. It enables people to regulate emotions, reduce distractions, stimulate physical action, and connect with others. However, with technologically facilitated ubiquitous listening now commonplace, new problems have emerged. The main problem is that of choice: how, given millions of songs to choose from, should providers curate listening experiences? To resolve this, many online platforms employ recommender systems, and there have been concerted efforts to orientate these systems in such a way that they are responsive to the short-term, dynamic needs of listeners in everyday situations. However, there is increasing scrutiny around the impact of automated recommender systems in terms of interpretability and data usage. To this end, researchers have begun exploring ways of integrating knowledge about user behaviours into the recommendation process, rather than through purely data-driven approaches.

This thesis aims to bridge these strands of intrigue by exploring an approach to generating situationally determined recommendations, based on an understanding of how and why contextual factors influence music selection in everyday life. This is achieved through three studies, in which contexts, functions, and content of listeners' music selections are triangulated to make inferences and estimates of situationally congruent musical characteristics. Firstly, a psychometric structure of the functions of music listening is generated. Secondly, this is triangulated with contextual factors and audio features of music selection. Finally, this is supplemented with an exploratory approach to generating recommendations through the explanatory model. These three studies result in both: a preliminary model of goal-orientated music listening that can be deployed by recommender procedures; and provides an exemplar methodology of how to construct behavioural models that can drive such systems. This thesis therefore holds relevance to both psychological research and those interested in music curation techniques.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I would like to thank my supervisors Professor Hauke Egermann and Dr Liam Maloney. I am grateful to you both for your continued support and guidance over the last few years. You have provided me with a safe and encouraging environment in which to work but have also challenged me to push myself and ensure the contributions I have sought to make in this thesis have been both meaningful and achievable. For that, I will be eternally grateful.

I have been fortunate during my PhD to have enjoyed the tremendous support of friends and colleagues in the Music Research Centre (MRC), to whom I will be always thankful for helping me improve my work, and for allowing me the space to vent on the more challenging days that inevitably come when undertaking a PhD. I am also grateful to innumerable friends outside of academia who have been equally kind and supportive.

Finally, I want to thank and ultimately dedicate this thesis to my family, without whose support this thesis would have been impossible, and so my gratitude and appreciation can never be fully expressed. I would like to thank my parents, Lucy and Stewart, in particular, who have always been exceptionally supporting and encouraging in allowing me to pursue my hobbies and interests.

Declaration

I declare that this thesis is a presentation of original work, and that I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

Aspects of the thesis (namely findings from Chapters 6 and 7) have been submitted for publication in a prospective journal article and partly disseminated in conference proceedings respectively. These are listed below and are correct at the time of submission (December 2023). It should be noted that the research article in question has not, at the time of writing, yet been accepted for publication (i.e., is under review). It is also intended that other aspects of this thesis (in particular Chapters 7 and 8) be published in other academic publications (e.g., journal articles and/or conference proceedings) in the future, if possible, but these have not yet been initiated at the time of writing.

Henry, N., Maloney, L., & Egermann, H. (2023). A Mixed-Method Exploratory Approach to Identifying the Utilitarian Functions of Music Listening. *Manuscript Submitted for Publication*.

Henry, N., Maloney, L., & Egermann, H. (2023). Functions of music listening mediate relationships between MIR-generated audio features of listeners' music selections and concurrent listening activities. In *Proceedings of the International Conference on Music Perception and Cognition 17-The Asia-Pacific Society for the Cognitive Sciences of Music 7*. Available at: https://www.researchgate.net/publication/373994427_Functions_of_music_listening_mediate_relationships_between_MIR-generated_audio_features_of_listeners'_music_selections_and_concurrent_listening_activities

1.1 Introduction

In the latter half of the twentieth century, portability became a significant design feature of consumer electronics (Weber, 2009). Portable audio players (e.g., handheld radios, Compact Disc [CD] Walkman) began to enable users to listen to music anytime and anywhere, providing a meaningful alternative to fixed appliances (e.g., record players). Ultimately, these technologies aimed to serve as aural companions for people on the move, enabling them to carry preferred material wherever they went (Bull, 2006). In the twenty-first century, digitised audio formats (e.g., MP3s) and the computational processing power of accompanying devices (e.g., iPod, Smartphones) afford users ever greater choice and autonomy when listening to music on the move. The progression of these technologies has culminated in cloud-based streaming services, online service platforms with repositories containing tens of millions of songs typically accessed via smartphones or other portable devices (Knees et al., 2019). As such, not only have the practical affordances of portable technologies continued to grow, but so has the variety of content available in everyday listening.

Whilst there are comparatively few barriers to accessing music in the twenty-first century given these developments, new technological challenges have nonetheless emerged in the global north. The primary challenge in question is that of curation. Online streaming platforms (e.g., Spotify, Apple Music, Deezer) may contain tens of millions of songs, and there is a pervasive need to ensure end users are able to find and retrieve desired content from these repositories. Also, there is a motivation to provide users with new content they have not yet engaged with but may like. This is because discovering new music, as well as having existing tastes effectively curated for, fosters satisfaction, and ultimately encourages user loyalty, conditions essential to cloud-based services in particular since they operate on subscription-based business models (Sinclair & Tinson, 2017; Schedl et al., 2018). In other words, online systems need to be effective at enabling users to discover content that is likeable to them so that they are encouraged to make continued use of the system. However, these functions must also be successfully achieved whilst simultaneously minimising the potential problem of choice-overload, given the sheer magnitude of options that comes with having tens of millions of songs to choose from. To navigate these issues, many online systems and repositories employ recommender systems.

In their broadest sense, recommender systems are a combination of software tools and computational techniques used to provide users with item suggestions likely to be of interest (Resnick et al., 1994; Ricci et al., 2015). Such systems are used in relation to a variety of online media contexts (e.g., news articles, movies, shopping items), but music recommender systems specifically are used to help listeners discover new music and to generate listening playlists containing recommended tracks. Listeners are now frequently reliant on playlists curated via recommender systems to discover and consume music from online libraries, forming the "backbone of how music is consumed" in modern settings (Saravanou et al., 2021; p. 602). There are varying methods by which recommender systems operate, however, they typically take large amounts of data on user interactions (Ricci et al., 2015). These data are then used, often in machine learning applications, to uncover patterns in users' interaction behaviours and construct predictive models intended to provide users with suitable recommendations (Kaminskas & Ricci, 2012).

Given the ease of access afforded by portable listening technologies, music may be used by listeners systematically in the stream of everyday life (i.e., in response to short-term, real-time needs in momentary episodes). In general terms, music may be used in everyday life to help people mitigate stress, elicit emotions, and facilitate social interaction and connectivity (Sloboda et al., 2001), and constitutes one of the most common leisure activities people employ (Schäfer et al., 2013). As such, when people listen to music, researchers argue that they do so for a particular reason, which is essentially to achieve a particular goal or function typically informed by the situation in which they are listening (e.g., Konecni, 1982; Krause et al., 2014; 2015; Lamont et al., 2016; Greb et al., 2018a; 2018b; Maloney, 2019). Although this topic will be expanded on in detail throughout this thesis, it has been subsequently argued that the autonomy in music listening afforded by portable technologies in everyday situations holds implications for the functions and utility that music may serve:

"…it is quite conceivable that the greater range of music available and the greater degree of control over it afforded by music technology has implications for the ways in which people might use music in everyday life" (Krause et al., 2014; p. 307).

These relative degrees of autonomy and choice allow people to use music in complex ways in the everyday situations in which they listen, with individuals considered *active* rather than *passive* in their listening practices (Krause et al., 2015). When it comes to providing recommendations via the platforms used by listeners to access music, therefore, there exists a motivation to maximise the efficacy of recommendations by generating systems that are responsive to how contextual or situational factors of the listening episode (e.g., location, activity, time of day) might influence the music listeners select or otherwise deem congruent in that moment (Wang et al., 2012; Takama et al., 2021). The proposed integration of context into recommender procedures are referred to as Context-Aware approaches (Dey, 2000), and Context-Aware Music Recommender Systems (CAMRSs) specifically are intended to generate recommendations according to the short-term needs of listeners, which may vary according to contextual factors (e.g., Wang et al., 2012).

To date, there have been several attempts to generate CAMRSs (e.g., Wang et al., 2012; Takama et al., 2021). Often, however, it has appeared to be the case that proposed CAMRSs are dependent on users' location data (as well as microphones and accelerometers) to infer context, which is seemingly at odds with recent critiques that recommender systems in general are too dependent on users' transaction or private data (Di Noia et al., 2022). Therefore, there appears to be a discrepancy between the intention to generate systems that respond to the contextual needs of users (i.e., to satisfy the short-term contextual needs of users) and the forms of data and methods that are used to achieve this. Given the increasing scrutiny on online platforms with regards to the kinds of data they use in recommender procedures in general (see Di Noia et al., 2022), there exists an opportunity to consider novel approaches towards achieving this in such a way that mitigates potential issues or concerns around the suitability, interpretation, and application of users' interaction data.

In recent years, researchers have expressed a general need to integrate research from fields such as music psychology and music education more proactively into the generation of recommender systems, citing an overreliance in Music Information Retrieval (MIR) as a field of data-driven approaches that are solely based on users' interactions with online streaming platforms (e.g., Knees et al., 2019; Lex et al., 2021). It has been suggested that such dependencies lead to a lack of nuanced understanding of the underlying cognitive and psychological motivations that underpin behaviour and experience (Lex et al., 2021). As such, it has been proposed that *psychology-informed* (also known as *knowledge-driven*) approaches be applied to expand the scope and expertise that are used in applied research, but also to improve the transparency and explainability of recommender systems as the black-box machine learning methods applied often reduce nuance and understanding of why certain recommendations are provided (Lex et al., 2021). This lack of explainability diminishes the trust users may ultimately have in systems (Schedl et al., 2014), and so there is an incentive from the perspective of providers to ensure there is some ability to explain the recommendation process. Psychology-informed approaches do not necessarily seek to do away with data-driven computational approaches entirely, but rather aim to make inferences as to the 'why' behind user interactions, rather than simply predicting a 'what'; for which collaboration between music psychologists and MIR researchers is needed (Kaminskas & Ricci, 2012). The aim of this is to ensure that the sum of knowledge and understanding that underpins what is ultimately an interdisciplinary endeavour suitably pools resources, knowledge, and skill sets across music research.

This feeds back into the discussion around increased *use* and *functionality* of music in everyday life. With the ever-increasing agency at the individual level afforded to listeners by technology, the data-driven vs psychology-informed discussion is particularly relevant. Consider that there is limited effectiveness in appropriate short-term music recommendations in cross-sectional settings since the pieces of music that make up these recommendations are limited in length, or typically consumed in listening sessions alongside other tracks, the relevance of which are highly context-dependent (Hansen et al., 2020). This thesis aims to contribute to the resolution of this problem, insofar as by triangulating the influence of contextual factors on music's *functionality*, it is argued that it may be possible to (exploratively) generate short-term, psychology-informed music recommendations responsive to relevant contextual factors. In this regard, the overarching aim of this thesis is to contribute to an understanding of the underlying psychological and/or cognitive processes that underpin situationally determined music selection and consider how such insights may be extended to assist music curation. Components necessary to achieving this are scattered throughout the literature, and this thesis hopes to integrate and consolidate these to provide a principled and considered approach to help mitigate this issue.

Moreover, relatively recent examples such as Lepa et al. (2020) and Herzog et al. (2020) illustrate some of the ways in which data from behavioural studies can be leveraged and operationalised by content-based methods of tag collection to provide recommendations. It is plausible that a similar approach could be applied to help achieve listeners' goals in given listening contexts. This is an area where a psychology-informed approach, utilising methodologies from the social sciences, has the potential to make a substantive contribution to knowledge.

In summary, the functions that music serves listeners in everyday life ultimately drives music selection, but these functions vary depending on listeners' context, which recommender systems are typically not dynamic in responding to (Wang et al., 2012; Greb et al., 2018a; Maloney, 2019). Therefore, it is interesting to consider the extent to which the functions of music listening (FML) may differ according to contextual variables, and to consider the ways in which this leads to changes in the music listeners select in a theoretical framework, which may then be used to formulate a psychology-informed approach to context-aware music recommendations. There are, therefore, two primary strands of interest considered: (1) the ways in which music choices may change under the influence of context and (2) how an understanding of such mechanisms might be used to synthesise recommendations for listeners in response to those relevant contexts. All in all, this is to generate an understanding of why FML and music selection changes between contexts, and how an understanding of this might be used to provide recommendations that are not dependent on datatypes that might diminish trust in systems or otherwise require additional data from users. To summarise the underlying motivations and goals contextualised above, a single overarching research question is presented:

> By what means might it be possible to curate everyday listening through a psychology-informed approach to music recommendations?

To address this question, this thesis needs to suitably capture and assess three primary constructs: (1) the contexts in which music listening occurs, (2) FML, and (3) the content of music deemed congruent (functional) during listening episodes. The aim here is not to generate an independent recommender system to test the accuracy of an identified model, but rather to establish an approach to model generation and prediction. This would serve as a theoretical or pre-processing basis upon which more complex recommender systems can be built, however, appropriate discussion and understanding of recommender systems will be covered in the literature review to further contextualise this proposed outcome. Next, the aims and placement of this thesis will be discussed in more precise terms before proceeding with a literature review central to this research.

1.2 Aims & Objectives

This thesis aims to bridge a gap within extant research by developing an approach by which psychological methods are used to associate audio-features with listening contexts via FML for the purpose of music curation. This is with a view to subsequently estimate congruent audio-features in everyday listening scenarios, for which short-term recommendations are made via behavioural modelling, rather than through data-driven models derived through machine learning with little to no intelligibility or longitudinal taste profiling. This holds obvious relevance in fields such as MIR, however, the primary focus of this research is to establish a theoretical framework that might follow, using methodologies from the social sciences. This is summarised through the overarching research question presented in the previous section, for which a series of aims and related objectives have been identified to help address:

1. To be able to validate a measure of FML from the utilitarian perspective

In considering the measurement of relevant constructs, an informed and reliable way of gauging FML in everyday life should be present, which may in turn be used to incorporate an understanding of why people listen to music in subsequent steps relating to the curation of listening experiences (i.e., by understanding or approximating *why* someone is listening to music in a given situation). For this, several objectives have been identified: (1) to understand what FML is, (2) to review existing measures and/or perspectives of FML and identify limitations to these approaches where relevant, and (3) to design and implement a robust methodological approach that ensures a valid measurement model is generated and applied in the subsequent steps throughout this thesis. These three objectives will enable comparative reflection of existing approaches to the measurement of FML and aim to address drawbacks of these.

2. To be able to associate a validated measure of FML with listening contexts and music content

Following the targeted outcome of the first aim listed above, the second aim alludes to the need to ensure any model generated holds ecological as well as content validity. Specifically, this relates to ecological aspects of the listening context (e.g., the location and/or activity of the listening episode) as well as the content or characteristics of the music selected by listeners. This is because there is evidence to suggest that contexts of music listening directly inform not only listeners' FML, but also the music they may select in turn (e.g., Greb et al., 2019). For this, it is intended to: (1) apply any measurement model generated during the development of the preceding aim in ecologically valid data and (2) to examine the ways in which FML relate to other factors informing the listening experience. Chapters 2 and 3 expand on this in greater detail, however, the second aim presented here can be succinctly considered as the intention to triangulate FML in an ecologically valid framework that considers additional aspects of the listening experience.

3. To be able to associate musical content with listening contexts

In addition to the second aim, it is not only intended to associate FML with aspects of the listening context and musical content respectively, but also to consider the ways in which these other integral constructs of the listening experience might also be associated directly and/or indirectly. For example, considering the hypothetical associations between the listening situation and FML, and also FML and musical content, it seems noteworthy to consider causal mechanisms linking these constructs as part of a larger behavioural system. It is seemingly plausible that an understanding of how aspects of the listening situation in conjunction with valid measurement of FML may be used to predict, or at otherwise estimate, what kinds of musical characteristics may be congruent for listeners between situations. This aim is therefore intended to be achieved by: (1) estimating relationships between relevant constructs of the listening context and musical content directly and (2) indirectly. In conjunction with the second aim, this modelling approach may therefore also be subsequently leveraged to provide an understanding of *how* music selection changes given contextual information and/or FML, which may in turn be applied in a psychology-informed approach to providing recommendations (i.e., by integrating knowledge about how situational factors and/or changes in FML may influence outcomes in music selection).

4. To be able to propose an actionable method of integrating knowledge generated in steps 1-3 into a recommendation procedure

Finally, given that the preceding three aims and associated objectives effectively aim to characterise and measure relationships between the situations in which people listen to music, FML, and the content of the music people select, it is lastly intended to explore a way of integrating knowledge about these relationships into the music curation process. Effectively, this summarises and incorporates the contributions made in the preceding steps to explore the research questions more directly. In this sense, it is proposed that by understanding the causal mechanisms between constructs, it may be possible to specifically estimate what kinds of music might be appropriate given certain information about a listeners' situation and/or FML. Using the information obtained from aims 1-3, therefore, we may be able to essentially reverse engineer the process from context to music selection. Hence, this would (1) facilitate the integration of knowledge into the recommendation process, and (2) provide responsive recommendations for the short-term needs of listeners. However, given intensive computational procedures associated with the generation of complex systems, it is likely more achievable within the scope of this thesis to highlight this as an exploratory proof of concept, rather than as a fully-fledged, independent system. In this sense, it is the motivation and approach of this final aim that is intended to contribute to knowledge further and address the overarching research question.

The four aims summarily suggest that by triangulating information about the listening context, FML, and musical features, it may be possible to curate everyday listening dynamically. In this sense, these aims look to bridge gaps in the literature by exploring complementary approaches that facilitate a psychology-informed approach to curating everyday listening episodes. Through these aims and objectives, it is hoped that the problem outlined in the introduction to this thesis is partially addressed. This work falls within the scope and interest of both music psychology and MIR, therefore, and as such may provide some useful insights for both. With these aims and objectives outlined, the following sections expand further on the position of this research in relation to these fields, followed by an outline of the thesis structure and definitions of key terms.

1.3 Situating this research

Advances in portable listening technologies have substantially impacted music engagement practices, to the point that music may now integrate with everyday life seamlessly with few to no barriers (Maloney, 2017). Such levels of flexibility and accessibility provides listeners with choice and has implications for the uses of music in people's everyday lives (Bull, 2006; Krause et al., 2014). For example, Brown and Krause (2020) noted that listeners are able to consciously mediate between formats according to their respective pros and cons, finding digital formats, for example, to be highly functional; "affording listeners convenience, accessibility and portability" (p. 95). Similarly, Sinclair and Tinson (2017) found that increased listening facilitated via digital formats leads to both music and the accompanying technology becoming more closely integrated into daily routines. Meanwhile, physical formats (e.g., CDs, Vinyl) can act as representations of identity by manifesting individuals' cultural capital into environmental spaces, which may serve "non-utility purposes" (Giles et al., 2007, p. 438). Such perceptions of identity and ownership do, however, also shift with technological affordances. For instance, the use of headphones allows individuals to assert environmental control in physical spaces which, whilst only being audible to the individual, may nevertheless serve to extend identity (Bull, 2006). As such, music has become increasingly integrated with everyday life through technological developments, yet affordances remain within and between formats (Brown & Krause, 2020).

Implications that formats provide listeners with relative affordances are consistent with the assertion that autonomous music listening is a goal-orientated process (Sloboda et al., 2001), and that when people listen to music, they are active consumers rather than passive listeners (e.g., Krause et al., 2015). Increased usage therefore implies increased utility, and research exploring the reasons for music listening according to the needs of listeners has grown in recent years. Yet, given the breadth and complexity of the subject of music utility, there have been historically different and conflicting perspectives and approaches. Maloney (2019), for instance, highlights that research has often viewed FML from regulatory perspectives (e.g., mood regulation). Maloney subsequently argues that whilst rationales are provided from these differing approaches, they are often unaligned and incongruent, and attempts to homogenise perspectives to provide a broadly congruent model of FML from the *utilitarian* perspective, which is broader in scope and more reflective of *functionality* than the narrower regulatory perspectives or approaches that fail to consider utility as a consequence of contextual factors. That is to say, to focus on the situationally determined goals of listeners in everyday life in ways that blend these varied approaches. This thesis aims to build on this work by leveraging this utilitarian perspective of FML to better encapsulate the situational uses that music provides listeners and gauge the ways in which contextual variables may influence FML. It is for this reason that this thesis specifically aims to build on these contributions by introducing quantitative measures to provide new opportunities for the assessment of FML in relation to contextual variables in everyday life. Broadly speaking, this constitutes the first strand of interest within this thesis.

In addition to the above, this thesis also aims to consider the role that such insights may have in the curation of music listening. This is because researchers in both music psychology (e.g., Maloney, 2019; Greb et al., 2019) and MIR (e.g., Kaminskas & Ricci, 2012; Lex et al., 2021) have proposed that substantive insights relating to the ways in which people use music in everyday life has profound implications for recommender systems development in particular. The streaming services that provide such portability and utility of music in the 21st century provide vast repositories available to listeners on demand, often containing tens-of-millions songs. So as to not overwhelm users, systems typically provide user interfaces for listeners to discover music through search and query and/or provide recommendations for them to discover new content (Kaminskas & Ricci, 2012). However, whilst these recommender systems are effective at curating long-term taste and preferences of listeners, they are less effective in handling short-term needs in particular (e.g., Wang et al., 2012). As such, researchers have sought to develop CAMRSs that curate short-term listening episodes. More pressingly, however, is that whilst these systems are being developed, there is a general lack of uptake in the development of psychology-informed recommendations, that are less dependent on user-interaction data with online platforms (Lex et al., 2021). Rather, it is suggested that systems leverage insights and knowledge from psychological research, to predict and inform recommendations, rather than train machine-learning models that are black-box in nature and restrict explainability and trust in the systems that curate everyday music listening (Schedl et al., 2014; 2022). Given the first strand of substantive interest, namely situationally rooted *functionality*, it is argued that there is an opportunity to not only contribute to knowledge in this domain, but also explore the applicability of such insights in subsequent music curation.

The themes discussed above are expanded on within the literature review of this thesis, but this short discussion is intended to provide a high-level overview, summarising two strands of research and opportunities for crossover. The former falls within the sphere of music psychology and can be considered contributory to the understanding of the utility that music serves people in everyday life, whereas the second strand considers the ways in which this knowledge may be applied to curate listening experiences (based on insights from the first strand). It is hoped that this helps bridge a gap between music psychology and the development of recommender systems.

1.4 Thesis Structure

What follows is a brief summary of this thesis' structure. This is to provide a clear and concise summary of each chapter's content, relevance, and value to this thesis as a whole.

Chapter 1: Introduction

This introductory chapter is intended to provide the reader with an understanding of the underlying motivations and aims of this thesis at large, summarised with a single research question (see section 1.1). This begins by providing some background literature on the topic at hand, which has then been used to inform the aims and objectives of this thesis at large and the topic area within which it sits. Before proceeding to a literature review, some definitions of key terms are provided to help the reader with some additional context as to the role and relevance of key terminologies within this thesis.

Chapter 2: Goals and Functions of Music Listening

This first chapter of the literature review serves as an overview of the role of music in relation to its functions and association with goal attainment. Initially, this chapter begins with a brief overview into the origins of music as a functional tool. This is followed by a discussion of music facilitated goals and formulations of musical preference and provides a comprehensive overview and baseline understanding of music's role in everyday life.

Chapter 3: Contextual applications of music

In continuing the literature review, this chapter aims to cover contextual factors that influence music use and selection. By reviewing literature relating to contextual demands and application of music listening, the aim is to address some of the myriad of listening locations and activities that are associated with music listening. This is attained by discussing key literature on contextual applications of music listening.

Chapter 4: Music Curation and Recommender Systems

As the final chapter of the literature review, this chapter pivots from the first to second domain of substantive interest. It aims to summarise and discuss relevant roles and applications of MIR in music recommendations and use, including the ways in which feature extraction is applied. This serves to contextualise the technological backdrop underpinning the extraction of audio features. This argues there is a need to apply psychology-informed approaches when developing recommender systems, especially as they increasingly look to curate listener's short-term goals in everyday life.

Chapter 5: Summary and Conceptual Approach

Chapter 5 consolidates the preceding three chapters of the literature review by summarising the extent to which research has been able to address FML, as well as the implementation of recommender systems to satisfy listeners' goals in everyday listening. This is used as a basis upon which a theoretical model is proposed, with an outline of the procedure hypothesised to provide insight by measuring variation in audio-features within listening data. This also outlines the relevant methodological approaches taken in this thesis, namely how a series of studies may address the thesis aims. This therefore effectively acts as the bridge between the secondary and empirical phases of the thesis.

Chapter 6: Study 1: Developing a utilitarian measure of FML

Study 1 is the first empirical study in this thesis. It serves as an exploratory structuring and quantification of the Consensus Functions Framework (CFF), put forward by Maloney (2019), into a psychometric construct of FML for further use. This involves the item-generation, measurement, systematic reduction, and structuring of 53 distinct FML. As the CFF is the most exhaustive model of FML available at the time of writing, it was deemed an appropriate model upon which to base a utilitarian measurement instrument. This serves to address the first aim listed in section 1.2.

Chapter 7: Study 2: Contextual applications of music: Repeated Measures

Study 2 aims to measure the temporal relationships and structures central to this thesis. A novel two-arm study design is applied using an online survey and Experience Sampling Method (ESM) study sequentially to capture individuals' listening behaviour in relation to concurrent listening situations. By utilising the psychometric construct generated in Study 1, Study 2 gathers data relating to the functions that music is serving in real time, as well as the music itself. Utilising an open-source Application Programming Interface (API), audio-features are extracted from user's music selections, affording analyses to take place between the three key constructs. These are assessed under a Structural Equation Modelling (SEM) framework to test for the presence of a causal, temporal structure amongst these constructs via hypothesised mediations between concurrent listening activities and MIR-generated audio features of listeners' selections via FML. This addresses the second and third aims outlined in section 1.2.

Chapter 8: Study 3: Design and Implementation of a psychology-informed recommendation procedure

Chapter 8 presents the third and final study carried out as part of this thesis. In this, a saturated structural path model is leveraged to estimate suitable values for audio features under different situational conditions. This is used to essentially reverse engineer the process fit in study 2, in which recommendations are targeted according to estimated audio features. This is then subject to a *user-centric* evaluation study to assess the efficacy of the approach. This addresses the fourth and final aim listed in section 1.2.

Chapter 9: Discussion and Concluding remarks

Chapter 9 summarises and discusses the findings of the three empirical studies and outlines the overall implications for research in this area. This includes implications for both FML and recommender systems research, as well as drawbacks and future directions and suggestions for further development. This conclusion serves to highlight conceptual approaches within this thesis that may be of use to other researchers, and so serves to consolidate the content of the thesis at large and address the overarching research question presented in section 1.1.

1.5 Definitions of key terms

The primary topic central to this thesis is FML in the pursuit of goal attainment. There are three key definitions within this that should be clearly stated before proceeding:

1. *Music* - Although an intersubjective understanding of what is being referred to by the term ‘music’ exists, it is conducive to provide a clear definition of what, practically speaking, constitutes ‘music’. Arguably the most notable and commonly used definition of music is that of Blacking (1973), who characterised music as “Humanly Organised Sound” (pp. 3-31).

When it comes to defining music, however, there are often ethnocentric limitations to definitions. In an effort to be pragmatic therefore, distinctions have not just been made about what is music, but also what it is not music. Nattiez (1990) delineates musical sound and noise, for instance, by using three descriptive levels: *poietic* (composer’s intentions/choices), *neutral* (physical definition, such as sound within the harmonic spectrum or not), and *esthetic* (perceptive judgements). Although this is useful in differentiating what may be considered music from non-music, this is still subject to ethnocentric biases (such as that noise is disturbing, unpleasant, or both). Nattiez (1990) acknowledges that “the border between music and noise is always culturally defined – which implies that, even within a single society, this border does not always pass through the same place; in short, there is rarely a consensus” (p. 48). By defining musical sounds as physical aspects of noise however, Levitin (2006) separates characteristic properties of music, namely *Pitch*, *Rhythm*, *Tempo*, *Contour*, *Timbre*, *Loudness*, *Reverberation*, *Meter*, *Key*, *Melody*, and *Harmony* (pp. 15-18). Additionally, Collins et al. (2014) ascribe music with the distinction that at the basic featural level, it is the structured organisation of auditory objects into temporally extended sequences. In wishing to combine these perspectives and emphases into a practical definition relevant to this thesis, music can therefore be considered a socio-cultural artefact referring to humanly organised sound in structured sequences of harmonic, melodic and/or rhythmic content, distinguishing it from noise. But this comes with a caveat that intersections between sound and noise are diverse and culturally varied, and so the point at which noise becomes sound is flexible according to additional contextual and/or cultural factors.

2. *Goal* – The Oxford English Dictionary defines a ‘goal’ as “An aim or outcome which a person, group, or organisation works towards or strives to achieve, the object of a person’s ambition or effort. Later also: (*Psychology*) an end result to which a series of actions, choices, events, etc., lead (whether consciously or unconsciously directed), the achievement of which brings reward or satisfaction” (“goal, *n*.”, n.d.). The latter half of this definition is especially important within the context of this thesis as it is acknowledged that at the psychological level, actions may be taken towards attainment of preferable psychological states.

Latham and Locke (1991) outline goal setting theory as a concept for future events or desired outcomes that individuals or groups envision, plan, and enact towards the attainment of. Humans’ capacity for reason enables goal conceptualisation and requires purposeful action to attain said goal (Latham & Locke, 1991). The components of a goal can be partitioned into two primary features: *content* and *intensity*. Goal *content* varies on a broad spectrum from vague to specific and may also vary according to degrees of difficulty; while *intensity* is an equally broad term that may refer to scope, clarity, and the mental effort required to achieve a goal.

Whilst Latham and Locke (1991) provide an extensive overview of goal setting, its relevance to the present thesis has been covered bar one significant point: "emotional responses are the result of automatic, subconscious value appraisals" (p. 231). The implication that the relationship between goal setting and affect is involuntary is important when considering music's use in facilitating goal attainment, as music's effect on emotion which, although by no means the sole interest to this thesis, is a central subject to the field in general. Emotions in this sense can be characterised as brief but intense affective reactions involving components such as subjective feeling, physiological arousal, expression, action tendency, and regulation. Emotions in music, or musical emotions, is the term used to describe emotions that were induced by music in some way (Juslin & Sloboda, 2011). More to the point, involuntary appraisal holds importance for our understanding of goal attainment in everyday life. Goal setting is conscious, but goal attainment is not.

Goal setting and attainment is therefore present in the regulation of emotional states in general (e.g., Tamir et al., 2020), and there is evidence that this may be facilitated via music listening (e.g., Juslin et al., 2008). Cooper (2018) notes that growing bodies of evidence support the notion that "goals and goal-related processes are fundamental to how positive and negative people feel" (p. 36), and hypothesises that emotions arise in the presence or absence of particular goal processes. This is shown in Table 1.

Table 1 Emotions hypothesised to arise in the presence or absence of goal processes (Cooper, 2018)

	Presence	***Absence***
Awareness of goals	Meaning, purpose, sense of direction, orientation, order	Meaninglessness, disorientations, chaos, directionlessness, despair

Perceived attainability of goals	Hope, optimism, control, order, excitement, expectation	Hopelessness, futility, fear, anger, shame, sadness
Progress/velocity towards goals	Hope, accomplishment, excitement, self-belief, expectation, control, flow	Frustration, failure, despair, disillusionment, lack of self-belief, anger
Achievement of goals	Satisfaction, accomplishment, fulfilment, experiencing of the desired state, per se (e.g., relaxation, physical pleasure)	Dissatisfaction, failure, sadness, loss, frustration, envy, anger

Note. Adapted from Cooper (2018).

Extant research surrounding goal attainment with regards to music listening have also identified several key concepts that highlight music's use as a tool toward the attainment of psychological goals more broadly (e.g., Greb et al., 2018a; Maloney, 2019), which underpin music listening in general (e.g., Sloboda et al., 2001). In this sense, goal-attainment via music listening need not refer solely to the regulation of mood, but other behaviours also. Examples of music-facilitated goal attainment may therefore include the use of music to obtain desired emotional goals (e.g., Scherer & Zentner, 2001; Juslin et al., 2008), but also to pass time (e.g., Heye & Lamont, 2010), accompany routine tasks (e.g., Greb et al., 2019), motivate physical action (e.g., Hallett & Lamont, 2015), and accompany social interactions with others (e.g., Cunningham & Nichols, 2009). The literature review of this thesis will expand on these goals as well as present research findings with regard to music-facilitated goal attainment, but this acknowledgement is intended to outline exactly what is being referred to by the term 'goal' at the cognitive level, and that this may be viewed through different lenses with varying points of emphasis.

3. *Functions of music listening* – Merriam (1964) distinguishes ‘uses’ of music from ‘functions’ by describing ‘use’ as the situation in which music is employed, whereas ‘function’ refers to the reasons for music’s employment and the purposes it serves. In a similar vein, Greb et al. (2018a) refer to music’s *functionality* as the intentional application of music to attain specific goals in specific situations. Building on theoretical approaches, such as those of Behne (1997), Schäfer (2016) asserts that over time an individual learns that music listening may assist in the attainment of goals in specific situations (*past functional experiences*). That is to say: “music listening can be a functional behaviour” (p. 3), highlighting that FML are subject to exposure to, and experience with, music. Merriam’s distinction between use and function has enabled researchers to view FML as something situation-specific, which may accordingly vary given the context within which music is being listened to (e.g., Greb et al., 2018a; Maloney, 2019). Characteristics of specific functions and relevant perspectives are discussed in the following chapter; however, this concise introduction serves to characterise FML as the purposeful application of music to attain contextually orientated goals.

With these characteristics outlined, the following chapter introduces the subject of goals and FML in greater detail, constituting the opening chapter of the literature review of this thesis.

2.1 Goals and functions of music listening

As mentioned in section 1.4, the literature review of this thesis is broken down into three sections: (1) the ensuing chapter, which outlines the goals and functions of music listening, (2) Chapter 3 which discusses contextual applications of music in everyday life, and (3) Chapter 4 which discusses music curation and recommender systems. These three chapters contextualise relevant theories, assumptions, and applications of existing knowledge that have cumulatively informed the aims of this thesis at large, as outlined in section 1.2. Their relevance, therefore, is that they provide an understanding as to the role music plays in everyday life, and the ways in which listening is curated. It is through a shared understanding of these key areas that the aims of this thesis may be fulfilled, and the central research question addressed.

Firstly, this chapter contextualises, in detail, what the goals and functions of music listening are. It is initially important to understand that regardless of whether individuals are fully aware of music's psychological, social, and physiological effects in everyday life, it constitutes a media that many apply on an everyday basis (Saarikallio, 2011; Maloney, 2019). People use music as a tool for goal attainment in everyday life by inducing desired cognitive, emotional, and/or physiological effects (Sloboda et al., 2001), which in turn formulates a given function (Greb et al., 2018a). This chapter therefore outlines existing research and subsequent theories of how *functionality* has emerged, the overarching purpose of which is to provide an overview of extant research relating to the FML and music-facilitated goals. It is this understanding and underlying theoretical bases that an informed approach to resolving the first aim of this thesis (see section 1.2) may be achieved.

2.2 Theoretical approaches to the functions of music listening

There are varied schools of thought through which FML have been considered. For instance, there are longstanding and diverse theories as to the evolutionary utility of music as a cultural, social, and behavioural phenomena. Such theories have received mixed reactions over the last two centuries, with Cross (2016) noting that in the late-nineteenth and early-twentieth centuries, musical theorists, historians, and anthropologists were heavily influenced by, and integrated, evolutionary theory to help explain musical change, difference, and value. However, from the mid to late-twentieth century onwards, such theories more or less vanished from musicological discourse and only began to resurface in the late-1980s. Others have instead viewed FML not through evolutionary lenses, but rather as an artefact applied to satisfy specific needs, an approach particularly prominent in recent years owing to music's increased integration and with everyday life (e.g., Lonsdale & North, 2011; Krause & Brown, 2021). Given these differences in approach and scope, this section serves to outline the differences in these theoretical orientations to understanding FML, beginning with the evolutionary perspective. This provides background to some of the differing theoretical approaches to understanding FML, contextualising the perspectives and viewpoints taken as this thesis progresses.

Evolutionary Approaches

One approach to understanding music's utility is to consider its evolutionary role in human development. With language, music is considered humanity's most distinctive behavioural trait, and the two are hypothesised to have coevolved (Benítez-Burraco & Nikolsky, 2023). Schäfer et al. (2013) describe a myriad of theories as to the evolutionary purposes of the production and listening of music. For example, it has been argued that music making may have developed to help facilitate sexual selection; whereby individuals with the time and energy to make music must be comparatively healthy to be able to spare time and energy. Others argue music developed out of a need for social and emotional communication between people, such as to coordinate cohesive group activities and reinforce social bonds, which helps establish hierarchies within groups. Schäfer et al. (2013) argue this can still be observed in familiar settings today, such as national anthems, work songs, and lullabies, which help bind together nations, groups, and families. Moreover, humming or singing is thought to have specifically occurred to maintain mother-infant attachment. It has been suggested that 'putting down the baby' allowed early humans to free up their hands for other tasks, and that humming or singing may have developed as a means of indicating proximity in the absence of physical touch (Schäfer et al., 2013).

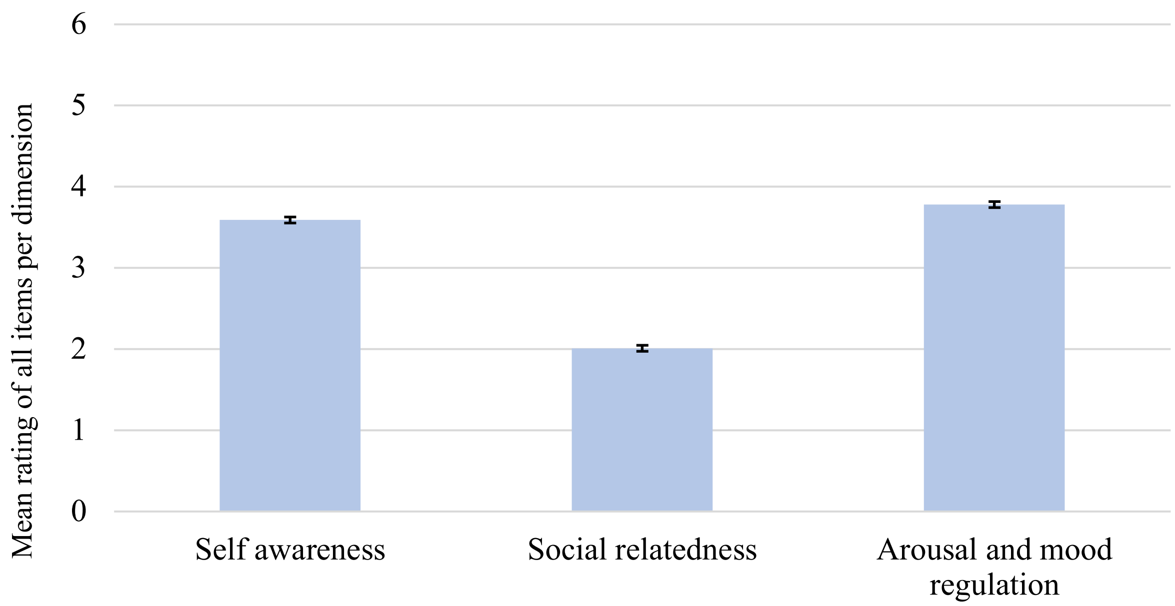
There may also have been less immediate functions that music may have facilitated, however. For instance, as humans became more effective hunters, less time and energy were required for the hunting and gathering of food sources, and we thus found ourselves with more free time. Subsequently, music making may have served to pass time, analogous to the way that lions sleep for large parts of the day (Schäfer et al., 2013). Additionally, more abstract theories have been put forward as to the initial functions of music making and listening, namely the use of music for transcendence. Music has been observed to induce flow states, beneficial for our psychological wellbeing, and its ability to provide listeners with a means of escapism and distraction may also have benefits (e.g., Saarikallio & Erkkilä, 2007). Relatedly, it is theorised that widespread music making acted as a form of escape for early humans from their harder and more rugged lives for the betterment of their mental wellbeing (Schäfer et al., 2013). Broad and somewhat anthropological in nature, evolutionary perspectives provide an interesting line of inquiry as to the music's evolutionary and/or biological purpose as a behaviour but say less about why people apply and engage in modern settings. The following therefore outlines a different theoretical approach, one more consistent with extant psychological research.

<u>*Uses and Gratifications*</u>

Away from evolutionary approaches, researchers have instead focused on the ways in which music is used in the stream of everyday life. *Uses and gratifications* approaches (e.g., Arnett, 1995; Krause & Brown, 2021) instead focus on the needs of users, and aims to explain how people select and use media (e.g., music) to attain those needs. Given this emphasis on the *utility* of media, empirical methods are used, where users are asked to report reasons that apply to them regarding use of a given media, and so involve very different methodologies than evolutionary approaches. *Uses and gratifications* assumes that listeners are active consumers, rather than passive recipients of media (Krause & Brown, 2021), and that people are self-aware and capable of reporting motives, or at least self-aware to the point of recognition of motives when asked to rate a given list of possible drivers (Lonsdale & North, 2011). Typologies of varying length and granularity have been presented in the literature, examples of which are discussed at later points in this review. Other non-evolutionary approaches include *experimental aesthetics*, in which subjective experiences of beauty and experience of pleasure are investigated (Schäfer et al., 2013), and may be applied to determine music preferences, for instance (North & Hargreaves, 2008), but are less closely related to FML as an indicator of user need, as per goal attainment.

As part of their overview, Schäfer et al. (2013) conducted an extensive literature survey to compile interpretations of FML and note that given the varied and complex nature of the FML, researchers often seek to identify structures or patterns in data obtained from real-world samples (e.g., via dimension reduction). Based on their literature review, they generated a pool of 129 items relating to functions that music may serve, that were subsequently rated by 834 participants according to the extent to which participants introspectively felt they used music for the described purpose (each of the 129 items were rated on seven-point Likert scales). From this, they discerned three dimensions of music listening through principal components analysis (PCA): *Self-awareness, Social relatedness* and *Arousal and mood regulation* (see Figure 1).

Figure 1 Mean ratings of Schäfer et al. (2013) three components of reasons for music listening



Note. Adapted from Schäfer et al. (2013). Self-awareness: *M* = 3.59 (*SE* = .037); Social relatedness: *M* = 2.01 (*SE* = .035); Arousal and mood regulation: *M* = 3.78 (*SE* = .032).

Self-awareness pertains to the use of music with regards to self-related thoughts, emotions and sentiments, absorption, escaping, coping, solace and meaning. Overall, this generalised FML considers private or personal relationships with music. Schäfer et al.'s (2013) second dimension of music listening, *Social relatedness*, includes the use of music for social bonding and affiliation to feel close to others, expression of identity and values, and to gather information about social environments. Social bonding has often been proposed as the main driver behind musicality as a mode of human behaviour and function. Savage et al. (2021), for instance, argue that musicality has evolved as a consequence of coevolution between genes and cultures. Through this coevolution, they argue that proto-musical behaviours spread as cultural evolution had feedback effects on biological evolution due to the impact of such behaviours on social bonding. The authors emphasise that there are deep links between production, perception, prediction, and social reward as a result of repetition, synchronisation, and harmonisation of rhythm and pitch, and consolidate this with the support of empirical evidence for such links in neurological networks, physiological mechanisms, and behaviours across cultural boundaries and even species. This suggests that music's association with social interaction, bonding, or *relatedness* is something of a 'mega' function. Additionally, Hansen and Keller (2021) suggest, with greater specificity, that oxytocin (a hormone released into the bloodstream by the pituitary gland) constitutes a 'socio-allostatic' agent that modulates senses, learning, prediction, and behavioural responses to physical and social environments to facilitate the kinds of social bonding music enables.

Finally, *Arousal and mood regulation* refers to the use of music for the purpose of background entertainment and diversion to generate positive mood and regulate psychological arousal. Arousal in this regard refers to the cognitive or physical stimulation music affords listeners, since differing degrees of stimulation may be required for certain tasks (Konečni, 1982; North & Hargreaves, 2000; Lamont et al., 2016). This implies that modulations in the level of arousal as perceived by listeners' results in music inducing modulations in physiological as well as cognitive faculties. Konečni (1982) argued that sources of arousal, namely music and context, are effectively summed by listeners, who then select music that brings about an overall level of desired arousal. This may move along cognitive as well as physiological lines, such as selecting music that is low in arousal when engaging in a highly arousing or cognitively intensive task and selecting more complex or arousing music when in a situation perceived as being boring (North & Hargreaves, 2008). Arousal as an aspect of listener experience has therefore often been a common lens, as this may relate to both cognitive and physiological research in response to music.

In subsequent research, Schäfer (2016) used these three dimensions to assess whether the strength of listeners' music preference are informed by the functions that music fulfils in their lives (e.g., to regulate emotions, moods, and physiological arousal, promote self-awareness, or advance social relatedness). Through a diary study, Schäfer (2016) found that those who had reported more intense past experiences of the functional use of music had stronger intentions to listen to music in the pursuit of goals in specific situations generally and showed greater strengths of musical preference overall. It is consequently theorised that over time, individuals learn that listening to music allows them to achieve goals according to situational needs, referred to as *past functional experiences*. *Past functional experiences* manifest to form specific intended effects in future music engagement (*listening goals*). The greater the extent to which music can fulfil these goals in a consistent manner results in increased music listening habitually, which subsequently raises the extent to which individuals enjoy and are involved in music listening (Schäfer, 2016). Counter-wise, if music is not conducive towards the attainment of goals, then it will not be reinforced psychologically to promote engagement, and preference for that music will remain weaker with regards to the situation (Schäfer, 2016).

The notion of music listening as a tool for goal-achievement in general is aligned with *uses and gratifications* approaches to FML, as the short-term dynamic relationships between situation, goal and music selection are aligned with the emphasis on the utility of media as an observable set of behaviours in everyday life, which can be empirically assessed as a triangulation between listeners, function, and music (e.g., Lonsdale & North, 2011; Krause & Brown, 2021). There is also substantial crossover with North and Hargreaves' (2000) arousal state-goal approach, which itself is derived from Konečni's (1982) contributions. This highlights conceptual overlap in aesthetic and utilitarian perspectives relating to music selection, but nonetheless speaks to an undercurrent of intentional or purposeful applications of music in everyday life. With this in mind, the following sections further explore the ways in which music elicits goals within listeners for the purpose of goal attainment, and in turn how these translate to the music selected.

2.3 Music-facilitated goals

"The *functionality* of music listening refers to the intentional use of music to accomplish specific goals in specific situations, such as eliciting personal memories, getting energised, or making time go by more quickly" (Greb et al., 2018a; p. 764). Music has been posited to be a utilitarian resource applied by listeners to enhance cognitive, emotional, behavioural, and physiological aspects of the self. Consequently, listeners interact with music to attain and achieve contextually orientated goals (Maloney, 2017). This is broadly reflected by Schäfer's (2016) diary study, for instance, in which dimensions underlying function were present in everyday listening. Music research has thus sought to provide cohesive typologies or models of goals of music listening. For instance, Lonsdale and North (2011) follow a *uses and gratifications* approach to measure the extent to which specific aims were present in listening habits. Like Schäfer et al. (2013), they applied dimension reduction (Exploratory Factor Analysis; EFA) to discern six latent factors: *Positive mood management* (e.g., to set the 'right' mood), *Diversion* (e.g., to pass the time), *Negative mood management* (e.g., to make me feel better), *Interpersonal relationships* (e.g., to have something to talk about with others), *Personal identity* (e.g., to create an image for myself), and *Surveillance* (e.g., to learn about how people think). Items/goals pertaining to each of these factors are shown in Table 2.

Table 2 Lonsdale and North's (2011) music-orientated goals

Factor	*Items*
Positive mood management	- To be entertained - To relax - To set the 'right' mood - To take my mind off things
Diversion	- To 'fill' uncomfortable silences - To pass the time - To relieve boredom
Negative mood management	- To help get through difficult times - To relieve anxiety - To relieve tension/stress - To express my feelings and emotions - To make me feel better - To alleviate feelings of loneliness

	- To escape the reality of everyday life
Interpersonal relationships	- To keep up with current events - To stay in-touch with current fashions and trends - To spend time with family - To have something to talk about with others - To spend time with friends
Personal Identity	- To construct a sense of identity for myself - To explore possible identities - To portray a particular image to others - To express my identity - To create an image for myself - To display my membership of social groups/subcultures
Surveillance	- To learn how to do things - To learn how to behave in future - To obtain useful information for daily life - To discover who I really am - To learn how other people think

Note. Adapted from Lonsdale and North (2011).

Of these six factors, mood-related functions were found to be of primary importance to listeners, whilst social FML were of secondary importance (Lonsdale & North, 2011). However, it has been noted that the methods used to attain this particular framework lacks one key quality central to the topic of discussion: that of context (Maloney, 2019). Lonsdale and North (2011) effectively treat music listening as a standalone leisure activity and compare it to other activities, rather than as an accompaniment to those activities (e.g., reading a book or playing video games). Konečni (1982) notes that music is enjoyed in the stream of everyday life and is critical of studies that treat aesthetic preferences of music being formed in a "social, emotional and cognitive vacuum, as if they were independent of the contexts in which people enjoy aesthetic stimuli in daily life" (p. 498).

Konečni (1982) conveys how properties of music, namely arousal, can affect listening choices when undertaking certain tasks, based on the complexity of said task. This is argued to stem from cognitive limitations that mean less complex or arousing music is more amenable when undertaking a more complex task requiring greater concentration, and vice versa. In other words, one piece of music selected in a given instance may be more beneficial to the listener in the given context over another piece, based on the levels of attention that an individual is able to give to multiple stimuli. With a more complex task, it may be more congruent to have less complicated music as selecting music that is overly complex may function as more of a distraction by subtracting cognitive processing ability (Konečni, 1982; North & Hargreaves, 2008). It is reasonable to expect that individuals may subsequently differ in perceptions of goal-congruence of music according to the context. Greb et al. (2018a) explore this further and describe how *functionality* may differ in situations at the individual level, as well as the situational level. This allows us to delineate between the relative levels at which FML may run and consider factors that may influence FML through these respective lenses. The following subsections explore this further by discussing the ways in which these respective levels may influence FML.

2.3.1 Individual level

Starting with the former of the two levels discussed by Greb et al. (2018a), the individual level refers to the idea that interindividual differences influence FML, for instance according to demographic variables like age and gender, as well as other variables like health, well-being, musical taste, and personality traits (Greb et al., 2018a). For example, Ter Bogt et al. (2010) formed a typology of music listeners based on engagement with music for the purpose of mood regulation. Three groups were identified: *High-involved listeners*, *Medium-involved listeners,* and *Low-involved listeners*. *High-involved listeners* used music most often for the purposes of raising mood and experienced the strongest emotional responses to music overall. *High-involved* and *Medium-involved listeners* experienced more negative emotions when listening to music than *Low-involved* listeners, however, the *Low-involved* group did still report the use of music for the purposes of mood regulation, but to a lesser degree than the other two groups (Ter Bogt et al., 2010). Those experiencing the strongest emotional responses to music subsequently used it more in response to events occurring in their lives. In other words, those that get the most from music listening subsequently apply it more readily (Ter Bogt et al., 2010). This is consistent with the findings of Schäfer (2016) where *past functional experiences* were argued to predispose music's presence in the future. The *High-involved* group described by Ter Bogt et al. (2010) were found to benefit the most from music's ability to enhance moods and formulate identities. Again, this is consistent with the finding that the extent to which music is used is associated with its perceived efficacy from the listener's perspective.

Other work, such as that of Greasley (2008) and Maloney (2019), effectively argues that individuals have a skillset, or repertoire, of FML. A sense of functional efficacy appears to be established within listeners depending on the extent to which music is applied during daily life. Those that apply music in a greater variety of settings and in response to a greater variety of stimuli are likely to have a wider variety of contexts within which music may be considered functional; compared to those who apply music to a lesser degree, as is reflected within Ter Bogt et al. (2010).

2.3.1.1 Music Preferences

In a similar vein, research has investigated relationships between FML and taste, which has shown relations between FML and the strength of music preference (Schäfer, 2016; Greb et al., 2018a). Indeed, individuals' music preferences have been described as "the very basis of musical behaviour" (Fricke et al., 2021; p. 372). It may be the case that different situations inform FML, but listeners do not unilaterally select the same music in response to those situations. Berlyne (1971) theorised that preference for artistic stimuli is linked to their arousal potential, essentially referring to the amount of activity they produced in the ascending reticular system. Music with an intermediate degree of arousal potential is liked the most, and this gradually decreases toward the extreme ends of potential, forming an inverted-U shape (North & Hargreaves, 2008). Moreover, empirical contributions have attempted to uncover latent dimensions of music preference, such as the work by Rentfrow and Gosling (2003).

Here, a series of six studies were conducted in which the groundwork was laid for a number of theories pertaining to the roots of music preference. This generated a structured view of preference which pooled lay beliefs about music, structures underlying preference, and associations between preferences and personality traits. As well as relationships between music preference and personality dimensions, additional associations were made between preference and the ways in which individuals view themselves (e.g., self-perceived physical attractiveness) and cognitive abilities. This contribution was significant given that prior to this, music research had indicated that there may have been links between genre-preference and personality traits but had not explored this at a deeper level. This study played an important role in laying the groundwork for research on musical preferences by generating the first cohesive structure (referred to the Short Test of Music Preferences, or STOMP). There are, however, limitations with STOMP that are worth noting. Namely, the sample used to generate the model exclusively contained undergraduate students in the United States. The wider generalisability of the resulting model can therefore be questioned with regard to how well it represents other demographic groups and/or cultures. Indeed, the localised nature of the construct means that genre classifications have had to be adapted for use in other cultural contexts elsewhere (e.g., Ferrer et al., 2013).

More recent research has refined and updated key principles relating to the understanding of music preference, whereby preference has not been viewed in social or cultural lenses so much as through preferences at the featural level. Such literature has suggested that feature-based preference can be viewed across five orthogonal dimensions: Mellow, Unpretentious, Sophisticated, Intense and Contemporary; MUSIC (Rentfrow et al., 2011). Specifically, it has been suggested that preferences are influenced by the extent to which an individual listener likes particular combinations of these musical attributes, with social and cultural connotations (e.g., genre preference) being of lesser importance (Rentfrow et al., 2012). Rentfrow et al. (2012) examined sound and psychological characteristics associated with each of the five dimensions of MUSIC. They found that of the described musical attributes, there was significant variance amongst the five factors. Musical features strongly associated with Mellow, for instance, included slow, quiet, not distorted, and acoustic; whilst the psychological characteristics were perceived as dreamy, romantic, warm, sensual, and inspiring, but not animated, enthusiastic, amusing, or fun (Rentfrow et al., 2012).

However, one limitation with this approach is that, again, it does not consider contextual factors and the ways in which these may affect how pieces of music are perceived. Nonetheless, this does contribute to an understanding of music preferences by distinguishing featural aspects across latent structures. Research prior to MUSIC was, however, constrained by technology, whereby researchers were limited to assessing structure of musical preference via self-reported genres, which are ill-defined and abstract (Fricke et al., 2021). Therefore, there is an argument for viewing preference through the lens of feature-based preference, rather than socially or culturally derived preferences.

In addition to the evidence that musical features play a key role in determining preference, links have also been identified between preference and personality traits, whereby higher levels of openness and agreeableness are associated with greater genre inclusivity and diversity (Bansal et al., 2020). Fricke et al. (2021) identified that three broad dimensions underlie music-feature preferences: arousal, valence, and depth (AVD), comparable to findings by Greenberg et al. (2016) that perceived musical characteristics reflect personality traits, which were in turn more important predictors of musical preference than demographic variables. This highlights the individuality of music listening experiences and interpretations, however, also reaffirms the subjectivity of such experiences. With this in mind, Barone et al. (2017) discerned that whilst personality traits may serve as predictors of preference for musical features, and by extension genre preferences, attributes found in other genres may additionally predict liking (i.e., how much a listener will enjoy music based on its similarity to music they presently like that holds similar features). Hence, preference of music genres and features are individual and may be influenced by personality traits, but preference for certain features may also be consistent even when music is unfamiliar.

Music preferences are difficult to place in relation to FML. On the one hand, they are related to individual, preferential applications of music based on its content and thus must play some role in music selection, regardless of context. On the other hand, they are by nature highly subjective (Barone et al., 2017), which makes them unreliable predictors in contextual settings. There remain, however, disparities as to whether such individual constructs influence music selection behaviours. For example, It has been argued that more open and intellectually engaged individuals use music in a more cognitive manner, whilst neurotic, introverted, and non-conscientious individuals are more likely to use music for mood regulation (Chamorro-Premuzic & Furnham, 2007); the latter of which is generally considered by some to be the most important FML (Groarke & Hogan, 2016). However, others have argued personality traits to be of little to no use in accounting for variance in music taste or preference (Schäfer & Mehlhorn, 2017; Greb et al., 2019), perhaps because measures like the Big-Five Inventory (BFI) are simply too broad to accurately predict complex behaviours like music selection (Greb et al., 2019). Therefore, whilst such constructs might influence music taste, these may not translate to meaningful differences in music selection.

Finally, Groarke and Hogan (2016) report that variables such as age have been observed to contribute to differences in FML; whereby younger adults emphasise affect regulation and social connection, while older individuals emphasise eudaimonic functions, such as transcendence and personal growth. However, these particular findings related to FML referring to wellbeing enhancement, implying that the variation was with reference to a specific functional domain (affect regulation), rather than situationally determined 'use'. Groarke and Hogan (2018) expanded on these findings by developing the Adaptive Functions of Music Listening (AFML) scale, which measures 11 FML dimensions in relation to well-being outcomes. Their findings reinforce the notion that individuals apply music for goal-attainment, however, are still limited to applications of music in the context of affect regulation, rather than in a broader set of everyday listening functions that assess broader utility (Maloney, 2019).

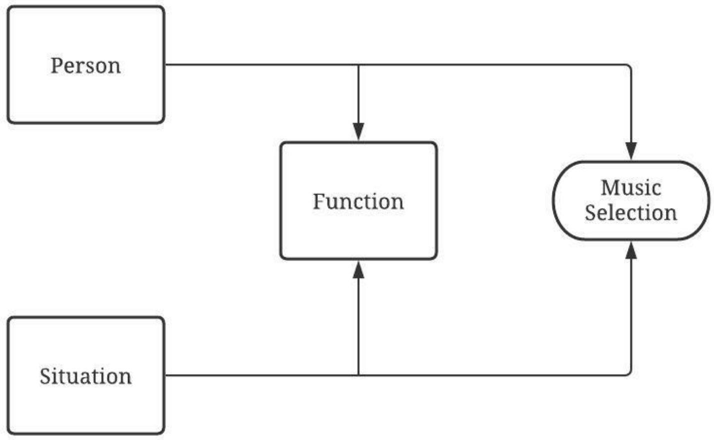
2.3.2 Situational level

The second dimension discussed by Greb et al. (2018a) refers to *situational* variables, the contextual aspects that accompany music listening. At the cultural level, we might associate certain pieces or genres of music with specific contexts (e.g., dance music being linked to nightclubs). Contexts of music listening are often social, such as in the case of nightclubs, and individuals gain an inter-subjective, social understanding of what music is appropriate for a specific situation in accordance with the social dynamics at play (Maloney, 2019). As such, there might be an assumption at a social level that, given the context, some music is more appropriate than others, given the nature of the event or activity. However, as these social elements fade and people listen to music alone, there is greater freedom and variety in music selection.

Greb et al. (2018a) discerned that listeners mostly engage with music during personal maintenance (such as housework), active leisure (such as exercise), and travel. This is generally for the purposes of enjoyment/entertainment, passing time, and enhancing mood. They identified five functions of contextual music listening: *Intellectual Stimulation*, *Mind wandering & emotional involvement*, *Motor synchronisation & enhanced wellbeing*, *Updating one's music knowledge*, and *Killing time & Overcoming loneliness*. Others have evidenced the frequency of activities like working/studying and social interaction during music listening but have also emphasised the role of intended (typically mood-related) outcomes in relation to these situations (e.g., Juslin et al., 2008; Randall & Rickard, 2017). These findings have explicitly suggested that emotional outcomes of music listening are determined by situational variables almost entirely (Randall & Rickard, 2017), highlighting not only music's role as an accompaniment to everyday life, but also as a means to achieve desired cognitive and emotional states in everyday situations.

Moreover, Greb et al. (2018b) argued that situation specific FML consistently explained the kinds of music listeners select, something often discussed elsewhere in the literature (e.g., Konečni, 1982; North et al., 2004; North & Hargreaves, 2008). Interestingly, they found that situational variables mostly accounted for differences in selection behaviour, whereas individual-level variables, like personality traits, had little effect on situation-specific music selection. This adds to evidence that situational factors are the main drivers of listening goals and subsequent music selection, supporting the idea that music is consciously and actively chosen in everyday life (e.g., North et al., 2004). Greb et al. (2019) expanded on this through a repeated measures study in which the notion that situational variables are the primary predictors of music's application in listening contexts was reaffirmed. They argue that research should consequently pivot focus from historically individual-level differences relating to FML and music selection, and rather focus on situational variables instead. Moreover, they outline a causal path model by which situational and individual level predictors may be used to predict changes in music selection directly and indirectly via FML. This model of a mediational relationship between situational and individual variables and music selection via FML is illustrated in Figure 2.

Figure 2 Greb et al. (2019) model of situation-specific music selection



Note. Adapted from Greb et al. (2019).

As has already been discussed, individuals at least partially determine music selection in specific situations according to learned associations based on *past functional experiences* with music in specific scenarios (Schäfer, 2016). Greb et al. (2018a; 2019) found that FML were predominantly reported in response to locales; for example: “creating the right atmosphere” was often reported for music listening in social situations, such as bars and clubs. Furthermore, it is also reported that there is contextual variety in the resulting intensity of music listening in particular situations. For instance, it was reported that there were more intense responses overall when listening to music whilst at the gym, compared to the role music played whilst in a restaurant (Greb et al., 2018a). This is broadly consistent with the perspectives of Konečni (1982) and North and Hargreaves (2000) who argued that contextual variables play a significant role in listeners’ goals and experiences, and that music selection is based on selecting congruent featural aspects of music, such as arousal.

Krause and North (2017) discuss predictors of music’s presence in given situations and subsequent judgements made about music amongst participants. They observed that location, activity, and individuals' perceptions of dominance were significant predictors of music use in everyday life. This was, however, contrary to previous research insofar as rather than considering situational pleasure and arousal variables as significant predictors of contextualised music listening, perceptions of dominance and control were indicated to be more important (Krause & North, 2017). Like Greb et al. (2018a), the two primary dimensions of consideration were variables concerning the listener and the context. With regard to music's presence in everyday situations, significant predictors amongst individuals included measurements pertaining to music's importance, average listening hours and music education level; that were all positively related to music's application in particular episodes (Krause & North, 2017). Contextual variables that predicted music's presence in everyday life gave specific attention to time of day, whereby individuals were significantly less likely to hear music throughout the day, as it went from morning to afternoon and on towards evening. Furthermore, location and activity were observed to be significant predictors of music's presence in episodes. Location variables indicated that listening episodes were more likely to occur in cars than at work, walking or shopping. Finally, it was also observed that episodes in which individuals had greater levels of environmental arousal and dominance were more likely to involve music's listening (Krause & North, 2017).

This latter point feeds into further areas of research on music's everyday roles and effects. Music's application for the purposes of environmental control has been observed to be positively associated with technological advancements increasing its integration into everyday life (Sinclair & Tinson, 2017). This has been argued to extend self-identity and perceptions of environmental autonomy, generating perceptions of control (Bull, 2006; Danckwerts & Kenning, 2019). This is a theme that holds consistency with literature regarding individual's music-facilitated goals, insofar dominance also plays a significant role for which environmental autonomy is key (e.g., Krause & North, 2017). Overall, Krause and North (2017) find that three dimensions are determinant in music's application towards listening contexts: pleasure, arousal, and dominance. However, they argue that the extent to which each of these dimensions are relevant across various contexts is understudied and should be subsequently considered in contextual applications of music listening. For example, dominance as a dimension may be particularly important in contexts in which it is expected to be high, whilst arousal may be more important in contexts where polarised states of arousal are central to the contextual application of music, such as relaxing or exercising in comparison to housework or whilst on public transport (Krause & North, 2017). This raises further questions, however, as to how autonomy plays a role in music selection insofar as music selection cannot be based solely on contextual *functionality* but nuanced individual experiences and tastes also.

2.5 Chapter Summary

To summarise this chapter, music serves emotional, cognitive, and social applications to attain intended psychophysiological outcomes amongst listeners. The *goals of music listening* are evoked according to contextual demands in which listeners find themselves, and subsequent selection is orientated towards the attainment of these goals. Listening contexts appear to be the primary variable that influences listener's identified goals in everyday interactions with music and by extension the function music is serving, whilst individual variables appear to be somewhat secondary in this respect. Individual variables such as personality traits may underpin contributory dimensions that influence the roles music plays in listening experiences, however, they do not provide extensive information about how contextual variables inform these individual responses since they are subject to longer-term traits, rather than the short-term states in which music listening readily occurs. In other words, they are not particularly useful in understanding differences in music selection in everyday life. Measures such as the AFML (Groarke & Hogan, 2018) and MUSIC scales (Rentfrow et al., 2012) are orientated towards individual variation and affect, but less about *utility*.

Preferences, however, may be associated with musical features, which have been used in psychological research as a measurement of music selection in everyday life, as well as preference itself (i.e., Greb et al., 2019; Rentfrow et al., 2012). This highlights that such measures can be applied to gauge features that convey the affective content of a piece of music. In turn, descriptors of the music that fulfils a given use could be leveraged to ultimately model its presence across contexts, hence drawing links between situational variables and features of selected music in everyday life. Moreover, since personality traits and music preference may both serve as predictors of preferred features, then it seems plausible that such attributes may in turn be utilised to predict efficacy in relation to desired listening outcomes via predicted liking for new music discovery according to short-term goals. It may be therefore possible to curate listening experiences by successfully predicting featural aspects of music according to situational variables (e.g., Greb et al., 2019).

Whilst this chapter has presented extant research to outline the ways in which music may facilitate psychological goals, and some of the factors that influence this at individual and situational levels, it has not fully delved into exactly how music may be employed across contexts themselves, and how FML may vary according to these aspects of the listening experience. The next chapter, therefore, aims to expand on this by focusing on the ways in which is employed across listening contexts in greater detail.

3.1 Contextual applications of music

It has been argued that situation-level variables are of greater importance than individual-level variables when it comes to determining FML and music selection in everyday life, and that research should shift focus from the latter to the former (Greb et al., 2019; Maloney, 2019). This is particularly relevant to the topic of utility as an affordance of technology (e.g., Hutchby, 2001), as there are seemingly endless possibilities for music engagement in the events, activities, and locations encountered in everyday life given the portability of modern listening technologies. Such factors influence the listening experience as, whilst the mechanisms involved in sound perception in the immediate environment are understood at the physiological level (e.g., Kostek, 2005; Lotto & Holt, 2011), listeners' cognitive interpretations, perception, and experience of music varies drastically between environments and situations. As Maloney (2019) argues:

> "…a piece of Mozart listened to within the concert hall environment is understood and interpreted very differently than the same piece heard over an elevator's speaker system" (p. 35).

This chapter therefore aims to delve further into context as an aspect of the listening experience, by addressing not only how listeners' goals and FML may change between situations but provide reference to the contexts in which music listening takes place and the affordances these provide. First, this aims to outline exactly what variables constitute the varied aspects of listening contexts.

3.1.1 The Contextual Triad

Researchers have considered context as an aspect of listening experiences in varied yet often inconsistent ways. Earlier work, such as that of Rentfrow and Gosling (2003) for example, used pre-determined categorisations of contexts in which music listening may occur, but that blur lines between aspects of context. This includes descriptions of context such as "alone at home", "getting up in the morning", and "hanging out with friends" (p. 1238). These categorisations conflate activities, locations, and temporal factors (e.g., presence of others, time of day), all elements of the situation that later work has segmented (e.g., Juslin et al., 2008; Greb et al., 2019). Similarly, North et al. (2004) described contexts accompanying music listening through such terms as "gym/while exercising" (p. 66), thereby conflating locations and activities in a similar way (e.g., listeners may exercise at home or in transitory spaces, such as during jogging). This highlights a relative lack of focus regarding the listening context in earlier work, whereby distinctions are not drawn between the relative aspects of a situation (i.e., location, activity, temporal factors). However, by giving attention to potential situations in which music listening may occur, this work seemingly acknowledges Konečni's (1982) argument that music listening does not take place in a vacuum, and therefore that contextual aspects should be considered influential in listener experiences.

In more recent work, it has been argued that listening contexts should not be defined by blended clusters of locations and activities (e.g., Rentfrow & Gosling, 2003; North et al., 2004), but should rather be thought of as a set of markers that may include physical locations, primary activities, and/or whether or not others are present (Juslin et al., 2008). Other variables may also contribute to the parameters of context, such as time of day, season, and weather, such that a combination of these features influence function and selection of music in everyday life (Maloney, 2019). Exceptions do exist of course; 'situationally rooted music' (such as music played during religious ceremonies) may be largely independent of these described factors due to social or cultural expectations taking precedence, however, such contexts are typically infrequent for most listeners day-to-day, given the autonomy provided by modern listening technologies (Maloney, 2019).

Variables like location and activity have often been treated by researchers as discrete indicators of the listening situation (e.g., Maloney, 2019; Greb et al., 2019). However, other approaches have instead viewed contexts through psychometric lenses, such as by measuring relative cognitive demands, rather than as discrete categorisations. For instance, Behbehani and Steffens (2020) investigated the relationship between the Situational Eight DIAMONDS and psychological characteristics of music: namely AVD. DIAMONDS is a taxonomy of situation characteristics identified by Rauthmann et al. (2014) which represent the key lines along which individuals perceive, describe, and evaluate psychological situations: Duty, Intellect, Adversity, Mating, pOsitivity, Negativity, Deception, and Sociality. These were applied by

Behbehani and Steffens (2020) according to situational characteristics of music listening, and their findings indicated that individuals adapt their listening behaviours according to the measures within the DIAMONDS model. During this process, they established a taxonomy of music listening scenarios which included, *positive-social*, *ambivalent-individual,* and *negative-demanding* situations. However, whilst this approach was able to predict certain applications of music listening according to situational goals identified through a taxonomic list, it did not consider other situational variables such as environmental context (Behbehani & Steffens, 2020). Nonetheless, this serves to contextualise that a variety of approaches may be applied in order to uncover and characterise the varied aspects that cumulatively formulate a listening situation.

Notwithstanding diversity in relative approaches, there remains the general consensus however, that music listening takes place in a triangulation between listeners, contexts, and music (Scherer & Zentner, 2001; Greb et al., 2018a). Maloney (2019) dubs this the 'Contextual Triad' (pp. 35- 36), and in particular focuses on a delineation between locations and activities. Greasley (2008) argues that activity is of particular importance as the main driver of function and by extension music selection, because it is the activity, rather than the location necessarily, that determines relative cognitive and/or social demands. Relatedly, Maloney argues that temporal aspects of the situation (e.g., time of day, weather) may augment listening experiences and functions, but display less direct impact than activities concurrent with music listening in particular. However, location has historically been a variable of strong importance when viewing contextual listening practices, and as such retains relevance and influence. To explore these further, therefore, the following provides a concise overview of these two central discrete dimensions of listening contexts.

3.1.2 Location

Location can be characterised as the place in or at which something occurs. Locations in which it is possible to listen to music are now more varied than ever before as an affordance of portable listening technologies, large-scale online repositories of music, and stable cellular networks (e.g., 4G). In particular, advancements in portable listening technologies have facilitated many common practices such as listening when travelling, when at home, or when at work (Krause et al., 2016; Maloney, 2019). Krause et al. (2016) found that listening in private locations was more common than listening in public, with portable devices being associated with positive responses that contrast strongly with responses to music being played aloud in public. This is consistent with earlier research that most listening occurs in private locales (e.g., the home), with a much smaller proportion occurring in overtly public places (North et al., 2004).

In general, however, research has been slow to interpret the relevance of music-selection into precise locations and vice-versa (Maloney, 2019). Notwithstanding a lack of in-depth analysis of music-selection within locations, attempts have nevertheless been made to identify broad themes. Maloney, for instance, groups together items from existing studies pertaining to particular locales of music listening (e.g., home, work, transitory spaces, gyms, urban environments, restaurants) and concludes "it would appear possible to allocate everyday behaviours relatively accurately to these locations" (p. 38). However, it is subsequently acknowledged that locations of music listening reveal little about the question of why people listen to music.

For example, Krause et al. (2014) found that perceived intensity in the effects of music listening may indeed vary between locations, for instance that music in the gym is perceived to be more motivating than music in a restaurant. However, this alone does not explain differences in the listener experience. Rather, there are situational demands that differ between the needs of a listener in the gym and a listener in a restaurant, and a change in location alone does not explain this. For example, certain locations appear more common in music listening; Maloney (2019) finds that home, transitory spaces, and work are the most common locations in which music listening occurs, yet it does not seem plausible that these factors alone evoke situational demands. Moreover, given portable listening technologies, it does not seem plausible that goals of music listening are constrained by locations either, but rather that listeners' needs may vary within the same locale (e.g., music listening may take place at home for the purposes of both relaxation when alone and to aid social interaction when listening with others).

Therefore, although location may serve as a contextual indicator, it cannot be argued to suitably explain the subsequent situational demands determined by listeners, and as such, Maloney's (2019) delineation of location and activity formulating the primary aspects of a situation is an important one, since it allows us to segment the environment in which music listening occurs from the actions being undertaken by the listener. Since it is the needs of listeners that determine engagement with music (e.g., Lonsdale & North, 2011; Greb et al., 2018a), it seems more plausible to consider that it is the concurrent activity that influences need and therefore function. At best, locations may be implicit of changes in FML, but are not the substantive driver of such changes. Rather, locations are accompanied by activities, and since these are argued to drive music selection, it seems plausible that that this is the primary driver underpinning changes in listeners' needs and selection, and that the location in which the activity and music listening occurs is generally more circumstantial than causal in this regard.

3.1.3 Activity

Like locations, activities that accompany music listening are more diverse today and hold fewer barriers than at any other point in the history of music curation. Unlike locations, however, activities are seldom fixed. In this sense, it has been argued that it is, therefore, cognitive goals informed by a concurrent activity that determines listeners FML and by extension music selection, rather than locations (Juslin et al., 2008; Maloney, 2019; Greb et al., 2019). Indeed, researchers have found that activities directly influence FML in everyday life, with common activities occurring alongside listening including being on the move, during housework, when working & studying, pure music listening, partying, relaxing & falling asleep, exercising, coping with emotions, making music, and social activities (e.g., Greb et al., 2018a; 2019).

Given the array of activities that plausibly have different requirements, congruency of the music plays a substantial role in determining whether listeners' experience of a piece of music alongside such activities is beneficial (Schäfer, 2016). For example, North and Hargreaves (2008) argue that music's congruence is subject to its ability to correspond to the listener's desired arousal (or arousal-based goals). They argue that listeners tend to prefer loud, fast music when in nightclubs, for example, whereas quiet, more gentle music is preferred when going to bed or sleeping. Though it has since been argued that featural aspects of music other than arousal help determine its congruence in everyday life (e.g., Behbehani & Steffens, 2020; Fricke et al., 2021), this nonetheless highlights that activity not only influences FML, but describable characteristics of selected music also.

Maloney (2019) identified 13 activity types that occur alongside music listening. Travel, working, chores, and relaxation were the most common to accompany music listening, and observes that "Activities, like locations, appear to display significantly associated functions that are somehow indicative of the activity or goal at hand …" (p. 236). Hence, if situational demands prompt a particular goal (e.g., to feel energised whilst running), then it is clearly the activity that determines the goal (and by extension function) and subsequent music selection. In other words, activities (like running) prompt goals and FML (like feeling energised). It would be practical, therefore, for researchers to standardise or be otherwise more consistent in the terminology used to describe listening contexts, since precise definitions of what is being referred to in research discussions are not always clear. Whilst this is a somewhat niche issue relating to the semantics of researchers' descriptions of context, it would be beneficial to have a more ubiquitous way of identifying what is and what is *not* being referred to. With this in mind, it should be considered that moving forwards in this thesis, the 'context' of music listening refers to a combination of situational variables. These may include locations and activities, however, as has been argued, locations are seldom useful indicators of FML, and whilst temporal features (e.g., weather) may enhance listening experiences, it is the activity of the listeners that is argued to prompt cognitive goals and by extension FML and music selection. When it comes to integrating context, therefore, this thesis is primarily interested in how these concurrent listening activities inform goals and FML.

In sum, activities, like music, are portable whereas locations are fixed. Rather, it is thought to be the activity that most plausibly determines the goals of listeners and thus selection and experience. For instance, although some locations (e.g., a gym) give rise to some activities (e.g., exercise), the activity itself is seldom constrained to the location unilaterally, although it should be acknowledged that "rarely does one attend the gym to perform chores or housework" (Maloney, 2019; p. 39).

3.2 Frequently Observed Situations of Music Listening

Aside from delineations between locations and activities, it is useful to consider some commonly observed situations in which music listening may occur from the literature. Research into contextual applications of music have generated positions that articulate broad contextual functions at varying levels of granularity. Lamont et al. (2016) discuss a series of contextual applications of music based on generalised groups of functions that include cognitive, physiological, and emotional effects of music listening. These broader groupings of FML are somewhat less well-defined than in other research, however, are practical in articulating insight into how some goals of music listening may be identified to generate listener aims in common situations that require certain effects. The following sections therefore outline some of these contextually orientated functional groupings, which typically refer to a combination of contextual requirements and activities accompanying music listening. This is with a view to further review contextual affordances of music listening in common scenarios that accompany music listening. For these purposes, the focus is primarily on why people listen to music according to certain situational needs, and so prior emphases such as distinguishing between locations and activities are more restricted for this purpose.

3.2.1 Travel

Travel is a situation that music accompanies more consistently than almost any other setting (Lamont et al., 2016). The technological advancements previously described have facilitated portable music technologies, enhancing music's integration into daily life quite generally with listening to music whilst on the move a particular benefit (Bull, 2006; Sinclair & Tinson, 2017). It is therefore no surprise that travel constitutes a major activity of heightened integration of music since there are few barriers to its employment in transitory spaces. For summary purposes, three subdomains of travel are briefly discussed: driving, public transport, and walking.

<u>Driving</u>

Music listening whilst driving has been possible for a far longer period of time than other scenarios of travel, with radios being a standard feature of cars since the 1930s (Loviglio, 2022). There is some evidence that when driving, music may enhance performance by assisting in the attainment or maintenance of appropriate levels of arousal and concentration, but it can also decrease driving ability as it makes attending to audio-visual signals (such as engine noise and warning signals) more difficult, whilst adjusting controls on audio systems also takes attention off the road (Lamont et al., 2016). Dibben and Williamson (2007) reported that, from a sample of 2,473 UK drivers, 87% selected the same music when driving as at home; with 62% reporting that music's ability to help them relax and feel calmer was the primary motivation for listening. A lower proportion (25%) said that music enabled them to concentrate better when driving, whilst drivers who had fewer accidents were those most likely to prefer quiet whilst driving (Dibben & Williamson, 2007). As such, music listening whilst driving may diminish the attention that drivers are able to focus on the road but may also serve distinct cognitive functions listeners perceive useful.

Public Transport

Another mode of transport in which music listening occurs is within public transit. Music listening on public transport is used to distract from a routine or "low-demand experience" (Lamont et al., 2016; p. 713). North et al. (2004) found that 60% of journeys on public transport utilised self-selected music listening through portable technology; with 85% of these listening episodes helping individuals to pass the time. Plausibly, given the time that has since passed since the time of North et al.'s (2004) study, music's application alongside public transport has only increased. Music listening through portable technologies facilitates isolation from external influences and other people (Bull, 2006; Kuch & Wöllner, 2021), which may help achieve certain positive effects from the listener's perspective, such as heightened senses of autonomy and environmental control, allowing them to become immersed in their own sonic sphere (Bull, 2006). However, there may also be adverse effects to this mode of listening, as research has indicated that iPod users, for instance, are less likely to engage in social interaction, acknowledge others or make eye contact (Garner, 2014). Overall, however, autonomous music listening during travel can help reduce anxiety and facilitate positive enhancements to the perceived environment (Skånland, 2011; Lamont et al., 2016).

Walking

Music has been reported to energise people more when travelling under their own steam (e.g., walking or cycling) than other forms of travel (Heye & Lamont, 2010). Compared to other modes of travel, walking may evoke slightly different sets of situational demands as the physical action it requires is different to say, public transport, which essentially involves either sitting or standing for an extended time-period without much (if any) physical action required. Conversely, Heye and Lamont (2010) observed no differences in music function between walking and public transport (via bus) but did find that time-passing was more important when on public transport than walking. This is plausibly due to more active elements of walking, such as cognitive and physiological demands, that are not required when on public transport (e.g., crossing roads, self-direction). In addition, there is evidence that music listening whilst walking may, as with driving, have negative effects as distraction leads to potentially unsafe walking behaviours (Mwakalonge et al., 2015). Finally, regarding music's psychophysiological effects when walking, Franěk et al. (2014) found that whilst musical beats and characteristics influence walking speed, they do not result in synchronisation between the listener and the beat itself. However, the authors did find that faster, more energetic music leads to increased walking speed, whilst slower, more relaxing music leads to slower walking speeds. This further highlights the impact of musical characteristics (e.g., tempo) on situational effect, and by extension, music selection depending on listeners' motivations.

3.2.2 Brain work

'Brain work' is a summary term used by Lamont et al. (2016) to refer to situations in which listeners engage with music whilst undertaking tasks that are cognitively demanding, such as concentrating during work or private study. Konečni's (1982) seminal argument, that music engagement and selection is influenced by tasks being undertaken externally is particularly relevant here, as more complex tasks require greater cognitive processing ability, and so music that is lower in arousal may be more conducive than music that is higher in arousal.

Specifically, some researchers have focussed on self-selected music listening in office spaces and during private study, as these are situations applicable to many people's daily lives (Lamont et al., 2016). Within office spaces, Haake (2011) found that 80% of office workers listen to music whilst at work, averaging some degree of music listening around 36% of the time spent at work. Respondents were found to have used music during routine individual tasks, such as word processing, internet-browsing, and emailing. Reasons that were cited included increased ability to focus on tasks (e.g., by blocking out unwanted distractions) as well as reduced stress levels, enhanced well-being, and a perception of a favourable working environment (Haake, 2011). Similarly, private study is a situation in which music is frequently used by both children and adults (Lamont et al., 2016). University students often use music to aid with study, although there are individual variations within this, as some find music beneficial for their productivity whilst others find it detrimental (Greasley & Lamont, 2011). However, listening episodes in which students engaged with music during private study were shown by Greasley and Lamont (2011) to be essential for maintaining focus and concentration. This was down to the application of music for the purposes of distraction and silence avoidance, which enabled removal of unwanted thoughts. Music listening for purposes such as this have been documented as effective strategies for mood regulation as well, as distraction may be conducive to overall mood (Saarikallio, 2008), and highlight music's comparative efficacy in cognitive regulation. Overall, music is often applied during day-to-day cognitive tasks like work and study, as this may help listeners concentrate and prevent distractions. Background affect regulation may be of particular benefit here also (Maloney, 2019), however, music must be efficacious in that it should prevent rather than become a new source of distraction (Greasley & Lamont, 2011), for which individual variations likely remain regarding content preferences (e.g., arousal).

3.2.3 Body work

Music has the ability to induce physiological as well as psychological states and may affect bodily processes such as coordination and motivation levels (Karageorghis & Terry, 2009; Laukka & Quick, 2013). Such physiological dimensions can be equally as important to listener experience as cognitive ones. For instance, North and Hargreaves (2000) found that different situations require different degrees of physiological arousal, which can be facilitated by music congruent to the levels of physiological arousal required (e.g., slow, and calming music attaining feelings of relaxation and energetic music resulting in physical stimulation). Moreover, listeners experience moderate bodily responses to music in everyday life, such as foot tapping and nodding. In addition to these external signifiers, many internal bodily processes also respond to music. These responses include heart and pulse rate, blood pressure, blood volume, blood oxygen, respiration, skin conductance, muscular tension, temperature, gastric motility, pupillary and startle reflex, and biochemical responses (Hodges, 2016). As such, physiological effects of music have an important role to play in the ways in which music is experienced by listeners. One limitation of studies measuring such physiological effects, however, is that many use researcher-selected music, whereas they should seek to recognise autonomy in music selection (Lamont et al., 2016).

Regarding physical activities in which listeners are in control of the music they listen to, music may be used to accompany exercise, relaxation, and pain management (Lamont et al., 2016). Other uses include employment alongside domestic chores such as washing, cleaning, cooking, and gardening, to which music is credited with providing physical motivation (Greasley, 2008; Greasley & Lamont 2011; Lamont et al., 2016). As with other uses, it has been argued that the congruence of music within physical activities is crucial, with appropriate energisation and entertainment levels being of key importance (Lamont et al., 2016). Researchers describing primary reasons for music use in such scenarios list arousal, emotion regulation, motivation, performance, and induction of flow states as essential factors that determine utility and congruence of music (Laukka & Quick, 2013). However, there have nevertheless been significant variances between groups as to what kind of music is appropriate during activities like exercise, with distinctions drawn between those wishing to use music for the purposes of distraction and those aiming for enhanced focus and motivation via music (Hallett & Lamont, 2015). Furthermore, like with other activities, the effects of music on users during exercise may not always be beneficial as it may enhance both positive and negative experiences (Hallett & Lamont, 2021).

In a broad sense, this highlights the fact that not only do FML vary between listening contexts, but also that the induced effects may not be exclusively desirable or beneficial towards the attainment of goals. Another example of the influence of variance in music's functional ability can be seen in the assertion that 'socialisers' may be distracted by music and or social interaction, whilst 'workers' aim to maximise their efforts during exercise and may select more focused approaches through music listening (Lamont et al., 2016). These mixed degrees of positive and negative outcomes reinforces the need to acknowledge music's variability in terms of not just goal-orientation, but also goal-attainment; in other words, disparity between the listener's intent and the actual effect that music may have on the situation.

3.2.4 Emotional work

Music's use for the purpose of mood regulation is one of the most common functions of self-selected listening (Saarikallio & Erkkilä, 2007; Saarikallio 2008; van Goethem & Sloboda, 2011; Groarke & Hogan, 2016; 2018; Karreman et al., 2017). At a basic level, people often choose to listen to music they like because it makes them feel good and experience positive emotions (Juslin & Laukka, 2004), and the relationship between music and emotion is a cornerstone of music psychology research. It is therefore useful to not only articulate how music's influence on emotion translates to its application in everyday life and experience, but to also provide a concise overview of the relevant theories and empirical research that underpins this understanding. Therefore, this section is divided between a concise overview of music and emotion in general terms, supplemented with relevant discussion into the role of mood management in everyday listening practices. Given the vast body of research on music and emotion, the following discussion is necessarily rather summary, however, readers interested in exploring this topic in greater detail may refer to Juslin and Sloboda (2011) for an accessible introduction.

3.2.4.1 Music and Emotion

Music's ability to induce emotional responses in listeners is perhaps one of the most intriguing effects it has on the human experience. Emotions belong to the field of *affect* (a catch-all term covering various affective phenomena) and their defining feature is *valence*; the evaluation of objects, people, or events as being positive or negative (Juslin, 2016). Congruence is the perceived fit between emotional and cognitive expressions of music on the one hand, and an expressed identity of a message on the other (Herzog et al., 2020). This is, however, subject to disparities between perceived and induced emotions. Other than through lyrics, music cannot express coded semantic content, but rather expresses 'connotative meaning' in terms of moods and emotions (Herzog et al., 2020). Researchers have aimed to capture both the perceived *expressive* and induced *emotional* qualities of music (e.g., Zenter et al., 2008), resulting in terminologies that describe *affective tuning* and *aesthetic character*, which are perceived as attributes of music itself. However, this does not mean that such features necessarily correspond to the *felt* emotional effects in the listener (Herzog et al., 2020). *Induced* emotions describe personal feelings that occur as a result of contact with music, but which are not necessarily identical to the expressed affective content of the music. Moreover, Juslin and Västfjäll (2008) identified a series of underlying mechanisms that may explain music's ability to elicit emotional responses: *brain-stem reflex*, *rhythmic entrainment*, *evaluative condition*, *emotional contagion*, *visual imagery*, *episodic memory*, and *musical expectancy*. The literature suggests therefore, that there is no one underlying mechanism that may predict the emotions elicited by music. Rather, a variety of mechanisms are at play, and that the presence of these mechanisms may vary depending on the individual, music, and contextual factors relating to the listening experience. This has led to debate, therefore, of whether affective responses to music predominantly stem from recognition or affective experience.

Broadly speaking, two camps emerge: cognitivist researchers, who argue that listeners recognise rather than feel affective responses to music, and emotivists, who contend that listeners have intrinsic affective experiences to music (Hill & Palmer, 2014). Lundqvist et al. (2009) conducted a study with the aim of resolving the debate between 'cognitivist' and 'emotivist' positions. Their study consisted of measurements pertaining to self-reported feelings, facial muscle stimulation, and automatic activity in participants whilst listening to contemporary-style pieces of music that either expressed 'happy' or 'sad' emotions to create ideal conditions for emotional contagion responses to occur. 'Happy' music was found to generate more zygomatic facial muscle activity, higher skin conductance, lower finger temperature, more experienced happiness, and less experienced sadness than the 'sad' music (Lundqvist et al., 2009). These findings were significant insofar as the emotional expressions of the music were consistent with the induced emotions within listeners. This supports the theory that music is able to evoke emotions through emotional contagion; supportive of the emotivist position. Furthermore, the notion that the music used in this study was specially composed and not previously known to the participants suggests that previous experience with a particular piece of music is not determinant of whether there are emotional responses to music, but rather that it is the characteristics of the music itself that help to determine the emotional response. This is highly relevant insofar as it opens up the possibility to predict emotional responses to pieces of music. On the other hand, it subverts the practical experience of listeners who hold preferences and *past functional experiences*, which contribute to listening experiences in everyday life.

It is commonly believed that affective states can be influenced by expressive characteristics that people are prone to infer onto inanimate objects within the environment, such as room colour and weather (Davies, 2013). More precisely and with regard to music, Juslin (2016) describes emotional contagion as a process in which an independent brain module responds to stimulus features as if they were coming from a human voice that expresses an emotion, which in turn results in an individual mimicking that expression internally. Contagion is, therefore, a process whereby emotions are evoked because the listener *perceives* the emotional expressions of the music and subsequently imitates this expression internally (Juslin, 2001).

As a subsection of research into emotional responses to music, contagion has generally been approached with regards to physiological responses to music such as facial expressions, however, it has been discussed that contagion may also occur in response to speech, as music features patterns of sound that resemble speech (Juslin et al., 2010; Juslin, 2016). This may be because humans become aroused by aspects of music akin to voices as a consequence of a process whereby a neural module responds automatically and rapidly to particular stimulus features of music, leading us to internally mimic the perceived emotion (Juslin, 2001). Moreover, brain images have shown that music activates brain regions that are associated with pre-motor representations of vocal sounds (Koelsch et al. 2006; Juslin, 2016). Contagion as a mechanism is therefore plausible since most music that is heard by present day listeners is highly vocal, but also that voice-like instruments (e.g., violins, cellos) may arouse emotional responses in individuals similarly to vocal content (Juslin, 2016). The significance of emotional contagion with regard to the present research is in the implication that music is able to induce emotions that correspond to its perceived emotional expression. Whilst it should be acknowledged that this is just one of a variety of emotion induction mechanisms that relate to music listening (e.g., Juslin and Västfjäll, 2008) it nevertheless demonstrates the association between musical qualities and the effects of music experienced by listeners.

3.2.4.2 Music and Emotion in Everyday Listening

With regard to music's application to everyday life, a range of mood regulatory strategies have been delineated in the literature. These include *entertainment*, *revival*, *strong sensation*, *diversion*, *discharge*, *mental work*, and *solace* (Saarikallio, 2008); as well as general maintenance and enhancement of happy moods and 'psyching up' for specific tasks or activities (Saarikallio, 2011; p. 312). Van Goethem and Sloboda (2011) found that active engagement with music made participants aware of the strategies and tactics by which they use music to regulate their moods. Generally speaking, older listeners are more likely to be self-aware of how music may fit situations and moods, with women being more likely to use music in mood regulation than men (Saarikallio, 2011; Lamont et al., 2016).

One counter-intuitive application of music listening is active engagement with sad music. It may seem that, as happy sounding music induces happiness in listeners, sad music may induce sadness. Deliberately listening to sad music does not, however, necessarily induce long-term negative affect and proves popular amongst listeners despite expectations this may lower mood (Garrido & Schubert, 2011). Huron (2011) proposed that sad music provides opportunities to feel positive over an extended period of time due to the causation of crying which leads to the production of prolactin, a hormone associated with feeling comforted that may leave people in more positive moods afterwards. Relatedly, Larsen (2000) presents the theory of delayed hedonic gratification, whereby certain actions may not immediately improve mood but promote positive emotions over extended periods of time. Sad music, for example, may increase melancholic feelings or negative experiences in the short-term, but overall may increase happiness by allowing listeners to reflect on negative experiences and gain understanding and clarification of such experiences (Saarikallio & Erkkilä, 2007). Van den Tol and Edwards (2015) reported that motivations amongst listeners for listening to sad music after experiencing adverse or negative events, and that music selection was linked to individual's identified self-regulatory goals and the expected effects of music listening. Additionally, it was found that if listeners intended to improve moods through listening, this was achieved by initially experiencing cognitive reappraisal or distraction (Van den Tol & Edwards, 2015).

It has been noted that using music to facilitate regulated mood states (that may be enhanced, maintained, or altered) is extremely common and widely practised (Saarikallio & Erkkilä, 2007; van Goethem & Sloboda, 2011). However, Maloney (2019) argues that "emotional regulation and self-regulation only accounts for a small portion of what music-facilitated goal attainment is capable of" (p. 30). Whilst music-facilitated goal orientations certainly include regulatory strategies, it is not limited to these strategies. 'Regulation' is one element of music's *functionality*, and such terms like *functionality* are appropriate when discussing the wider concepts of goal attainment through music listening as it includes regulatory strategies, but is not limited to them (Maloney, 2019).

3.2.5 Live events

The contexts and FML discussed so far focus on listener engagement with recorded music. However, individuals also report strong experiences when listening to music in live settings (Lamont et al., 2016). The primary reasons individuals engage with music in live settings is to affirm, or challenge, existing taste; as well as enhance personal and social dynamics to feel part of a community (Lamont et al., 2016). However, whilst live events are a core dynamic of music engagement, it is separated from contextual *functionality* with regards to this research. The present focus is toward perceived autonomy or appropriateness of music listening in particular scenarios. Although live settings are a context in which music listening is of course central, it is not the aim of this particular research to explore what manner of live performances are most effective in achieving situationally determined goals, as live experiences are accompanied by a myriad of underlying factors and dimensions specific to that context (e.g., O'Neill & Egermann, 2022). For this reason, live events are set to one side, as this mode of music engagement falls outside of autonomous selection of recorded music. This raises the question, however, of what autonomous music listening actually means. Not in the sense of just being able to select music in a manner facilitated by personal listening technologies, but also at the perceptual level.

3.3 Autonomy in music selection

It is worth noting the influence that degree of control may have over users' daily listening experiences, because the intentional selection of music inherently requires agency in the listener (Krause et al., 2014). Krause et al. (2015) observed that the three most common means of users accessing music in daily life are through radio, mobile MP3 players, and computers, all of which offer control regarding the presence (or lack of presence) of music. They note that devices that allow personal input are often met with more positive responses from listeners, suggesting that the greater the perceived autonomy afforded by technology, the greater the extent to which complex patterns of everyday use are observed.

Such findings are congruent with Krause et al. (2020), who in turn argued that music's use in everyday life roughly corresponds into Mehrabian and Russell's (1974) Pleasure-Arousal-Dominance model, which states that interactions and interpretation of an individual's surrounding result from variation in the three factors: pleasure, arousal, and dominance. With regards to music listening, Krause et al. (2020) propose that pleasure is operationalised by how much an individual enjoys the music they are listening to, arousal as the extent to which the music arouses the individual, and dominance as being characterised by the individual's control over the music that is heard. Krause and Brown (2021) furthered this by identifying eight underlying dimensions that fulfil the uses and gratifications that listening formats serve: *usability and intention to use*, *discovery*, *functional utility*, *flexibility*, *connection*, *social norms*, *value for money*, and *playback diversity*. Once again, this reinforces the prominence of technology on perceived use and satisfaction regarding music listening in the 21st century. It is the affordances of such technologies to provide autonomy that has led researchers to explore FML in recent years (e.g., Greb et al., 2018a; 2018b; Maloney, 2019). As such, this thesis considers the ways in which, in modern settings, situational factors influence listeners' FML and by extension music selection as an affordance of technologically facilitated listener autonomy.

3.4 Chapter Summary

This chapter has aimed to outline how contextual factors influence listeners' FML in everyday life, and how these in turn may influence music selection as a function of goal-attainment. Researchers have argued that situational variables are more important determinants of FML than individual-level constructs, implying that contexts of music listening primarily determine uses of music in everyday life. In general terms, the primary variables that constitute 'contexts' may be delineated by locations and activities (Maloney, 2019). In particular, different activities require diverse degrees of cognitive and physiological stimulation to induce desired effects. Such effects are determined by listeners' situationally determined goal during said activity, and music attaining desired effects is (typically) perceived as being a beneficial accompaniment to the activity. These may include emphasised states of physiological arousal, for instance, in which regard the listener may aim to feel energised during exercise or subdued when relaxing. From the cognitive perspective, some listening activities require alternate levels of concentration, and so listeners' goals may shift according to the degree in which they are seeking to focus or concentrate, for example. In this regard, FML may change according to whether the aim is to prevent distraction and to act as background stimuli (such as during work or private study), or to entertain and raise the emotional state of an activity (such as within social settings).

Locations, on the other hand, appear to be related to the activities during which music may be applied, but are of somewhat limited use when it comes to explaining variability in FML. This is because single locations may accommodate multiple uses of music depending on the activity being undertaken in that locale. For instance, when looking at the situational features of listening practices, Maloney (2019) reported that the most common listening location was in the home, whilst some of the common listening activities included: Working, Chores, Relaxation, and Exercise. It is plausible that any of these activities may occur within the same location, which is to say that activities are not exclusive to or limited to one single location. As such, it is more conducive to focus on concurrent activities as drivers of FML and music selection, rather than locations (Maloney, 2019). Research has additionally highlighted other factors affecting listening practices that fall outside of these dimensions. For instance, being with others, degree of control over music selection, time of day, and weather can also play a role in situational music selection (Greasley & Lamont, 2011; Greb et al., 2018a; Krause & Brown, 2021). These temporal features may enhance information about contexts but are somewhat secondary when it comes to determining FML and music selection in the first place, for which activities are argued to be of primary importance (e.g., Maloney, 2019).

The first two chapters of this literature review have provided an overview of music's role in everyday life, and the ways in which different situations influence these roles. Next, this review somewhat pivots into the second avenue of interest in the current project, which is that of music curation. Since technological advancements have greatly changed the ways in which music may be used in everyday life, there is substantive value in understanding how and why such technologies operate to influence such a complex set of behaviours as personalised music listening. The next chapter, therefore, provides an overview of recommender systems as artefacts of music curation in modern listening practices, and conceptualises applications of psychological research in such systems for the purposes of real-time music curation based on contextual information.

4.1 Music Curation and Recommender Systems

The variety in the literature regarding the ways in which people interact with music is dependent on means of access, specifically the technologies that curate music listening (e.g., Krause et al., 2015). Consistent with *uses and gratifications* approaches, Krause and Brown (2021) note that when listeners select music, their decisions are subject to particular affordances of media formats that are applied to satisfy the needs of the listener (i.e., *uses*). In turn, *gratifications* refer to the perceived fulfilment of those needs. Other domains of psychological research on music listening also highlight the crossover between music listening and listening technologies. For example, Greenberg et al. (2016) and Bansal et al. (2020) both express implications from their research on music preferences for the development of MRSs. However, given that recent research, such as that of Greb et al. (2019), has found that individual-level variables (like preferences) are of lesser importance on listeners' music selection compared to situational variables, there remains an underexplored gap in the literature on the crossover between music selection as determined by the situation the listener is in, and effective means of curation that maximise the functional efficacy of listening technologies.

The chapter, therefore, aims to introduce and articulate the relevance of music curation (in particular MRSs) to the uses and functions of music in everyday life. This is primarily motivated by the implied relationship between *functionality* and technology and serves to contextualise practical means of curation as well as avenues of opportunity to which psychological research into FML may be of benefit. Given that this subject requires some prior knowledge of computational music research, however, this chapter is structured to first provide a concise overview of the field of MIR in generalised terms, which is then used to inform an overview of what recommender systems exactly are and how they operate, as well as present limitations and future directions. This thesis ultimately argues that there are opportunities for integration and collaboration of computational and psychological research in this area moving forward, enhancing the efficacy of music curation via recommender systems by integrating knowledge from the social sciences. These outlined steps essentially funnel down from the high-level overview provided, to the granular area of relevance to this thesis just hinted at.

Firstly, this chapter characterises the field of MIR as a research domain relating to music curation via listening technologies.

4.2 Music Information Retrieval: Overview

An exhaustive discussion of each technological development that has led to the incremental integration of music in everyday life is not practical within this literature review, given the scope of such a discussion. Rather, it is more pragmatic to highlight the most relevant technological advancements, namely the encoding and compression of audio signals into file formats (such as MP3s) which, with the increased use of the Internet as a means of communication and distribution, initiated profound changes to the ways in which music is accessed and used by listeners beginning in the 1990s (Knees et al., 2019). Such advances have provided listeners with means to engage with large-scale music repositories and access individual pieces of audio on demand. Repositories have continued to expand over time, culminating in the widely used cloud-based streaming services now familiar to many (Schedl et al., 2014). This holds significant relevance to the prior discussion in Chapters 2 and 3 because it is only through means of technological advancement that prior barriers to portable music listening have been significantly reduced. Moreover, as previously outlined, modes of access to music influence its *functionality* (i.e., Krause & Brown, 2021).

The trajectory from the encoding of raw audio signals to cloud-based repositories has had many notable phases. Knees et al. (2019) note that the most popular and intuitive interfaces for navigating early music repositories utilised metadata (e.g., titles and artist names). When repositories were small, such interfaces were effective and usable, however, as collections grew, user interfaces had to adapt to continue curating music discovery, requiring retrieval, classification, and organisation of music. In the early 2000s, researchers in MIR shifted from prior focuses on symbolic representations of music (e.g., digital representations such as Musical Instrument Digital Interface, or MIDI) to the processing of raw audio signals as an affordance of increased computational power enabling readily available applications for signal processing (Schedl et al., 2014). Because of this paradigm shift, definitions of MIR may vary in their emphases. Kostek (2005), for instance, refers to MIR as the process of extracting and retrieving data from musical databases found on the Internet, whilst Downie (2004) defines it as a “multidisciplinary research endeavour that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world’s vast store of music accessible to all” (p.12). Hence, although points of emphasis and access might be inconsistent, MIR may generally be considered as being concerned not only with automated learning about music, but also with curation and discovery.

From the computational perspective, music is considered a multimodal human artefact that may take the form of audio, symbols (e.g., scores), text (e.g., lyrics), images (e.g., photographs and album covers), and gestures (i.e., of a performer), with it often experienced as a combination of these aspects (Schedl et al., 2014). This information may be subsequently used in applications, including for the purposes of curation and recommendation. Also, our perception of music, and of music similarity, is influenced by these factors, as well as by diversity of lyrics, beats, perceptions of performers, and the mental states of users (Schedl et al., 2013). Computational techniques are applied in many MIR applications to learn about such features and describe music by aspects of various categories (e.g., music content or features, context, and user properties; Schedl et al., 2014). The practical use of such information prompts analysis of various data sources (e.g., web pages, blogging) to generate tags that collate information about music for later use, such as in search and retrieval and recommender systems. MIR is therefore often concerned with the extraction and inference of this meaningful content, which can be used to index music and develop search and retrieval mechanisms to facilitate navigation to desired content (Schedl et al., 2014).

Moreover, increased computational power in signal processing has enabled automated-tagging of features to digitised pieces of audio directly, which in turn may be shared and distributed through file-sharing and music libraries or repositories. Thus, stages of development within MIR for the purposes of music curation can be summarised in four notable advancements: (1) development of digitised audio compression in the 1990s; (2) the ability of users and applications to extract audio features from music in reasonable timeframes; (3) the widespread availability of mobile music players; (4) the emergence of streaming services (Schedl et al., 2014).

By analysing and attributing features to audio directly, the information gathered has proved useful in making inferences and predictions about the ways in which listeners engage with music, and the effects this will have (Kaminskas & Ricci, 2012). As has already been discussed, music serves different functions in different situations (e.g., Greb et al., 2018a; Maloney, 2019), and music with different affective content may fulfil or satisfy different contextual needs, such as by eliciting the appropriate amount of physiological stimulation during exercise, or reflecting emotional cues (e.g., North & Hargreaves, 2000; Lamont et al., 2016; Barone et al., 2017; Greb et al., 2019). Relatedly, MIR researchers have, in recent years, argued that the listening context should constitute an essential focus in developing responsive and dynamic MRSs in light of increasing accessibility in everyday situations accompanying short-term demands (Wang et al., 2012; Takama et al., 2021), for which different formats have differing degrees of suitability (e.g., Krause & Brown, 2021). Therefore, there is substantial crossover in *functionality* from the listeners' perspective, and curation from the perspective of providers. It is at this intersection to which the discussion now proceeds.

4.3 Music Recommender Systems

In the last two decades or so, a central aim of MIR researchers has been to develop technology that assists consumers in finding music in light of increasing digitisation (e.g., Downie, 2004). MIR engages with the ways humans interact with technology through user interfaces and aims to satisfy accuracy of music curation and discovery amongst listeners (Knees et al., 2019). A common means by which this is realised, as has been alluded to, is by providing listeners' with music recommendations via automated processes.

As outlined briefly in the introduction to this thesis, recommender systems can be thought of as a combination of software tools and computational techniques used to provide users with item suggestions likely to be of interest (Resnick et al., 1994; Ricci et al., 2015). An 'item' refers to any product, service, or content that a system suggests (e.g., news articles, shopping items, films, music). This is typically useful in circumventing overwhelming repositories of items that might be available on a given platform, and/or in locating items users need or are otherwise more likely to be interested in (Ricci et al., 2015). This has become particularly necessary for MRSs in recent years as online streaming services now host tens of millions of pieces of music, and so filtering this abundance by providing recommendations from the repository limits choice overload (Schedl et al., 2018). To complete computational tasks such as this, recommender systems often use advanced machine learning techniques to make predictions about preferences for, and utility of, the items they suggest (Schedl et al., 2022). Such approaches use data provided by users either explicitly (e.g., ratings and reviews of certain products or user surveys) or are inferred by the interpretation of user interactions with the system or repository; referred to as *explicit* or *implicit* feedback respectively (Ricci et al., 2015; Lex et al., 2021). Machine learning techniques typically partition datasets containing this feedback data on which a predictive model is first trained, before testing the accuracy of those predictions on a subset of data outside of the training model to assess the prediction accuracy of the trained model (Shmeuli, 2010; James et al., 2017). Such methods are applied in recommender systems to estimate or predict items that users may like but have not yet engaged with. Given the high-dimensionality of such endeavours, however, this typically requires large sets of data to generate accurate predictions and prevent overfitting (when the training model is too dependent on the data on which it was trained and generalises to out-of-sample predictions poorly; Rajput et al., 2023).

The need for recommender systems, and by extension predictive accuracy, follows the simple observation that individuals often rely on the recommendations of others when making routine, everyday decisions (Resnick et al., 1994). To provide users with effective suggestions, a recommender system essentially needs to be able to assess whether an item is worth recommending to a user or not. To do this, "the system must be able to predict the utility of some items, or at least compare the utility of some items, then decide which items to recommend based on this comparison" (Ricci et al., 2015; p. 10). By way of example, Ricci et al. (2015) describe a simple, non-personalised system that recommends the most popular songs from a given repository. If a new user interacts with the repository, and there is an absence of precise information about the new individual's preferences, a popular song (i.e., one liked by many users) will be considered most likely to appeal to the new user, at least when compared with the likelihood of other, less popular songs. As a result, the utility of the most popular song(s) is predicted to be higher for the generic user compared to less popular songs.

As stated, however, this is the case when recommendations are non-personalised, or there is no information about the user's preference or taste. Non-personalised recommendations are, at best, effective in generating broadly appropriate recommendations to large numbers of users, but are not likely to provide high-resolution, perfect recommendations for any single user. To generate personalised recommendations, systems require more precise user data, be that *explicit* and/or *implicit*. The resulting system may then operate according to information about three object types: *Items*, *Users*, and *Transactions* (the relationships between *Users* and *Items* gathered through direct human-computer interaction; Ricci et al., 2015). Similar to data-logs, *Transactions* may include references to *Items* engaged with, descriptions of context (e.g., a search query), or through *explicit* and *implicit* feedback. Data such as this help recommender systems curate the suggestions made to users according to the approach of the system.

When it comes to MRSs in particular, common goals include *accuracy* (recommendations matching music preferences), *diversity* (in contrast to *similarity*, since users tend to be more satisfied with recommendations when they are somewhat diverse), *transparency* (the trust users have in systems when it is understood why certain pieces of music have been suggested), and *serendipity* (a measure of how unexpected a recommendation is; Schedl et al., 2014). To this end, two broad MRS classes are used: *collaborative filtering* (whereby if *Listener A* has similar music preferences to *Listener B*, then songs liked by *Listener A* not yet considered by *Listener B* will be recommended to *Listener B* and vice versa) and *content-based* methods (whereby if *Listener A* likes song *S*, then songs with similar features to *S* will be recommended to *Listener A*). *Hybrid* methods are also often used, which blend *collaborative filtering* and *content-based* approaches (Wang et al., 2012).

4.3.1 Uses and methods of collating audio features

Collaborative filtering closely resembles the novel conceptualisation of the need for recommender systems provided by Resnick et al. (1994), whereby individuals require the input of others in everyday life for discovering new items of media (e.g., music, television, films). In *content-based* systems, however, characteristics of musical content (e.g., descriptive audio features, tags, metadata) are leveraged, from which resulting comparisons with user preferences are made (Bogdanov et al., 2013). Specifically, *content-based* recommender systems rely on the item or user descriptions obtained via extraction methods to build representations of items, and also user profiles to suggest items similar to those a user previously liked or engaged with (de Gemmis et al., 2015).

Algorithmic analysis of audio (or symbolic) data is used to extract and infer meaningful features of musical content (Schedl et al., 2014). Feature extraction methods are based on the computation of time and frequency representations of audio signals (see Schedl et al., 2014 for a comprehensive overview; pp. 145-173), and are employed to address a varied range of problems, such as beat detection, automatic music transcription, artist recognition, and genre classification, as well as music recommendation (Lamere, 2008). The latter of these, of key relevance to this thesis, leads to the indexing of music using extracted features and becomes integrated into search and retrieval schemes for subsequent recommendation purposes (Schedl et al., 2014). Miotto and Lanckriet (2012) characterise audio features, or tags, as "keywords or short phrases that capture relevant characteristics of music pieces, ranging from genre and instrumentation, to emotions, usage, etc" (p. 1096). The utility of audio features stems from the need to operate the search and query systems mentioned above, but also from the dependence of *content-based* recommender systems in particular on audio-tagging via semantic content to operate (de Gemmis et al., 2015).

It should be understood that music conveys cognitive meaning within individuals whom in turn seek to classify and describe it in linguistic terms, albeit in varying ways (Swain, 1996). Put simply, music evokes thoughts which can be articulated through speech or some other form of communication (e.g., writing), and 'semantics' is the term used to describe the associated meaning. Koeslch et al. (2004) evidenced that at the neurobiological level both music and language are able to prime the cognitive meaning of words and that music, like language, determines physiological indices of semantic processing in the brain. This is demonstrative that music influences the processing of words, but also represents meaningful concepts in abstract or objective terms, independent of the emotional content of such concepts (Koelsch et al., 2004). However, it is also noted that music and language do not necessarily share identical semantics, since individuals do not seem to have the vocabulary to convey "thoughts and intentions musically as well as they do linguistically" (p. 306). Nonetheless, Jentschke (2016) notes that both music and speech are naturally sequential auditory signals that unfold in real-time "according to the rules of syntax and harmony" (p. 349), implying that there is a distinct relationship between music and language which furthers the notion of conveyed meaning in descriptions of music as well as music itself (even if there may be discrepancies between the two).

Notwithstanding such discrepancies, researchers have pragmatically utilised syntax as indicators of meaningful content and thought elicited by music, albeit with the caveat that such descriptions may not fully encompass individuals' full experience. It has been discussed in the literature for quite some time that MIR has taken particular interest in music databases which are based on systems that, beside music, provide machine-processable semantic descriptions which may be implemented in recommender systems. As Kostek (2005) put it, "The semantic description is becoming a basis of the next web generation, i.e., the Semantic Web" (p. 281).

MIR may follow several orientations in extracting musical features, and it has been proposed by Grosche et al. (2012), for instance, that tagging systems may be classified according to *specificity* (with high levels of specificity intended to identify a given audio signal and low levels of specificity to generate statistically or categorically similar pieces of music) and *granularity* (with large granularity retrieving complete pieces of music and small granularity locating specific time locations or fragments). Such parameters have been inferred onto commonly used MIR methods (or tasks) such as *Audio identification*, *Audio alignment*, *Cover Song identification*, and *Query by humming* (or *tapping*). Tags may be expressed linguistically (e.g., "happy" or "rock"), in which respect music is one of the most commonly referred to phenomena. The semantic-based retrieval systems used in this process are dependent on the accuracy of these methods in estimating labels from pieces of audio that match human experiences. Often, these are "characterised by a low specificity and long-term granularity" (Schedl et al., 2014; p. 134).

Methods by which semantic terms are selected for audio tagging, however, are varied and ambiguous to say the least. One example of a semantic search engine is *SearchSounds* (Celma et al., 2006) which utilises user-generated content from blogs to identify music through text queries, the results of which are expanded through audio features. In other research, however, the terms used in content analysis may be initially identified by music "experts" (e.g., Rentfrow et al., 2012; Lepa et al., 2020) which could then be extracted by algorithms using audio as an input (Kostek, 2005). One limitation with this is that the term "expert" is seldom defined clearly within the relevant literature and may be subject to biases and subjective interpretations of the semantic descriptors used in tag acquisition.

That semantic terms are often subjective is a recurring theme, and as such collecting high-quality tag data is difficult with a lack of vocabulary standards making it hard to define a cohesive tag acquisition and quality assessment stratagem (Kaminskas & Ricci, 2012). This constitutes a general problem, and circumventing the "semantic gap", mentioned by Schedl et al. (2014), is something that may be eventually achieved by implementing reliable methods to gather semantic data.

Such steps are important in generating consistent content-based recommendations in particular, since poorly defined methods of gathering semantic descriptors influences the output of one model compared to another, making it difficult to assess which is most effective at representing users' taste and intentions. In any case, the complexity and nuance of variations between regions, cultures, and languages remains, leaving a unified approach to tag acquisition a complex issue. Though there are clear limitations to tag acquisition, audio features are often applied as far as is practical for search and retrieval purposes and are also implemented for the purpose of music curation via recommender systems. Regarding example approaches, Turnbull et al. (2008a) present five methods to collect tags for music, respective the pros and cons of which are shown in Table 3.

Table 3 Turnbull et al. (2008a) strengths and weaknesses of tagging methods

Approach	*Strengths*	*Weaknesses*
Survey	custom-tailored vocabulary, high-quality annotations, strong labelling	small, predetermined vocabulary, human-labour-intensive, time-consuming approach lacks scalability

Social Tags	collective wisdom of crowds, unlimited vocabulary, provides social context	create and maintain popular social website, ad-hoc annotation behaviour, weak labelling, sparse/missing in long-tail
Annotation Games	collective wisdom of crowds, entertaining incentives produce high-quality annotations, fast paced for rapid data collection	"gaming" the system, difficult to create viral gaming experience, listening to short clips rather than entire songs
Web Documents	large, publicly available corpus of relevant documents, no direct human involvement, provides social context	noisy annotations due to text-mining, sparse/missing in long-tail, weak labelling
Auto-tagging	not affected by cold-start problem, no direct human involvement, strong labelling	computationally intensive, limited by training data, based solely on audio content

Note. Adapted from Turnbull et al. (2008a).

It is suggested that whichever method is applied, extensive vocabularies are preferable in generating tags since fixed vocabulary limits retrieval to a small set of predetermined tags. Additionally, vocabulary is preferably structured since ontological relationships (such as genre hierarchies) between tags encode further semantic content useful in retrieval practices (Turnbull et al., 2008a). The ability of vocabularies to be dynamic is important in semantic retrieval, where semantic content is related to aspects of audio content, such as instrumentation or genre. These may largely be agreed upon by listeners, implying that computational models can learn the relationship between the two (Turnbull et al., 2008b). In any case, systems often automatically annotate songs by modelling the characteristic acoustic patterns of each song, that are in turn associated with the tags that formulate a vocabulary (Miotto & Lanckriet, 2012). These principles have prompted a shift towards concept-based representations of items and users within content-based recommender systems, which integrate techniques from Natural Language Processing (NLP) as well as Semantic technologies (de Gemmis et al., 2015).

With regard to recommender systems, Miotto and Lanckriet (2012) note there are two primary methods to applying semantic tags: (1) keyword searches (e.g., "mellow rock songs with acoustic guitar") and (2) example-based retrieval based on semantic representation (e.g., generating playlists based on songs with similar annotations). The latter of these is more pertinent to this discussion since it is a direct implementation of semantic retrieval in recommender systems. Ferrer and Eerola (2011) found that semantic structures in music (such as affects, and instrumentation) have certain timbral characteristics, which were then found to be associated with perceived timbral qualities. It is implied that it is possible to derive useful semantic structures that transcend genres and can be linked to acoustic features, which can in turn be computationally operationalised (Ferrer & Eerola, 2011). Such computational processes often refer to automatic annotation based on audio content, which is used by auto-taggers by associating acoustic patterns with tags in a given vocabulary. Auto-taggers generate a vector of tag weights when annotating a new song, which may be interpreted as a *semantic multinominal* (SMN; Turnbull et al., 2008b; Miotto & Lanckriet, 2012). Miotto and Lanckriet (2012) expanded on previous applications of SMNs by capturing tag correlations and implementing a model whereby each tag is considered to define a broader context (e.g., as referring to genres, such as "rock"). This can be used to make inferences about broader properties of a piece of music, in an aim to mitigate the noise from descriptors. An example of the relationships between correlated tags is shown in Table 4.

Table 4 Top-5 Co-Occurring Tags for a Sample of CAL500 Tags (Miotto & Lanckriet, 2012)

Tag	**Top-5 co-occurring tags**
Hard rock	Angry, aggressive vocals, unpleasant, negative feelings, male lead vocals
Acoustic guitar	Acoustic, not exciting, folk, mellow, light beat
Happy emotion	Festive, positive feelings, optimistic, carefree, catchy

Very danceable song	Fast tempo, using at a party, cheerful, awakening happy
Going to sleep	Calming, tender, mellow, slow beat, low energy

Note. Adapted from Miotto and Lanckriet (2012).

Notice, however, that 'tag' categories vary in ways that alternative perspectives have otherwise distinguished. For instance, Rentfrow et al. (2012) differentiate between *sound-related* features (such as instrumentation and auditory features) and *psychological* features (such as emotional and physical effects). The underlying caveats therefore remain, in that tags are extremely noisy, may be structured in a myriad of ways, and often contain a great deal of irrelevant information (Turnbull et al., 2008a; Lamere, 2008), something Miotto and Lanckriet (2012) were trying to circumvent.

The lack of standardisation in tagging makes it difficult to discern the most effective way of utilising semantic content in relation to music recommendations. As has been alluded to, *content-based* recommender systems are based on the similarity of features associated with items that imply user preference, and these are often implemented through these described semantic vocabularies (Ricci et al., 2015). Nonetheless, for recommender systems that leverage such descriptors to be effective, they require vocabularies that are reflective of user experiences, which requires a consistent and representative audio-tagging procedure.

4.3.2 Context-Aware Music recommender systems

A consistent limitation with both *collaborative filtering* and *content-based* recommender systems is their ability to only model listener's long-term preferences (Wang et al., 2012). Given the increased integration of music into everyday life, with accompanying situational demands that accompany these developments, attention has been given to the efficacy of user-aware, personalised, and multimodal recommendations (Schedl et al., 2014; 2018). Some of these applications hold tangible orientations toward context-specific recommendations, such as the *InCarMusic* system (Baltrunas et al., 2011), as well as systems aimed to satisfy short-term listening goals according to situation-level data (Wang et al., 2012). Given the previous discussions highlighted regarding the fluidity of music following technological advancements, it is interesting to note motivations in the MIR literature that seek to integrate this awareness.

The integration of contextual information in computing applications was first introduced in the mid-1990s, however, was not applied in MIR until much later (Kaminskas & Ricci, 2012). ‘Context’, with regards to computing systems, can be defined as:

> “any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between user and an application, including the user and application themselves” (Dey, 2000; p. 4).

Such factors are typically beyond the influence of a given recommender system, but nonetheless influence users’ experience (Knijnenburg et al., 2012). With specific regard to music, however, Kaminskas and Ricci (2012) infer a simpler classification relating to contextual information: *environment-related context* (e.g., time, location), *user-related context* (e.g., activity, demographic information), and *multi-media context* (other types of information the user is exposed to other than music, such as text and images). This integration of such factors has culminated in the development of *Context-Aware Music Recommender Systems* (CAMRSs), which specifically aim to utilise contextual information to satisfy the short-term needs of listeners.

In computing, ‘Context-Aware’ refers to systems that are not always well defined, however, generally make inferences about daily events experienced by human beings based on the data available to them. Dey (2000) outlines three categorical features that are generally consistent with context-aware applications: (1) *presentation* of information and services to a user, (2) automatic *execution* of a service for a user, and (3) *tagging* of context to information to support later retrieval. In addition, Schedl et al. (2022) distinguish item-related context, user-related context, and situational context; where the former refers to the position of a track in a playlist or listening session, user-related context refers to variables relating to demographics, cultural backgrounds, activities, or mood of the listener, and situational context refers to the characteristics of a given listening event and may include time and location among other variables.

Examples include the system developed by Wang et al. (2012), which follows a novel approach by circumventing the methods used by more traditional recommender systems that rely solely on *collaborative filtering* or *content-based* methods aiming to satisfy listeners' long-term taste. Instead, primary attention is given to the short-term needs of listeners, which are primarily influenced by the context they are in (e.g., location, activity, and emotional states). Wang et al. (2012) demonstrate that automated music content analysis is able to determine whether songs are suitable or not in given daily activities computationally. They argue that by using data from sensors on mobile phones (e.g., acceleration, ambient noise, time of day) the user's current activity can be inferred; "we expect that a system that *combines* activity inference with music content analysis can outperform existing systems when no rating or annotation exists, thus providing a solution to the cold-start problem" (p. 99). They used a Bayesian framework to integrate context-aware activity classification and music content analysis. First, their model identifies two variables: a set of songs (*S*) and a set of context categories (*C*). The model assumes the listener is in one single context category at any given time (e.g., walking) and that they are carrying their mobile phone which may record a sensor data stream (e.g., time of day, accelerometer data, and microphone audio). This is then divided into a sequence of frames, each of which are attributed a vector (*f*) of *features* from the extracted data.

From this, the recommendation problem is codified as a two-step process: (1) infer the listener's context category (*c*) from *f* and (2) identify a song (*s*) matching *c* best; "We call the first step *context inference* and the second step *music content analysis*" (p. 100). This was operationalised in a probabilistic model that combined automated activity classification and music content analysis. Three datasets (one set of playlists from the internet, one of 1,200 annotated songs and a set of sensor data captured from user's daily activities) were gathered from which results indicated that the subsequent system was effective in providing context-specific user recommendations in the absence of pre-existing user ratings or annotations, satisfying listener's short-term needs (Wang et al., 2012).

Though a significant step in the development of CAMRS, there were some limitations to the methods applied by Wang et al. (2012). The first is with the annotation process, in which 10 students were hired to manually annotate the contextual features of each of the 1,200 songs that were included in the study. It is acknowledged by the authors that such a process may be subjective, yet participants generally agreed the annotations were appropriate. This re-establishes the general limitation of using manually determined annotations or tags found in more traditional systems, but more pressingly limits the array of music that can be provided since they are dependent on these manual annotations. Although 1,200 songs is a reasonably large corpus for the purpose of methodological development and exploration, this is not a pragmatic method to ascribe audio features in significantly larger repositories, such as those of cloud-based streaming services containing tens of millions of songs, and also limits personalisation and *serendipity* as the potential array of music is limited when considering variety in taste and preference to those of the sample dataset. This example nonetheless highlights the merits of CAMRS as an endeavour, as providing a method to generate contextual music recommendations is both achievable and effective. As such, despite the contributions of this approach, the relative scalability remains limited.

Elsewhere, Takama et al. (2021), proposed a CAMRS that is instead based on implicit feedback, and utilises context/content information to predict appropriate items based on contextual features. The researchers employed a Factorisation Machine (Rendle, 2010) which treated contextual information as features, and flexibly considered interactions between users, items, and these features to generate recommendations (Adomavicius & Tuzhilin, 2015). Similarly, Pichl et al. (2015) outlined a hybridised CAMRS that leverages a *k*-Means clustering method to identify and categorise contextual groups based on existing playlists in Spotify. They outline a method of identifying contextual clusters based on playlist titles and descriptions via NLP, identifying 34 contextual clusters of playlist types (from a total of 143,528 unique playlists). The authors then applied a collaborative filtering approach to each cluster in turn, generating recommendations based on users' interactions with other playlists in that cluster.

However, the systems highlighted hold limitations in the extent to which they may be applied. Specifically, these are (1) their dependence on real-time data extracted from a mobile device (e.g., accelerometer and microphone audio), which requires a large amount of access and user-permission, and (2) a dependence on machine learning algorithms to generate recommendations. The former of these points refers to the proposition that it is plausible that forms of data collection perceived as being more intrusive may not sit comfortably with many everyday listeners outside of academic or research environments/contexts. Note the prior specification made by Schedl et al. (2014), that *transparency* is a key part of developing successful recommender systems, since users react positively to such systems when they trust and understand how recommendations are being made to them. Reliance on mobile phone data poses potential limitations on the effectiveness of large-scale systems as users become increasingly aware and protective of personal data (Di Noia et al., 2022).

In addition, and regarding the latter point, machine learning algorithms partition datasets which are first trained and then tested to make predictions about user interactions and preference (Shmeuli, 2010; Yarkoni & Westfall, 2017). Typically, these prediction models work by detecting low-level linearities in high-dimensional data, generating models that, although may hold predictive accuracy in many applications, are unintelligible to humans and operate as an automated process (Shmeuli, 2010). Dependence on such methods can be therefore problematic as it is subject to the characteristics of the data used to train models and requires large amounts of data to detect low-level patterns (particularly in complex behaviours, like music listening). Moreover, the ability to explain recommendations is similarly related to *transparency*, as users are more likely to trust systems when they are able to understand why they have received the recommendations they have (Knijnenburg et al., 2012; Schedl et al., 2014). Aside from ethical concerns relating to data sources and data-dependent modelling processes, it has been particularly noted that context-aware recommender systems do not integrate situational signals well enough to understand listeners' needs and intents in given situations (Schedl et al., 2018), practical as well as methodological limitations with existing approaches.

Elsewhere, however, substantial progress has been made into understanding the ways in which people apply and use music when interacting with systems. For example, Hansen et al. (2020) analysed a dataset from the popular music streaming service Spotify and found that recent song consumption and ‘session-level’ contextual variables (such as time of day or listening device) are predictors of track selection, more so than static music preferences. Similarly, Gillhofer and Schedl (2015) find that contexts in which users consume music has a significant effect on the ways in which predictions about music listening can be made and note that mood classifications are significantly less effective than genre or artist classification in predicting contextual music listening behaviour. However, they highlight the importance of context in generating accurate recommendations, arguing that integrating such knowledge holds strong implications for recommender accuracy. Given the discrepancies in the ways in which recommendations are provided to listeners via machine learning, as well as concerns about the volumes and kinds of data required to provide recommendations in contextual settings, it may of substantive use to explore other approaches that may mitigate such issues, by integrating knowledge about situations for the purpose of music recommendation. To that end, the following section expands on the nature of these highlighted issues and outlines an alternative approach that integrates knowledge about behaviour to mitigate dependency on these existing approaches.

4.3.3 Psychology-Informed recommender systems

As mentioned, the increasing availability of social networks and services (e.g., cloud-based streaming) has resulted in a greater abundance of information and items/content than was previously available to listeners. Most systems attempting to curate for these circumstances are data-driven and predict the utility of items by relying on machine-learning algorithms, which are often leveraged to cluster and/or classify items in a given set of observations for predictive purposes. Common examples of prediction-based machine learning algorithms include decision trees, neural networks, and *k*-nearest neighbours (e.g., Gershman et al., 2010; Van den Oord et al., 2013; Ludewig et al., 2018), as well as Factorisation Machines and Support Vector Machines (Rendle, 2010). Although these algorithms can indeed be effective when provided with enough training data (and yield effective recommendations), they often lack interpretability and fail to incorporate any understanding of the data, behaviour, or phenomenon in question (Lex & Schedl, 2022). These ‘black-box’ models fail to elicit trust and provide limited autonomy to listeners (Millecamp et al., 2018). In turn, this lack of interpretability and explainability may lead to a failure in *transparency* and *control* (Di Noia et al., 2022).

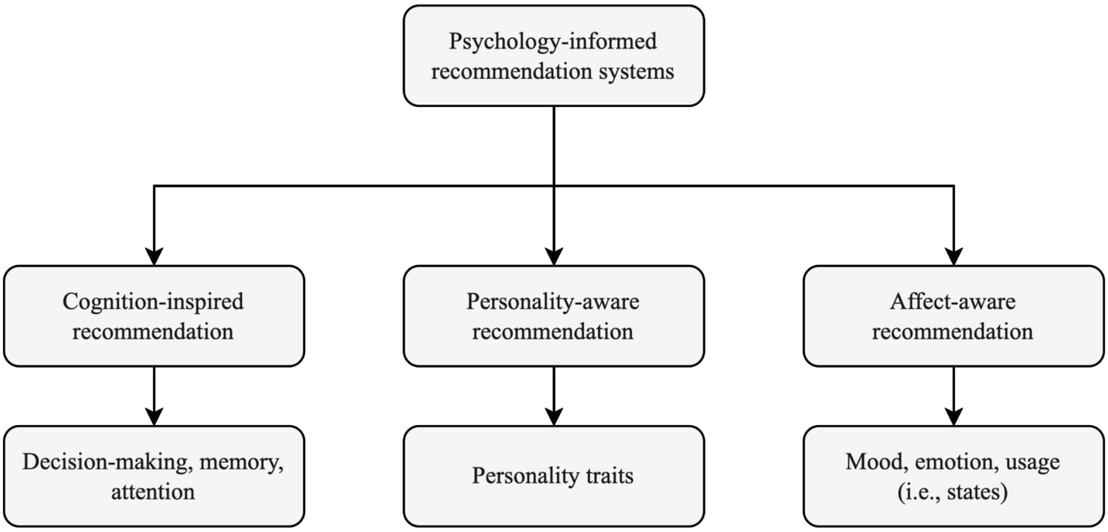
To mitigate dependencies on data-driven tools, therefore, researchers have suggested leveraging psychological and behavioural research to model and predict user needs and improve the recommendation process. Recommender Systems that integrate such information sources are dubbed psychology-informed recommender systems (Schedl et al., 2018; Lex et al., 2021). To be clear, these systems are motivated by the desire to mitigate data dependencies in recommender procedures by integrating knowledge and understanding from the social sciences in the generation of recommender systems, which reduces reliance on black-box machine-learning models.

Such impetuses intersect with an area of debate in the wider methodological literature, whereby psychological research typically seeks to explain phenomena without making predictions, whereas big data applications applied in machine learning seek to make accurate predictions at the expense of (intelligible) explanations (Yarkoni & Westfall, 2017). Psychological research emphasises the explanatory approach, in which hypotheses are derived from subject-specific theory which are then reflected in a suitable model to measure or test the effect of one variable on another. On the other hand, predictive modelling applications "mimic the outputs of the true data-generating process when given the same inputs, without caring *how* that goal is achieved" (Yarkoni & Westfall, 2017; p. 1101). It is this lack of interest in *how* processes happen that result in the data dependency of big data applications for which psychology-informed recommendation procedures are proposed. This is because machine learning techniques require (very) large amounts of data on which to train data for future predictions, whereas psychological methods apply explanatory modelling to explain *why* phenomena operate the ways in which they do. However, psychologists seldom predict, and rather focus on explanation. There are, however, opportunities to take lessons from (e.g., regression-based) machine learning methods to generate alternate predictive methods (e.g., Yarkoni & Westfall, 2017; Fokkema et al., 2022). In other words, explanatory modelling needs to embrace predictive utility to circumvent data-dependent predictive approaches implemented via machine learning, which implements a *what* without explaining the *why*.

To this end, differing approaches to making predictions based on accurately modelling phenomena are proposed, and psychology-informed recommender systems can be considered to fall into three main categories: (1) cognition-inspired, (2) personality-aware, and (3) affect-aware (Lex et al., 2021). The first of these categories employ models from cognitive psychology to calibrate recommendations in which cognitive processes like decision-making, memory, or attention are modelled. Personality-aware recommendations integrate information about an individual's personality traits, considered to be stable characteristics and thus not context-dependent (i.e., trait rather than state). Relevant examples include the integration of music preference into recommendation procedures, a paradigm upon which most music recommender systems operate regardless of whether they are psychology-informed or inferred via machine-learning algorithms (Ferwerda et al., 2017). These may also consider behavioural preferences within personality dimensions, for which users' desire for diversity, popularity, and serendipity may differ according to personality traits and modelled in the recommendation procedure (Nguyen et al., 2018). Finally, affect-aware recommendations integrate psychological (e.g., mood and emotion) affect into recommendation decisions. Like personality traits, these effects are integral human characteristics and the subject of intensive research. Unlike personality traits, however, they are not traits but states in response to short-term, contextual factors. It is for this reason that emotion, for example, is a prominent contextual factor in CAMRS (e.g., Zheng et al., 2013; Lex et al., 2021).

There are effectively two ways in which affective content can be described, referred to by Lex et al. (2021) as *categorical* and *dimensional* models. *Categorical* models refer to the description of emotional content in vocabularies of terms derived from a selection of universal and innate basic emotions (e.g., happy, sad), or through secondary emotions that are reactions to these primary emotions (Song et al., 2016; Lex et al., 2021). On the other hand, *dimensional* models describe values by situating them in a continuous space, commonly spanned by *arousal*, *valence*, and occasionally *dominance* (Russell, 1980; Schubert, 2007; Lex et al., 2021). This highlights the importance of the ability to effectively annotate music using either method, such as with *content-based* recommendations as has been applied in previous CAMRSs (e.g., Wang et al., 2012). By inferring such categorical or dimensional parameters into recommendation procedures, it may be possible to provide listeners with the means of satisfying short-term goals, based on their contextual indicators, provided reliable and consistent descriptors are applied. These are particularly effective in *context-aware* recommendation scenarios. Figure 3 illustrates these domains of psychology-informed recommender systems.

Figure 3 Categories of Psychology-Informed Recommender Systems



When considering the topics highlighted up to and including this stage, it is interesting to note the following three points:

1. A desire in the MIR literature to provide short-term, contextually oriented music recommendations (e.g., Wang et al., 2012)
2. The integration of emotion and mood in affect-aware recommender systems, whereby state variables are used to provide recommendations based on psychological modelling or knowledge (e.g., Lex et al., 2021)
3. Evidence in the music psychology literature that such factors lead to changes in audio content the music listeners select (e.g., Greb et al., 2019)

There is, therefore, a clear gap in the literature, whereby a procedure that effectively integrates knowledge about music selection from psychological and behavioural research in response to contextual factors may help synthesise an affect-aware CAMRS. This may help mitigate data-dependency and the prevalence of black-box models currently applied in existing music recommender systems, improving *Transparency* and bases for incorporating understanding of listening behaviour by understanding users' experience (e.g., Schedl et al., 2018). These elements are kept in mind throughout the remainder of this thesis. However, there is one more key issue that needs to be addressed, and that is the ways in which an MRSs are evaluated. Any advancements made with regard to MRSs must be appraised to substantiate any improvements in user experience or accuracy deriving from such approaches. To that end, the next section outlines evaluation methods of recommender systems.

4.3.4 Evaluation of Recommender Systems

Broadly speaking there are three primary ways in which the effectiveness of recommender systems are assessed: *offline evaluation*, *online evaluation*, and *user studies. Offline evaluation* is the most commonly applied method in academic settings and relies on existing (often public) datasets of user-item interactions, conducted without the involvement of application users (Schedl et al., 2022). This approach retroactively analyses data for model-based predictions, providing quantitative insights into algorithmic performance (Schedl et al., 2018; Lex et al., 2021). An example of *offline evaluation* with regard to context-based recommendations is the system developed by Pichl et al. (2015), in which existing transaction data were leveraged from users into training and test datasets to assess the efficacy of a collaborative filtering approach to CAMRS. Such evaluations fail, however, to provide sufficient information about the perceived quality of recommendations, or their efficacy as perceived by users (Schedl et al., 2022). The second method, *online evaluation*, is more common in industry settings and involves the use of data taken from users – specifically *implicit* forms of data. *Online evaluation* typically applies A/B testing, in which comparisons are made between two or more competing algorithms and assess each systems' performance by looking at metrics such as user retention, click-through rate, and amount of music streamed (Schedl et al., 2022). Limitations associated with *Online evaluation* are that it does not seek to practically assess or consult users to understand their experience, again implementing a *what* without seeking to understand *why*.

Finally, there are *user studies* (also known as *user-centric evaluation*), where systems are assessed through means of user engagement and satisfaction (Schedl et al., 2022). User-centric studies are advantageous compared to competing approaches as they allow inferences to be made from users' *explicit* feedback, uncovering intrinsic characteristics of the user experience not attainable in *offline evaluation*, and ignored in *online evaluation*. Gathering more nuanced response data in this way allows researchers to interpret the efficacy of a system from users' perspective, rather than through automated comparison between systems based on retention rates (as during *online evaluation*, for example). By addressing user experience directly, however, real-world perceptions and judgement of system quality can be generated (Schedl et al., 2018). Intuitively, it can be argued that there exists an alignment between psychology-informed recommender systems and user-centric approaches as a means of evaluation. It seems plausible, therefore, that it would be beneficial to not just integrate psychological inferences into recommendation procedures, but also during the evaluative process to assess the quality of recommendations explicitly, informing further refinement and improvement. User studies are, however, notoriously difficult to implement due to difficulty in gathering a large enough number of participants to draw significant conclusions due to intensive effort on the user's side, with many limited to tens or a few hundred participants at best (Schedl et al., 2022).

To maximise the efficacy of user-centric evaluation, therefore, researchers have developed frameworks intended to address the most important aspects of system performance from users' perspective. Examples include that presented by Pu et al. (2011), referred to as *ResQue* (*Re*commender *s*ystems' *Q*uality of *u*ser *e*xperience), which includes aspects relating to the perceived quality of recommendations (e.g., attractiveness, novelty, and diversity), interface adequacy (e.g., sufficiency of information, clarity of the layout), interaction adequacy (e.g., preference elicitation and revision), perceived usefulness, ease of use, user control, transparency, explicability, and trust. Although Pu et al. (2011) do not validate this framework with respect to an MRS (rather using YouTube videos as a working example), the principal of the method can be reasonably applied to other forms of media (Schedl et al., 2022). Another example is that of Knijnenburg et al. (2012), who measured dimensions of the user experience via latent psychometric structures (see Loehlin & Beaujean, 2017) of perceived quality, effectiveness, and variety, as well as choice satisfaction, intention to provide feedback, general trust, and privacy concerns. To assess the relationships between such constructs, the authors use SEM as a statistical framework, highlighting a pertinent cross-over between quantitative methods applied in the social sciences to understand user experience empirically. Schedl et al. (2018) and Lex et al. (2021) highlight this particular framework as holding a high-level of abstraction also, earmarking it as a useful tool in user-centric evaluation. Moreover, the evaluative techniques applied to this particular framework (SEM) is a widely used approach in psychometric modelling in general (see Kline, 2016), and so it follows well that psychology-informed approaches may be well extended by applying such methods in the evaluation stage of a system. This would serve to incorporate the strengths of such theory-based models throughout the lifetime of a system, encompassing conception, prediction, implementation, and evaluation.

4.4 Chapter Summary

This chapter has aimed to summarise key principles relating to present issues and motivations in the development of MRSs. Both situational and individual variables may serve as predictors of music selection and have been implemented in differing ways within recommender systems; albeit to varying extents (Wang et al., 2012; Hansen et al., 2020; Takama et al., 2021). This is broadly consistent with theoretical underpinnings of *functionality* (e.g., Greb et al., 2018a; 2019), in that music selection is reflective of cognitive goals which have also been shown to be influenced by contextual variables (Greb et al., 2018a), reaffirming the notion that MRSs should consider such variables (Schedl et al., 2018). To some extent, this has been implemented through CAMRS, however, some of the effective CAMRS developed, such as those of Reddy and Mascia (2006) and Wang et al. (2012), often utilise real-time data from listening devices such as microphones and accelerometers. It is plausible to consider, therefore, whether establishing relationships between affective music content and situational variables would be conducive to generating context-based recommendations which do not require contextual inference via such forms of data but may rather follow a psychology-informed (specifically affect-aware approach) to music recommendations. This is proposed to be attainable by integrating knowledge from psychological research, specifically relating to FML in everyday life.

The motivations behind such an approach are that when considering how systems may be effectively implemented at scale, there are broad inconsistencies with the principle that a recommender system holds *transparency*; referred to by Schedl et al. (2014) as users' understanding of how recommendations are made, which elicits trust. Users are hesitant to integrate new technologies that are perceived to compromise privacy into their lives (Lobera et al., 2020), and Afchar et al. (2022) note that the complexity in existing music recommendation processes hinders the ability to explain recommendations to users, which is extremely important in the eventuality that it is unexpected or inappropriate. Moreover, trust and privacy concern are integral factors to gauging the efficacy of a recommender system (e.g., Knijnenburg et al., 2012).

Interestingly, Di Noia et al. (2022) address this issue in a report for the European Commission, where they argue that artificial intelligence (AI) algorithms should be more considerate of the data they acquire and use for the purpose of providing recommendations. It is stated that "We require AI systems to be fair, secure, and privacy-preserving, and interpretable" (p. 73). The authors address several issues concerning recommender systems' placement in society, such as the fact that there is a commonly believed trade-off between the accuracy and interpretability of recommendation, which is not inherently true. They also cite *transparency* and *control* as being core principles that may guide the generation of future recommender systems that are effective but also consistent with reasonable privacy standards for everyday users. The issues highlighted therefore hold relevance for both industry and academia since there is increasing scrutiny regarding the methods applied by existing systems and platforms.

Millecamp et al. (2018) further note that the black-box nature of 21st-century recommender systems leads to a lack of trust and limited autonomy, and that providing users with the autonomy to steer the recommendation process may be conducive to eliciting a greater sense of trust. Meanwhile, Born et al. (2021) note that one of the most pressing issues relating to the ways in which recommender systems currently operate is the way that extraction of personal data has been "privatized and corporatized by curation platforms" (p. 2). They go on to say that there is little public interest in this issue, which leaves debate and regulatory intervention in terms of accountability and transparency lagging behind the development and implementation of such systems. As a consequence, more of the curation and consumption of media in the everyday lives of individuals is ever more reliant on the AI algorithms that industry often relies on; this does not just include music streaming services of course, but also refers to film streaming services, news media outlets, and social media also. With regards to audio streaming services, however, Darmody and Zwick (2020) note that streaming services like Spotify personalise listening experiences through algorithmic individuation. They view Spotify, and other companies using Big Data (e.g., Netflix, Amazon, Google, and Apple), through the lens of surveillance capitalism, a term coined by Zuboff (2015) to describe the reliance of such companies on AI in digital marketing and brand performance. Zuboff's (2015) surveillance capitalism comprises unanticipated, indecipherable mechanisms of extraction, commodification, and control that exiles the individual from their own online behaviours, whilst generating new modes of behavioural prediction and modification. It is through this lens that many companies reliant on Big Data, like Spotify, have been critiqued (e.g., Darmody and Zwick, 2020; Born et al., 2021). This reiterates the scrutiny with which systems are being increasingly viewed, and as such there is an incentive to consider competing approaches that mitigate the issues highlighted both in academic as well as legislative settings.

Whilst the scope of the issues outlined above are infinitely more complex than this thesis can reasonably address, it serves as a useful reminder of just how and why integrating an informed understanding of the ways in which people interact with media, that distrust in certain technologies could be mitigated. It is to this end that this thesis considers exploring alternative methods to help address these issues, specifically by utilising conceptual approaches to psychology-informed recommendations that emphasise retaining explainability and interpretability (e.g., Schedl et al., 2018; Lex et al., 2021; Schedl et al., 2022).

When considering such broad concerns, it is difficult to assess and distinguish each one to the point of circumvention due to the sheer complexity of this issue. However, one summary suggestion is to explore and utilise alternate methods of data collection to help estimate user-transactions with content, rather than automate the process via unsupervised machine learning models. There is no shortage of methods that may be applied to this end. For instance, semantic retrieval has been applied to annotate audio with meaningful content, though there remains

some difficulty in acquiring high quality tags through common vocabularies which affects comparability between systems and methods. This has limited the extent to which content-based recommendations can be fully exploited in general (Kaminskas & Ricci, 2012).
Studies such as those by Ferrer and Eerola (2011) have, however, reaffirmed the links between semantic structures and perceived timbral qualities of music, thus implying that semantics can be linked to acoustical features, which may in turn can be used to represent music selections at the listener level (e.g., Greb et al., 2019). Furthermore, Miotto and Lanckriet (2012) demonstrate that tag correlation may be able to effectively infer and predict broader contextual information about a piece of music which may be operationalised. More recently, Wang et al. (2018) generated a model that utilises users' general and contextual preferences from listening records to meet real-time requirements, however, this is within data-driven frameworks which Kaminskas and Ricci (2012) describe as having received much of the focus within CAMRS research, as opposed to psychology-informed approaches that seek to generate links between music and contexts theoretically prior to algorithmic training. They recommend that researchers look towards psychology-informed approaches to first gain a better understanding of associations between music content and listening context, however, there appears to be limited progress in this regard.

Others too have argued there is a need for robust pre-processing phases. For instance, Knees et al. (2019) address future directions of user interfaces according to three evolving phases of listening culture over time. The first phase structures and visualises small scale music collections (such as personal collections or early digital sales repositories) which are driven by content-based algorithms. Phase two refers to web-based interfaces with strong focus on textual descriptors of music, represented through tags and descriptive audio features. The third and current phase in which we find ourselves, is "shaped by lean back experiences driven by automatic playlist algorithms and personalised recommender systems" (p. 49). To put it another way, MIR research has been moving towards ever-increasing personalisation alongside more widely available consumer technology that affords listeners greater agency. According to Knees et al. (2019), the implications of the third and current phase of music curation in a general sense is shifting towards the exploitation of user interaction data with a focus on integrating content-based methods, community metadata, user information, and contextual information. Meanwhile, Lex et al. (2021) highlight that most recommendation algorithms are data-driven and based on interaction data which, although produces effective recommendations, are often impenetrable, black-box models that do not integrate the underpinning cognitive reasons for behaviour in user interactions.

One way of mitigating this relative dependency and associated trade-offs, however, may be circumventing issues relating to *transparency* and data-dependency, particularly in relation to CAMRSs. This may be achieved by modelling relationships between music content with situational and/or listener-level constructs (e.g., activity, locations, and cognitive goals). Since this is part of the current, evolving phase of music recommendations, utilising cognitive indicators to inform contextual inferences may serve to strengthen the base upon which future systems are developed. Future processes may expand on such motivations, for instance by inferring variables such as music preferences or personality traits to provide higher resolution recommendations for intended users, however, the aim of this thesis at large is to explore how FML is situationally determined and an influencer on music selection in turn, and that information about these relationships can be used to generate recommendations that result in perceived congruency between music and short-term listening scenarios.

This concludes the literature review of this thesis. The following discussions in the next chapter serve to summarise the key discussions of the review and ultimately formulate and articulate the central aims of this thesis at large. The following chapter therefore effectively acts as a bridge to consolidate the secondary research discussion that has taken place thus far, followed by proposed empirical approaches to addressing this issue as conducted as part of this thesis.

5.1 Summary and Conceptual Approach

From the review of the literature conducted, it appears that listeners' context-orientated goals are of greatest influence when it comes to determining *functionality* in everyday listening situations. Listeners' responses to musical affect formulate learned associations in response to different situations, and music may become associated with being congruent when applied alongside particular events, locations, or activities. Such experiential phenomena, such as emotion contagion and *past functional experiences*, play a role in influencing the experiences of listeners, for instance by stimulating emotional qualities consistent with the emotions that are expressed by a piece of music deemed congruent with the listener's situationally determined cognitive goal.

Alongside psychological research that explores the functions, goals, emotions, and situational variance associated with music listening; other, interdisciplinary fields have sought to operationalise such behaviours for the purpose of curation in an increasingly digitised world. MIR researchers, for instance, have sought to provide listeners with the ability to curate and discover music within their everyday lives. As discussed in Chapter 4, one of the ways this is implemented is via recommender systems. The development of these systems over time has largely run parallel with the increasingly seamless integration of music in daily life, and to overcome choice overload, as has been afforded by widely accessible technologies such as mobile phones and cloud-based streaming services in particular. One of the more recent outcomes of this endeavour has been the development of systems that aim to generate recommendations for the short-term, cross-sectional needs listeners have when engaging with music, as motivated by observations in psychological research that music is typically listened to achieve short-term, situational goals.

Previous systems, such as those presented by Wang et al. (2012) and Takama et al. (2021), raise questions as to long-term viability of the methods that underpin their operation; and Schedl et al. (2014) have noted that transparency is a key element in eliciting widespread trust in recommender systems. Moreover, Kaminskas and Ricci (2012) note that most CAMRS specifically relate music to contextual conditions through 'data-driven' approaches, which is to say a combination of contextual parameters in an algorithm without aiming to understand relations between music and contextual conditions explicitly. They recommend that psychology-informed approaches, whereby recommendations are made based on knowledge of the relationships between core constructs, are used to inform estimates of desirable content. This is further supported by assertions in the literature that recommendations could be driven by knowledge rather than data to sustain the *transparency* and interpretability of recommendations more broadly, in a manner that is consistent with aims surrounding data protection, user privacy, and explainability (e.g., Lex et al., 2021; Lex & Schedl, 2022; Di Noia et al., 2022).

5.2 Proposed approach

As such, it is proposed within this thesis that alternative methods should be explored to formulate associations between music and contextual variables, with a view to specifically estimating musical characteristics that may be congruent given such indicators. This is because features of music, as expressed via semantic content and tags are implemented in content-based recommender systems, have been used to generate recommendations based on predicted effects on musical content (Kaminskas & Ricci, 2012). Although, it should be acknowledged that within tagging methods there remains the semantic gap; a disparity between users' and systems' perceptions of music (Lew et al., 2006; Kaminskas & Ricci, 2012; Schedl et al., 2013). Nevertheless, studies have indeed shown that music descriptors can be implemented in order to induce sought effects of music based on psychological indicators (e.g., Lepa et al., 2020), implying that, hypothetically, a similar rationale may demonstrate a triangulation between listening context, music function, and music content.

It is this precise gap that this thesis aims to narrow by establishing a deeper understanding of content-preference in response to the situationally determined functions. A psychology-informed approach is therefore proposed, in which an understanding of the relationships between situational variables (namely listeners' concurrent activity), FML, and audio features are leveraged to inform content-based recommendations for everyday listening situations. This intends to reflect the aim of maintaining intelligibility through a psychology-informed approach that factorises functional music content in accordance with theory. This proposes an alternative to data-driven approaches, which have thus far often relied on a user's willingness to allow their listening context to be inferred by allowing access to mobile phone data such as microphones and accelerometers, and a dependency on black-box machine learning techniques to detect low-level linearities in high-dimensional data, which although may produce accurate prediction models, have limited to no explainability or intelligibility (Shmeuli, 2010).

Moreover, it is worth considering the broader motivations and implications that have led to this motivation. Diversity in methodologies may help areas of digital curation mitigate against the need for pure data-driven approaches that have negative algorithmic impact (Zuboff, 2015). Psychology-informed approaches to generating recommendations are just one of these alternate approaches, through which understanding is first built empirically, which is then used to inform recommendations (Lex et al., 2021). This is not to say that this thesis will be able to address all issues associated with data dependent MRSs, but it is intended to draw attention to the need for such transitions and hopes to contribute by first exploring methods to understand listeners' behaviour in everyday music selection, before exploring ways of implementing subsequent recommendations based on such insights. The next section will discuss the generation of an empirical process through which to explore and test the theory and models outlined above.

5.3 Empirical approach

How then, is it possible to evaluate this proposition? As has been stated, listeners determine their goals of music listening according to a given situation. The situation is primarily characterised by the activity being undertaken and the location in which it occurs, and this subsequently determines the listener's music-orientated goals. Activities, however, are the primary drivers of situationally determined *functionality* (Juslin et al., 2008; Maloney, 2019). The goals, and their associated functions, are facilitated by music with the appropriate features (content) that realise the contextual demand. When music with the necessary context-specific featural requirements is applied, it fulfils the function (thus realising the goal) and is therefore congruent. This identifies three key variables to generate a predictive model: *Activity*, *Function*, and *Content*, illustrated in Figure 4.

Figure 4 Suggested variable model for context-orientated goal attainment through music listening

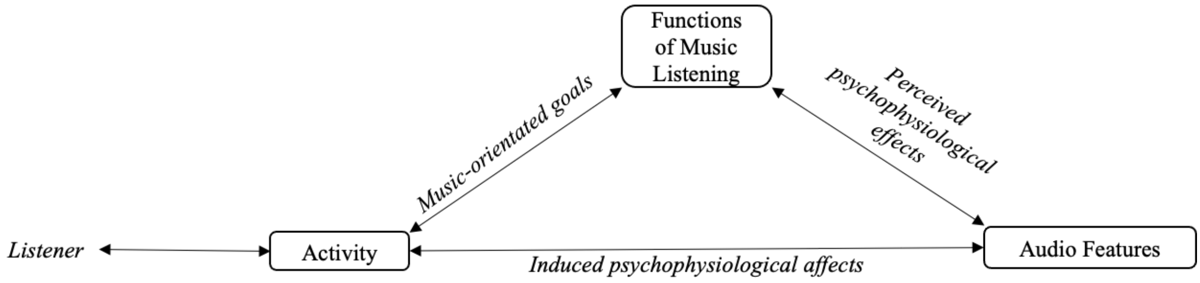


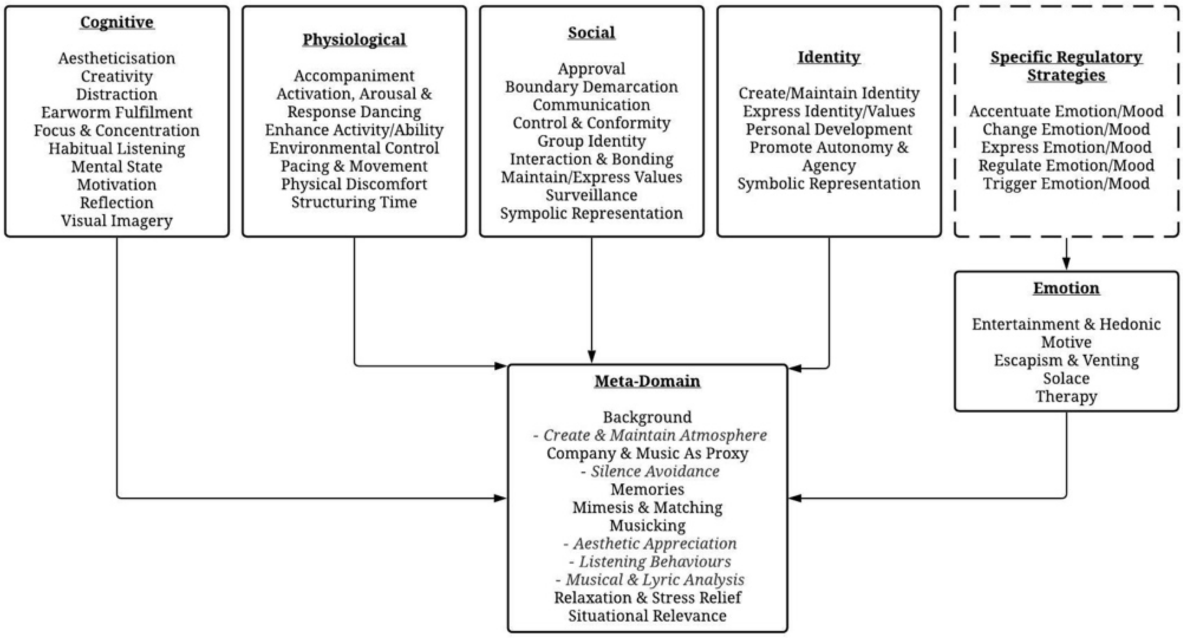
Figure 4 effectively illustrates the behavioural process that is hypothesised to inform contextual music listening selection and holds relative similarity to that proposed by Greb et al. (2019) but is distinguished by the fact that it focuses on the situation-level, and also that it focuses on listeners' activity. By focusing on the three primary variables that constitute daily listening practices, this is the triangulation that the proposed research aim of generating a methodology for psychology-informed CAMRS is aiming to operationalise within this thesis. This model therefore provides a framework to establish a psychology-informed approach to contextual music recommendation.

To this end, it is imperative that each of these three constructs be appropriately measured, so as to ensure validity, reliability, and replicability. Maloney (2019), however, notes that many frameworks and measures relating to FML are either focussed on regulatory uses of music, such as mood regulation (e.g., Groarke & Hogan, 2018), or that context is effectively ignored in the development of FML measures (e.g., Lonsdale & North, 2011). Rather, Maloney argues that *functionality* necessitates a broader perspective than narrower regulatory perspectives to FML, and as such that focus should be placed on music's utility to maximise the breadth of understanding of potential functions of music in everyday life. This perspective is therefore reflective of Merriam's (1964) founding argument, where function is considered the underlying reason for music's employment. The utilitarian perspective described, therefore, considers music listening a goal-orientated behaviour, as per *functionality*, wherein lies the use of music to achieve goals in everyday life. In line with Maloney's argument, it is important to consider how FML is measured and viewed as part of the triangulation depicted in Figure 4.

Evaluative frameworks and measures that broadly follow this utilitarian orientation of FML, however, are scant in the literature. Given this issue, Maloney (2019) conducted a comparative bibliometric analysis and ESM study of utility in music listening, which provides an exhaustive taxonomy of FML from the utilitarian perspective: the Consensus Functions Framework (CFF). 53 identified functions were qualitatively generated in this work, blending ecologically valid data with strong theoretical foundations informed by extant research emphasising the utility of music in everyday listening, illustrated in Figure 5. It is hypothesised that by using the content of this qualitative framework as a grounding, it would be possible to generate a utilitarian psychometric model of FML that could be applied in future work. The contributory value of this derives from the observation that FML are dependent on the utility of music in given situations. For any evaluative framework to gauge *functionality*, therefore, it should seek to measure FML according to the utility with which music may be applied in everyday situations and must therefore be broader in nature than regulatory models (e.g., Saarikallio, 2008; Groarke & Hogan, 2018), or those that fail to consider context as a driver of utility (e.g., Lonsdale & North, 2011).

Moreover, such an approach (i.e., using a qualitative framework as a grounding) is consistent with exploratory mixed methods (Punch, 2014). Exploratory design follows the general logic that quantitative investigations are best applied when qualitative methods have built strong prior foundational theories or groundings of the phenomenon that researchers are interested in (Punch, 2014). This has useful implications in identifying the scope of theoretical constructs, with quantitative approaches subsequently leveraged to apply that construct in later work (Fetters et al., 2013). It therefore seems prudent, to leverage the scope and theoretical strength of Maloney's (2019) framework as a grounding from which to uncover a well-informed psychometric structure. Moreover, a structure of FML derived from a qualitative utilitarian framework, may also enable subsequent steps to be taken to utilise audio content in relation to the identified latent constructs. This is because audio content has been attributed to the effects of music in relation to musical preferences or branding scenarios for instance (Rentfrow et al., 2012; Lepa et al., 2020), so it seems plausible that a psychometric structure, informed by the CFF, may serve to facilitate associations between music content and listening contexts directly and/or as a mediator.

Figure 5 Visualisation of the CFF (Maloney, 2019)



In addition to the 53 functions, Maloney (2019) identified 13 individual activity types (e.g., *Working*, *Travel*, *Socialising*) formulating situational features of music listening across the same ESM study. Qualitative analyses were also conducted on the frequency of listening locations, identifying 11 themes (e.g., *Home*, *Work*, *Transitory Spaces*, *Restaurant*/*Bar*). The relevance of this is that the 53 musical functions identified in the CFF were observed to occur within these discrete categorisations, reinforcing the notion that the broad array of functions identified take place within the stream of everyday life.

It therefore seems plausible that it is not only possible to structure and reduce the 53 identified FML following a procedure of psychometric construction and dimension reduction as applied elsewhere in the literature (e.g., Lonsdale & North, 2011; Schäfer et al., 2013; Groarke & Hogan, 2018), but also possible to apply that identified construct in relation to real-world listening contexts as part of the triangulation outlined in Figure 4. This would in essence provide a means of measurement for FML as a latent construct, which may be used to operationalise the structure and realise the first of the four aims outlined in section 1.2.

Moreover, in conjunction with subsequent steps, this may begin a process that ultimately addresses the overarching research question by linking with subsequent aims via a considered empirical approach, through which relative aims are addressed. For example, validating a latent structure informed by the CFF would enable exploration to identify which utilitarian FML are associated with which activities, which in turn, may be associated with content of listeners' music selection (the second and third aims outlined in section 1.2). It is hypothetically possible, therefore, to subsequently use such insights, to essentially reverse engineer this process, by estimating desirable content parameters given indicators of context and FML to realise the fourth aim. Such estimations in turn may be used to help develop psychology-informed music recommendations, as per associations with everyday activities and FML ultimately addressing the overarching research question. To this end, three empirical studies are considered to realise the four aims in sequence:

1. To address the first aim, it is proposed that by leveraging Maloney's CFF, a utilitarian model of FML may be generated and applied in subsequent work. The utilitarian approach of Maloney's framework provides a theoretical grounding form which to draw a theoretically and statistically parsimonious structure that may be used to measure utility in everyday music listening. This would therefore address the first thesis aim. For this, a process of psychometric development is proposed, in which the CFF is used as a structural grounding to generate an item pool representative of FML from the utilitarian perspective. This would, however, likely be a large pool of items, and as such a process of dimension reduction would be necessary to uncover a reduced latent construct of the items which explains the observed variance in the items' ratings, and may be used in subsequent work, hence realising the first aim.

2. The second and third aims outlined in section 1.2 effectively refer to the triangulation of key constructs of interest (listening context, FML, music content/features), by exploring the causal relationships between these constructs. The second study considered therefore aims to operationalise the model shown in Figure 4, whereby listening activities are proposed to ultimately lead to changes in music content directly and/or indirectly via FML. To this end, the second study will measure these three

constructs in ecologically valid settings (i.e., real-world music listening episodes), and model these data accordingly. This would also serve as a means of cross-validation of the model generated in the first study, since FML may be measured via an appropriate latent construct to assess the utility of music in listeners' concurrent situation. This would provide insights into the relationships between activities, FML, and music content/features empirically, addressing the second and third aims.

3. Finally, while the first two studies serve to effectively model FML and music selection processes in everyday life, the third and final study will seek to effectively reverse engineer the modelling processes in the second study. With appropriate information regarding context and FML, it may be possible to estimate the appropriate featural content for listeners dynamically, using the behavioural model previously generated as a structural grounding. Hence this would realise the fourth thesis aim by bringing together the inferences from the prior studies and ultimately help address the overarching research question. Though this may be somewhat explorative in the context of this thesis, as a proof of concept, this may serve to highlight how information and knowledge about everyday listening may be used to provide recommendations to users.

These three studies will ultimately address the four research aims and address the overarching research question. This leads us to the first empirical research in this thesis. The following chapter will broaden the discussion of the CFF and report a study conducted to rate each function and reduce these to underlying dimensions through a quantitative approach.

6.1 Study 1: Developing a utilitarian measure of FML

In wishing to triangulate the three structures of interest (i.e., the activity of the listener, FML, and the content of the selected music) suitable measurement techniques are essential. For this, there are particular challenges to measuring FML. Many phenomena measured in behavioural research (including FML) are derived from theory and/or prior empirical research, which often relate to latent (unobserved) phenomena, such as cognitive or behavioural processes relating to experiences or decision making (Loehlin & Beaujean, 2017). The latent nature of such phenomena holds especially strong implications for psychological measurement, in which cognitive and/or behavioural processes are not observed to exist directly, but whose presence is inferred through correlations amongst sets of observable indicators (DeVellis, 2017). It is through such indicators, which might be considered symptomatic of some underlying trait of the phenomena in question, that latent constructs are measured. Phenomena may be unidimensional (containing one latent variable) or multidimensional and contain multiple indicators as these are considered more reliable than single indicators or univariate constructs (Brown, 2006). In this sense, the latent phenomena are assumed to influence the ratings of indicator (sub)sets, which correlate as a result.

In quantitative research relating to FML, psychometric structures have been developed with the aim of measuring FML through latent structures in this exact way (e.g., Lonsdale & North, 2011; Groarke & Hogan, 2018; Greb et al., 2018a). However, structures generated by extant research to date are primarily concerned with specific aspects of FML that do not necessarily apply well to context (e.g., AFML; Groarke & Hogan, 2018), or are developed without consideration for context-orientated utility (e.g., Lonsdale & North, 2011). Though such structures have their uses, these typically reflect aspects of function as an outcome, and seldom view use through the lens of *functionality*. But the utilitarian viewpoint previously described characterises music listening as a goal-orientated activity determining the utility of music in everyday listening. Maloney's (2019) theoretical framework, the CFF, contains 53 distinct FML generated through this lens and "describes the most exhaustive approach to the functions of music currently available" (p. 264). Moreover, it has the distinct advantage of being informed by ecologically valid observations of FML via an ESM study cross-referenced with bibliometric data from the broader literature. The CFF therefore provides a strong theoretical grounding derived from qualitative work, upon which to build a utilitarian measure of FML. Therefore, this first study explores how the CFF may be used as a basis upon which a utilitarian measure of FML is generated, which is posited to hold particular uses and application to quantitative studies relating to FML in everyday life.

6.2 Study Aim

This study aimed to gauge the extent to which people engage with certain FML through a questionnaire encapsulating the broad theoretical scope of utility. More specifically, the aim was to use the list of 53 functions identified in Maloney's (2019) CFF to generate a latent structure of FML by applying a process of item generation and dimension reduction. This was to extend the availability of measurement instruments that have been derived from the utilitarian standpoint, which is focused on listeners' use. This was achieved by utilising the content validity of the CFF, which acts as a theoretical framework to provide suitable scope from which to derive a proposed utilitarian measure. With regard to the thesis more broadly this sought to help address the first aim outlined in section 1.2., referring to the need to generate a utilitarian measure of FML.

6.3 Methods

6.3.1 Item generation

The process of generating a set of items representing the definitions of functions described in the CFF began by qualitatively interpreting all 53 functions according to their attributed content and definitions that were adapted from Maloney (2019; see Appendix A). In each case, a tentative set of items was generated to represent the content of each function. Multiple items were generated for each function in the CFF, with each item reflecting the aspects of each function as attributed (DeVellis, 2017), an approach similarly applied in other cases of psychometric development in the music psychology literature (e.g., Groarke & Hogan, 2016; 2018). In most cases this yielded two items per function, however, functions with extended or more nuanced definitions were ascribed with as many items as relevant to suitably encapsulate the given definition. Following this, items were reviewed by the research team (including the original author of the framework) to assess whether the items' content was reflective of each function's scope and definition.

Following this review and relevant amendments, a finalised set of 114 items (see Appendix B) was generated, which were assessed to be reflective of the 53 functions. This broad set of items was consciously generated to avoid construct under-representation, "which is when a scale does not capture important aspects of a construct because its focus is too narrow" (Boateng, 2018; p .6).

6.3.2 Materials and procedure

To administer the 114 items to participants and uncover a latent structure, an online survey was generated and distributed using Qualtrics. In this, the 114 items were presented over 11 matrix tables, of which 10 contained 10 items, and one contained 14 items. Each table presented a randomised subset of items for each participant. Prior to taking part, participants were given access to relevant study information (e.g., study aims and purpose, right to withdrawal, and anonymity), were required to confirm being at least 18 years of age, that they had read and understood the available study information and provide informed consent to participating via checkboxes embedded in the survey. Ethical approval for this study was granted by the Arts and Humanities Ethics Committee (AHEC) at the University of York.

The items were assessed on a 5-point Likert scale, ranging from 0 (*Never*) to 4 (*Very Often*). Within the questionnaire, items were presented with the prompt: "On a scale from Never (being you do not recall ever using music for that purpose) to Very often (being you use music for that purpose very frequently), to what extent do you use music to...". This mode of assessment was deemed practical to effectively gain an overall rating of items' perceived frequency in day-to-day life, but it should be acknowledged that in the current setting this does not achieve complete ecological validity with reference to everyday listening episodes, for which 114 items were deemed unwieldy. This reflects the primary study aim of dimension reduction to a practical subset of items that may be applied in ecologically valid settings as part of future work, but which is cognisant of a utilitarian view of FML.

6.3.3 Participants

Participant recruitment for this study was carried out online, partially gathered through Prolific[1], with additional recruitment taking place via other forms of internet distribution (e.g., social media, emailing lists). Participants taking part via the Prolific platform were compensated £1.35 for taking part, with those from other avenues taking part on a purely voluntary basis without compensation. In total, n = 327 complete responses were recorded, of which n = 208 were recruited via Prolific and n = 119 participants via other forms of internet distribution (51.4% female, 47.4% male, 0.9% non-binary/third gender, 0.3% prefer not to say). Age was recorded via coded bands, the frequencies of which are shown in Table 5.

Table 5 Age Distribution of Study 1 sample

Age	*N*	*%*
18-24	78	23.9%
25-34	106	32.4%
35-44	61	18.7%
45-54	42	12.8%
55-64	34	10.4%
65-74	3	0.9%
75-84	3	0.9%
85+	0	0%
Total	327	100%

Note. Age was captured via reported groups coded 1-8. 0 responses were recorded for the eighth band (85 and older).

[1] https://www.prolific.co/

6.3.4 Approach to data analysis

This study sought to generate an underlying structure of the initial pool of the generated 114 items that may be reapplied in subsequent research. For this, content should be optimised, and dimensionality reduced through appropriate procedures to ensure statistical as well as theoretical parsimony. Methodologically speaking, it is common and recommended to reduce dimensionality and uncover latent structures using factor analysis (Worthington & Whittaker, 2006; DeVellis, 2017). Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) are both commonly applied methods for this purpose (Matsunaga, 2010). Here, it was deemed prudent to first identify the latent structure via EFA and to then supplement this with CFA, a method imposed on a factor structure to assess model fit in restrictive settings (Schmitt et al., 2018).

Note that this would not constitute cross-validation, but rather enable assessment of content and discriminant validity as well as model fit in the present study, and also enable comparisons with any future uses or applications of the resulting model. Any cross-validation would, however, need to come from a new sample or data (Brown, 2006). Also, the use of a common factor model as a method of dimension reduction is consistent with other researchers in uncovering latent variables of FML (e.g., Lonsdale & North, 2011; Groarke & Hogan, 2018), as it was hypothesised that underlying dimensions influence the ways in which items correlate and is thus reflective in nature. This is in contrast to formative approaches that apply PCA (e.g., Schäfer et al., 2013; Kuch & Wöllner, 2021), whereby latent components are uncovered by maximising explained variance across eigenvectors (see Widaman, 2018). The utility of the applied methods therefore allows a structure to be identified that subsequent work within this thesis is able to reapply as a potential measure of utilitarian FML. This is essential to realising the first goal of this thesis, outlined in section 1.2.

6.4 Results

6.4.1 Initial item reduction

Before proceeding to EFA, inter-item correlations were inspected to mitigate the presence of highly correlated items, as per Field's (2018) recommendation that bivariate correlation coefficients at the .80 level may indicate substantial overlap between the items in question.

This was the case for the items 97 (*To reduce feelings of being lonely when social interaction is not possible*) and 99 (*To reduce feelings of loneliness when you are alone*; see Appendix B), measuring *Company and Music as Proxy* (r (325) = .80, p<.001, n = 327). Regarding interpretation, Field notes that this requires the researcher to qualitatively assess whether one of the items should be dropped, and if so which. In this instance, item 99 was removed from subsequent analyses as it was deemed that item 97 held greater clarity in terms of expressing an absence of social interaction (e.g., being 'alone' may be conflated with potentially emotional experiences), and was hence more closely aligned with the function's definition (see Appendix A).

Additionally, Watkins (2018) notes that the variables submitted to factor analysis should adequately represent the domains thought relevant, and that unrelated variables from theoretically divergent domains should not be included. With this inference in mind, it was considered that two further items should not be included in the factor analysis, namely items 104 (*To experience music whilst you are making it* yourself) and 105 (*To perform or generate music*; see Appendix B), which were initially included to encompass the full scope of Maloney's (2019) framework. Both of these items relate to *Musicking* in the sense of the creation or performance of music itself (such as through playing an instrument) which theoretically diverged from the intended scope and application of these analyses which are orientated towards autonomous music listening, rather than other forms musical exposure, and thus on reflection were considered to be inconsistent with study intentions. With these items removed, it was deemed appropriate to proceed to EFA with the remaining 111 items.

6.4.2 Factor Analysis

EFA was conducted in JASP (version 0.16.1) to uncover an underlying structure of the items. The EFA used Parallel Analysis, in which factors are extracted by comparing eigenvalues observed in the correlation matrix to eigenvalues generated in a simulated matrix using a Monte-Carlo method, producing more reliable results than Scree tests or Kaiser's eigenvalue greater-than-one rule (Ledesma & Valero-Mora, 2007; Matsunaga, 2010; Lim & Jahng, 2019). This was deemed to be the most appropriate factor extraction technique in the absence of an a priori structure that would manually determine the number of factors. This was used in conjunction with an Oblique (oblimin) rotation to allow for inter-factor correlation and a Maximum Likelihood (ML) estimation method, which is generally recommended where possible and suitable for ordinal data with five or more response categories (Dolan, 1994; Beauducel & Herzberg, 2006; Rhemtulla et al., 2012; Robitzsch, 2020).

For the initial EFA iteration, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser & Rice, 1974) indicated the data were suitable for factor analysis (KMO = .95). Bartlett's Test of Sphericity was then inspected to ensure the correlation matrix of the items suitably diverged from the identity matrix (Watkins, 2018), the result of which was significant ($\chi^2(6105) = 27416.079$, $p<.001$), further indicating that the data was suitable for factor analysis. In the initial solution, a seven-factor structure was implied to fit the data best ($\chi^2(5349) = 8136.109$, $p<.001$, CFI = .869, TLI = .848, RMSEA = .040), however, a substantial number of items ($n = 61$) failed to load onto any factor, and the removal of such items is recommended to generate stable, parsimonious solutions (Worthington & Whittaker, 2006). Moreover, fit indices for the initial model were below acceptable thresholds (CFI and TLI $\geq .90$; RMSEA $\leq .08$; Brown, 2006). Also, it should be acknowledged that the χ^2 statistic should (ideally) be non-significant ($p > .05$), however, is sensitive to a number of biases including sample size (both large and small n). It often rejects models for trivial misspecifications and is accompanied by a stringent assumption that the predicted variance-covariance matrix (Σ) is equal to the sample variance-covariance matrix ($S = \Sigma$), something rarely achieved in real-world data (Brown, 2006). These limitations result in inflated type II errors (Brown, 2006; Perry et al., 2015), for this reason alternative fit measures more robust to these sensitivities are utilised during factor analyses (e.g., CFI, TLI, RMSEA). A cut-off value for factor loadings of .50 was used as, given the large number of items, it was deemed suitable to retain items that load more strongly than the lower thresholds applied with smaller numbers of variables (e.g., .40; Maskey et al., 2018).

An iterative process was subsequently used to generate a stabilised solution in which items were retained if they strongly loaded onto one factor at the $\geq .50$ level, did not cross-load onto multiple factors, and demonstrated adequate communalities (Worthington & Whittaker, 2006; Eaton et al., 2019; Güvendir & Özkan, 2022). The iterative process of removing items failing to meet these criteria was conducted with theoretical as well as statistical considerations based on an understanding of the literature previously laid out in this thesis to generate a theoretically as well as statistically sound factor structure (Beavers et al., 2013; Güvendir & Ozkan, 2022). This process resulted in the removal of 70 items in all, retaining 41 in a stable solution. The final iteration of the EFA implied a six, rather than seven, factor solution fit the data best ($\chi^2(589) = 901.358$, $p<.001$, CFI = .963, TLI = .948, RMSEA = .040), explaining 61.1% of the observed variance. Hinkin (1998) suggests that approximately 60% is appropriate in this regard, however, also notes that there are no strict guidelines for EFA as such. Table 6 shows the result of the stabilised EFA.

Table 6 Stabilised EFA solution of CFF items

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Identity and Social Bonding						
To identify with others through your shared values and/or culture	**.784**	-.114	-.034	.020	.099	.073
To help express your identities and values to others	**.769**	.040	.088	.006	-.079	.003
To feel that certain artists, pieces, or genres of music are central to your social group's culture and sets you apart from others	**.763**	.012	-.082	-.030	-.002	.007
To help you bond with others, and to subsequently feel a sense of belonging with those individuals	**.745**	.065	-.050	.042	.067	.001
To act as a topic of discussion with others and ease communication or interaction	**.739**	-.084	.013	.087	.051	-.023
To match a group's dynamic so you are able to bond with group members when listening with others	**.725**	-.073	.027	-.014	.106	.067

To help to bond and/or interact with others	**.718**	-.006	-.080	.060	.122	.044
To identify or feel connection with others who share your taste in music	**.716**	.044	.015	.042	.025	-.096
To differentiate yourself from others in order to stand out	**.700**	.016	.103	-.091	-.018	-.040
To act as a reference point for your social groups to maintain your shared culture (e.g., feelings of nostalgia with others)	**.693**	-.053	-.048	.114	.060	.091
To allow others to survey your music taste and gather information about you	**.678**	.006	.040	-.023	-.093	-.031
To help you and your social group to express your culture or values	**.670**	.087	-.049	.008	.100	.026
To feel validation or approval as a part of a group	**.664**	-.098	.012	.099	-.047	.021
To establish and maintain a part of your personal identity	**.647**	.177	.113	.051	-.181	-.069
To foster and develop new personal relationships	**.645**	.078	-.002	-.032	.101	.065
To act as a point of symbolic representation of who you are	**.642**	.222	.117	-.128	-.148	-.037
To help you maintain your identity as it reflects who you are as a person	**.614**	.198	.011	.037	-.146	.067
To share content with others by sharing (i.e., sharing playlists or mixtapes)	**.564**	-.040	.077	.015	.068	.083

Emotion Regulation						
To distract yourself from negative or stressful situations	-.075	**.828**	.015	.083	.044	.015
To relieve stress and negative emotions associated with negative events or situations	-.014	**.723**	.032	.076	-.006	.056
To manage emotions that you may be experiencing despite external influences, whether they are positive or negative	.072	**.701**	.129	-.061	-.027	.040
To find meaning within music that allows you to reduce negative emotions or moods	.213	**.700**	.021	-.061	-.005	.000
To feel certain specific emotions, such as joy or sadness	.099	**.696**	-.015	-.089	.092	-.006
To help you reverse your emotions or moods	.013	**.678**	-.015	.109	.064	-.046
To act as a therapeutic tool to help you reduce negative emotions	.007	**.678**	-.004	.074	.137	.055
To help you escape stressful events or situations	-.051	**.668**	.054	.190	.045	.048
To distract yourself from unwanted thoughts and/or feelings	-.098	**.644**	.099	.166	.015	.086
To help you feel certain specific emotions when feeling 'neutral' (e.g., neither happy nor sad)	.218	**.553**	.066	.019	-.037	.092
Focus and Concentration						
To help you focus or concentrate on tasks	-.062	.021	**.829**	.080	.025	.011

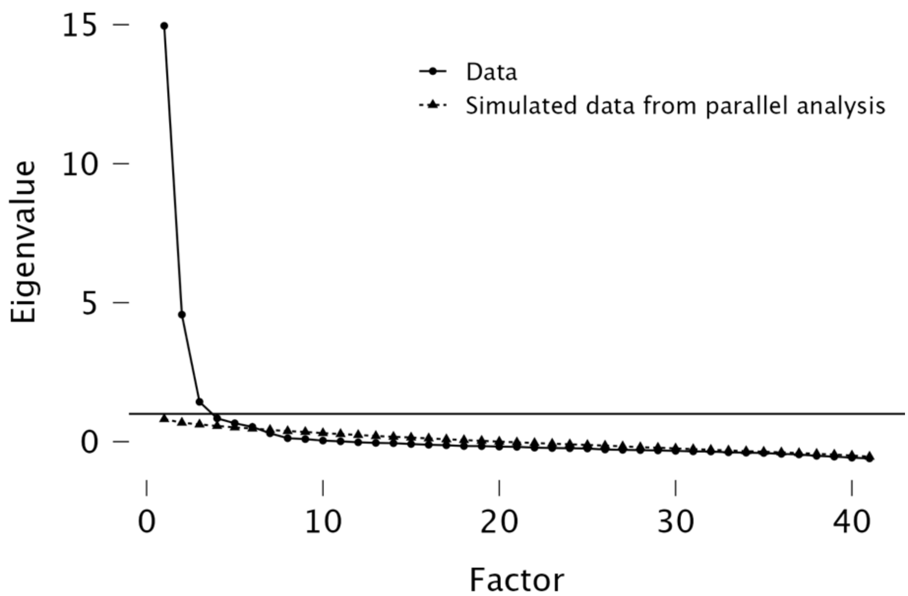
To help you ‘flow’ when trying to concentrate on something	.057	.043	**.813**	-.048	.035	.038
To stop external factors from distracting you when trying to concentrate on a task	.007	-.026	**.704**	.118	.050	.054
To help you attain the necessary mindset to working on certain tasks	.069	.082	**.557**	.004	.188	.033
Background and Accompaniment						
To avoid silence when you’re alone (e.g., playing music when nobody else is home)	.028	.051	.047	**.818**	.010	.002
To feel a sense of company in the absence of others (e.g., playing the radio when home alone)	.078	.066	.053	**.749**	-.021	-.021
To reduce feelings of being lonely when social interaction is not possible	.062	.274	-.030	**.619**	.057	.068
To provide background noise and remove silence	.033	-.054	.327	**.534**	.015	.044
Physiological Arousal						
To help you maintain pacing during physical activities, such as yoga, walking or whilst in the gym	.047	.011	.010	.035	**.853**	.003
To help physically stimulate you to carry out physical tasks, such as exercise or sports	.015	.092	.045	-.008	**.764**	.041
To help you achieve goals by motivating you to further action	.019	.094	.217	-.047	**.648**	-.035

(such as increased effort during exercise)						
Earworm Fulfilment						
To remove songs that are 'stuck' in your head to prevent distraction	-.009	-.018	-.007	.012	-.009	**1.008**
To satisfy or clear songs that are 'stuck' in your head	.051	.107	.078	-.078	-.014	**.698**
Cronbach's α	.95	.94	.88	.88	.86	.86
McDonald's ω	.95	.94	.89	.89	.87	.86

Note. Applied rotation method is oblimin.

A scree plot was examined to visualise the number of factors to extract, which in the case of Parallel Analysis can be visually assessed by observing at which point the observed and simulated data intersect. Although Parallel Analysis is generally considered more reliable than standalone inspections of a scree plot (Lim & Jahng, 2019), it can still be useful to visualise the simulated intersection of the data (DeVellis, 2017). In this case, the six-factor solution was consistent with the interpretation of the scree plot's intersection, shown in Figure 6.

Figure 6 Scree plot of parallel analysis



Items loading onto Factor 1 (*Identity and Social Bonding*) contained a broad set of 18 items relating to the use of music to bond, interact, and identify with others. These included items representing the CFF functions: *Group Identity*, *Express Identity and Values*, *Interaction and Bonding, Communication, Approval and Cultural Capital, Maintain and Express Cultural Values, Symbolic Difference, Express Identity and Values, Control and Conformity, Create and Maintain Identity*, *Surveillance*, and *Situational Relevance*.

Factor 2 (*Emotion Regulation*) also consisted of a fairly broad selection of CFF functions through 10 items: *Escapism and Venting*, *Change or Shift Emotions, Trigger or Elicit Emotions*, *Therapy, Regulate and Maintain Emotions*, and *Relaxation and Stress Relief*. The relative importance of these two factors is theoretically consistent with the model presented by Lonsdale and North (2011) insofar as mood management and social functions were similarly found to be the two most important factors underpinning music engagement, albeit in the reversed order in the present case.

Factor 3 (*Focus and Concentration*) contained four items, of which three were attributable to the function *Focus and Concentration*, whilst one item represents *Mental State*. Similarly, Factor 4 (*Background and Accompaniment*) contained four items – three of which represent *Company and Music as Proxy*, and one *Background*. Factor 5 (*Physiological Arousal*) contained three items, with one relating to *Motivation*, *Activation Arousal and Response*, and *Pacing and Movement* respectively. Finally, Factor 6 (*Earworm Fulfilment*) exclusively contained two items relating to its namesake. These six factors were deemed to be theoretically sound and practically distinct when considered with the wider literature.

6.4.3 Subsequent iteration with reduced subset of items

These six factors constitute theoretically consistent FML and are practically distinct when considered with the wider literature. However, upon inspection and interpretation of the 41 items across the six factors, it was considered that there was room for further reduction and simplification for the sake of brevity and reduced conceptual overlap. It was noted that the first two factors (*Identity and Social Bonding* and *Emotion Regulation*) retained a notably larger number of items than the others (18 and 10 respectively). This theoretically runs the risk of exacerbating participant fatigue when considering reapplying the prospective structure, which may not be necessary given a large number of items for the two constructs. As such, it was considered whether a subset of the highest-scoring items from the stabilised EFA within these two factors would suffice, reducing conceptual overlap and unnecessary length. Worthington and Whittaker (2006) note that models can be adapted by researchers on theoretical and practical grounds based on stable EFA solutions and suggest retaining a subset of the highest scoring items in sub-scales with a larger than the desired number of items. Similarly, Robinson (2018) notes that structures should be as concise as possible and that longer than desired structures may be adapted by using the highest loading items of a set of factors.

Furthermore, we considered that following this reduction, it would be beneficial to consequently inspect model fit and discriminant validity in a constrained setting using CFA (Worthington & Whittaker, 2006; Matsunaga, 2010; Schmitt et al., 2018), for which it is preferential to have at least three items per factor (e.g., Brown, 2006; Schmitt et al., 2018). In the case of the sixth factor, *Earworm Fulfilment*, this is not met and as such, it was considered pragmatic to also remove these two items as future cross-validation would require at least three items for each factor. To be clear, CFA is typically applied in scale development or construct validation after the underlying structure has been tentatively established by prior analyses or on theoretical bases (Brown, 2006; Matsunaga, 2010). Supplementing an EFA with CFA allows researchers to demonstrate model fit in a restrictive setting, allow for the comparison of model fit with future research, and establish the extent to which results differ based on model specifications (Schmitt et al., 2018). However, Worthington and Whittaker (2006) suggest that any adaptations made are subject to a further EFA iteration, to ensure that modifications did not affect model parsimony and result in cross-loadings or poorly associated items. As such, this further reduction first verifies that the structure of the remaining items remains consistent with the previous model, conditions which if are met would then lead to constrained fit and inspection via CFA.

With these motivations in mind, a subset of the six most strongly associated items were taken from the two constructs with the largest number of items in the stabilised EFA solution (see Table 6) as these were assessed to maintain minimal conceptual overlap and would suffice in representing the relevant constructs of *Identity and Social Bonding* and *Emotion Regulation*. Using the initial stabilised solution as a grounding, an EFA was run where these highest loading items were retained in both *Identity and Social Bonding* and *Emotion Regulation* and all other items were left to load freely, thus assessing whether these reduced sets of items would generate a stable structure across five factors as expected. As mentioned, the two items relating to *Earworm Fulfilment* were also omitted. This reduced subset of 23 items maintained the relative five-factor solution ($\chi^2(148) = 213.614$, $p<.001$, CFI = .985, TLI = .975, RMSEA = .037), albeit in a reduced subset, explaining 64.1% of the variance, shown in Table 7.

Table 7 EFA factor loadings of reduced subset of items

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Identity and Social Bonding					
To identify with others through your shared values and/or culture	**.813**	-.076	.003	.008	.051
To help express your identities and values to others	**.750**	.079	.078	.022	-.095
To act as a topic of discussion with others and ease communication or interaction	**.736**	-.066	.003	.088	.025
To feel that certain artists, pieces, or genres of music are central to your social group's culture and sets you apart from others	**.733**	.075	-.082	-.029	-.033
To match a group's dynamic so you are able to bond with group members when listening with others	**.725**	-.016	.046	-.026	.076
To help you bond with others, and to subsequently feel a sense of belonging with those individuals	**.705**	.098	-.024	.035	.026
Emotion Regulation					
To distract yourself from negative or stressful situations	-.102	**.822**	.017	.100	.045
To relieve stress and negative emotions associated with negative events or situations	-.042	**.743**	.046	.090	-.010
To find meaning within music that allows you to reduce negative emotions or moods	.206	**.705**	.035	-.046	-.032
To manage emotions that you may be experiencing despite external influences, whether they are positive or negative	.072	**.696**	.145	-.048	-.023
To feel certain specific emotions, such as joy or sadness	.073	**.687**	-.017	-.064	.088

To help you reverse your emotions or moods	.005	**.674**	-.025	.117	.041
Focus and Concentration					
To help you focus or concentrate on tasks	-.052	.029	**.861**	.054	-.014
To help you 'flow' when trying to concentrate on something	.066	.049	**.817**	-.055	.031
To stop external factors from distracting you when trying to concentrate on a task	.010	-.001	**.739**	.086	.017
To help you attain the necessary mindset to working on certain tasks	.054	.097	**.543**	-.001	.200
Background and Accompaniment					
To avoid silence when you're alone (e.g., playing music when nobody else is home)	.010	-.008	.010	**.880**	.015
To feel a sense of company in the absence of others (e.g., playing the radio when home alone)	.039	.036	0.022	**.782**	-.023
To reduce feelings of being lonely when social interaction is not possible	.047	.228	-.050	**.672**	.081
To provide background noise and remove silence	.020	-.082	.319	**.561**	.020
Physiological Arousal					
To help physically stimulate you to carry out physical tasks, such as exercise or sports	-.010	.027	-.022	.008	**.867**
To help you maintain pacing during physical activities, such as yoga, walking or whilst in the gym	.030	-.036	-.022	.033	**.854**
To help you achieve goals by motivating you to further action (such as increased effort during exercise)	-.001	.056	.159	-.046	**.687**
Cronbach's α	.89	.90	.88	.88	.86
McDonald's ω	.89	.90	.89	.89	.87

Note. Applied rotation method is oblimin.

Given that the factor structure was not theoretically altered by the further item reduction (Worthington & Whittaker, 2006), a CFA on the structure was conducted in R (version 4.1.0; R Core Team, 2021) using the *lavaan* package (Rosseel, 2012) to estimate the model using a robust maximum-likelihood (MLR) estimation method, and the *semTools* package (Jorgensen et al., 2022) to calculate the average variance explained (AVE) of each latent factor. To scale the latent variables and ensure the model was identified, factor variances were fixed to 1 as this allows each indicator to be freely estimated and removes the need to select an arbitrary reference variable (although either method is generally adequate in CFA settings; Brown, 2006; Kline, 2016). The CFA indicated good model fit indices ($\chi^2(220) = 367.506$, $p < .001$, CFI = .963, TLI = .957, RMSEA = .049), implying the factor structure was able to replicate the characteristics of the data. CFA results are reported in Table 8, whilst a plot of the constrained model can be seen in Figure 7.

Table 8 CFA loadings for 23-item structure

	Item Reference	Unstandardised Estimate (λ)	Standardised Estimate (β)	*SE*	*p*	AVE
Identity and Social Bonding						.574
To identify with others through your shared values and/or culture	S1	0.898	.787	.054	<.001	
To help express your identities and values to others	S2	0.908	.781	.051	<.001	
To feel that certain artists, pieces, or genres of music are central to your social group's culture and sets you apart from others	S3	0.870	.716	.060	<.001	
To help you bond with others, and to subsequently feel a sense of belonging with those individuals	S4	0.824	.771	.051	<.001	
To act as a topic of discussion with others and ease communication or interaction	S5	0.830	.740	.053	<.001	
To match a group's dynamic so you are able to bond with group members when listening with others	S6	0.842	.751	.051	<.001	

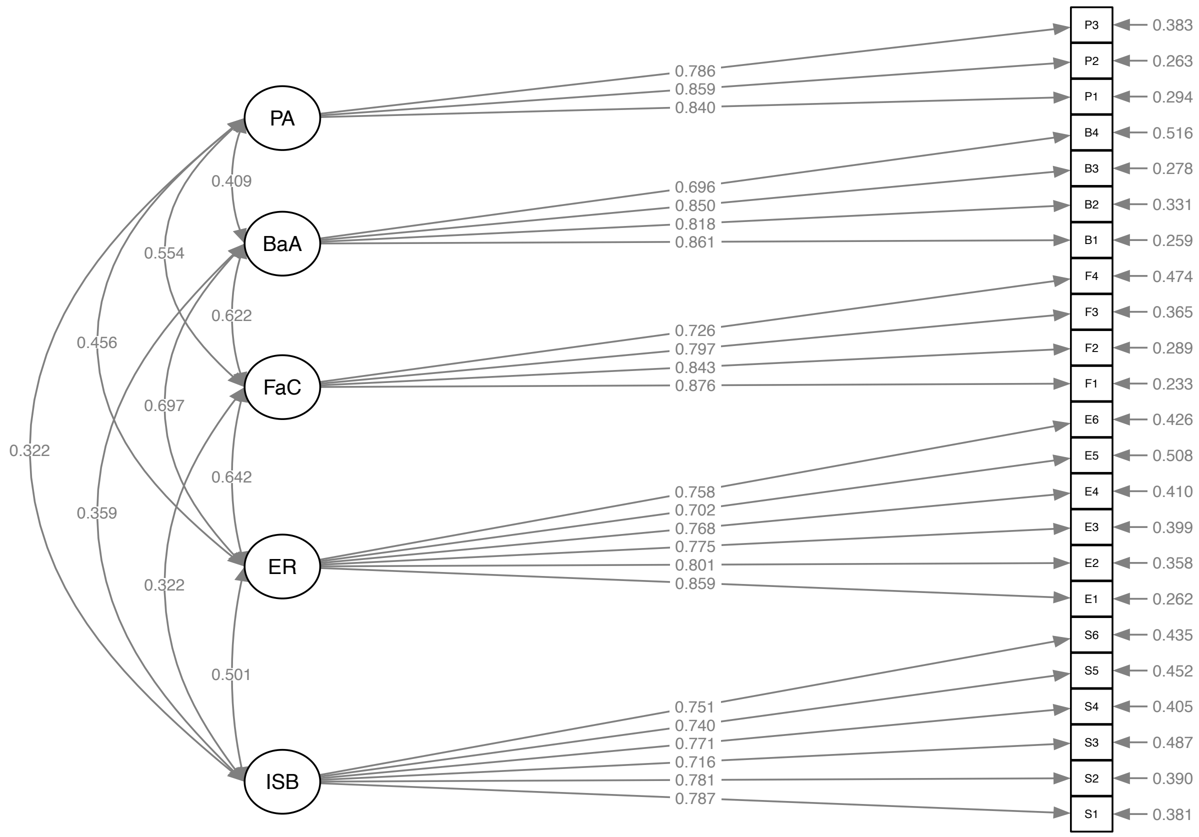
Emotion Regulation						.609
To distract yourself from negative or stressful situations	E1	1.009	.859	.047	<.001	
To relieve stress and negative emotions associated with negative events or situations	E2	0.922	.801	.053	<.001	
To manage emotions that you may be experiencing despite external influences, whether they are positive or negative	E3	0.972	.775	.051	<.001	
To find meaning within music that allows you to reduce negative emotions or moods	E4	0.964	.768	.053	<.001	
To feel certain specific emotions, such as joy or sadness	E5	0.756	.702	.055	<.001	
To help you reverse your emotions or moods	E6	0.878	.758	.052	<.001	
Focus and Concentration						.666
To help you focus or concentrate on tasks	F1	1.019	.876	.048	<.001	
To help you 'flow' when trying to concentrate on something	F2	1.039	.843	.046	<.001	
To stop external factors from distracting you when trying to concentrate on a task	F3	0.933	.797	.049	<.001	
To help you attain the necessary mindset to working on certain tasks	F4	0.786	.726	.055	<.001	
Background and Accompaniment						.663

To avoid silence when you're alone (e.g., playing music when nobody else is home)	B1	1.027	.861	.051	<.001	
To feel a sense of company in the absence of others (e.g., playing the radio when home alone)	B2	1.047	.818	.055	<.001	
To reduce feelings of being lonely when social interaction is not possible	B3	1.100	.850	.053	<.001	
To provide background noise and remove silence	B4	0.766	.696	.064	<.001	
Physiological Arousal						.688
To help you maintain pacing during physical activities, such as yoga, walking or whilst in the gym	P1	1.090	.840	.056	<.001	
To help physically stimulate you to carry out physical tasks, such as exercise or sports	P2	0.960	.859	.051	<.001	
To help you achieve goals by motivating you to further action (such as increased effort during exercise)	P3	0.886	.786	.056	<.001	

Note. *SE* = Standard Error. AVE = Average Variance Explained. *Item references pertain to Figure 7.*

A trade-off of with this process is that the internal reliability statistics of the reduced subset is lower than that of the 41-item version (see Table 7). This is, however, expected, as more items tend to yield higher internal consistency values regardless of the theoretical or practical implications (Worthington & Whittaker, 2006). The internal reliability statistics were nevertheless well above the .70 threshold (Hair et al., 2014), thus providing confidence of construct reliability for each factor. This process has therefore reduced the number of items subsumed in these subscales with practical motivations following theoretical and statistical justifications. The generation of the more concise (23-item) measure is motivated by a desire to reduce overlap and mitigate the length of the structure, particularly conducive for more intensive study designs. Moreover, standardised parameter estimates (i.e., factor loadings; β) were all >.50 at the $p<.001$ level and AVE was also >.50 for each latent factor, meeting thresholds recommended by Hair et al. (2014). These results provide additional confidence that the structure and its latent factors hold suitable convergent reliability as well as model fit.

Figure 7 CFA model of reduced factor model



Note. Bi-directional arrows indicate factor covariances. Arrows to items indicate standardised parameter estimates (factor loadings). ISB = *Identity and Social Bonding*, ER = *Emotion Regulation*, FaC = *Focus and Concentration*, BaA = *Background and Accompaniment*, PA = *Physiological Arousal*.

Moderate factor covariances (Φ) in the CFA model were observed with coefficients ranging from .322 to .697. To ensure discriminant validity holds between constructs (i.e., that they are distinct), the AVE of each factor can be compared with the square of each covariance between them (Φ^2; Fornell & Larcker, 1981; Hair et al., 2014). Table 9 shows Φ^2 between factors in relation to each factor's AVE. Each squared coefficient between latent factors was lower than the AVE of the factors themselves, providing evidence of discriminant validity across the construct because more of each factor's variance is explained by its item measures than it covariance with other factors (Hair et al., 2014). This provides evidence that the structures in the specified model are distinct from one another, providing evidence of discriminant validity. This is in addition to content validity arising from the theoretical consistency and delineation within and between factors' items, and construct validity arising from AVE, reliability coefficients and standardised factor loadings. These are the primary criteria to assess model validity in SEM (Bollen, 1989).

Table 9 Factors' AVE in relation to squared covariance

Factors	Φ	Φ^2
Identity and Social Bonding (.574) & *Emotion Regulation* (.609)	.501***	.251
Identity and Social Bonding (.574) & *Focus and Concentration* (.666)	.322***	.104
Identity and Social Bonding (.574) & *Background and Accompaniment* (.663)	.359***	.129
Identity and Social Bonding (.574) & *Physiological Arousal* (.688)	.322***	.104
Emotion Regulation (.609) & *Focus and Concentration* (.666)	.642***	.412
Emotion Regulation (.609) & *Background and Accompaniment* (.663)	.697***	.486
Emotion Regulation (.609) & *Physiological Arousal* (.688)	.456***	.208
Focus and Concentration (.666) & *Background and Accompaniment* (.663)	.622***	.387
Focus and Concentration (.666) & *Physiological Arousal* (.688)	.554***	.307
Background and Accompaniment (.663) & *Physiological Arousal* (.688)	.409***	.167

Note. AVE of each factor is in brackets. (***p<.001).

6.5 Discussion

This first study sought to generate a psychometric structure of FML by identifying a latent structure of items derived from the CFF (Maloney, 2019). Using the definitions attributed to all 53 functions, a list of 114 items was generated which were assessed on a five-point Likert scale rated through an online survey (n = 327). EFA was used to uncover an underlying factor structure of the items, using both statistical and theoretical criterion to generate a parsimonious model (Beavers et al., 2013). A model containing 41 items (with 61.1% of variance explained) was generated through this process, at which point it was considered that the structure could be reduced further due to conceptual overlap in the content of the remaining factors, specifically the first two factors which retained a larger number of items than the other three. A reduced 23-item structure containing the six highest loading items of each of these factors, minus two items forming the sixth factor (*Earworm Fulfilment*), was found to retain the relative five-factor structure, with 64.1% of variance explained, mitigating conceptual overlap and unnecessary length. CFA was then applied to constrain and inspect the model in restricted conditions (Schmitt et al., 2018), in which the structure satisfied relevant conditions in restrictive settings (i.e., model fit, AVE, standardised factor loadings, and discriminant validity).

Whether taking this model forward in its current form or not depends on whether the identified factors are theoretically sound to facilitate possible contingencies in future studies. Indeed, cross-validation is needed in new data, however, it is prudent to discuss each of the identified factors according to their relevance and placement with wider literature. Such inferences are important to assess whether there is confidence that the retained structure(s) can be applied as suitable measures of FML from the utilitarian perspective.

6.5.1 Discussion of implied factors

1. ***Identity and Social Bonding***

Factor 1 contains six items indicative of social and identity-based function. This dimension is consistent with findings that music acts as a means of self-socialisation (whereby individuals actively engage with their developmental needs in everyday life), identity building, and communication (Arnett, 1995; Rentfrow & Gosling, 2006; Cunningham & Nichols, 2009;

Egermann et al., 2013). There is value in items constituting part of a psychometric instrument gauging such dimensions of *functionality*. As was discussed earlier in this thesis, music's ability to facilitate social bonding by serving multiple adaptive functions has been proposed as an overarching "super function" (Savage et al., 2021; Hansen & Keller, 2021). This would imply that music as a means of facilitating social interactions and bonding is perhaps one of its most important uses, yet the extent to which social *functionality* has been captured, classified, and applied as a distinct FML is thus far limited.

Meanwhile, music preferences have been seen to carry individual identities, which may be communicated in social settings to share values, identity, and culture (Rentfrow & Gosling, 2003). In social settings, perceptions of shared culture or values help individuals perceive similarity with others which leads to liking; in turn, this allows individuals to infer further similarity onto others which reaffirms said liking (e.g., Caprara et al., 2007). With regards to music, shared preferences between individuals positively relate to perceptions of both similarity and liking due to music's ability to convey identity and culture, and to also facilitate communication between individuals (Lonsdale & North, 2009; Boer et al., 2011; Cross, 2014). At the individual level, it has been noted that music provides a structural framework within which reminiscence and positioning of the self may be facilitated, which is to say music may function as a means of autobiographical reminiscence (Lamont et al., 2016).

Furthermore, the use of music as a vessel for the presentation or affirmation of identity holds a number of sociological as well as psychological implications. For instance, differences have been seen in the ways in which people report their own musical preferences publicly in comparison to those they privately endorse (Finnäs, 1989), and in Rentfrow and Gosling's (2003) work to generate of a cohesive model of music preference, they noted that young adults held beliefs that their preferences revealed information to others about their personalities. Moreover, it is in the subscription to a musical culture that individuals associate themselves with a wider set of values that form connotations about an individual from a shared perspective (Russell, 1997; Lamont et al., 2016).

Additional viewpoints have considered different stimuli as tangible, touchable embodiments of individual identity and values. For instance, Dittmar (2008) suggests that physical objects and stimuli (such as possessions and cultural phenomena; like music) act as self-extensions of individual identity. When such stimuli are projected into the immediate environment, then the identity of the individual is inferred, enhancing feelings of autonomy and control. The phenomenon of psychological ownership is notably characterised by Pierce et al. (2003) in three main parts: efficacy and effectiveness, self-identity, and sense of place. Possessions or stimuli may be used for the purposes of communication and the sharing of control, culture, identity, and values; however, this may vary from context to context (Pierce et al., 2003). This is consistent with key concepts of cultural capital, described by Bourdieu (1986) as an individual's accumulated knowledge and skillset. In social contexts, an individual's capital operates as an aggregate of resources that are associated with "a durable network of more or less institutionalised relationships of mutual acquaintance and recognition—or in other words, to membership in a group" (p. 247).

Pierce et al. (2003) further note that the sharing and interaction of music is contextually varied, reaffirming a link between social uses of music and goal-attainment. Contextual variables may be subject to additional structural and cultural dimensions that inform behaviour, values, and beliefs that may be shared within or between cultures. Hofstede (1980) defines 'culture' as "the collective programming of the mind which distinguishes members of one human group from another" (p. 25), consistent with research relating to music preference (e.g., Rentfrow & Gosling, 2003; Lonsdale & North, 2009). Moreover, perceptions of psychological ownership regarding music have been related to both physical and digital music formats; and listening to one's own music may manifest in perceptions of environmental control and consolidation, as well as identity formulation (Giles et al., 2007; Sinclair & Tinson, 2017; Danckwerts & Kenning, 2019), thus contextualising music as a form of cultural capital that serves social communication and interaction in varied ways. However, whilst there is a body of research that identifies social dimensions of music listening and music-related expression of identity as important factors characterising listening behaviour, research also found such perceptions to be significantly less important in day to day listening experiences. Kuch and Wöllner (2021), for instance, evidence that when it comes to mobile music listening, mood related and cognitive functions are most prevalent, whereas social dimensions were the least important functions that were observed.

With these inferences in mind, we may consider the content of the factor's indicators as an expression of communicative *functionality* between individuals using music as a basis, which additionally suggests that identity-related functions can be contextually orientated and somewhat based on an exchange of cultural capital to feel a sense of kinship or ease. Music is a contributing factor to self-identity which furthers agency but also facilitates interaction. The implied function may act as a representation of the self; perhaps to ease social interaction and/or feel agency, security, and comfort. However, the contextual reasons that inform the nuances within this factor are not immediately clear, highlighting the need to associate it with contextual data. This would make it clearer as to whether *Identity and Social Bonding* predominantly occurs alongside social and individual listening activities, or both. Overall, the implied function of *Identity and Social Bonding* is reflective of environmental projections, concurrent social perceptions, and communication, as well as personal identity formulation and consolidation.

2. *Emotion Regulation*

Factor 2 also contains six items, reflecting dimensions of mood regulation. It has been suggested that music's ability to manage and enhance mood is its single most important function, and that its entertainment value may help facilitate this ability to regulate, change, or enhance emotions (Ter Bogt et al., 2010). Karreman et al. (2017) note that individuals may influence regulatory FML by applying emotion regulation strategies when listening, whilst Saarikallio and Erkkilä (2007) documented an association between regulatory strategies applied during music listening with regulatory strategies employed in the wider emotion regulation literature. Furthermore, Saarikallio (2011) evidenced that a variety of regulatory goals and strategies remain consistent throughout life, whilst Lamont et al. (2016) outline that listeners may use music for a range of regulation strategies; evidenced by observations that management of both positive and negative mood states serve as tangible goals of music listening (e.g., Lonsdale & North, 2011).

Music selection has been associated with self-regulatory goals and expected effects of music listening. The intention to achieve mood enhancement is often initially achieved by experiencing cognitive reappraisal or distraction (Van den Tol & Edwards, 2015). Moreover, therapeutic applications of music listening have been documented to serve as a coping mechanism (Ter Bogt et al., 2010) which has in turn been associated with mood regulation strategies that are informed by stressful life events (Silverman, 2020; Henry et al., 2021).The aforementioned AFML scale (Groarke & Hogan, 2018) refers to a variety of emotion regulation strategies, highlighting the expansive nature of mood regulation, however. Similarly, the Music in Mood Regulation (MMR) scale developed by Saarikallio (2008) is broad, ranging from factors relating to *Diversion* to general *Entertainment*. As such, a potential limitation with the present model may be in its ability to provide comparable nuances in terms of sub-domains of mood regulation, due to the broad nature of this function in its own right. This may be an apparent trade off with a utilitarian approach, in that broader domains of *functionality* may be captured, but that the regulatory subdomains may be better represented in designated models.

One interesting observation within this factor, however, is that there is seemingly limited attention towards positive mood enhancement, and more towards negative mood management. Whilst there are items that allude to the mitigation of negative emotions and reversing of emotions, there are no explicit statements that translate explicitly to positive mood enhancement. Researchers have previously separated music-based mood regulation into positive and negative strategies. Lonsdale and North (2011), for instance, distinguish between positive and negative mood management as separate functions altogether. Whilst particular items may allude to positive regulatory outcomes (such as items relating to feeling 'specific' emotions and to 'reverse' emotions) there is not an explicit set of statements that reflect positive mood management. Hypothetically speaking, a possible explanation for this may be that positive feelings as a consequence of music listening may be something of a by-product of different contributory factors contained elsewhere in the model. Though difficult to assess, cross-validation of the structure may provide insights as to the consistency and reliability of the present items regarding measurement, as well as enable greater exploration of covariances between this and other factors to explore this further.

3. *Focus and Concentration*

Factor 3 contained four items, of which three relate to *Focus and Concentration* and one to *Mental State*. The content of this factor is implicit in music's use in regulating cognitive processing ability during tasks. As has been discussed throughout this thesis, the heightened integration of music in everyday life has afforded listeners opportunities to make increasingly self-guided decisions about when and how to listen to music, of particular relevance to this factor due to implied task-orientated *functionality*. As mentioned in section 3.2.2, music is often employed to accompany both work and private study (Lamont et al., 2016). With regard to workplace environments, people use music to help improve mood generally, but also to help improve concentration and focus during tasks requiring little to no interaction with colleagues (Haake, 2011). When at home, music is frequently used during private study (e.g., Greasley, 2008). For instance, university students often report listening to music to help with focus during study (Greasley & Lamont, 2011), with the caveat that whilst some find music's presence beneficial others may find it detrimental. Nevertheless, Greasley and Lamont (2011) argue that students generally report music to be an essential tool in enabling concentration. This may be for a number of other underlying reasons such as distraction and silence avoidance, which is also partially seen in the *Background and Accompaniment* factor. Interestingly, however, it has further been noted that music with particular characteristics may be optimal during such tasks. For instance, dance music may be congruent when typing up notes, whilst the avoidance of music with lyrical content is key during study or revision (Avila et al., 2012). This supports an underlying assumption of this thesis that music serves optimal levels of *functionality* when there is congruence between the listener's goal and the affective content of the music being employed. This permeates more than just this one mode of *functionality*, however, since it broadly reflects the idea that music's ability to stimulate affective experiences is a key component of utility, which may also be key during emotion regulation, for example (e.g., Juslin et al., 2008; van Goethem & Sloboda, 2011). As such, FML should be treated as a continuous rather than discrete measure, in line with extant research (e.g., Greb et al., 2019).

4. *Background and Accompaniment*

Of the four items contained within Factor 4, three relate to *Company and Music as Proxy* and one relates to *Background*. This function seemingly refers to feeling company in the absence of others, reducing feelings of loneliness, and silence avoidance. Hypothetically, this may serve multiple purposes such as music's ability to serve as a social surrogate to reduce loneliness and raise mood (Schäfer et al., 2020). Within the implied factor structure, *Focus and Concentration* was separately identified as an underlying construct in the observed data. Theoretically speaking, both of these functions serve purposes relating to environmental perception, whereas *Background and Accompaniment* is about perceiving an absence of loneliness whilst *Focus and Concentration* refers to focused cognitive ability. The separation between these perceptual characteristics may be theoretically viewed as divergence in the role of background auditory stimuli (also referred to as irrelevant sound effect, or ISE; Threadgold et al., 2019). ISE has limited beneficial effects on cognitive performance, which further corroborates the notion that such an aspect of music as a broader background function is seemingly unlikely to occur in instances requiring creativity, focus or concentration (Perham & Vizard, 2010; Threadgold et al., 2019). Rather, it should perhaps be considered that music listening in this regard is orientated towards perceptual effect (i.e., feelings of company) as opposed to cognitive efficacy. In other words, *Background and Accompaniment* and *Focus and Concentration* may both refer to perceptual control, but their delineation implies that the role of music alters depending on the desired perceptual background effects. This is further implied by the presence of an item relating to *Background*, in which music's use as an aural filtering tool (i.e., a means of perceptual environmental regulation) is further implied. Finally, although emotion-based regulatory affect is more closely reflected by *Emotion Regulation*, *Background and Accompaniment* seemingly refers to the more specific perceptual process that facilitates feelings of accompaniment. As before, this may only become clearer through further applying the measure.

5. *Physiological Arousal*

The three items attributed to Factor 5 reflect motivation, pacing, activation, and arousal alongside concurrent activities. As has been previously discussed, physiological stimulation and regulation is a common application of music listening, especially alongside portable listening technologies that make interaction with music highly autonomous. Listeners might use music in this regard to move to its rhythm and use it to enhance exercise or physical work (Williams, 2007), the congruence of which is largely predicated on music's ability to induce desirable levels of energy and entertainment in the listener (Lamont et al., 2016). Arousal plays a key role in the identification of the appropriateness of music more generally, however, such as the notion that listeners prefer music with high-levels of arousal during exercise and music with lower-levels of arousal when relaxing (North & Hargreaves, 2000). This is supported by the idea that with regard to the effects of music on listeners during exercise, there are consistent themes; two of which are physiological arousal and subjective experience. According to Clark et al. (2016), for example, music's application can promote behavioural change that leads to increased exercise adherence and participation (additionally, cortical and subcortical stimulation is observed as an additional theme in health-orientated studies, but this is less relevant to the present discussion).

It is difficult to tell from the present model, however, whether this factor translates to both high-level physical activities, such as exercise, and low-level physical activities, such as domestic chores, which require low interest or attention (Greasley, 2008; Greasley & Lamont, 2011; Lamont et al., 2016). Whilst there is no particular reason to assume that this is not the case, it nevertheless draws attention towards the need for not just cross-validation of the construct, but specifically cross-validation in ecologically valid settings (i.e., during exercise and/or relaxation). There are, however, additional consistencies that are reflected between features of the implied factor and other research. For instance, the presence of motivation and the desire to control arousal (e.g., to physically stimulate but also to maintain pacing) is consistent with Laukka and Quick's (2013) observations, however, they also observe the presence of emotion-based regulation alongside such functions. Therefore, it might be interesting to note the extent to which *Physiological Arousal* covaries with *Emotion Regulation* in ecologically valid settings.

6.5.2 Limitations

Limitations of this study include an absence of reversed items, despite some recommendations on their use (e.g., DeVellis, 2017). On the one hand, cognitive biases may predispose individuals to overestimate self-assessed responses, resulting in recall biases or otherwise leaning towards agreeable or positive responses (Gove & Geerken, 1977; Porta, 2014). In this instance, that may have meant that some participants overstated the frequency of proposed functions. On the other hand, it has been argued that reversed items are difficult for study participants to understand, do not reduce response bias, and should generally be avoided (van Sonderen et al., 2013). Moreover, reverse items are seldom consistent with positivist models, such as many of those relating to FML, including the present structure and others (e.g., Saarikallio, 2008). Groarke and Hogan (2018) retained just two of just 38 reversed items in an initial item pool, for example, highlighting that reverse items are inconsistent to say the least. It may be the case that this did not have a drastic impact on results, however, this is difficult to assess further.

Furthermore, the size of the item pool that was used may have had a detrimental effect on results due to fatigue. Participants rated 114 items over 11 matrix tables that, whilst intended to be as manageable as possible, nevertheless required a great deal of nuanced and focused attention that could not have been avoided. Items were presented in a randomised order for each respondent, however, which was intended to mitigate fatigue by distributing items from various domains to mitigate repetitiveness. The length of the item pool may have contributed to fatigue and also additionally put further participants off or otherwise prevented individuals from completing the study.

The final sample size of 327 may be limited in the extent to which it may be applied to larger populations. Various 'rules of thumb' have been proposed to assess whether a sample size is appropriate for factor analysis, which may arbitrarily range from between two and ten observations per variable. It is contended, however, that there are no absolute thresholds for a minimum sample size and that recommendations may range from less than 50 observations to numbers well into the hundreds (De Winter et al., 2009). Kline (1994) suggests that samples below 100 "could produce misleading results" (p. 180) but does not specify if there are ideal sample sizes as such. Field (2018) implies that sample sizes >300 may produce more generally stable results, however, where adverse effects of smaller sample sizes are reduced. This is also suggested by Comrey (1973), who states that sampling adequacy in factor analysis may "be evaluated very roughly on the following scale: 50-very poor; 100-poor; 200-fair; 300-good; 500-very good; and 1000-excellent" (p. 200). Whilst a larger sample size may have contributed to a higher degree of confidence in the present model, the present sample is implied adequate both theoretically (e.g., Field, 2018) and statistically (e.g., the previously reported KMO score). Spector (1992) states that sample sizes between 100 and 200 are suitable for item analysis using Cronbach's Alpha, and that scales may be incrementally reduced using this method. Such an analysis would, unfortunately, fail to produce an underlying structure or model of such items; neither would it have been appropriate to use this prior to factor analyses. However, internal reliability analyses of each identified factor did generate high alpha scores, indicating high internal reliability for each dimension of the latent construct.

A further consideration to make is the wider social context within which this study was conducted. Data was collected in the United Kingdom between April and May 2021, and further data was gathered in January 2022 when a need for more data was identified. At the time of the study, the COVID-19 pandemic was ongoing. The pandemic has resulted in social distancing and 'lockdown' measures being enforced in many countries. Research on music listening during the pandemic has indicated that particular applications of music and listening behaviours have been accentuated, reduced, or otherwise affected by the pandemic (Gibbs & Egermann, 2021; Henry et al., 2021; Krause et al., 2021). This raises the prospect that participants' interpretation of the study's framing (i.e., 'generalised' listening behaviour) may be affected by the wider cultural context that has impacted daily life since the start of the pandemic in 2019, compared to prior data without this contextual backdrop. Plausibly, certain FML (such as exercise) have been more limited by periodic inabilities to employ music within accompanying contexts during lockdown periods. As a result, functions hypothetically subject to or otherwise linked with particular contexts may also be limited in their prevalence through limited application. However, the extent to which this has affected the present study is not an immediate worry but rather a contextual aspect of the study itself since it is not a contributing factor that could have been plausibly circumvented during data collection. However, it may nonetheless be interesting for future research to replicate this study to see if a comparable factor structure is identified when this circumstance is absent.

Finally, it is likely that most participants were also either fluent or native English speakers, probably based in the United Kingdom (due to online reach, study location, and so on). The sample therefore carries a Western bias, for which cross-cultural validity cannot be reasonably asserted. In considering the ways in which this relates to music listening, it would be interesting to consider whether the identified structure would cross-validate in other cultural settings, or whether further data on the initial item pool gathered in those populations may produce different latent constructs.

This set of limitations should undoubtedly be considered when discussing the present model. However, as has been discussed, the five-factor solution is largely consistent with literature relating to FML. When coupled with considerations regarding factor analyses (e.g., Kelloway, 1995; Worthington & Whittaker, 2006; Nye & Drasgow, 2011; Field, 2018; Watkins, 2018) there is justification to assert that on balance the present model sufficiently reduces the scope of the initial CFF scale into a set of theoretically consistent factors, subject to cross-validation.

6.5.3 General Discussion

Overall, the implied model holds consistencies with existing literature, and each of the five implied factors are consistent with identified functions of autonomous music listening. The generated factors hold internal reliability and discriminant validity and given that parameter estimates of each item are >.50, convergent reliability is present also (Hair et al., 2014). This study has additional strengths in its theoretical approach. The CFF is in-depth, holistic, and broad, whereas prior research has been perhaps less well-defined. For instance, Ter Bogt et al. (2010) anecdotally assess that there are "at least" four primary FML (p. 148), and in their discussion, Clayton (2009) also proposes that behaviours of music listening fall into one of four dimensions: *Regulation of an individual's emotional, cognitive, or physiological state*, *Mediation between self and other*, *Symbolic representation*, and *Coordination of action*. The five underlying functional dimensions put forward in the present study hold consistencies with these competing frameworks but are drawn from the comparative framework generated by Maloney (2019). Conversely, however, whilst the generated model holds five theoretically consistent factors, they should not be considered to be the exclusive FML, but rather the primary utilitarian indicators as to the most common applications of music in everyday life.

For instance, background affect regulation specifically during journeys is a common and well-documented FML (e.g., Bull, 2006; Lamont et al., 2016; Maloney, 2019), however, items relating to music listening in transitory spaces or alongside journeys did not suitably load onto any factor. Rather, items relating to silence avoidance and background regulation loaded onto Factor 4 (*Background and Accompaniment*), however, these were not specifically related to music listening within transitory spaces, which could have theoretically been a more explicit, separate factor given its prevalence within the FML literature. On the other hand, this may be reflected within the functions that have been associated within remaining factors. For instance, when applied in a longitudinal, ecologically valid study design (e.g., ESM), it may become apparent that *Background and Accompaniment* serves as a function alongside the activity of travelling (which would make sense hypothetically), rather than music listening *whilst* travelling being identified as its own distinct dimension. As such, it may be difficult to tell until the identified model is measured alongside specific listening episodes.

The function, *Focus and Concentration* was additionally uncovered during this study. Kotsopoulou and Hallam (2010) observed that music has been observed to accompany work during private study in adult populations, however, there are some distinctions between studying tasks in this regard. For instance, general studying and writing are more likely to be accompanied by music listening than other study practices such as revision for examinations and memorisation. Such notions may be translated into music selection, such as the avoidance of music with lyrical content when studying (Avila et al., 2012). Such task differences should be considered with any future applications or measurements using this scale, especially when related to listener contexts. It does, however, further the notion that it is the accompanying activity that is the primary context-based variable by which the application of music may be considered congruent or not.

Overall, the point here is that within previous literature there is some occasional blurring of lines between the function that music serves, and the accompanying activity or context. By distinguishing the underlying function from context completely, however, the results of this study may be beneficial in identifying underlying FML in everyday life, in which context is treated as a predictor of FML rather than as a constituent feature of FML. On the other hand, this may be where difficulties appear in effectively translating qualitative findings into quantitative models and outputs. The qualitative nature of Maloney's (2019) study generates an expansive structure of listening functions yes; however, it is difficult to reliably infer a scale from said structure that is consistently parsimonious until it is applied in ecologically valid settings. Hence, there remains work to be done and it is difficult for now to assess the extent to which this dimension reduction of an expansive qualitative model has been effective in capturing and structuring the core distinct FML from the utilitarian perspective.

6.5.4 Conclusion

The present study sought to generate an item pool reflecting the theoretical content of the 53 FML identified in the CFF, which were then rated and reduced to a latent variable structure. 327 participants rated the preliminary 114-item pool. Following removal of theoretically divergent and highly correlated items, EFA was used to uncover the underlying structure, which was shown to have good fit indices, generating a six-factor solution containing 41 items and explaining 61.1% of the variance. Following further reduction to reduce conceptual overlap and sustain brevity, EFA and CFA on a reduced iteration containing 23-items indicated good fit indices across the five factors included, with all items holding standardised parameter estimates >.50, and each latent factor holding AVE >.50 also. Theoretically, the model may be best thought of as a representation of the utilitarian domains of FML. These five domains are: *Identity and Social Bonding*, *Emotion Regulation*, *Focus and Concentration*, *Background and Accompaniment*, and *Physiological Arousal*. Hypothetically, these may in turn represent the use of music as a means of social, emotional, task-orientated, cognitive, and physiological regulation.

On balance, whilst this model requires cross-validation, theoretical consistencies with other research on FML have been identified, and this study has successfully reduced the breadth of dimensions relating to *functionality* within the structure. Furthermore, it should be noted that the present structure is a substantial reduction, dimensionality-wise, from the initial 114 items across 53 functions, and does not account for every known FML, but rather a subset applicable for average listeners. Any future applications of the measure could, therefore, consider maintaining a qualitative element in order to facilitate some degree of flexibility and nuance; especially with the ability to cross-reference such outcomes with the initial 53 functions identified by Maloney (2019). The hope is that this may afford strengths of both qualitative and quantitative research methods and thus make some contribution towards an integration of the two by drawing awareness to respective strengths and limitations where relevant. More pressingly, however, the model is indicative of the broader utilitarian focus on FML in everyday life with which this thesis is concerned. The model builds on previously generated structures by firstly aligning with wider utilitarian perspectives of music listening, rather than purely regulatory ones (e.g., Groarke & Hogan, 2018). Through the use of the CFF as a theoretical grounding (Maloney, 2019), it is also cognisant of the scope of FML in everyday situations, something prior research has not consistently considered (e.g., Lonsdale & North, 2011). The resulting structure is therefore broad enough to consider a wide range of FML in everyday life and is aligned with Merriam's (1964) founding distinction between *use* and *function*. Hence, this helps address the first aim outlined in section 1.2. As a consequence of this study and its outcomes, subsequent work may: (1) cross-validate the structure in ecologically valid settings and (2) triangulate this with contextual variables (i.e., activities) and listeners' music selection. Therefore, the second study of this thesis applies this structure model within everyday life.

7.1 Study 2: Contextual applications of music: Repeated Measures

Following Study 1, a latent variable structure derived from Maloney's (2019) CFF has been presented and is hypothesised to serve as a useful measure to assess FML from a utilitarian perspective; however, a need for cross-validation (ensuring the reliability of the generated model by fitting it to a new sample of data) remains. Although this latent variable structure may not be definitive and is not exhaustive of all potential FML, it may nevertheless serve to identify utility in everyday listening episodes. Subsequently, an approach can be considered to assess the extent to which these functions are present within listeners' daily listening habits by applying them to real-world instances of music selection in everyday life, rather than general listening habits as in the prior study. This may be achieved by measuring the construct of FML in conjunction with relevant variables relating to the listening activity and content of selected music (i.e., audio features/characteristics), as was suggested in section 5.3.

This process may therefore enable the identified factors to be measured in relation to specific listening episodes, thus operationalising the psychometric structure in a manner that extends ecological validity as well as construct validity through cross-validation. However, because FML was just one of the constructs considered in the triangulation described in Chapter 5, it is considered that by operationalising FML as part of a network of three constructs that it may also be possible to model music selection in everyday life, with concurrent listening activities and FML determining music selection as per *functionality* (e.g., Greb et al., 2018a; 2019). Therefore, by orientating a second study toward listeners' self-selected listening episodes, it is conceptually possible to observe which functions are associated with listening situations through a cross-sectional lens to operationalise the hypothesised temporal structure of the three focal constructs. This would help address the second and third aims of this thesis at large, referring to the need to triangulate these three constructs (see section 1.2).

To implement this, several considerations were necessary. Firstly, a large sample across a broad variety of listening activities was needed to achieve adequate statistical power, due to the steep variability and individuality that inevitably accompanies a universal yet diverse phenomena like music listening. Consequently, this raises questions about which surveying method(s) should be used to attain appropriate power but optimise ecological validity. When it comes to the assessment of music listening in everyday life, one common method applied in the literature is ESM, which has been applied in a large number of studies on music listening and selection in everyday life (e.g., Sloboda et al., 2001; North et al., 2004; Bailes, 2007; Juslin et al., 2008; Greasley & Lamont 2011; Krause et al., 2014; Randall et al., 2014; Randall & Rickard, 2017; Maloney, 2019; Greb et al., 2019).

The strength of ESM, as noted by Hektner et al. (2007) is that, because environmental and cultural factors influence an individual's daily behaviours and experiences, repeated measurements "provides glimpses into the real-life habits of a human community that are otherwise difficult to detect in retrospective studies or through single administration research instruments" (p. 4). Additionally, the use of ESM has become easier to facilitate in recent years, as smartphone technologies afford researchers with "an unprecedented opportunity to collect complementary data about how people live and experience their daily lives" (Lathia et al., 2013, p. 191). For instance, Randall et al. (2014) utilised a mobile ESM (m-ESM) approach to generate real-time data regarding music's application as a mood regulator. This was attained through event-based Experience Sampling Reports (ESRs), administered when listeners engage with music on their own device. Various tools have been developed to administer ESRs, such as the application MuPsych[2] developed by Randall and Rickard (2013).

On the other hand, there remain several limitations to ESM that are worth considering. Firstly, although ESM offers an ecologically valid way to gather information about everyday life, the proposed and presently most widely used implementation method (i.e., personal mobile devices) can limit samples in their diversity. Krause et al. (2014) note, for instance, that whilst it is possible to achieve diverse age groups in m-ESM studies, older individuals are less likely to engage with mobile technologies than younger individuals. Furthermore, due to the intensive nature of ESM, they tend to have sample sizes considerably smaller than those of some other

[2] https://www.mupsych.com

surveying methods, such as online surveys. This is especially true when researchers are unable to compensate participants when resources are limited, and as such are prone to participant drop-off (Hektner et al., 2007). Conversely, and although more susceptible to recall biases, online surveys are more effective in gathering larger and diverse samples than intensive methods like ESM (e.g., Evans & Mathur, 2018). Online surveys can therefore gather a wider array of individuals than may be solely possible with ESM, which allows researchers to draw inferences from a larger, more diverse sample, highlighting relevant strengths and weaknesses of competing surveying methods.

With this in mind, it was considered whether taking a novel approach could afford the relative strengths and weaknesses of both of these methods in providing two arms of data in a single study. It was hypothesised that relative benefits and limitations of each surveying method may be mitigated, provided the complexity of the data structure is suitably accommodated in statistical analyses. Moreover, by integrating a multi-method quantitative approach (see Morse, 2003), it was hypothesised that a larger pool of data could be gathered from which wider inferences could be made. It was considered that through an initial online survey, a large sample of individuals could first be surveyed regarding their most recent autonomous listening episodes. Following this an ESM would follow with interested participants from that initial survey, which would serve to increase the variety of listening situations reported, proximity of observations to listening episode, and total number of observations. For there to be parity between the two arms, however, they would need to be orientated in a similar way (e.g., use of equivalent measures).

As such, the reports in both study arms would be treated as cross-sectional in nature, as longitudinal hypotheses are not considered since this research is primarily interested in music selection. This is because music selection is considered to be effectively instantaneous, with staggered measurements being impractical (Greb et al., 2019). Therefore, both study arms would serve to collate a large set of cross-sectional observations, which could then be pooled into a partially nested data structure with analyses carried out at the observational level.

In applying these two surveying methods to collate a large pool of cross-sectional observations, the limitations of both methods are somewhat mitigated, and by treating all observations as one collated set of pooled cross-sectional observations, the subsequent test power to make statistical inferences is maximised, as seen in other pooling methodologies like integrative data analysis (Curran & Hussong, 2009; Hussong et al., 2013) and in multimethod quantitative designs employing more than one avenue of data collection (Morse, 2003). Although not the first study to precede an ESM with an initial survey (e.g., Moreno et al., 2012; Jelenchick et al., 2013) this is, however, to the best of the author's knowledge the first to focus on a data pooling approach using both an initial survey and ESM as sequential study arms with simultaneous analysis to maximise statistical power and sampling diversity. This general approach is similar to quantitative sequential multi-method designs put forward by Morse (2003), however, the approach in the current study places an emphasis on increased test power through cross-sectional pooling, rather than an exploration of multiple research outcomes.

Moreover, such an approach has additional affordances because it allows multiple constructs to be assessed according to cross-sectional listening episodes. Chapter 4 outlined the use and role of MIR-generated audio features as indicators of music selection, and as a means by which recommender systems may suggest songs to users (i.e., content-based methods). Therefore, utilising such measures would provide characteristic insight into the role features have when applied in psychological work. This is supported by extant literature, as Greb et al. (2019) argue that studies considering everyday music selection behaviours should integrate an aspect of objective measures of musical features via MIR research tools. Also, in doing this, these variables may in turn be used directly to estimate or otherwise make inferences in hypothetical recommendation tasks, by providing a raw metric for MIR-generated audio features in ecologically valid data, of potential use to CAMRSs. This may therefore also help in the achievement of the fourth aim of this thesis (see section 1.2).

As such, it was also considered how MIR-generated audio features could be collated, and that it was prudent to integrate self-report measures to draw comparisons between perceived musical characteristics and computationally attributed featural ratings. The benefit of this is that it will be possible to observe the extent to which self-reported and MIR-generated audio features correlate. If there are theoretically consistent correlations between self-reported and MIR-generated audio features, then it will provide confidence that the MIR-generated features are at least roughly consistent with listeners' perceptions of pieces of music. This extends the applicability of such features beyond that of MIR tools and provides future work the opportunity to integrate such computational tools.

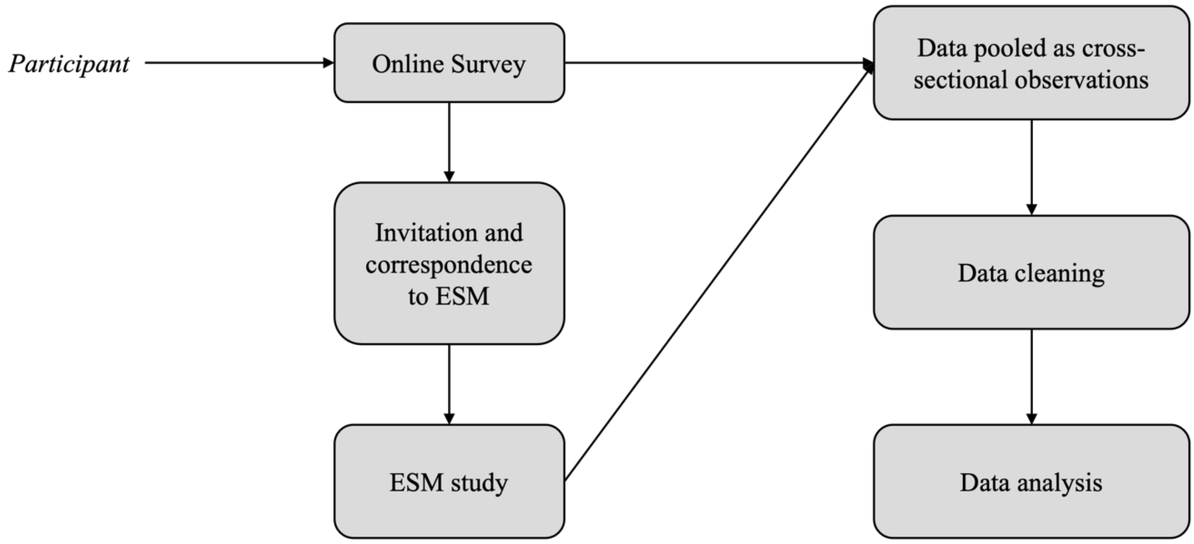
With these motivations in mind, this study aims to specifically address three research questions:

1. Can the previously identified FML measure be cross validated in ecologically valid data of everyday listening episodes?
2. To what extent do MIR-generated audio features correspond to listeners' perception of audio content?
3. Do activities concurrent to music listening lead to changes in the audio content of music selected by listeners directly or indirectly via FML?

7.2 Methods

As mentioned in the discussion above, this study comprises two arms: an initial survey intended to capture a large pool of participants' most recent listening episodes, and an ESM study to capture listening episodes over the course of an extended period of time. This synthesises a multi-method approach, which can be considered a sequential QUAN → quan approach according to Morse's (2003) paradigms of multimethod designs (p. 204). In addition to operating as a route of data collection in itself, the initial survey is also used as a recruitment tool for the subsequent ESM study, thus assisting by gathering participants who have some previous experience or interest in the research. This procedure is summarised in Figure 8.

Figure 8 Summarised design of Study 2



Based on this design, ethical approval was granted by the Arts and Humanities Ethics Committee (AHEC) at the University of York to carry out this study.

7.2.1 Measures

In both study arms the three target variables of listening activity, function, and content were gathered. These will now be discussed in turn and should be considered to apply in both the online survey and the subsequent ESM unless stated otherwise.

Listening Contexts

Context was assessed by first separating location and activity variables, as discussed in section 3.1. Specifically, location was captured via a predefined list of nine locations informed by Maloney's (2019) study, and an 'Other' option with an open text-field. Activity, on the other hand, was treated as a qualitative variable and captured solely via an open-text field. This afforded greater distinction in subsequent analyses and categorisation of activities as, given the complexity and variety of activities that may be reported, this would hold greater nuance in the representations of concurrent activities. Given evidence in prior research (e.g., Juslin et al., 2008; Greb et al., 2018a; 2019; Maloney, 2019) that activity is the determinant of music use in the attainment of cognitive goals, then providing nuance is key to avoid pre-determining a narrower definition subject to researcher biases. Subsequent qualitative assessment of reported activities allows for this nuanced assessment, hopefully generating more precise activity variables where possible.

This comes with a caveat, however, in that there would inevitably be a certain amount of noise in these qualitative assessments, such as participants naming unclear or multiple activities simultaneously. In turn, individual descriptions may lead to a large number of activity groups when inductively coded, where there is not an equal distribution in terms of the number of participants assigned to each group (for instance, if a large number of participants report that they are travelling, but a small number report studying). This could affect subsequent analyses since predictor groups would ideally be roughly equal where possible to mitigate sampling biases. With that said, this does of course not mean that there would be any guarantees of equal groups if the Activity variable was pre-coded either, rather that qualitative assessment could plausibly exacerbate this risk due to the researcher's subjective interpretation of a written response. To put it another way, concurrent listening activities are likely to naturally vary in their frequencies amongst listeners since certain activities are simply more common than others. Nevertheless, the multimethod approach could mitigate this risk somewhat by casting as wide a net as possible in the gathering of cross-sectional observations at the observation level. Once gathered, activities would be categorised through thematic analysis via deductive coding (Braun & Clarke, 2006), as per other music listening studies (e.g., Maloney, 2019). Using existing literature as a framework (Lamont et al., 2016; Greb et al., 2019; Maloney, 2019) is a useful way of contextualising these findings within the wider literature.

Function

Function was measured via the model identified in Study 1, measuring 5 dimensions of utilitarian FML through the reduced set of 23 items reported in section 6.4.3 (used in the present study due to a reduced burden on ESM participants in rating an excessive number of items). As before, items were assessed on a 5-point Likert-type scale rated according to the degree of functional importance (*Not very important* to *Extremely important!*) according to the individuals' most recent listening episode (see Appendix D), rephrased from a frequency report in Study 1 (i.e., *Never* to *Very Often*). This rephrasing was important since the first study related to an overall perception of presence in described FML, whereas this study was interested in the items' prevalence in one specific situation. This 5-point scale was assessed to be useful, however, in gathering information about the magnitude of relevance participants' perceived in the items with reference to their listening experience. In addition to this, participants were also presented with an open-text field in which they were able to describe their functions or purpose of music listening, if they felt that their reasoning was not fully encapsulated by the preceding FML measure.

Audio Content

To measure audio content in the collated cases, several options are available. Namely these are to use self-reports as in other psychological studies (e.g., Greb et al., 2019), or to gather features from an MIR tool. Self-reports are subjective perceptions to characterise music that the listener experiences, whereas MIR features are those typically attributed automatically through signal processing. Given the interest in this thesis in the role of audio content in recommender systems, it was decided that MIR features would be preferable since this would enable future comparison and application as part of a content-based system, and also help address RQ2 by allowing comparisons to be made between self-reports and MIR-generated audio features. Audio content data was therefore generated via the publicly accessible Spotify Application Programming Interface (API[3]). An API can be thought of as the user interface of a library of functions to programmers, in which services or data are provided by a software application through predetermined resources, methods, and objects (Stylos et al., 2009). These resources allow other applications and users to access data or services without needing to implement the original objects and procedures (Meng et al., 2018). In the case of the Spotify API, this affords users a range of functionalities, of which just one is extraction of the audio features of tracks, as encoded by Spotify's systems.

[3] https://developer.spotify.com/documentation/web-api/

In order for a user/programmer to do this, track identifiers (IDs) are needed, which were therefore gathered when participants were able to name the track they had most recently listened to in their respective experience, and this song could be found in a manual Spotify search. Participants were asked to name both the song name and an artist. Once a named track was identified and found on Spotify (again depending on the varying degrees of clarity reported by participants), its ID could be extracted from the Uniform Resource Locator (URL). Audio features can be gathered from the Spotify API directly, using an internal client URL (cURL) command, in which the API prints out each track's audio features directly in a web interface[4]. Each track is attributed ratings across 12 featural dimensions: *Danceability*, *Energy*, *Speechiness*, *Acousticness*, *Instrumentalness*, *Liveness*, *Valence*, *Key*, *Mode*, *Tempo*, *Time Signature*, and *Duration* (see [4] for definitions of each feature).

The existing API interface only allows users to gather one track at a time, however, which would be a laborious task to repeat potentially hundreds of times at scale. As such, a Bash Script (see Appendix F) was used to call a cURL from the API in which an array of track IDs could be provided. The script was run on a local terminal and returned audio features were collated in a JavaScript Object Notation (JSON) file. From here, they could be converted into a comma-separated values (CSV) file and imported in the main dataset. Not every audio feature gathered is especially useful in testing the hypothesis that activities effect music selection. For instance, the measure of Mode, a binary coding of 0 or 1 attributing a track with a tag of major or minor, is not likely to be a strong predictor of music selection from a theoretical perspective. It is worth, however, gathering these variables where relevant all the same since it may be useful to explore any unforeseen relationships or otherwise be able to eliminate these variables on statistical as well as purely theoretical grounds. To enable comparability with self-reported musical features, the measures provided by Greb et al. (2019) were also included in the questionnaire and are shown in Table 10.

[4] https://developer.spotify.com/console/get-audio-features-track/

Table 10 Greb et al. (2019) Self-Reported Musical Features

Calming	1-2-3-4-5-6-7	Exciting
Slow	1-2-3-4-5-6-7	Fast
Sad	1-2-3-4-5-6-7	Happy
Unfamiliar	1-2-3-4-5-6-7	Familiar
Less melodic	1-2-3-4-5-6-7	Very melodic
Less rhythmic	1-2-3-4-5-6-7	Very rhythmic
Simple	1-2-3-4-5-6-7	Complex
Peaceful	1-2-3-4-5-6-7	Aggressive
Less intense	1-2-3-4-5-6-7	Very intense
Instrumental	0-1	Vocal

7.2.2 Participants: Initial Online Survey

Recruitment for the initial survey took place using social media, institutional distribution at the University of York, survey sharing platforms (e.g., SurveyCircle), and through other emailing lists (e.g., MUSICOLOGY-ALL). This survey asked participants about their most recent autonomous listening experience, which was clarified via a prompt at the beginning of the survey asking participants to acknowledge that they should respond from the perspective of the last time they chose to listen to, and were in control of, the music. Participants were asked to confirm that they had understood this prior to proceeding to the remainder of the survey, thus providing confidence that each report relates to their most recent autonomous listening experience. Aside from the broader intentions of this study, the survey itself provided a broad view of the context-specific music selection and *functionality* from a larger sample of participants than would otherwise be attainable in a more intensive design (e.g., ESM). This was administered through Qualtrics, in which study information and informed consent were included. Participants were required to acknowledge they had read and understood the terms of the study, consent to partaking via embedded checkboxes, and were also required to confirm that they were at least 18 years of age. Of the 436 responses recorded in this phase, 373 had named tracks that were found on Spotify and submitted for feature extraction via the API. Cases where tracks could not be found were due to a combination of lack of clarity, naming multiple tracks, or not remembering.

7.2.3 Participants: ESM Study

For the ESM phase of data collection, the Smartphone Ecological Momentary Assessment (SEMA3[5]; version 1.3.2) smartphone application was used (Koval et al., 2019). SEMA exclusively operates to facilitate ESM studies, using both iOS and Android operating systems. Mobile ESM tools are relatively common in recent research using ESM, especially those on music listening (e.g., Randall et al., 2014; Greb et al., 2019). SEMA is free-to-use and made available under the auspices of the University of Melbourne.

7.2.3.1 Design

At the end of the initial survey, participants were presented with an invitation to take part in the ESM. If interested, participants were able to provide an email address for further correspondence and invitation in the remaining sign-up form. The sign-up form, as embedded in the initial survey, included a participant information form, followed by the attainment of informed consent and confirmation of accessibility to a Smartphone with either an Android or iOS operating system. Participants were able to opt-in to receive a personalised listening report at the end of the study, provided they held a compliance rate of 50% or more (these were disseminated to participants after data analyses were completed). From the email addresses provided, 157 potential participants were contacted with a follow-up email that provided study information and guidance on the use of SEMA.

In the ESM, participants were presented with a survey form nearly identical to that of the initial online survey (see Appendix E). However, SEMA does not allow researchers to present matrix tables of items in which all have to be responded to, as was used with the initial survey. Instead, SEMA either allows one item per page or 'select and slide' lists, in which a series of items or statements can be presented in a short list, much like a matrix table, to which participants only select the items they deem relevant to their most recent experience and rate them accordingly. In essence, this meant there was a choice to be made between presenting the 23 items of the

[5] https://sema3.com/about.html

FML measure individually or grouping them into subsets to present on the select and slide model. 23 pages containing one-item was deemed to be impractical, repetitive, and labour intensive; sub-optimal conditions in ESM that are likely to exacerbate drop-off rates and prevent the completion of forms through fatigue (Eisele et al., 2022). Consequently, the latter was deemed the most pragmatic of the options available. In this case, participants were presented with three tables of five items and two of four (totalling 23 items) in a random order. They were asked to select and rate the items they deemed relevant (i.e., above the null value) on a reduced scale from 1-4, with unselected items presumed to be therefore irrelevant and thus attributed with lowest value of the psychometric measure (i.e., 0), inferring the fifth (null) scale point.

Though not a perfect solution to this issue, this was assessed to be a reasonable way of treating non-selection amongst item values. The non-obligation to select all items likely increases the risk of non-response, rather than conscious non-selection, however. Therefore, whilst it was intended to pool the data between the initial survey and the ESM to maximise test power, preliminary assessments of the validity of the psychometric structure will need to be verified in each separate study arm prior to this pooling process. Should the psychometric structure of the measure satisfy assessments conditions of reliability and validity in both study arms, then data pooling for the purpose of simultaneous analysis will be deemed suitable.

7.2.3.2 Procedure

157 (36%) participants from the online survey indicated an interest in taking part in the ESM and were contacted via email in the weeks leading up to the study. Correspondence prior to the study involved instructions on what to do once an invitation email triggered via SEMA, as well as a link to further study guidance and researcher contact details if participants had any questions or concerns. The ESM study was conducted over the course of 10 days, with participants receiving four notifications each day. An embedded link remained available in SEMA throughout the duration of the study for participants to refer back to which included instructions, the original study information as provided upon sign-up in the initial survey, and researcher contact details. Prompts came in equally spaced two-hour intervals anytime between: 09:00 & 11:00, 12:30 & 14:30, 16:00 & 18:00, and 19:30 & 21:30. Participants would

receive notifications randomly within these two-hour time frames and had a one-hour window to respond. They were also able to submit ad-hoc submissions in the SEMA app, if they wished. The intention of this was to give participants the ability to submit a listening report as close to their listening episode as possible.

In total, 3,012 listening prompts were received by participants of which 1,107 were responded to, giving an initial compliance rate of 36.75% for scheduled responses. Regarding ad-hoc responses, there were a further 515 ESRs submitted. In total, this generated 1,622 ESRs in which participants responded with either 'Yes' or 'No' as to whether they had listened to music since the previous response. This gave a final compliance rate of 53.85% in relation to the number of notifications received. From these ESRs, 819 logs indicated that participants had listened to music since their last notification, thus triggering the rest of the ESM survey. The remaining 803 ESRs indicated that participants had not listened to music, at which point the survey terminated. From the 819 logs that indicated music listening, 742 ESRs were complete. Of these, 515 (69.40%) were scheduled responses and 227 (30.60%) were ad-hoc. The remaining 77 ESRs were incomplete at various stages of the survey form and removed from study data.

Overall, the 742 fully completed ESRs were completed as relevant from a sample of 81 participants between the ages of 18 and 70 (M = 31.73; SD = 12.22). Of this sample, 49 (60.5 %) were female, 27 (33.3%) male, 4 (4.9%) non-binary/third gender, and 1 (1.2%) who preferred not to state their gender identity. Of the 742 completed reports, participants named 571 individual tracks that were found manually within Spotify. Sample sizes vary in ESM studies within the systematic musicology literature, and a comparison with other ESM studies assessing music listening in everyday life can be seen Table 11.

Table 11 Sample sizes in systematic musicology research using ESM

Publication (*see Reference List*)	*N*
Sloboda et al. (2001)	8
Bailes (2007)	11
Greasley and Lamont (2011)	25
Juslin et al. (2008)	32

Maloney (2019)	71
Present Study	*81*
Krause et al. (2014)	101
Greb et al. (2019)	119
Randall and Rickard (2017)	195
Randall et al. (2014)	327

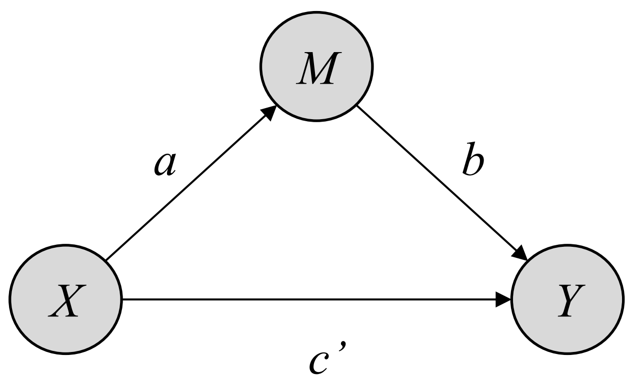
The aims and methods of the studies mentioned in Table 11 do of course vary, such as whether they are qualitative or quantitative. Such contextual aspects can further affect the suitability and validity of study samples, in which case qualitative analysis may supplement more moderate sample sizes (van Berkel et al., 2017). If researchers are looking to make statistical inferences, however, then Bayesian statistics can be used to yield robust and meaningful results in small sample settings (Kay et al., 2016; van Berkel et al., 2017), a consideration kept in mind should the outcomes of the study be limited. This nonetheless provides a useful comparison point with other research in this area applying comparable methods.

7.3 Approach to data analysis

When combined, there were 1,178 cross-sectional listening episodes reported, of which 436 were responses to the initial survey and 742 were from the ESM. From this initial pool of observations, 945 reported tracks were found on Spotify, with 944 (n = 373 from the initial survey, n = 571 from the ESM study) yielding audio features (n = 1 track failed to return audio features regardless of extrapolation method, likely due to a glitch internal to the Spotify API). However, the data collected constitute a large pool from which inferences may be drawn relating to the relationships amongst the focal constructs of interest. To this end, this section contextualises and outlines to modelling approach taken in this study to explore these relationships.

Two experimental designs were employed: an initial online survey followed by an ESM study with a subset of participants from the initial survey. In both of these study arms, participants have reported their most recent music listening experiences, and though slight differences exist where relevant to each study design, the method of measuring the relevant constructs was not affected since the study designs were consistently oriented towards participant's most recent listening episodes, with both data sets aiming to capture the same phenomena via the same measures. To assess the general hypothesis that FML mediate relationships between listening activities and audio features indicative of music selection, a mediation analysis was hypothesised to be the most prudent statistical test to assess the proposed temporal structure of these three key focal constructs. Mediation refers to a set of statistical procedures that are used to test whether an independent variable affects a dependent variable indirectly as well as directly. Researchers applying mediation analysis hypothesise that the indirect effect takes place through an intervening process resulting from the mediator variable (Baron & Kenny, 1986). Ultimately, the aim of mediation analysis is to understand through which mechanism(s) variables relate to each other (Fairchild & McDaniel, 2017). As such, inferences are made about these interrelationships, for example, that the independent variable (*X*) affects the mediator variable (*M*), which in turn affects the dependent variable (*Y*) (Hayes & Preacher, 2014). Figure 9 shows the simple hypothesised structure upon which mediation analysis is based.

Figure 9 Standard Trivariate Mediation (Iacobucci, 2008)



However, researchers often wish to extend models such as that shown to facilitate the use of mediation models that include multiple indicators or variables in any (or all) of the three constructs. This is where a SEM approach to testing mediation is recommended to allow for a robust statistical procedure to be used in a wider array of contexts than that shown (Iacobucci, 2008). Though difficult to define in concise terms, broadly speaking SEM is a set of high-powered multivariate techniques that utilise conceptual models, path diagrams, and systems of sequential regression-style equations to capture complex relationships as part of a network of observed and/or latent variables (Gunzler et al., 2013). In the present context, it allows a researcher to test for mediated relationships amongst constructs but, unlike regression techniques, is also effective when multiple items or variables are measured to capture the focal constructs and can adjust error terms as a consequence (Iacobucci, 2008).

Baron and Kenny (1986) were the first to propose a method of testing mediation, in which they described a series of regression equations to test the hypothesised causal paths, however, this has two key limitations: (1) it does not allow the measurement of multiple measure constructs; (2) conceptually, mediation tests both causality and for the presence of a temporal structure amongst three variables (i.e., intervention, mediation, and outcome/response). As variables in the hypothesised causal relationship can be both causes and effects standard regression paradigms are poorly suited for modelling these relationships because there remains an a priori assignment of each variable as either cause or effect as each step becomes segmented (Iacobucci, 2008; Gunzler et al., 2013). Gunzler et al. (2013) therefore argue that SEM "provides a more appropriate inference framework for mediation analysis and for other types of causal analysis" than regression techniques (p. 391). As a result, SEM approaches are widely recommended to extend mediation models, and are generally favoured over earlier regression techniques (i.e., Baron & Kenny, 1986) to test hypothesised indirect relationships as part of a network of multivariate focal constructs.

Interpretation of mediation analysis

The following discussion addresses and highlights the conceptual groundings upon which the proposed mediation analysis is interpreted. This is intended to assist the reader with an understanding of the rationales and understandings subsequently applied to the key analysis of this study, and the assessment of statistical effects in assessed models. This is deemed both necessary and appropriate due to ambiguity and debate regarding the assessment of mediation hypotheses, which stretch all the way back to its inception by Baron and Kenny (1986). These refer to three key points of interpretation: (1) mediation with cross-sectional data, (2) the interpretation of indirect, direct, and total effects, (3) the use of saturated models in path analyses.

1. First, a note should be given to the caution inferred by researchers on the use of mediation analysis with cross-sectional data (e.g., Fairchild & McDaniel, 2017). It has been argued that mediation analysis with cross-sectional data necessarily compromises the longitudinal nature of the analysis itself since each construct is not assessed at different time points (i.e., Y is not measured after M, M is not measured after X). However, there are instances in which this may be appropriate, as is argued to be the case here. Existing literature supports the hypothesised temporal ordering of the three focal constructs, namely that listening activity evokes listener's goals and by extension FML (e.g., Lamont et al., 2016), and that this affects the content of the music subsequently selected (e.g., Greb et al., 2019). Moreover, there is precedence in the literature, which has identified mediated relationships in cross-sectional observations of music listening (e.g., Greb et al., 2019). Also, music selection behaviours are not reasonably observable in longitudinal forms.

   Conceptually, music listening is a behaviour limited to observation and not subject to experimental designs under lab-based conditions, for instance, in which the entire contextual application of music in everyday life is removed. As such, a method that necessitates separate measurement at all three time-points is not well-suited to addressing present research questions as the phenomenon under consideration is, pragmatically speaking, an instantaneous process for which staged measurement would detract ecological validity. It is for these reasons that mediation is believed to be an appropriate procedure given the aims of this study, and although it holds conceptual limitations, these are no more or less limited to any other form of statistical modelling using cross-sectional data, since each is exposed to the same assumptions.

   Strong theoretical groundings have been cited as reasonable conditions under which cross-sectional mediation may be applied (Fairchild & McDaniel, 2017), and this is argued to be one such instance since the relationship between concurrent activity, FML, and the audio features of music selection is effectively instantaneous. Fairchild and McDaniel (2017) do, however, place an onus on researchers to articulate the reasoning of its use in cross-sectional data, and the points made above are intended to acknowledge and take responsibility for this.

2. Next, it is worth addressing the approaches used to interpreting the results of the mediation analysis itself. Mediation, as a hypothesis, is not especially complicated, and its underlying aims have been articulated. However, its interpretation and practical application is rather less straightforward (Bullock et al., 2010). This is because there are divergent schools of thought with regard to mediation, that impact the criteria to infer causal effects within models (Agler & De Boeck, 2017). At its heart, there are two key interactions to be interpreted under the mediation framework: the direct effect (*DE*) and the indirect effect (*IE*). In addition, the total effect (*TE*) may also be interpreted, which is the sum of *DE* and *IE*. Moreover, there is some divergence of opinion in the mediation literature regarding which paths ($X \rightarrow M$, $X \rightarrow Y$, $M \rightarrow Y$) should be considered important when interpreting the results of the analysis, and the debate is centred around the interpretability of these effects.

   In a mediation analysis, several hypotheses are typically being assessed simultaneously; that is that an exogenous variable (X) has a significant effect on a mediator variable (M), which in turn has an effect on an outcome variable (Y). In addition, another hypothesis is tested, which is that X has at least some effect on Y whilst controlling for the effect of M (i.e., $X \rightarrow Y$), the *DE*. Some argue that the *TE* (the sum of the *DE* and all *IE*s) should hold statistical significance, as it should show an overall relationship between the exogenous and outcome variables, before *IE*s are interpreted (Baron and Kenny, 1986).

   However, others argue this is irrelevant as the *IE* is of greater importance than the *TE* when it comes to interpreting mediation, which is often ignored when the latter is required to be significant (Zhao et al., 2010; Rucker et al., 2011; Nitzl et al., 2016; Agler & De Boeck, 2017). Agler and De Boeck (2017) explain that many tests of the *IE* are significant when tests of the *TE* are non-significant, and argue that given a mediation hypothesis, there is no real need to require a significant *TE* as it is irrelevant to the presence of an *IE* as they are estimated by different statistical models (in other words, mediation refers solely to *IE*). As such, whether mediation is present or not is not contingent on the significance of the *TE*, but rather on the context of all estimated effects. For example, a significant or non-significant *DE* in the presence of a significant *IE* characterises a model between full (also known as complete) and partial mediation, where full mediation refers to instances where the effect of *X* on *Y* is entirely borne through the mediator, whereas a partial mediation refers to instances where the *DE* between *X* and *Y* remains statistically significant in instances where the *IE* is also significant (Fairchild & McDaniel, 2017). The *TE* rather sums each of these specific effects to formulate an overall indication of whether *X* effects *Y* but should not be required to hold statistical significance since this may prevent useful insights (e.g., if the *DE* and *IE* pull in opposite directions).

Therefore, when mediation analysis is applied in this study, interpretation of the causal paths between the three focal constructs will be assessed in turn. Emphasis on significant *TE* is not considered to be essential in light of extant guidance (e.g., Rucker et al., 2011; Agler & De Boeck, 2017), and as such each term will be assessed according to its contextual merits (for example, in the presence of a significant *IE*, the *DE* and *TE* will be interpreted to contextualise the manner of the effect, and vice versa). It is hoped that this will provide a nuanced understanding of the specific relationships amongst all constructs and variables assessed.

3. Finally, this analysis is conceptually an exploratory one that fits all causal paths in a structural model, and each of these will be interpreted where relevant. This is because there is a general hypothesis that FML mediates relationships between concurrent activities and the audio content of listeners' music selection. Because this hypothesis is general, there are no degrees of freedom (*df*) in the structural model and so a saturated model will be specified (Raykov et al., 2013a; 2013b). This is because all activity variables in the model will be fit to all mediators and outcome variables, and all mediators will in turn be fit to all outcome variables.

As such, any *df* effecting model estimation and fit would stem exclusively from the measurement model, if the latent variables were simultaneously estimated (e.g., the factor structure of the FML model). Given, however, that there are no degrees of freedom in the structural model, it is considered to be more parsimonious with the present theoretical framework to take a two-step approach and segment the measurement and structural models. In short, this means that a measurement model (i.e., the estimation of a factor structure) would take place prior to the estimation of the structural model. This is pragmatic because it effectively facilitates cross-validation of the psychometric structure estimated in Study 1, prior to estimation of the saturated structural model. Subsequently using factor score regression to generate estimated scores for the latent variables can be used as proxies in the structural part of the model, increasing parsimony of the structural (path) model (e.g., Devlieger & Rosseel, 2017; Andersson & Yang-Wallentin, 2021) as well as mitigating the risk of misspecifications and improving model convergence.

Saturated models cannot provide meaningful fit statistics as the data is interpolated by the model itself, yielding 0 *df* and resulting in model fit statistics attaining 'perfect' values (e.g., CFI = 1, RMSEA = 0). These can be useful in cases where researchers wish to assess the temporal structure of a path analysis (Raykov et al., 2013a; Raykov et al., 2013b), but do not have cause or specific hypotheses to integrate model constraints to introduce positive *df* and thus facilitate comparison of the observed and simulated variance-covariance matrices of an underlying χ^2 distribution (upon which advanced fit measures are based). Introducing model constraints in the absence of specific hypotheses, however, can bias parameter estimates and confound results (Agler & De Boeck, 2017), and would not be appropriate in this case.

It is reasonable to fit a path analysis through a SEM approach when the model is saturated as this nonetheless facilitates the integration of multivariate constructs and provides more robust error terms, however, care should be taken to acknowledge the limitations of doing so with regard to assessing model fit. With this caveat in mind, an *effects-focused* approach will be used in interpretation, described by Agler and De

Boeck (2017) as the interpretative approach to mediation that primarily explores all possible effects to highlight the ones showing significant relationships, be that *IE*, *DE*, or *TE*. This is considered appropriate since there is an absence of specific hypotheses that would integrate parameter constraints into the analysed model, hence all paths are fit. To put it colloquially, the focus of model interpretation will be on the trees rather than the forest.

The context described here is intended to provide the reader with an a priori understanding of the ways in which the path model in question will be interpreted as part of this study. This is because there remain discrepancies in interpreting *IE* in particular, and so providing this perspective at an early stage is intended to enable ease of access as the results of this study are reported. These three points here are worth considering, therefore, as the discussion now proceeds to the reporting of data analyses in this study.

7.4 Results

Analyses were conducted in SPSS (version 28.0) and R (version 4.1.0). A number of steps were taken to assess the temporal structure between the concurrent listening activities of cross-sectional listening episodes, FML, and the audio features of identified tracks. Ultimately, this temporal structure was assessed using an SEM path analysis (mediation), however, earlier steps were necessary. Namely, these included the corroboration that the previously identified five-factor structure of FML adequately fits the data of both the online survey and the ESM studies (RQ1), and that MIR-generated audio features are comparable to self-perceived featural counterparts (RQ2).

RQ1 is addressed by fitting CFAs to both the online survey responses and the ESM responses separately, and then together in a pooled dataset (should CFAs demonstrate good model fit in respective study arms). RQ2 is to be assessed through bivariate correlation analyses between self-perceived and MIR-generated audio features extracted from the Spotify API. Should these two sets of audio features correlate where theoretically relevant, it can be argued that the API generated features approximately align with listener's perceptions of the music they encounter, thus providing greater confidence the API measures reflect listeners' experiences. Finally, and to address RQ3, the use of the aforementioned mediation analysis is applied, to explore whether activities influence audio features directly and/or indirectly via FML.

7.4.1 Thematic Analysis of Activity Variables

Once all data had been pooled, thematic analysis was used to interpret and deductively the code qualitatively reported listening activities. In this method, codes and sub-categories are generated according to systematic interpretation of the data (Barbour, 2014). Pre-defined themes were qualitatively gathered from comparative studies or characterisations of concurrent listening activities, namely those of Lamont et al. (2016), Maloney (2019), and Greb et al. (2019). Thematic analysis based on existing activity frameworks were used, however, consideration was also given to unclear or otherwise ambiguous responses. If new themes emerged in this, they would be accordingly assigned to their own category to extend the framework of activities. In cases of high specificity (e.g., getting a haircut, giving blood), responses were assigned to an "Other" category whereas if they were ambiguous or unclear, they would be placed into an "Unknown" category. After an initial walkthrough of participants' responses, 15 codes were initially generated, reported in Table 12.

Table 12 Initial Deductive Coding of Activity Variables

Activity	*N*	%
Work/Study	335	28.4%
Travelling	220	18.7%
Relaxing	172	14.6%
Personal Maintenance	135	11.5%
Chores	67	5.7%
Recreation/Leisure	64	5.4%
Other Media	40	3.4%
Exercising	28	2.4%
Other	28	2.4%
Eating	24	2%
Unknown	24	2%
Pure Music Listening	15	1.3%
Socialising	13	1.1%
Musicking	9	0.8%
Dancing	4	0.3%

These initial activity groupings were useful in identifying underlying themes within the cumulative pool of cases, however, to generate a more practical set of variables from which to conduct further analyses, further reduction was needed to minimise conceptual overlap whilst considering the wider implications of each initial category. The activity variables Chores, Personal Maintenance, and Eating were grouped to form one new group: 'Routine Activities'. The content of the initial groups maintained variations in the subsequent grouping (e.g., Chores was reflective of home tasks such as cleaning, whilst personal maintenance refers to everyday ablutionary acts such as Showering and Getting Dressed). It was deemed, however, that these three groups collectively formulate a broader set of comparatively routine daily tasks such as shopping, cleaning, washing, cooking, and eating, and as such that as a set of activities would elicit comparable modes of *functionality*.

Also, Exercising and Dancing were grouped on the rationale that both are indicative of active physiological output (such as the requirement to feel motivation in order to move to music or otherwise exert physical exertion), with dancing often treated as a sub-category of exercise in other literature (e.g., Gerber et al., 2014). Theoretically, such physiological affordances are benefited by more stimulating music, and thus it was hypothesised that the initial groups would be indicative of comparative stimulation. Finally, Pure Music Listening and Musicking were grouped to match Small's (1998) definition of musicking more closely as a reference to all music-centric behaviour, which includes active listening and performance which were previously separated by the initial groupings. Other Media, which refers to music listening alongside playing video games and watching television, was grouped with the Recreation/Leisure group which previously contained activities such as drinking alcohol and engaging with personal hobbies. Finally, the previously described 'Other' category responses were segmented between the remaining groups taking more individual responses into account to cluster them into conceptually broad groupings where possible. If this was not possible, then remaining responses were grouped together with the Unknown category. This process reduced the initial group of 15 activity groupings to nine (including one Other/Unknown category), shown in Table 13.

Table 13 Reduced Activity Groupings

Activity	*N*	%
Working/studying	335	28.4%
Routine Activity	234	19.9%
Travelling	220	18.7%
Relaxing	172	14.6%
Recreational Activity	108	9.2%
Unknown/Other	39	3.3%
Exercising	31	2.6%
Musicking	26	2.2%
Socialising	13	1.1%

<u>Case Selection</u>

Next, it was important to consider whether all of the retained 1,178 cases in the data pool were consistent with the aim of assessing autonomous cross-sectional music selection. First, to remain consistent with the exploration of autonomous music selection and retain the most theoretically robust observations possible, cases from the ESM in which participants were not in control of the music were filtered from analyses (n = 113), as were cases in which participants had responded beyond 24 hours in proximity to their most reason listening episode (n = 18), or if they could not remember when they last listened to music (n = 3). Of the remaining cases, 30 remained within the Unknown/Other Activity group. These were also removed as they did not form a clearly identified grouping of concurrent listening activities, thus retaining 1,014 observations in which participants had: (1) control of the music selected, (2) responded within a 24-hour time period from when they last listened to music, and (3) were part of an identified activity group.

Next, it was considered whether the activity group assigned to 'Musicking' should be retained, which by this stage had 24 cases remaining (n = 2 removed through prior processes outlined above). Musicking as a category presents complex theoretical issues insofar as its perceived meaning permeates several different aspects of musical engagement or interaction. Under the rationale presented by Small (1998), cases in this category contained cases effectively relating to music listening for the sake of listening itself (e.g., pure music listening), but also cases relating to musical performance or engagement (e.g., conducting, playing an instrument). This is because Small (1998) characterises Musicking as a catch-all parameter underpinning all forms of music-centred activity, which includes both music listening and performance. Small (1998) argues that such music-driven engagement falls under one banner, and although pertinent to more abstractive arguments in its own context, this does not yield clarity from a qualitative perspective in the current study, upon which reliable inferences could reasonably be substantiated in subsequent statistical modelling. This is because Musicking does not conceptually encapsulate one particular activity, but several, in that both active listening engagement and music performance (e.g., practicing, rehearsing, etc.) are subsumed, thus not providing nuance on different underlying motivations or goals, a rationale central to this study. Elsewhere, Greb et al. (2019), for instance, include the activity "Pure Music Listening" (p. 7), and although its definition is not discussed, presumably segment music listening as the concurrent activity, and music performance is not considered.

This conceptually diverges from Small's (1998) characterisation of Musicking, however. Since the aim of this study is to observe the ways in which concurrent listening activities influence FML and audio features of music selection, then music listening itself sits in a somewhat paradoxical position as listening for the sake of listening is not as internally harmonious as other categorical groups of activities. It should be conceded that there are of course plenty of instances where listeners engage with music for no other reason than to enjoy the music itself, but that this categorical positioning is muddled, and as such the presence of a Musicking category is questionable. This lack of clarity led to the decision that Musicking should be removed for the purpose of subsequent analyses, due to this conceptual issue regarding its internal consistency and broader interpretability. Other researchers may wish to further consider and engage with the ways in which this activity may be associated with music listening in everyday life, and subsequently consider how it can be parsed into discussions such as this, however, cases falling into this categorisation were removed on practical grounds.

Finally, it was noted that a substantial number of remaining cases (n = 115) did not have tracks that were found on Spotify, and thus no MIR-generated audio features were attributed. To further maximise the consistency of the sample and remove problems arising from missing values, these were also removed, retaining 875 complete observations in which all the described conditions were satisfied with no missing values. This was possible due to the large number of observations that remained, affording a high degree of specificity regarding retained cases. It is this final dataset of 875 cross-sectional observations that were used for all subsequent analyses, of which 335 were from the initial survey (M = 33.39; SD = 13.79) and 540 from the ESM, nested within 74 participants (M = 31.38; SD = 11.87). Initial survey responses were assigned matching identifiers (IDs) with observations in the ESM arm, so that participants with retained responses across both study arms were identified by the same ID and the partial clustering arising from this was reflected. This final dataset was therefore partially nested within a total 347 participants between the ages of 18 and 72 (M = 33.49; SD = 13.86). Of these, 138 (39.8%) identified as male, 195 (56.2%) as female, and 10 (2.9%) as non-binary/third gender. Four (1.2%) participants preferred not to state their gender identity.

7.4.2 Descriptive Statistics of Listening Episodes

Next, it is useful to report the descriptive statistics regarding music listening in the remaining data, as this contextualises the nature of the listening episodes used in subsequent analyses in terms of engagement style and circumstances. Table 14 illustrates the recency of ESRs to their respective listening episodes, as reported by participants.

Table 14 Recency of ESR to listening episodes

	Online Survey		ESM		Total	
	N	%	*N*	%	*N*	%
At time of response	71	21.19	243	45	314	35.89
Last hour	60	17.91	137	25.37	197	25.51
Last 2 hours	30	8.96	59	10.93	89	10.17
Last 2-3 hours	15	4.48	24	4.44	39	4.46
Last 3-4 hours	14	4.18	15	2.78	29	3.31
Last 4-12 (4+) hours	37	11.05	62	11.48	99	11.31
Last 12-24 hours	108	32.24	*NA*	*NA*	108	12.34
Total	335	38.29	540	61.71	875	100

The largest category overall was that listening episodes were concurrent to the response (35.89%). This is notably larger for ESM responses (45%) than the online survey (21.19%), which makes comparative sense given the different designs. Although there are no particular hypotheses related to this question in the present study as all data is viewed as cross-sectional, it is useful to contextualise the proximity of reports to the listening episode. Due to the varying designs, extended options were given for the online survey, which was limited to 4+ hours in the ESM arm due to the repeated notifications received by participants, hence the 12–24-hour time band is not applicable (*NA*) in the ESM arm. In hindsight, it would have been more consistent to increase the bands for the ESM study arm, but given the ecologically valid setting of the study, this is thought to have a relatively minor impact. Cases in the 4+ hours band of the ESM study are therefore treated as occurring within the last 24 hours and added to the last 4-12 hours band.

This recency measure is useful insofar as online surveys often suffer from fading affect biases, and as such provides some measure of how recent the listening episode being reported was (although, this does admittedly hold potentially arbitrary or inaccurate assessments on the part of participants). Though by no means a perfect system to assess this, it does nevertheless provide a greater degree of confidence that cases subject to recollection biases are somewhat mitigated. Next, the frequencies of participants' reported location are shown in Table 15.

Table 15 Location of listening episodes

	Online Survey		ESM		Total	
	N	%	*N*	%	*N*	%
Work	37	11.90	57	10.56	94	10.74
Home	206	60.60	353	65.37	559	63.89
Friend's Home	1	0.20	6	1.11	7	0.80
Gym	3	0.90	0	0.00	3	0.34
Transitory Space	76	21.10	112	20.74	188	21.49
Urban Location	0	0.00	4	0.74	4	0.46
Restaurant/Bar	1	0.50	1	0.19	2	0.23
Cultural Location (e.g., Place of Worship)	1	0.50	0	0.00	1	0.11
Musicking Location (e.g., Rehearsal Studio)	1	0.50	0	0.00	1	0.11
Other	9	3.90	7	1.30	16	1.83
	335	38.29	540	61.71	875	100

It is clear that Home was by far the most common location for music listening, accounting for 63.89% of all cases. Although listening to music at home has always been extremely prevalent in the literature (e.g., Maloney, 2019), its frequency here may be somewhat confounded by circumstances stemming from the COVID-19 pandemic, such as home working. Data collection for this study took place in the United Kingdom during spring 2022, which may go some way to explaining why the frequency of this particular category was quite so large in this instance given the growth in home working in countries like the United Kingdom (e.g., Felstead & Reuschke, 2023). Next, Transitory Spaces (i.e., whilst being on the move) was the second most common category overall (21.49%), which is again consistent with the affordances of listening technologies (Bull, 2006; Lamont et al., 2016; Maloney, 2019). The portability afforded by modern technologies was highlighted in the introduction to this thesis and corroborates the notion of music serving as a tool when users are travelling. The third location that will be specifically addressed here is the frequency of Work (10.74%). Lamont et al. (2016) note the prevalence and benefits of music listening whilst in office spaces for instance, and as such its relatively common occurrence is no surprise, although substantially lower music listening in the home remains. Next, Table 16 presents the frequencies of audio formats across all cases.

Table 16 Format of Audio-Content

	Online Survey		ESM		Total	
	N	%	*N*	%	*N*	%
Streaming Service	229	68.36	397	73.52	626	71.54
Radio	32	9.55	3	0.56	35	4
Physical Format	14	4.18	33	6.11	47	5.37
Digital File (e.g., mp3)	18	5.37	39	7.22	57	6.51
Audio-Visual Content (e.g., YouTube)	37	11.05	65	12.04	102	11.66
Other	5	1.49	3	0.56	8	0.91
	335	38.29	540	61.71	875	100

Consistent with both literature and industry findings (e.g., Brown & Krause, 2020; Friedlander, 2019; 2020; 2021), streaming services dominate audio consumption and form the majority of cases (71.54%). In addition, audio-visual content (e.g., YouTube) formed the second largest category (11.66%), whilst Radio formed 4% cases overall, although it was not inherently clear by what manner this is accessed (e.g., home device, car, smartphone). Physical formats accounted for 5.37% of cases, whilst other digital files (e.g., mp3s saved onto a device) was 6.51%. The final 0.91% of cases were made up of 'Other' formats. These frequencies corroborate the dominance that streaming services in particular have in accessing music in the twenty-first century. Brown and Krause (2020) note that formats may be selected by users for the affordances they may provide users' in the accomplishment of certain goals. As such, although formats are not inherently related to particular research questions, these frequencies are largely consistent with the share of format's usage in everyday life according to industry. This may have psychological implications, such as with regards to the goals and, by extension, FML (e.g., Brown & Krause, 2020). Next, Table 17 highlights the relative frequencies of Medium during listening episodes.

Table 17 Medium of Audio-Content

	Online Survey		ESM		Total	
	N	%	*N*	%	*N*	%
Specific Song/Track	70	20.20	107	20	177	20.23
Playlist	158	43.30	240	44.44	398	46
An Album	63	18.30	170	31.48	233	26
Other	44	18.10	23	4.26	67	7.66
Radio	*26*	*9.40*	*5*	*0.93*	*31*	*3.54*
An Artist	*6*	*2.50*	*7*	*1.30*	13	*1.49*
	335	38.29	540	61.71	875	100

'Medium' here refers to the form of music content selected by listeners. Playlists (46%) formed the largest category in this instance, likely tied to the prevalence of streaming services, although this is not explored further as part of this study. The selection of albums was the next largest category (26%) whilst specific songs or tracks accounted for 20.23% of cases. This is conducive to further contextualising the manner of interactions with music. Like other descriptive statistics outlined, this is not directly related to any particular research questions, however, the manner by which music is engaged is nonetheless conducive in formulating an understanding of the ways in which listeners engage with and select music. Note that although there was an 'Other' category, there was a large number of qualitative responses specifically naming Radio and specific artists. As such, these were retrospectively deductively coded in addition to the 'Other' category to illustrate their relative frequency.

Since activity variables are to serve as the predictors in subsequent analyses, a summary table of the frequencies of activities in this finalised set of observations is shown in Table 18.

Table 18 Activity group sizes in final data

	Online Survey		ESM		Total	
	N	%	*N*	%	*N*	%
Work/Study	120	35.82	152	28.15	272	31.10
Travelling	75	22.39	98	18.15	173	19.80
Relaxation	53	15.82	86	15.93	139	15.90
Routine Activities	52	15.52	124	22.96	176	20.10
Exercise	13	3.88	13	2.41	26	3.00
Socialising	2	0.60	5	0.93	7	0.80
Recreation/Leisure	20	5.97	62	11.48	82	9.40
	335	38.29	540	61.71	875	100

Group sizes clearly vary, which should be acknowledged. Whilst it made hypothetical sense that certain applications of music listening would be more common than others, it is clear that Work/Study is by far the most prevalent. At the opposite end of this consideration, is the category, Socialising. This small subsample of observations would ideally be larger in order to increase test power, as confidence intervals within this category will presumably be substantially wider than other categories. Nevertheless, it was deemed prudent to proceed with this reduced dataset to maximise consistency of cases in subsequent analyses.

7.4.3 Cross-validation of Psychometric Structure

Next, it was considered how to best address RQ1, regarding whether the 23-item FML measure generated during Study 1 can be cross validated in the present data; and if so, how best to incorporate these latent variables into subsequent analyses. For this, it was deemed appropriate to apply CFA using the 23 items as indicators of the five latent variables as in the prior study. However, because the data structure includes multiple responses from the same individuals, the assumption of independence of observations is violated and the data is clustered, or nested, due to these repeated assessments (Galbraith et al., 2010; Moen et al., 2016; McNeish & Harring, 2017; Bauer et al., 2020). If dependency were to be ignored (known as disaggregation), incorrect parameter estimates, standard errors, and fit statistics may be generated and lead to inflated Type I errors (du Toit & du Toit, 2008; Cheong & MacKinnon, 2012). In ESM studies, it has been often argued that data should be aggregated to a higher level (e.g., averaging responses to the level of each individual; Hektner et al., 2007). Such aggregation is, however, problematic in that it can lead to a loss of information and leave analyses open to ecological fallacies (Pollet et al., 2015). It was therefore considered to conduct intended analyses in a clustered framework, which could take a multilevel approach to utilise adjusted margins of error to account for the clustered data structure where needed. Specifically, this was deemed applicable due to the cross-sectional rather than longitudinal design of both studies, gathering situation-level observations with subject-level differences outside of the present scope of interest.

In a SEM framework, it is preferable to take the multilevel sampling scheme into account using either design-based or model-based approaches (Wu & Kwok, 2012). Though applied in other instances of quantitative analysis of ESM data in music listening studies (e.g., Greb et al., 2019), data aggregation was not therefore considered to be the optimal approach to handling this issue, and design-based or model-based approaches were considered as alternatives. In design-based approaches, the multilevel data, or dependency, is accounted for by adjusting standard errors based on the study design, also referred to as cluster-robust standard errors (Rosseel 2017; Huang, 2018; Bauer et al., 2020). This works by first specifying a disaggregated model under the normal assumption that all observations are independent, for which corrections are then applied to properly account for dependencies present in the data by specifying its complex structure (Oberski, 2014; Bauer et al., 2020). The upshot is that standard errors and *p*-values in the adjusted model are (unsurprisingly) larger than the disaggregated model, with the applied corrections maintaining the nominal Type I error rate (Bauer et al., 2020).

In model-based approaches, the multilevel data is accommodated by specifying level-specific models for each level of data, which adjusts for clustering at each level, generating a multilevel

SEM (MSEM; Wu & Kwok, 2012). MSEM has been applied elsewhere in the music psychology literature with ESM data clustered within individuals (Randall & Rickard, 2017), or following aggregation (Greb et al., 2019). However, such studies held higher level hypotheses, whereas the present study is only interested in observation-level responses, rather than response variance across sampling units. This is more consistent with a design-based approach, preferred when researchers are only interested in examining the marginal estimates, the precision of which are adjusted to reflect the dependency in the data (Wu & Kwok, 2012; McNeish & Harring, 2017; Huang, 2016; 2018).

MSEM may also be applied when researchers are only interested in a single level of the model, however, software packages require the researcher to specify high-level models regardless (for instance, by specifying a saturated model at Level 2 if Level 1 is of sole interest). Applying model-based designs further complicates the analysis and leads to researchers making additional assumptions about their data, which may not always be appropriate or known (McNeish et al., 2017). In other words, when the clustering of the data is considered by the researcher to be more of a nuisance than it is related to any particular research questions or hypotheses, design-based approaches may be preferable as they avoid unnecessary assumptions about the data (Rosseel, 2017; Bauer et al., 2020), although it should be noted that there is some divergence of opinion regarding this, and that both methods carry relative advantages and disadvantages (Wu & Kwok, 2012; Stapleton et al., 2016; McNeish & Harring, 2017). It is important to point out, however, that either design-based or model-based approaches can be applied when the nature of the clustering in a dataset is the consequence repeated assessments within participants, as is the case here, and that both are valid ways of accommodating nested data (McNeish & Harring, 2017).

In wishing to balance the guidance in the methodological literature with the needs of the present study, the following decision was made. Since all observations in this study are at the within-level and have no higher-level hypotheses, a design-based approach was the most appropriate method of accommodating the complex data structure. This is in contrast with other music listening ESM studies that have gone down the model-based (i.e., MSEM) route (e.g., Randall & Rickard, 2017; Greb et al., 2019). Given the sole interest is the observation level in this study, however, it was decided to conduct these analyses at the disaggregated level, and adjust parameter estimates and standard errors based on the clustered, cross-sectional data structure by following a design-based approach to account for violation of test independence (Skinner & de Toledo Vieira, 2007; Wu & Kwok, 2012; Oberski, 2014).

To this end, preliminary CFAs were first fit to both study arms separately and model fit was inspected to assess whether the constrained model was reflective of the characteristics of the observed data in respective study arms. Robust maximum likelihood (MLM) was applied in all instances. These were fit in R (version 4.1.0), using the *lavaan* package to fit initial models (version 0.6-12; Rosseel, 2012), and the *lavaan.survey* package (version 1.1.3.1; Oberski, 2014) to specify the complex survey design and accordingly adjust standard errors and model fit (i.e., scaled χ^2, robust fit statistics). *lavaan.survey* provides a means of making design-based adjustments according to clustered or nested structures in SEMs and addresses biases in estimators of covariance matrices as a consequence (Oberski, 2014; Rosseel, 2017).

A CFA of the 23-item structure was first conducted on the retained set of online survey responses (n = 335), which showed the model was a good fit for the observed data ($\chi^2(220)$ = 344.450, p <.001, CFI = .968, TLI = .964, RMSEA = .044). This model was not adjusted as there were no repeated assessments (i.e., clustering) within individuals. Next, a CFA of the structure was fit to the ESM observations (n = 540), in which the aforementioned design-based adjustments were made to account for the violation of test independence (using participant IDs as the clustering variable). This showed the model was a good fit for the observed ESM data also ($\chi^2(220)$ = 374.473, p <.001, CFI = .927, TLI = .916, RMSEA = .045). These preliminary assessments indicated that the 23-item psychometric structure was a good fit for the data in separate study arms. Therefore, it was deemed appropriate to conduct a CFA on the pooled dataset of both initial survey and ESM observations for simultaneous analysis to maximise test power for subsequent analyses.

This third and final CFA was fit using participants' IDs as the clustering variable on the pooled dataset of all observations (n observations = 875; n participants = 347), with robust fit statistics again indicating the model was a good fit for the pooled dataset ($\chi^2(220)$ = 512.408, p <.001,

CFI = .962, TLI = .957, RMSEA = .047). As previously mentioned, initial survey responses were matched to ESM participants' IDs to accordingly identify individuals with retained submissions in both study arms. Remaining online survey responses (i.e., those without submissions amongst ESM observations) were assigned unique IDs to act as single-subject clusters (e.g., Sanders, 2011), hence each ID represented one participant when modelled. As in Study 1, standardised factor loadings, AVE, and covariances were then inspected to aid judgement of validity and reliability, shown in Tables 19 and 20.

Table 19 Loadings, AVE, and reliability of third CFA iteration

	Item Refer-ence	Unstandar-dised Estimate (λ)	Standar-dised Estimate (β)	*SE*	*p*	AVE	ω
Identity and Social Bonding						.571	.888
To identify with others through your shared values and/or culture	S1	0.736	.734	0.047	<.001		
To help express your identities and values to others	S2	0.778	.759	0.047	<.001		
To feel that certain artists, pieces, or genres of music are central to your social group's culture and sets you apart from others	S3	0.680	.732	0.045	<.001		
To help you bond with others, and to subsequently feel a sense of belonging with those individuals	S4	0.750	.815	0.046	<.001		
To act as a topic of discussion with others and ease communication or interaction	S5	0.673	.745	0.044	<.001		
To match a group's dynamic so you are able to bond with group members when listening with others	S6	0.615	.748	0.047	<.001		

Emotion Regulation						.590	.893
To distract yourself from negative or stressful situations	E1	1.218	.835	0.040	<.001		
To relieve stress and negative emotions associated with negative events or situations	E2	1.214	.824	0.041	<.001		
To manage emotions that you may be experiencing despite external influences, whether they are positive or negative	E3	1.096	.777	0.040	<.001		
To find meaning within music that allows you to reduce negative emotions or moods	E4	1.081	.728	0.042	<.001		
To feel certain specific emotions, such as joy or sadness	E5	1.027	.656	0.040	<.001		
To help you reverse your emotions or moods	E6	1.074	.794	0.047	<.001		
Focus and Concentration						.669	.891
To help you focus or concentrate on tasks	F1	1.378	.863	0.042	<.001		
To help you 'flow' when trying to concentrate on something	F2	1.322	.835	0.044	<.001		
To stop external factors from distracting you when trying to concentrate on a task	F3	1.234	.785	0.056	<.001		
To help you attain the necessary mindset to working on certain tasks	F4	1.23	.784	0.054	<.001		
Background and Accompaniment						.552	.818

To avoid silence when you're alone (e.g., playing music when nobody else is home)	B1	1.166	.706	0.078	<.001		
To feel a sense of company in the absence of others (e.g., playing the radio when home alone)	B2	1.309	.834	0.067	<.001		
To reduce feelings of being lonely when social interaction is not possible	B3	1.216	.817	0.087	<.001		
To provide background noise and remove silence	B4	0.956	.606	0.084	<.001		
Physiological Arousal						.683	.869
To help you maintain pacing during physical activities, such as yoga, walking or whilst in the gym	P1	0.990	.808	0.055	<.001		
To help physically stimulate you to carry out physical tasks, such as exercise or sports	P2	1.222	.887	0.047	<.001		
To help you achieve goals by motivating you to further action (such as increased effort during exercise)	P3	1.074	.777	0.045	<.001		

Note. *SE* = Standard Error. AVE = Average Variance Explained.

Table 20 Third CFA iteration AVE in relation to covariance squared

Factors	Φ	$Φ^2$
Identity and Social Bonding (.571) & *Emotion Regulation* (.590)	.545***	.297

Identity and Social Bonding (.571) & *Focus and Concentration* (.669)	.219***	.048
Identity and Social Bonding (.571) & *Background and Accompaniment* (.552)	.276***	.076
Identity and Social Bonding (.571) & *Physiological Arousal* (.683)	.538***	.289
Emotion Regulation (.590) & *Focus and Concentration* (.666)	.349***	.122
Emotion Regulation (.590) & *Background and Accompaniment* (.552)	.466***	.217
Emotion Regulation (.590) & *Physiological Arousal* (.683)	.625***	.391
Focus and Concentration (.666) & *Background and Accompaniment* (.552)	.297***	.088
Focus and Concentration (.669) & *Physiological Arousal* (.683)	.494***	.244
Background and Accompaniment (.552) & *Physiological Arousal* (.683)	.391***	.153

Note. AVE of each factor is in brackets. ***p<.001.

As seen in Tables 19 and 20, and following the criteria Hair et al. (2014), the model demonstrated good validity and reliability as standardised factor loadings of each item was >.50, AVE for each factor was >.50, and squared covariances were below the AVE of any factor. This provides strong evidence of validity and reliability according to such criteria across both studies 1 and 2, constituting cross-validation.

As this model was a good fit for the pooled set of cross-sectional observations when clustering was accounted for, factor scores for the latent variables were generated using the regression method, extracted for each case, and added to the dataset to act as proxies for the latent variables in the structural path model (Devlieger & Rosseel, 2017; Loehlin & Beaujean, 2017). It was decided to segment the measurement and structural models in this way because it firstly reduces the dimensionality of the variables in an already complex model, and also because the factor score regression (FSR) approach has been shown to perform well in sufficiently large samples (Andersson & Yang-Wallentin, 2021).

7.4.4 Bivariate Correlations of self-reported and MIR assessments of audio features

This study sought to observe how objective measures of audio features as provided by the Spotify API are affected by listener's concurrent listening activities and FML. This builds on research conducted by Greb et al. (2019), in particular, insofar as this prior study used self-reported characteristics of audio features, thus limiting computational uses, a subject of interest to this thesis (i.e., in content-based recommender systems). Therefore, it was considered that if MIR-generated measures of audio features share theoretically consistent correlations with self-perceived musical characteristics, then there is support that these features approximately align with listener's experiences. This would allow subsequent analysis within this study as well as future research to use such measures with a greater degree of confidence that they are (at least approximately) reflective of listeners' perceived experiences. Although not all aspects of the self-reported and MIR-generated features are paralleled between measures, there are a significant number of variables that one would hypothesise to see correlate (e.g., higher levels of perceived excitement positively correlating with the API measure, *Energy*).

To gauge this, bivariate correlation analyses were conducted between listeners' self-reports of music characteristics (Greb et al., 2019), and the measures extracted from the Spotify API. Due to the descriptive nature of this analysis, all 875 cases were used. All variables were standardised prior to analysis, in which Spearman's rho (ρ) was used as the correlation coefficient (Dancey & Reidy, 2017). This analysis was conducted using SPSS (version 28.0), the results of which are shown in Table 21.

Table 21 Bivariate Correlations of Self-Perceived Musical Characteristics and Spotify Audio Features

Liveness	Instrumentalness	Acousticness	Speechiness	Mode	Loudness	Key	Energy	Danceability	
ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	ρ	
.133**	-.157**	**-.427****	.275**	-.062	**.407****	.049	**.499****	.143**	Calm-Exciting
.150**	-.114**	**-.445****	.283**	-.025	**.415****	.073*	**.526****	.184**	Slow-Fast
.129**	-.138**	-.193**	.196**	.043	.195**	.025	.244**	.255**	Sad-Happy
-.045	-.119**	-.008	-.043	.049	-.010	-.025	-.012	.040	Unfamiliar-Familiar
-.070*	-.042	.148**	-.140**	.021	-.115**	-.040	-.165**	-.073*	Less melodic-Very Melodic
.104**	-.036	-.216**	.205**	.001	.169**	-.016	.256**	.130**	Less Rhythmic-Very Rhythmic
.024	.055	-.023	.001	-.039	-.005	.037	.016	-.121**	Simple-Complex
.101**	-.141**	**-.405****	.278**	-.036	**.364****	.053	**.433****	.054	Peaceful-Aggressive
.042	-.013	-.253**	.172**	-.094**	.211**	.037	.277**	-.110**	Less Intense-Very Intense
-.017	**-.490****	-.107**	-.028	.091**	.272**	.019	.168**	.146**	Instrumental

Time Signature	Duration (ms)	Tempo	Valence
ρ	ρ	ρ	ρ
.128**	-.095**	.122**	.262**
.174**	-.078*	.160**	**.313****
.122**	-.111**	.034	**.339****
.008	.014	.033	.057
-.016	.049	-.017	-.026
.040	.002	.060	.194**
-.025	.158**	.041	-.058
.088**	-.051	.137**	.149**
.020	.058	.079*	-.029
.050	-.187**	.070*	.227**

Note. Spotify API features are along the *x* axis; self-reported characterisations are along the *y* axis. ** Correlation is significant at the .01 level (2-tailed). * Correlation is significant at the .05 level (2-tailed). Statistically significant coefficients ±.30 highlighted in bold.

These coefficients show a number of theoretically consistent correlations, such as positive correlations between the self-perceived bipolar variable *Calm-Exciting* and the Spotify features *Energy*, ρ (873) = .50, p <.001 and *Loudness,* ρ (873) = .41, p <.001, and a negative correlation (i.e., closer to perceptions of calmness) with *Acousticness*, ρ (873) = -.43, p <.001. *Slow-Fast* was positively correlated with *Energy*, ρ (873) =.53, p <.001, and *Loudness*, ρ (873) = .42, p <.001, and again negatively correlated with *Acousticness*, ρ (873) = -.45, p <.001. *Sad-Happy* was positively correlated with *Valence*, ρ (873) = .34, p <.001, although magnitude of this correlation was relatively weak given the hypothetical strength of associations between the two. Additional consistencies can be seen between perceptions of *Peaceful-Aggressive* and *Energy*, ρ (873) = .43, p <.001, *Loudness,* ρ (873) = .36, p <.001, and *Acousticness*, ρ (873) = -.40, p <.001. Finally, the standardised binary scoring for *Instrumental-Vocal* (i.e., where Instrumental was coded with a 1 prior to standardisation, and Vocal with a 2) was negatively correlated with *Instrumentalness,* ρ (873) = -.49, p<.001, thus indicating correlation with the lower value representing *Instrumental*, rather than *Vocal*.

These examples highlight consistencies one would expect to see between the bipolar self-report measures and audio features obtained from the Spotify API, implying that the MIR features have correspondences with self-perceived scores of musical characteristics. Though individual variation undoubtedly impacts listeners' perceived description of the music, this nevertheless provides a higher degree of confidence that the proposed objective measures are approximately congruent with self-perceived scores, which further rationalises their use in further analyses with the intention to identify which objective featural parameters vary according to a listeners' context.

One noteworthy absence, however, is the lack of any meaningful correlation relating to the measure for *Tempo* and *Slow-Fast*. This is plausibly due to beat, or onset, detection being a wide-ranging issue in the MIR literature (e.g., Vinay et al., 2021). Signal processing algorithms are able to detect beat transients quite well, however, are inconsistent when attributing tempo measures (i.e., Beats Per Minute; BPM) reliably since other factors may well contribute to the perceptual experiencing and pacing of a piece of music (e.g., a listener may feel a beat at 100 BPM, but a signal processing algorithm may often double or halve this). In this sense, identifying a 'correct' answer is complicated since there could plausibly be more than one correct BPM score. Therefore, although there is no support here for a strong correlation between the MIR-attributed *Tempo* measure and self-perceived *Slow-Fast* rating, this may be a limitation associated with a broader issue of beat detection in signal processing. Overall, however, these correlations provide evidence that the extracted audio features from the Spotify API are roughly correlated with listener's theoretically comparable perceived audio features where relevant. This provides a higher degree of confidence that these particular measures are suitable representations of listener experiences, and thus conducive to subsequent analysis.

7.4.5 Dimension Reduction of Spotify Audio Features

Six outcome variables of broad interest were identified within the audio features extracted from the Spotify API (*Danceability, Energy, Speechiness, Acousticness, Instrumentalness*, and *Valence*). These selected audio features were assessed to encapsulate perceptual features of music, encompassing acoustical features such as loudness and tempo within perceptual measures such as *Energy*. There is, however, high dimensionality arising from the large number of outcome variables, and it was hypothesised that through dimension reduction the structure and interpretability of the model would be simplified. A PCA was deemed to be the most prudent approach to this reduction, the aim of which is to transform a set of variables by generating a set of uncorrelated linear components that encapsulate as much of the variance in the original variables as possible (Jackson, 1991; Suhr, 2005). This is in contrast with factor analysis in which the primary aim is to explain the correlations among observed variables, of which explained variance is more of a by-product than a central aim (Widaman, 2018).

PCA was considered preferable, therefore, as common factor models like EFA assume that there is an underlying dimensionality that influences a set of variables or items (latent 'factors' can be thought as a function of that set of items), whereas a component model generates composites out of the original variables themselves in a (preferably) rotated space, which in turn reduces the number of dimensions amongst that set of variables (Jolliffe, 2002; Widaman, 2018). This distinction is important conceptually as, in the present study, a PCA would serve to reduce the six outcome variables into a set of smaller, formative measures to which those variables contribute, but not go so far as to assume that these are inherently functions of latent factors, something often confused in the literature (Matsunaga, 2010). Moreover, a common factor model (i.e., E/CFA) would not be appropriate in the present study given that the specific methods of computation of the Spotify API features is proprietary knowledge and remains unpublished, although these are likely calculated from smaller items that are not individually accessible (Maloney et al., 2021).

Based on this rationale of dimension reduction to improve model interpretability whilst retaining as much of the observed variance as possible, three iterative PCAs (SPSS; version 28.0) were used to reduce these variables into a smaller set. An initial PCA was conducted on the six selected Spotify audio features (*Danceability, Energy, Speechiness, Acousticness, Instrumentalness*, and *Valence*), which used the Eigenvalues >1 rule and an orthogonal (Varimax) rotation, to reduce the number of outcome variables (Varimax provides particular affordances in PCA, such as that the solution yields components with a small number of large loadings and simplifies interpretation; Abdi & Williams, 2010). Component extraction was based on the correlation matrix of the variables, for which the variables were standardised (i.e., $M = 0$; $SD = 1$) prior to submission (Jolliffe, 2002).

The KMO measure of sampling adequacy was adequate (KMO = .628), and Bartlett's Test of Sphericity ($\chi^2(15) = 1499.962$, $p<.001$) was significant, indicating that dimension reduction analyses were suitable. The initial extraction of the standardised audio features yielded two components, in which five of the six selected audio features loaded meaningfully with at least one component (≥.40). *Speechiness*, however, failed to demonstrate any particularly strong association to either of the components and, as a consequence, the PCA was re-run with *Speechiness* removed. In the second iteration, the five remaining variables again yielded two components, structurally identical to the first. Component 1 primarily indicated higher levels of *Energy* and lower levels of *Acousticness*, whilst Component 2 consisted of higher levels of *Danceability* and *Valence*, with lower levels of *Instrumentalness*. The results of these first two iterations are shown in Table 22.

Table 22 Initial Component Extractions for selected Spotify Audio Features

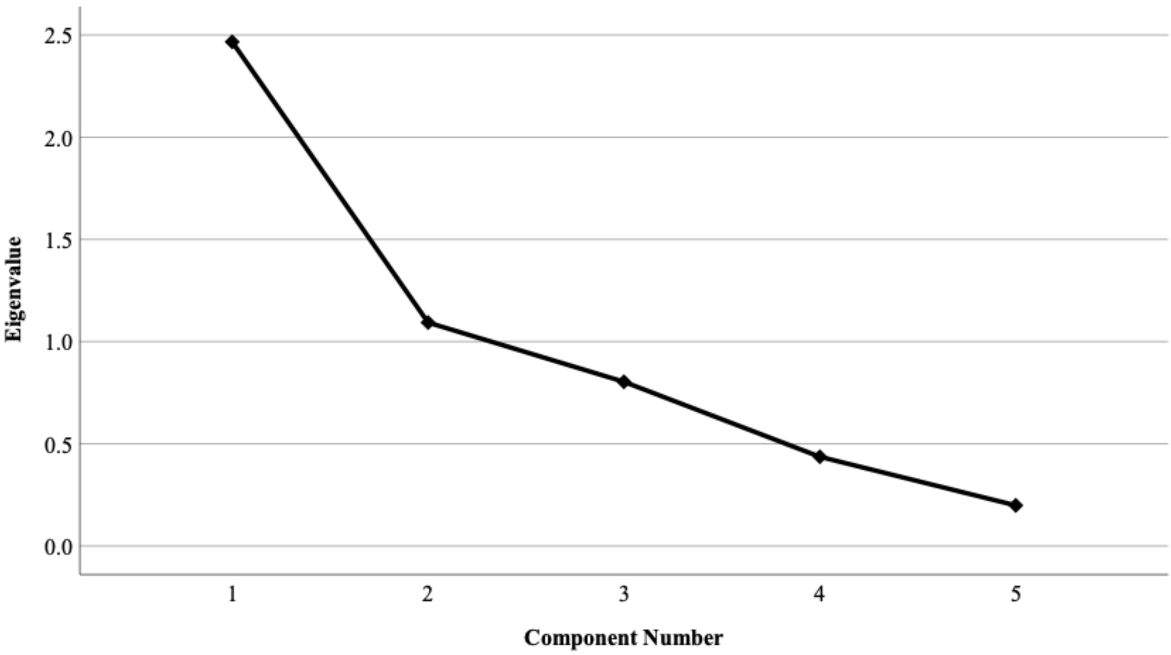
	Component	
Audio feature	1	2
Iteration 1		
Energy	**.923**	.165
Acousticness	**-.913**	-.143
Danceability	.027	**.878**
Valence	.286	**.789**
Instrumentalness	-.273	**-.485**
Speechiness	.285	.159
Speechiness removed and PCA re-run		
Iteration 2		
Energy	**.921**	.196
Acousticness	**-.919**	-.176
Danceability	-.004	**.878**

Valence	.272	**.800**
Instrumentalness	-.260	**-.495**

Note. Applied rotation method is Varimax with Kaiser Normalisation.

The two extracted components from the second iteration explained 71.21% of observed variance. This was, however, based on the eigenvalues >1 rule, which infers a rule-of-thumb cut-off point of which there is no formal numerical value. Rather, assessments of the relative eigenvalues can be used according to interpretation and assessment of the proportions of variance explained (i.e., the magnitude of distances between eigenvalues; Jolliffe, 2002). This can be interpreted visually through a scree plot, which is shown in Figure 10 for the second iteration, to help researchers assess the number of latent variables to retain.

Figure 10 Scree Plot of Second PCA Iteration



The implied third component from this extraction had a large eigenvalue (λ = .804) and explained a further 16.07% of the variance. The scree plot was interpreted to support the implication that a third component explained a substantial amount of variance, although it should be acknowledged that visual interpretations of scree plots remain subjective. Given that the aim of PCA is to encapsulate as much of the explained variance as possible and given that the third component explained a large amount of this, it was nevertheless decided to run a third iteration of the PCA in which three components were forcibly retained. Table 23 shows the results of this third iteration.

Table 23 PCA iteration retaining third component

	Component		
Audio Feature	1 (*Arousal*)	2 (*Valence*)	3 (*Instrumentalness*)
Iteration 3			
Energy	**.928**	.167	-.097
Acousticness	**-.920**	-.137	.119
Danceability	.035	**.904**	-.096
Valence	.303	**.806**	-.129
Instrumentalness	-.142	-.153	**.978**

Note. Applied rotation method is Varimax with Kaiser Normalisation.

The results indicated that the first component (*Arousal*) was unchanged. The primary contributing variables to this component were *Energy* and *Acousticness*. The negative loading of *Acousticness* can be interpreted to be negatively correlated with higher values of the component. Components extracted via a Varimax rotation can be interpreted from the opposition of variables with positive loadings to variables with negative loadings (Abdi & Williams, 2010). As such, higher values of the component indicate higher levels of *Energy* and lower levels of *Acousticness*, and vice versa. The second component, meanwhile, changed in that it partitioned *Instrumentalness* into a third (essentially standalone) component. This makes theoretical sense since the extent to which a piece of music is instrumental does not seem to theoretically play a particularly large role in subsequently low valence, as implied by the second iteration. These three components: (1) *Arousal*, (2), *Valence*, and (3) *Instrumentalness* were interpreted to make substantive theoretical sense and explained 87.28% of the observed variance in total.

Thus, the six selected audio features from the API were halved and *Speechiness* excluded altogether; a subset of three components explaining 87.28% of the total variance was deemed sufficient in light of the theoretical consistencies. Component scores, linear combinations of observed variables weighted by eigenvectors (Suhr, 2005), were computed, and added to the dataset via SPSS (version 28.0) for subsequent analysis, in which they would be used as outcome variables in the described mediation model.

7.4.6 Mediation Analysis

Prior analyses corroborated that: (1) the previously identified psychometric FML structure fits the observed data in both study arms, as well as the pooled data, (2) the audio features extracted from the Spotify API correlate with theoretically relevant counterparts from a self-perceived measure of audio content (Greb et al., 2018a; 2019), and (3) that these variables can be reasonably reduced into a smaller subset explaining a large amount of variance through PCA. Next, the mediation hypothesis discussed earlier was conducted to assess the structure amongst the three focal constructs. This was intended to assess whether listening activities affect FML in cross-sectional data, and whether this modulates the content of the music they selected, directly and/or indirectly. For reasons previously outlined, a SEM framework was used to fit the hypothesised model which, as with the iterative CFAs reported in section 7.4.3, was conducted in R (version 4.1.0) using the *lavaan* package (version 0.6-12; Rosseel, 2012) to specify and fit the model, and the *lavaan.survey* package (version 1.1.3.1; Oberski, 2014) to model dependency in the data and adjust standard errors and p-values according to dependency in the data structure.

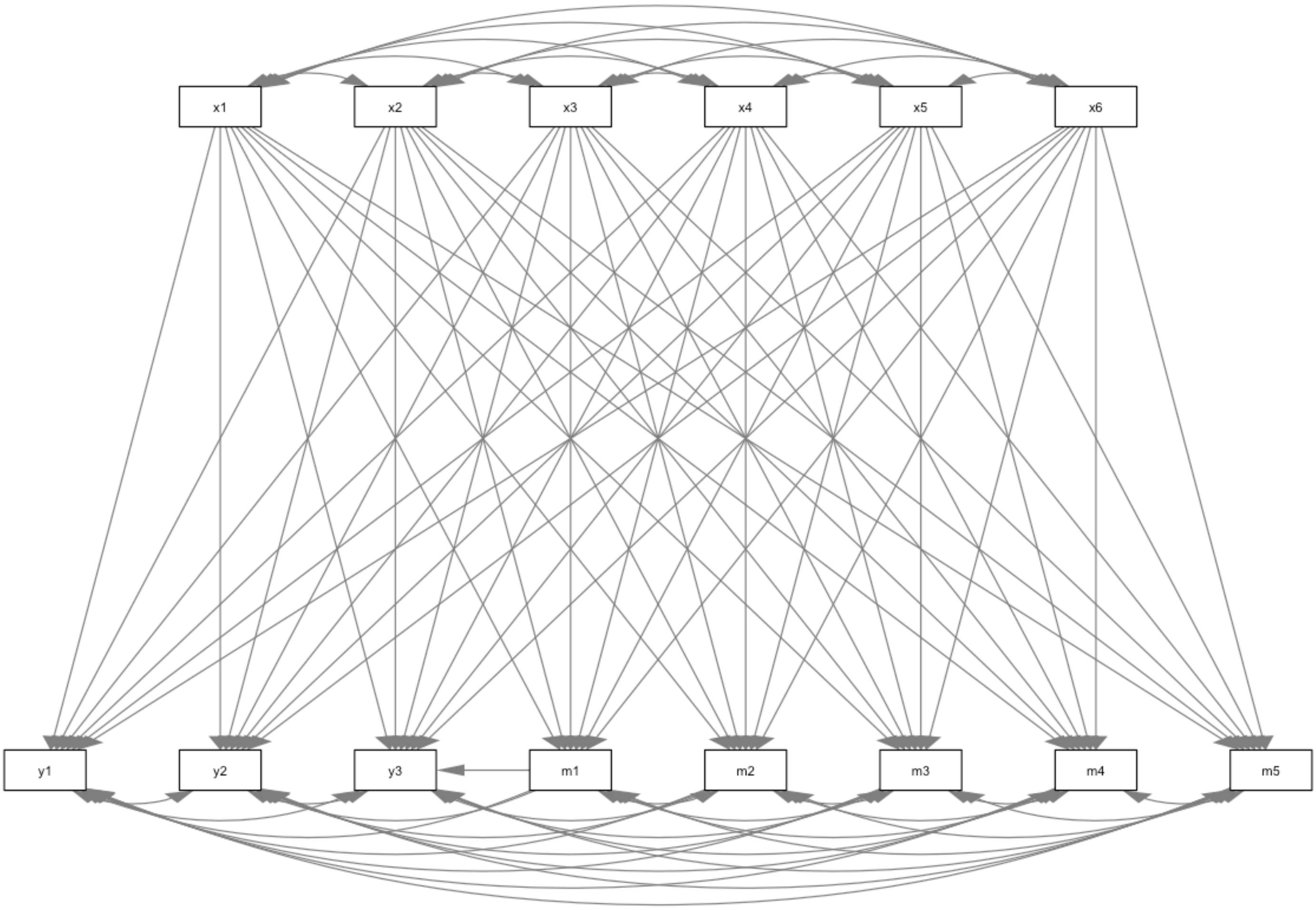
Since the exogenous construct, listening activity, was dummy-coded, one category had to be left out of the analysis as per the $K - 1$ rule (where $K = n$ groups) applied in classical regression (Fox, 2016) and recommended in the case of multicategory exogenous variables in SEM (Edwards et al., 2012). This presents a theoretical challenge as there is no clear reference group amongst the activity variables. It was decided, however, to use the category 'Routine Activities' ($n = 176$), comprising a broad range of reported everyday activities such as cooking, cleaning, showering, and getting dressed. This was selected as the reference group on the rationale that the everyday tasks contained within this category are generally more diverse and less specific than other activity categories. This category was therefore selected to be the reference against which the other categories, consisting of more specific tasks, would be compared. Whilst this solution is a qualitative assessment and the selection of reference categories is essentially an arbitrary decision in these settings, it was nevertheless considered that more specific tasks would yield meaningful interpretations of the relationships between the remaining listening activities in comparison to the reference group.

7.4.6.1 Model Estimation

To fit the model, the remaining six dummy coded activity variables and FML mediators were first regressed onto the components extracted from the Spotify audio features, fitting the *b* and *c'* paths of the path model. Next, the exogenous variables were regressed onto the mediator variables (fitting *a* paths), thus fitting all effects between exogenous and endogenous variables. Note that in the context of mediation, *DE* exclusively refers to *c*'; that is, the direct effect of an exogenous variable (*X*) on an outcome variable (*Y*) when controlling for the mediator(s) (*M*). Once these paths had been specified, the indirect effects could be defined as decomposition effects, which involves specifying each mediated interaction (e.g., $x_1 \rightarrow m_1 \rightarrow y_1$; the product of each given *a* and *b* path) as well as the total indirect effect (*TIE*), the sum of all specific indirect effects, which can then be used to calculate *TE* (*TE* = *TIE* + *DE*). The specified structural model is illustrated in Figure 11, whilst Appendix G provides the R syntax for this analysis.

Finally, and to account for the clustered data structure, *lavaan.survey* (Oberski, 2014) was again used to adjust standard errors, *p*-values, and confidence intervals according to the dependency in the data structure. As before, these adjustments were made using participant IDs as the clustering variable, thus addressing the violation of sampling independence to maintain the nominal Type I error rate by treating repeated measures as being nested within individuals (Oberski, 2014; McNeish & Harring, 2017; Bauer et al., 2020).

Figure 11 Estimated structural model



Note. *X* = Listening Activities, *M* = Listening Functions, *Y* = Audio Feature Components.

7.4.6.2 Results of Mediation Analysis

The mediation analysis examined how concurrent listening activities affected the content of listener's music selection both directly and indirectly via the five FML factors generated in Study 1, with Routine Activities serving as a reference category. A Robust Maximum Likelihood (MLM) estimator was used with cluster-robust standard errors (SEs), thus applying the design-based approach previously applied through the CFAs of the ESM and pooled datasets. Following the comprehensive discussion put forward by Agler and De Boeck (2017) on the interpretations and applications of mediation models, an 'effect-focused' approach (that is to say an interpretative approach that focuses on the individual effects between constructs, rather than the global model) is used to orientate the interpretation of any effects within this model.

Activities as predictors of FML

The regressions of the mediators onto the respective activity variables showed FML vary according to concurrent listening activities, thus providing evidence of significant *a* path effects in the mediation model. The results of the regressions of the five factors onto the exogenous predictors are shown in Table 24.

Table 24 Path a: Regressions between exogenous and mediator variables

	Exogenous Variable	Path Term	Parameter Estimate	*SE*	β	*p*	CI Lower	CI Upper
Identity and Social Bonding (m_1)	Work (x_1)	$a11$	.150	.103	.073	.147	-.053	.353
	Travel (x_2)	$a12$	.110	.107	.046	.306	-.100	.320
	Relaxation (x_3)	$a13$	.150	.112	.058	.180	-.069	.369
	Exercise (x_4)	$a14$	.364	.227	.065	.109	-.081	.809
	Socialising (x_5)	$a15$	1.459	.701	.137	**.037***	.085	2.832
	Recreational Activity (x_6)	$a16$	.141	.169	.043	.404	-.190	.471
Emotion Regulation (m_2)								
	Work (x_1)	$a21$	.160	.102	.077	.119	-.041	.360
	Travel (x_2)	$a22$	.438	.123	.183	**<.001*****	.197	.679
	Relaxation (x_3)	$a23$	.276	.122	.105	**.024***	.037	.515
	Exercise (x_4)	$a24$	.559	.238	.099	**.019***	.091	1.026

	Socialising (x_5)	*a*25	-.022	.369	-.002	.953	-.746	.702
	Recreational Activity (x_6)	*a*26	.123	.157	.037	.435	-.186	.431
Focus and Concentration (m_3)								
	Work (x_1)	*a*31	.857	.128	.419	**<.001*** **	.606	1.107
	Travel (x_2)	*a*32	-.033	.128	-.014	.797	-.285	.219
	Relaxation (x_3)	*a*33	-.253	.135	-.098	.061†	-.518	.012
	Exercise (x_4)	*a*34	.488	.213	.088	**.022***	.071	.906
	Socialising (x_5)	*a*35	-.478	.325	-.045	.141	-1.114	.158
	Recreational Activity (x_6)	*a*36	-.026	.149	-.008	.861	-.317	.265
Background and Accompaniment (m_4)								
	Work (x_1)	*a*41	-.122	.125	-.061	.327	-.366	.122
	Travel (x_2)	*a*42	.102	.133	.044	.444	-.158	.362
	Relaxation (x_3)	*a*43	-.209	.129	-.082	.106	-.461	.044
	Exercise (x_4)	*a*44	.104	.217	.019	.632	-.321	.528
	Socialising (x_5)	*a*45	-.690	.301	-.066	**.022***	-1.280	-.100
	Recreational Activity (x_6)	*a*46	.006	.232	.002	.979	-.448	.460

Physiological Arousal (m_5)								
	Work (x_1)	$a51$	.089	.106	.044	.398	-.118	.297
	Travel (x_2)	$a52$	.204	.127	.086	.109	-.045	.454
	Relaxation (x_3)	$a53$	-.138	.121	-.053	.253	-.374	.098
	Exercise (x_4)	$a54$	1.597	.372	.288	**<.001*****	.868	2.327
	Socialising (x_5)	$a55$	-.041	.369	-.004	.911	-.764	.681
	Recreational Activity (x_6)	$a56$	-.226	.152	-.070	.136	-.523	.071

Note. *SE* denotes standard errors. CI denotes confidence intervals. *$p \leq .05$, **$p \leq .01$, ***$p < .001$, † denotes non-significant trend. β = standardised parameter estimates.

It was seen that Socialising yielded significantly higher levels of *Identity and Social Bonding* than other activities (see *a*15), thus generating a close theoretical alignment between the activity Socialising and this FML factor. The confidence intervals for this were, however, wide as a likely consequence of the relatively small subset of observations within this category (n = 7). Thus, whilst this supports the hypothesis that Socialising elicits higher levels *Identity and Social Bonding* as a function, greater test power would be conducive to increasing the confidence of this overall and narrow the margins of error to around the point estimate. Travel (*a*22), Relaxation (*a*23), and Exercise (*a*24) were all observed to elicit higher levels of *Emotion Regulation*, thus implying that mood management is greater during these activities in comparison to the reference category. It makes substantive theoretical sense given that during each of these activities, mood management is important in affect regulation (e.g., Bull, 2006).

Work/Study (*a*31) led to higher levels of *Focus and Concentration*, as did Exercise (*a*34), albeit to a lesser degree, holding further consistency with wider literature in that work, private study, and exercise all require suitable background regulation to maintain concentration and goal congruency (e.g., Lamont et al., 2016). Additionally, Relaxation (*a*33) showed a non-significant trend in a negative direction, which may imply that the desire to focus and concentrate is lower during relaxation. Whilst this cannot be affirmed by the present study, it makes sense that when relaxing, listeners typically hold fewer cognitive demands requiring them to focus on a specific task. However, as the confidence intervals suggest, there is variability around the point estimate implying wide-ranging inconsistencies with regard to this measure. Finally, Socialising ($a45$) had a negative effect on *Background and Accompaniment*, and Exercise ($a54$) yielded higher levels of *Physiological Arousal*. Like the remaining $X \rightarrow M$ effects, this is consistent with theoretical considerations and expectations as social actions plausibly to not evoke feelings of loneliness and exercise obviously requires a desired level of physical motivation and stimulation.

Recreational Activities was the only exogenous variable not observed to influence any FML, with all others indicating a significant effect on at least one of the mediators. These results provide overall evidence that listeners' concurrent activities influence FML, however, there were no observable differences between the reference and recreational activity groups. Following this, the remainder of the model, which tests whether these effects subsequently influence the audio content of listener's selected music, can now be interpreted to explore direct and indirect effects.

<u>*Direct Effects on Audio Features*</u>

Next, the direct effects between the exogenous and outcome variables (c' path, also known as the direct effect) will be reported, as will the effects of the regressed paths of FML on the outcome variables (b paths). This is shown in Table 25.

Table 25 Paths b and c': Regressions between exogenous and mediator variables to audio features

	Exogenous/Mediator Variable	Path Term	Parameter Estimate	*SE*	β	*p*	CI Lower	CI Upper
Arousal (y_1)								
	Identity and Social Bonding (m_1)	$b11$	-.064	.049	-.061	.190	-.160	.032
	Emotion Regulation (m_2)	$b12$	-.031	.057	-.029	.594	-.143	.082

	Focus and Concentration (m_3)	*b*13	-.153	.063	-.145	**.015***	-.277	-.030
	Background and Accompaniment (m_4)	*b*14	.031	.062	.029	.616	-.090	.152
	Physiological Arousal (m_5)	*b*15	.176	.063	.166	**.005****	.053	.299
	Work (x_1)	*c*11	-.092	.118	-.043	.435	-.324	.139
	Travel (x_2)	*c*12	.041	.111	.016	.715	-.177	.258
	Relaxation (x_3)	*c*13	.031	.129	.011	.811	-.222	.284
	Exercise (x_4)	*c*14	.124	.213	.021	.560	-.293	.542
	Socialising (x_5)	*c*15	.157	.323	.014	.628	-.477	.791
	Recreational Activity (x_6)	*c*16	.121	.130	.035	.355	-.135	.376
Valence (y_2)								
	Identity and Social Bonding (m_1)	*b*21	-.012	.056	-.012	.824	-.122	.097
	Emotion Regulation (m_2)	*b*22	-.082	.057	-.078	.154	-.194	.030
	Focus and Concentration (m_3)	*b*23	-.116	.054	-.110	**.033***	-.222	-.009
	Background and Accompaniment (m_4)	*b*24	.149	.050	.139	**.003****	.050	.248
	Physiological Arousal (m_5)	*b*25	.164	.075	.155	**.029***	.017	.311

	Work (x_1)	*c*21	-.131	.111	-.061	.236	-.348	.086
	Travel (x_2)	*c*22	-.239	.115	-.095	**.037***	-.464	-.014
	Relaxation (x_3)	*c*23	-.207	.130	-.076	.112	-.462	.048
	Exercise (x_4)	*c*24	-.261	.240	-.044	.276	-.731	.209
	Socialising (x_5)	*c*25	.085	.337	.008	.800	-.575	.746
	Recreational Activity (x_6)	*c*26	.060	.149	.017	.688	-.233	.352
Instrumentalness (y_3)	Identity and Social Bonding (m_1)	*b*31	.049	.057	.047	.392	-.063	.162
	Emotion Regulation (m_2)	*b*32	-.019	.052	-.018	.721	-.120	.083
	Focus and Concentration (m_3)	*b*33	.057	.062	.054	.358	-.064	.178
	Background and Accompaniment (m_4)	*b*34	-.165	.058	-.153	**.004****	-.277	-.052
	Physiological Arousal (m_5)	*b*35	-.084	.062	-.080	.174	-.206	.037
	Work (x_1)	*c*31	.253	.112	.117	**.024***	.033	.473
	Travel (x_2)	*c*32	-.033	.100	-.013	.738	-.229	.162
	Relaxation (x_3)	*c*33	.093	.099	.034	.348	-.101	.288

Exercise (x_4)	*c*34	.275	.223	.047	.218	-.162	.711
Socialising (x_5)	*c*35	-.556	.280	-.050	**.047***	-1.105	-.006
Recreational Activity (x_6)	*c*36	.060	.146	.018	.679	-.225	.345

Note. *SE* denotes standard errors. *$p \leq .05$, **$p \leq .01$, ***$p < .001$, † denotes non-significant trend. β = standardised parameter estimates.

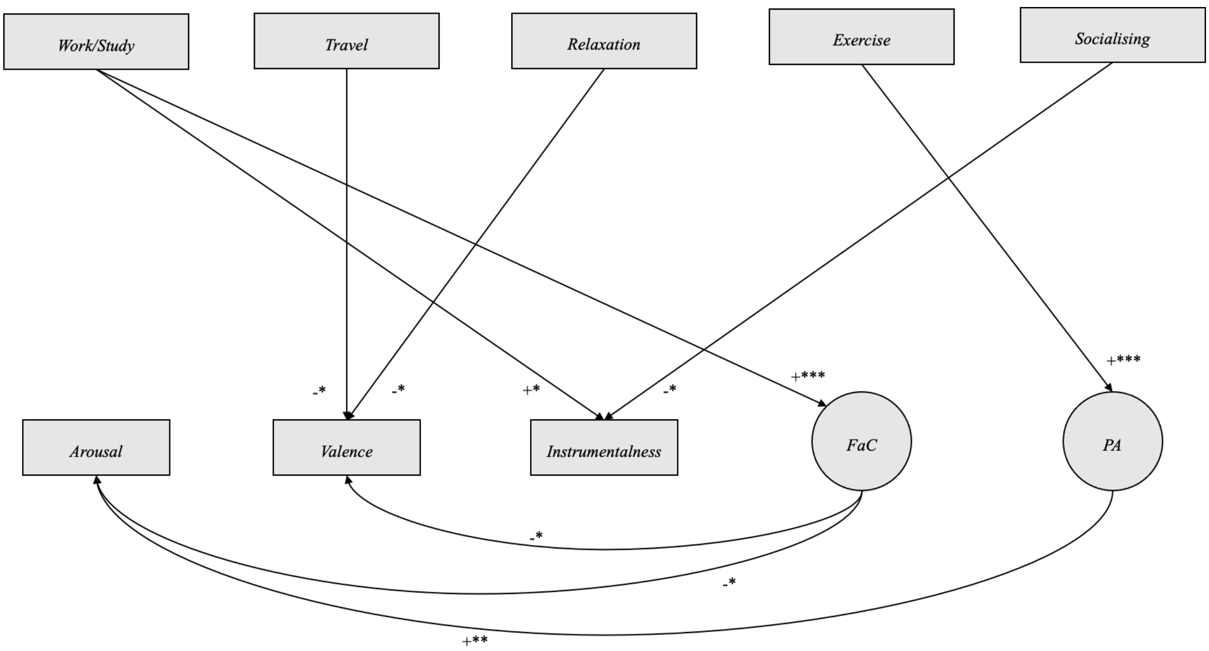
Table 25 effectively reports the direct effects of the exogenous (*X*) and mediator (*M*) variables on the outcome variables (*Y*). In the mediation model these are the *b* and *c'* paths, thus indicating the effect of each preceding variable in the structure on the three Spotify feature components. Regarding direct effects between *X* and *Y* variables, it was observed that Travel had a significant negative effect on *Valence* (*c*22), whilst Work/Study (*c*31) and Socialising had significant positive and negative effects respectively on *Instrumentalness* (*c*35). Regarding *M* → *Y* regressions, *Focus and Concentration* had significant negative effects on *Arousal* (*b*13) and *Valence* (*b*23), whilst *Physiological Arousal* had a statistically significant positive effect (*b*15; *b*25). *Background and Accompaniment* also had a significant positive effect on *Valence* (*b*24), and a negative effect on *Instrumentalness* (*b*34). Of the remaining factors, neither *Identity and Social Bonding* nor *Emotion Regulation* were observed to have any effect on the *Y* variables.

<u>*Relationships between Activities and Audio Content*</u>

Given the scale of the specified model, it would not be helpful to the reader to set out all relationships in a series of tables with full results here. Rather, a model of each significant path from *X* leading to a change in *Y* will be presented in a series of diagrams, which will then be discussed accordingly by the modes of their relationship (i.e., *DE*, *IE*, *TE*). A full table of results is available in Appendix H. In the meantime, however, the following models illustrate significant paths between the focal constructs. Additional non-significant trends have been included where relevant (e.g., where one path is significant and an additional is non-significant). These present a combination of paths, each of which is calculated and interpreted differently (these will be discussed later on). All other effects not reported in-text should be assumed non-significant.

First, to summarise effects representing changes in *Y* in a single model, Figure 12 illustrates all significant decomposition effects. This does not illustrate all significant *a* or *b* paths as shown in Tables 24 and 25, rather those paths leading from *X* that led to changes in a *Y* variable. Subsequent figures illustrate individual relationships in more precise terms to assist with interpretability and contextualise the manner of these effects. Although these decompose effects to one outcome variable at a time, readers should note that all effects constitute part of the same model and that this manner of presentation is intended to ease interpretability, rather than indicative of the specification of sequential/multiple smaller models.

Figure 12 Summary diagram of changes in Y leading from X



Note. *$p \leq .05$, **$p \leq .01$, ***$p < .001$. + or – indicates direction of effect. *FaC = Focus and Concentration. PA = Physiological Arousal.*

Figure 12 summarises the decomposition effects of the paths in the model with regard to *Y*. The following figures decompose these further. In each case, at least one of *IE*, *DE*, or *TE* is statistically significant. In cases where no mediator was found to have a significant effect on the relationship between the exogenous and outcome variables (i.e., *IE*), these have been left out entirely. In all cases, however, *DE* and *TE* are reported, with a discussion of their interpretations ensuing. First, Figures 13 and 14 show the indirect effects leading from the Work/Study and Exercise categories to changes in the *Arousal* in listeners' selected music.

Figure 13 Work/Study → Arousal

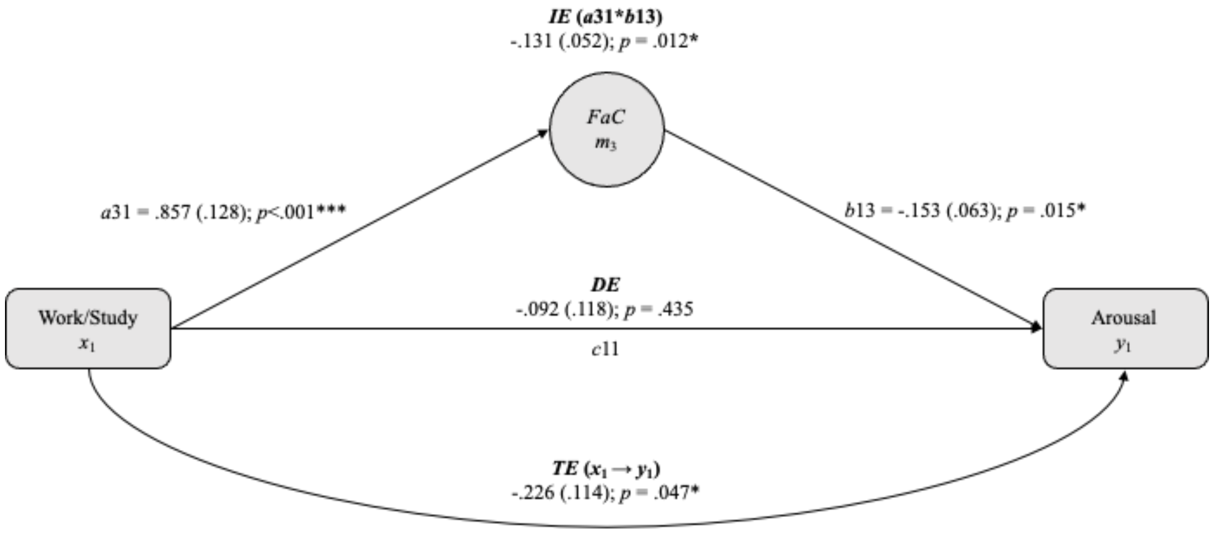
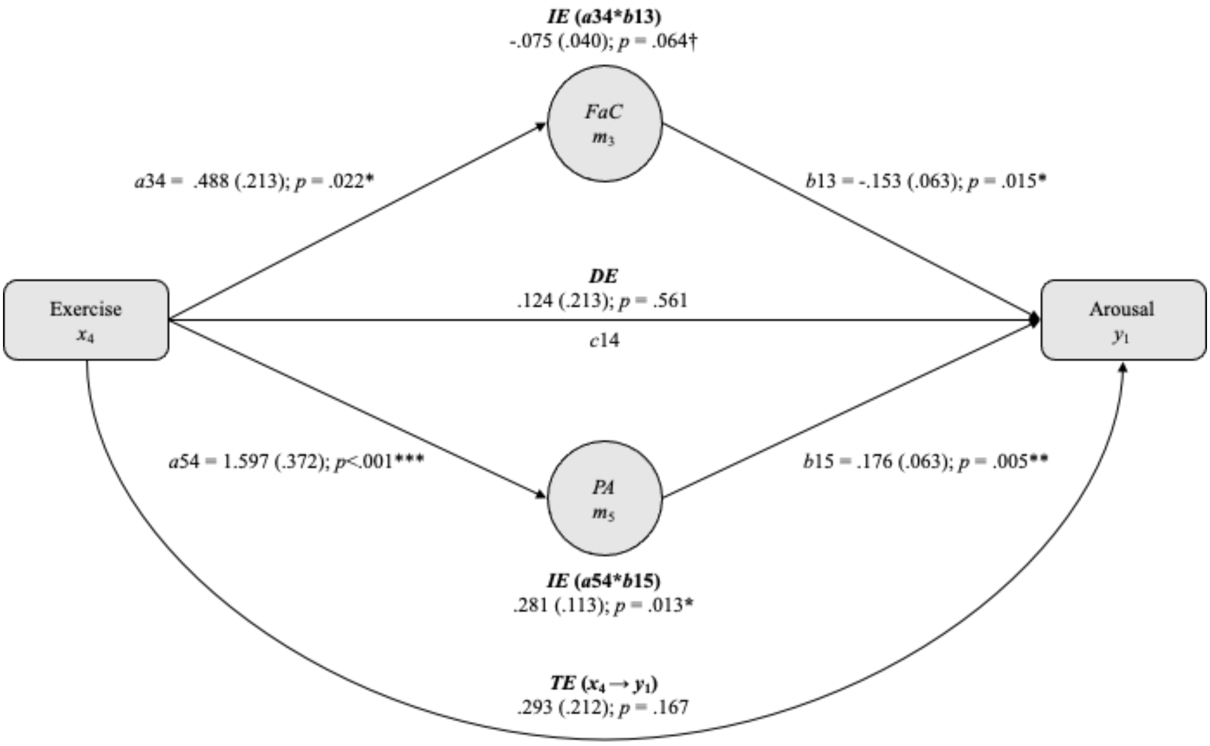
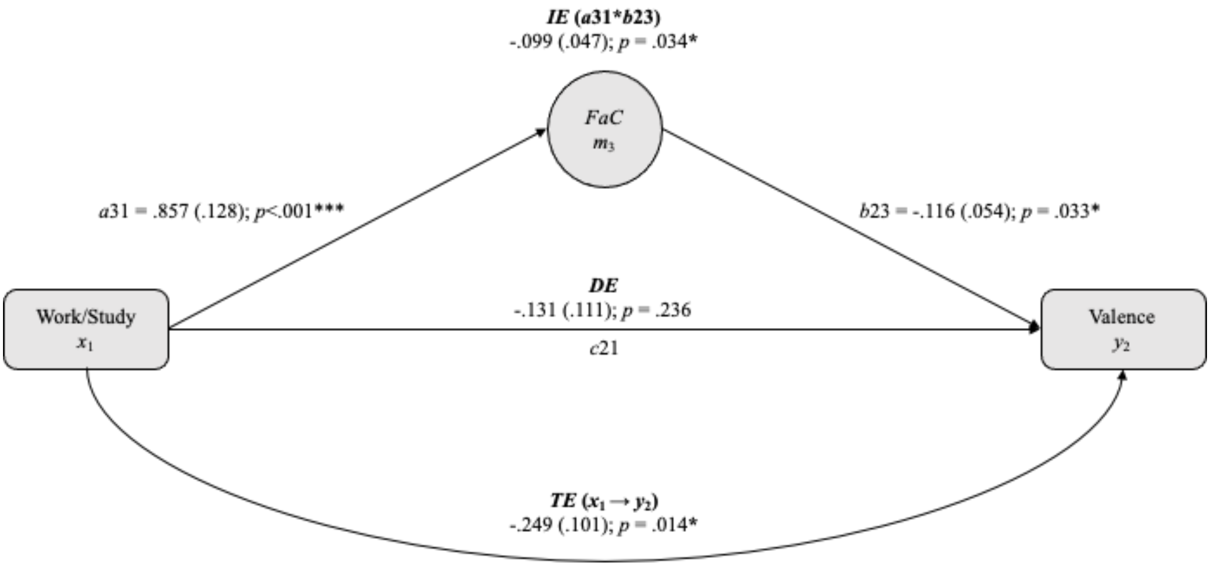


Figure 14 Exercise → Arousal



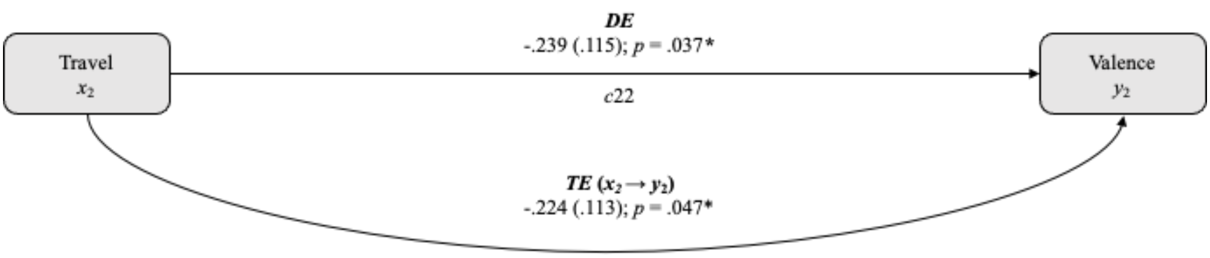
It can be seen that *Focus and Concentration* mediates a relationship between Work/Study and *Arousal*. This was the only significant *IE* between these variables which, coupled with a non-significant *DE*, indicates that the relationship between these two variables is fully mediated by *Focus and Concentration*. Thus, Work/Study elicits higher levels of *Focus and Concentration*, which in turn results in lower levels of *Arousal*. It was found that *Physiological Arousal* mediates a relationship between Exercise and *Arousal*. This indicates that during Exercise, *Physiological Arousal* increases, which in turn leads to higher levels of *Arousal* in the audio content. As with the previously mentioned effect, the lack of a significant *DE* suggests this relationship is fully mediated. An additional non-significant trend is also highlighted, the effect of Exercise on *Arousal* via *Focus and Concentration*. The reason for this is that although the *IE* is non-significant, the respective *a* and *b* paths are. It is interesting to consider whether despite Exercise leading to both increased *Focus and Concentration* and *Physiological Arousal*, the respective outcome on *Arousal* pulls in opposite directions.

Figure 15 Work/Study → Valence



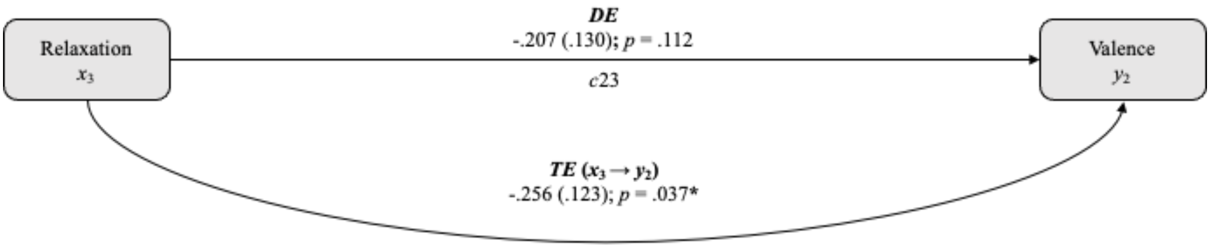
As in Figure 13, Figure 15 shows an indirect effect was observed between Work/Study and the outcome variable, in this case *Valence*, where *Focus and Concentration* again fully mediates the relationship subsequently leading to a reduction in *Valence*. As before, this was complete mediation in the absence of a significant *DE*, indicating it is via *Focus and Concentration* that Work/Study leads to lower *Valence*.

Figure 16 Travel → Valence



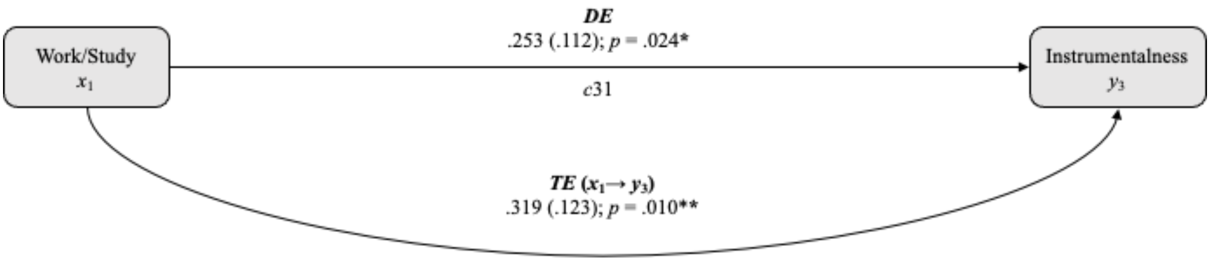
Additional effects included a significant *DE* between Travel and *Valence*, illustrated in Figure 16. The implication of this is that no mediator plays any role in this relationship, and that it is as a direct consequence of Travel that music lower in *Valence* is selected. It is interesting to note that despite Travel influencing the *Emotion Regulation* factor, the relevant mediator had no influence on *Valence*. This is in contrast with the prior models that have thus far indicated full mediation, providing evidence that there may be some mixed cases of *IE*s and *DE*s that lead to changes in audio content. Moreover, it is indicative that whilst certain activities may predispose mood regulation as a FML, that the measure is unable to ascertain as to whether this leads to positive or negative outcomes (at least as a function of an indirect effect).

Figure 17 Relaxation → Valence



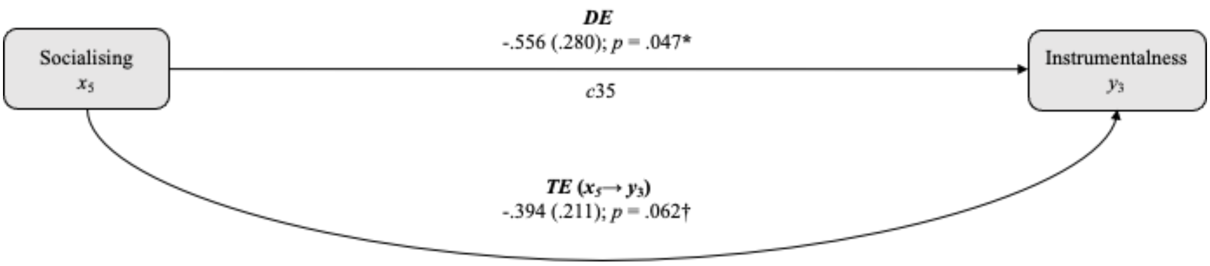
As Figure 17 shows, Relaxation was also found to have a significant *TE* on *Valence*, again in a negative direction. However, in this case all specific *IE*s and the *DE* were non-significant, making the interpretation of this effect difficult. There is seemingly little to no guidance in the literature on how to interpret this circumstance, however, given the negative point estimate of the *DE* and (despite being non-significant) a *p* value well below 1, there is likely some low-level effect influencing *Valence* during cases of Relaxation. General point estimates of *IE*s are also negative and hold *p* values in a similar range as the *DE* (see Appendix H). It therefore seems that there is an overall negative effect of Relaxation on *Valence*, but that from the present structure it is not possible to partition this amongst the specific effects. It may be some combination or further interactions that lead to the cumulative observed outcome, but the interpretation of this remains somewhat constrained.

Figure 18 Work/Study → Instrumentalness



As shown in Figure 18, Work/Study was found to have a significant positive *DE* and *TE* on *Instrumentalness*. The lack of any significant *IE*s, however, shows that in this instance, there is no mediated effect unlike with the prior two components between Work/Study and the outcome variable, and that the observed effect is a direct consequence of the listening activity, rather than through any FML. This is consistent with the idea that music with lyrics is less beneficial during work than music that is instrumental due to limited cognitive processing ability (e.g., Konečni, 1982), but it was interesting to note *Focus and Concentration* seemingly played no role in eliciting lower levels of *Instrumentalness*.

Figure 19 Socialising → Instrumentalness



Finally, Figure 19 shows the *DE* of Socialising on *Instrumentalness* was also significant, and in a negative direction. This again implies that none of the hypothesised mediators played a role in this relationship. The *TE* indicated a non-significant trend which, given the relatively low number of observations for this category, may be a result of limited power given the sum of *TIE* and *DE*. Nonetheless, it is in the same direction as the *DE* and so it seems probable that there is an overall negative effect on the levels of *Instrumentalness* stemming from Socialising.

7.5 Discussion

This study sought to assess the temporal relationship of three focal constructs: activities concurrent with music listening, FML, and MIR generated audio features extracted from the Spotify API (following dimension reduction). A multi-arm study design was used in which an initial online survey gathered data on a large pool of participant's most recent listening episodes, whilst an ESM was then used to gather further observations of participants over an extended time period. Each response between the study arms was treated as a cross-sectional observation of listeners' most recent autonomous listening episode, and as such, these two sets of observations were pooled to maximise statistical power. Observations were filtered prior to primary analyses to ensure theoretical consistency with regard to the assessed measures, retaining 875 cases of the original pool of observations in the final dataset, which were partially nested within 347 participants.

***RQ1:* Can the previously identified FML measure be cross validated in ecologically valid data of everyday listening episodes?**

To address the first RQ of this study, iterative CFAs were used to assess whether the 23-item psychometric structure generated during Study 1 was a good fit for the observed data in each study arm respectively, as well as in the pooled data. Robust Maximum Likelihood (MLM) was applied in each instance, and design-based adjustments were made when the assumption of test independence was violated (i.e., repeated assessments within participants; McNeish & Harring, 2017). Segmented CFAs indicated the model was a good fit for the data gathered in each surveying method. It should be noted, however, that both assessments of internal reliability and model fit were stronger for the initial survey responses than the ESM responses (though adequate nonetheless). This could be due to some inherent differences between the implementation of the study arms as afforded by different research tools.

A third CFA pooled the sum sets of cross-sectional data together, using cluster-robust standard errors to accommodate the partially clustered data structure, and indicated that the previously presented 23-item structure was a good fit for the observed data, which additionally satisfied described conditions relating construct reliability and validity. This provides cross-validation of the five-factor model identified in Study 1 as a measure of FML and is of substantive importance to this project as a whole. There remain, however, opportunities for further use, such as other in cross-cultural studies, as both studies 1 and 2 have been conducted with largely Western samples. As such, although the structure appears valid, further use and extension would corroborate this or highlight shortcomings in more diverse samples and/or settings.

***RQ2:* To what extent do MIR-generated audio features correspond to listeners' perception of audio content?**

Bivariate correlation analyses indicated that, in general, MIR-generated audio features extracted from the Spotify API were correlated with comparable measures of listeners' self-perceived audio features, such as that music perceived by listeners as being more 'exciting' is positively correlated with API measures like *Energy* (see section 7.4.4). These results provide confidence that further analyses employing these measures are roughly reflective of users' perceptions of musical characteristics. Again, these may be limited insofar as these features do not all share obvious theoretical counterparts with the self-perceived measure applied by Greb et al. (2019), however, the open access to the Spotify API as a means of gathering audio content might be an area that other researchers could utilise to assess how audio content varies alongside a multitude of other research contexts. This mitigates cognitive or perceptual biases in participant responses also, generating a more consistent measure of audio content than may be gathered from self-reports alone. It should be acknowledged that these measures hold limitations, however, such as their own calculation is not in the public domain (Maloney et al., 2021). Nonetheless, this does mitigate certain perceptual biases within listeners' assessments of audio content by utilising a more objective, consistent measure of audio content. Indeed, the application of such audio-features in real-world listening technologies further provides the basis of their use since these are applied in music curation in practice.

***RQ3:* Do activities concurrent to music listening lead to changes in the audio content of music selected by listeners directly or indirectly via FML?**

Of the Spotify-generated audio features, six were selected for analysis on grounds of theoretical relevance and the subsummation of acoustic features into perceptual ones (e.g., measures such as *Tempo* and *Loudness* sharing cross-over with *Energy*). It was these six audio features, represented through 0-1 scaled *float* scores, that were selected to serve as outcome variables in the proposed mediation analysis. Since there were a large number of outcome variables conceptually, it was decided to simplify the fitted model by reducing these through PCA. This yielded three components retaining five of the six initial features: *Arousal*, *Valence*, and *Instrumentalness*. These were utilised as outcome variables in the mediation analysis, thus simplifying the model whilst encapsulating optimal variance of the original variables (explained variance = 87.28%). The component *Arousal* indicated high levels of *Energy* and low levels of *Acousticness*, whilst *Valence* implied high levels of *Danceability* and *Valence*, and *Instrumentalness* consistent with higher levels of its namesake. Following this, a mediation analysis using a SEM framework was conducted.

The aim of the mediation analysis was to assess the temporal structure of the three focal constructs to test the general hypothesis that concurrent listening activities lead to observable changes in the audio features of users' music selection, either directly or indirectly via FML (integrated via factor scores extracted from the pooled CFA iteration). This was fit using a SEM framework, which is generally deemed preferable to regression paradigms (e.g., Iacobucci, 2008), however, model fit statistics could not be calculated since the model was saturated (i.e., no *df* left in the data; Raykov et al., 2013a; 2013b). Therefore, an effects focused approach was taken, in which relevant model pathways were explored to uncover effects occurring between the three constructs (e.g., Agler & De Boeck, 2017).

As for the paths in the model, a large number of *IE*s were estimated, however, relatively few of these were statistically significant. Nonetheless, these results do show a number of theoretically meaningful relationships between the exogenous variables and reduced components of the Spotify audio features. Namely, these were *IE*s between *Focus and*

Concentration and Work/Study on both *Arousal* and *Valence*, which were found to be negative. The lack of a *DE* in both instances, which in the presence of significant *TE*, implies fully mediated relationships (Kline, 2016). In addition, a significant *DE* and *TE* was observed between Work/Study and *Instrumentalness*, with no significant indirect paths. This implies that there exists a *DE* of Work/Study on music selection that elicits higher levels of *Instrumentalness*, with no mediating relationship. Overall, lower levels of *Arousal* and *Valence*, and higher levels of *Instrumentalness*, are consistent with findings that when working, listeners seek music that prevents distraction and enables concentration by optimising cognitive arousal (Haake, 2011; Greasley & Lamont, 2011).

Physiological Arousal was found to mediate a positive relationship between Exercise and *Arousal*. An additional negative non-significant trend was observed on the mediated effect of *Focus and Concentration* between Exercise and *Arousal*. This is highlighted partly because the *TE* was not statistically significant, which may imply that the effect of the two *IE*s may nullify the *TE* (i.e., because they pull in opposite directions). It may, for example, be the case that there is something of a dichotomy, whereby some forms of exercise require *Focus and Concentration* and others *Physiological Arousal* (for instance, consider the difference between yoga and jogging). Alternatively, it may be the case that whilst Exercise influences multiple dimensions of *functionality*, it is the notably more substantial effect of higher *Physiological Arousal* that leads to greater *Arousal* in the music content. Since the null hypothesis could not be rejected in the case of *Focus and Concentration*, the interpretability of this is limited; however, the presence of significant *a* and *b* paths between Exercise and *Focus and Concentration*, and in turn between *Focus and Concentration* and *Arousal*, does call for further exploration. In any case, the indirect effect of *Physiological Arousal* is nonetheless consistent with literature, such as that referring to the use of music during exercise to feel stimulated and motivated (Hallett & Lamont, 2015; Lamont et al., 2016). This specific effect also corroborates that Exercise influences the theoretically aligned FML (i.e., *Physiological Arousal*), which in turn affects the audio content. Further stratification of types of Exercise may enable further exploration of these differences, however, the means by which to carry out this is limited with the current data which holds significantly less observations for low-arousal exercise activities (e.g., yoga). In any case, future research may wish to explore these relationships further to better understand how different kinds of physical activity elicit different levels of audio content via relative FML.

Travel was observed to have a negative *DE* on *Valence*. It appeared, therefore, that during Travel, the audio content of the music tends to hold lower levels of *Valence* than the reference category, however, FML did not play a role in this. This is also true of the relationship between Socialising and *Instrumentalness*, whereby the degree of *Instrumentalness* was observed to be lower. Conversely, Relaxation was found to have a significant *TE* on *Valence*, again in a negative direction, however, no specific *IE*s or the *DE* were significant. It therefore seems that it may well be the case that during Relaxation listeners engaged with music that is lower in *Valence* overall, yet this effect could not be partitioned between any of the specific effects.

This is interesting since although the predictor-mediator paths found that Relaxation led to higher levels of *Emotion Regulation*, this did not translate in turn to changes in audio content. Yet, there were lower levels of *Valence* in terms of audio content overall. Others have indeed argued that listeners engage with diverse music types during relaxation and have shown little in the way of music preference (North & Hargreaves, 2000), and so perhaps other factors not measured in this study, such as present mood states, may also play a role in terms of determining music selection, but this is difficult to assess with the present data. It may be the case, however, that whilst situational factors influence uptake in *Emotion Regulation*, the predictors of music selection here are not determined by *functionality* so much as by personal characteristics (e.g., episodic memories, present mood states, desired mood states). This may in part be down to certain limitations in the applied measure, whereby it seemingly is able to detect the presence of mood regulation as an FML, but that in being of more limited nuance that broader regulatory measures (e.g., AFML, MMR), it does not fully encapsulate all dimensions of mood regulation. It would be interesting, therefore, to explore whether more nuanced measures of mood regulation (Saarikallio, 2008; Groarke & Hogan, 2018) are able to detect more tangible changes in audio characteristics, as the present measure rather subsumes mood regulation in more general utilitarian terms. This would enable more fine-grained comparison of regulatory techniques, however, may constrain the ability to explore other, more varied dimensions of *functionality*. In any case, this should be acknowledged as a potential drawback of a broader utilitarian model.

The paths identified make general theoretical sense and are aligned with observations that FML may mediate relationships between contextual variables and music selection (measured via audio content). All bar one of the activity variables, Recreational Activities, was seen to influence FML and/or at least one measure of audio content. However, although exogenous variables were observed to affect relevant mediators, these relationships rarely influenced outcome variables indirectly, and when it did, was limited to a small number of activities. The implication that these factors do not lead to observable changes in *Y* is interesting as it perhaps reflects that although FML may be prompted by contextual variables, they do not always lead to observable changes in selected audio content. There may be several reasons for this, some theoretical, some methodical.

This study specifically put hierarchical individual variables, such as music preference, to one side. This was because the aim was to exclusively see how the activity affects listeners' FML, which in turn modulates audio content of selected music. This is to remain consistent with the previously discussed separation of personality and affect-aware recommendation approaches to psychology-informed recommendations (Lex et al., 2021). Individual variation may still play a role, however, in the relationship between utility and music selection. By implication, this indicates that some activities directly influence the audio content of the music that will be selected by listeners, whilst other activities are less dependent on specific content, or where the situational characteristics are of reduced importance. This is where personal preferences may be influential, since although listeners may well listen to ease *Identity and Social Bonding* in social situations, the content of the music itself may depend on additional factors, such as those present, rather than any specific audio content. In other words, although listening activities do appear to influence music's function overall, not all of these functions require specific audio content across listeners. Rather, during Travel, for example, although *Emotion Regulation* was found to be positively affected, this did not translate to meaningful changes in audio features. This may be because music selected in certain scenarios is plausibly not so much dependent on the content of the music, but its meaning to or on the listener (e.g., episodic memories, personal taste). This may be somewhat specific to certain FML, as for example, people tend to listen to music that has deeply personal meanings during mood regulation (e.g., Gibbs & Egermann, 2021). In other words, the music selected may be chosen as it is enjoyed or liked by the listener, and that the content of the music is somewhat secondary. Rather than selecting music that has certain perceptual features, it may be the case that listeners select music according to its specific meaning to them in certain situations, for which audio content is diverse and not necessarily directly influenced by the situation.

In short, some activities may directly and/or indirectly influence the audio content of music selected. This adds to evidence that contextual variables influence FML, and also that these constructs may ultimately influence the music that listeners choose to engage with during everyday listening (Greb et al., 2019). However, this study's findings suggest that this may be the case for some but not all activities. In part, this may be due to some limitations in the data (e.g., unequal group sizes in the discrete predictors), but it also leaves open the possibility that other/additional contextual or cognitive factors may play an additional role in determining music selection in certain contexts. Given that the aim of this study was to test a conceptualised method, however, there are several important outcomes. Firstly, most listening activities were not found to affect listeners' subsequent music selections, however, there are several specific examples of this. Primarily, Work/Study was found to influence all three components of audio content in some way, yielding lower levels of *Arousal* and *Valence*, and higher levels of *Instrumentalness*. The former relationships were mediated by *Focus and Concentration*, specifically implying that higher levels of this function during Work/Study leads to these outcomes. This is consistent with a long-held view that alongside work in particular, desirable levels of cognitive stimulation may be sought, to balance between preventing distraction and eliciting a sense of flow (Konečni, 1982; Haake, 2011; Lamont et al., 2016).

Exercise was observed to lead to higher levels of *Arousal* via *Physiological Arousal*. This is again consistent with literature insofar as music may be employed alongside physical activities in order to elicit stimulation, for which music with higher levels of *Energy* in particular makes theoretical sense. Thus, there is support for mediated relationships amongst the three focal constructs, although this is limited in its ubiquity. This is hypothesised to be the case since although other activities largely led to some indication of effect on *functionality*, this did not translate to changes in audio features.

In sum, the findings of this study provide limited but meaningful evidence of a both direct and indirect relationship between music and selected audio content. However, it is hypothesised that under *some* circumstances, this relationship is crucial, whereas under others it is much less so, with listener's individual variations reducing ubiquity of selected content under the same circumstances. Synthesising this into this thesis at large, it may be the case that for a limited subset of listening activities, it is possible to calibrate content-based recommender systems to concurrent activities, but that under other circumstances, this may need to be preceded by further exploration of higher-level constructs, to better understand how these factors influence music selection alongside activities. Further research will therefore be needed to expand on and synthesise more refined ways of assessing this problem of prototypicality, however, it is hoped that this study has conveyed a first step in doing so, for which there is now basis to consider how MIR generated audio features can be directly linked to music selection under certain circumstances.

7.5.1 Limitations

As with all research, there were limitations in this study, some of which has been touched on in the prior section. The use of dummy coded Activity variables was limited insofar as this requires the selection of a reference category, which in the absence of a clear control group, is essentially an arbitrary decision. Whilst the deductively coded activity group Routine Activities was considered to constitute a comparatively broad set of listening activities less specific than others, this selection is the consequence of a qualitative decision based on the researcher's interpretation of that category's content. This is therefore subject to two qualitative decisions: (1) thematic coding of activity variables according to responses, and (2) identifying a suitable reference category based on those groups. Though systematic and informed by prior research, alternative approaches are available, such as NLP as employed by Pichl et al. (2015) to identify activity categories of existing Spotify playlists. In suitably large samples, such approaches may provide an alternative means of categorising written strings, enabling the continued use of qualitative elements where possible. Alternatively, contrast or effect coding, whereby arbitrary reference categories need not be selected, could provide an alternative option in future work (te Grotenhuis et al., 2017).

The use of m-ESM holds limitations because younger individuals are more likely to engage with online listening studies using ESM than older participants (Krause et al., 2014). This may be a general limitation with accessibility to the tools used for the study, which may in turn lead to a high proportion of users' already comfortable with using smartphones, which may hypothetically bias perspectives towards users of streaming services in particular. Since formats and modes of accessing music may be tied to the goals/aims of the listener (Brown & Krause, 2020), this may further bias towards listeners in situations where engagement via portable technology is more expedient that other formats potentially favoured by other users, although this is speculative.

Additionally, gathering participants' self-reports of specific pieces of music is not without its drawbacks. It is almost certainly the case that most listening episodes involve more than just one piece of music, and as such requesting participants to name one track may reduce the nuance of selected music across the full listening episode in which a broad array of pieces may have been selected. This is important to acknowledge as the audio features could therefore be considered indicative of the music the listener named, rather than a representation of the entire episode. With that said, this does not necessarily invalidate the utility of the reported pieces themselves, since it also seems plausible that listeners named tracks they remembered most clearly listening to, and which were therefore most strongly associated with, or reflective of, the listening experience in question. Methodologically speaking, it may not be practical to request participants to name each and every piece they listen to, and so this trade-off may just be the consequence of taking a more pragmatic approach. Indeed, this is not inconsistent within the literature as others too have also requested participants report the features from individual pieces of music (e.g., Greb et al., 2019).

It was not surprising to find that certain listening activities were more prevalent than others in cross-sectional observations. It was not possible to assess, however, how these would be distributed prior to data collection, and, as such, it was unknown how large group sizes would be. Certain activities carried a large number of observations in the sample, with the extremes being Work/Study (n = 272) and Socialising (n = 7). These imbalances in the sample sizes of observations may yield underpowered results for certain categories despite the large sample size. As such, this limitation is accordingly acknowledged, and it is suggested that in future a stratified sampling approach may be useful in data collection. In this, researchers may find it beneficial to stage sampling to gather (at least roughly) equal observations for each category to mitigate these issues. However, this would have to overcome two issues: (1) maintaining ecological validity by not forcing such cases to occur through study design, and (2) the time and resources that would be needed for this process.

Finally, and as with Study 1, the socio-economic context of the sample cannot be reasonably extrapolated to other cultures. This is especially true given the dependence of this study and its results on online modes of access, particularly including the availability of a smartphone in the case of the ESM study arm. It would be interesting to consider how and whether it would be possible to repeat this study or similar in different populations to explore how and why people listen to music in different cultures in the stream of everyday life.

7.5.2 Conclusion

This study has found a limited but meaningful number of mediated relationships between everyday listening activities and the audio features of listeners' music selection, via FML. It was found that for five of the six identified listening activities, statistically and theoretically meaningful influences on FML were observed. However, the extent to which these mediated a relationship to a reduced set of principal components encapsulating audio features was more limited. Additional meaningful *DE*s provided evidence of further context-based effects about the content that listeners select alongside everyday activities. Overall, only two activities demonstrated the presence of significant indirect effects, Work/Study and Exercise. In all of these instances, their *DE* on *Y* variables were non-significant, thus implying fully mediated relationships. Further *DE*s were observed between Travel and *Valence* and Socialising and *Instrumentalness*, demonstrating an absence of mediation but an effect nonetheless, whilst a significant *TE* was observed between Relaxation and *Valence*.

Importantly, these results provide evidence that FML are indeed affected by the context of the listening episode. However, only under certain activities do these functions practically mediate the relationship between that context and the audio content of users' music selection. In other instances, these functions do not mediate relationships, but rather activities directly influence differences in the audio features observed in subsequent music selection. These mixed results articulate the presence of mediation in some circumstances, but not all. This supports the general hypothesis that activities concurrent with music listening will lead to observable changes in the audio content of music selection, but that depending on the context, these may be direct or indirect. This helps to address the second and third aims of this thesis also, by providing a triangulated structure associating the three constructs of interest empirically.

In conclusion, although a limited number of mediated relationships have been observed, the applied method has found evidence that FML, assessed through a model generated in Study 1, mediates relationships between concurrent listening activities and the audio content of music selection, as represented by audio features gathered from the Spotify API. Additional research may benefit from applying a stratified sampling method to provide more equitable group sizes across activities which may in turn improve test power across the model. Since the aim of this thesis at heart is to propose an approach to generating psychology-informed CAMRSs, an understanding of situational demands and their relationship to audio-content itself acts as a useful basis and is argued to be a meaningful first step in triangulating the key variables identified. Although the results are subject to limitations and are by no means exhaustive, it is hoped nonetheless that the rationale and results provide meaningful evidence that methods from the social sciences can represent theoretically consistent understandings of the ways in which music is engaged with in everyday life according to contextual variables, with implications for parameters of selected audio features within those activity frameworks.

8.1 Study 3: Design and implementation of a psychology-informed recommendation procedure

Earlier in this thesis, it was argued that activities occurring alongside music listening lead to changes in the audio content of the music listeners select, directly and/or indirectly via FML (see Chapter 5). It was also argued that an empirical understanding of the nature of these relationships could be used to help inform an approach to generating psychology-informed CAMRSs. To this end, the two studies presented in Chapters 6 and 7 respectively have (1) provided a utilitarian measure of FML and (2) assessed the hypothesised relationships between the core constructs of interest through path modelling (mediation analysis). With regard to the second study, the use of MIR-generated audio features (obtained from the Spotify API) provides an opportunity to directly apply knowledge obtained during the study and from the data, since the measures used are compatible with other uses of the API tool. In other words, this provides an opportunity to fulfil the fourth and final aim of this thesis as laid out in section 1.2, which was to be able to propose an actionable method of integrating knowledge generated obtained in preceding phases into a recommendation procedure. To reiterate the motivation behind this, there is a drive in the literature to mitigate existing dependencies on data-driven approaches in the development of systems through black-box machine learning techniques (Lex et al., 2021), as well as provide users with recommendations based on short-term (i.e., situationally determined) needs, rather than solely long-term representations of taste (Wang et al., 2012).

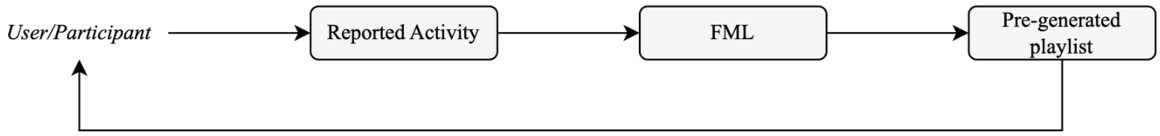
Explanatory models may help provide targeted recommendations for given situations without being dependent on the ability of algorithms to detect low-level linearities in large volumes of data. Though the predictive accuracy of explanatory models is, methodologically speaking, an area that requires development in the long-term (e.g., Yarkoni & Westfall, 2017), there nonetheless remains scope to explore approaches to providing recommendations that future work may build on. To be clear, this will extend the substantive value of the explanatory modelling approach taken thus far, by providing estimates which can be applied in a recommendation procedure. The overarching aim of this third and final study is, therefore, to consolidate the substantive contributions already made in this thesis as a proof-of-concept. In this sense, this study aims to address the fourth and final aim outlined in section 1.2., which was to present an actionable method of integrating knowledge into the recommendation process.

To be clear, however, this study is not intended to provide an exhaustive system or method, but rather to contextualise and highlight a principled approach to incorporating knowledge about music listening as a set of behaviours into the implementation and evaluation of a recommendation process. It is for this reason that the study is framed in broadly conceptual terms, and its relevant processes and approach will be contextualised in the following sections. Conceptually, the proposed method is an affect-aware approach (see section 4.3.3), which orientate recommendations according to short-term effects influenced by modulating features of item content (Lex et al., 2021). For this reason, audio features are useful indicators of predicted or desired levels of music content, for which content-based systems approximate values accordingly. What follows is an explanation of the considered approach from prediction to implementation to evaluation. First, the design and implementation of the procedure is outlined. The implementation and evaluation of this procedure formulates the third and final study of this thesis.

8.2 Design and Implementation of a recommendation procedure

The approach taken in this study is to use *contextual prefiltering*, in which information about users' context is provided to deliver recommendations that are pre-determined within a system or delivery tool (Baltrunas et al., 2011; Schedl et al., 2022). Specifically, the intention is to use input parameters informed by the results of the data collated during Study 2 and calibrate recommendations according to a given predictor through an explanatory path model like that reported in section 7.4.6. This effectively reverse engineers the conceptual process through which music selection occurs. In using contextual variables (i.e., activities), it is hypothesised that listeners may be provided with suggested tracks in the form of a playlist directly that is prefabricated (i.e., generated prior to engagement). In this sense, the proposed system effectively operates as a decision tree, in which relevant audio features are upweighted and/or downweighted to filter recommendations according to contextual information provided by the prospective user, illustrated in Figure 20.

Figure 20 Conceptual Approach to Proposed Recommender Mechanism



For this, the previously employed audio features from the Spotify API were considered practical measures that may be used to both estimate and implement these content-based recommendations. This process therefore integrates methodological principles of music recommendation and curation with existing psychological data; implemented through an explanatory model that leverages features from the previously applied Spotify API. It is worth acknowledging in advance, however, that the personalisation of the proposed system is limited in comparison to more advanced systems, which accumulate user taste profiles over extended time-periods of use and have larger volumes of data on which to make inferences. Also, there consequently remains the *cold-start* problem, for which *seed* data is required (Liu et al., 2012). Setting aside any assumptions about the wider suitability of this rudimentary system is therefore practical, as this study does not aim to generate a fully-fledged independent system, but rather extend the practical relevance of the exploratory method as a whole based previously collated ecologically valid data.

8.2.1 Preliminary set-up

In Study 2, the Spotify API was used to retrieve audio-features of participants' self-reported music selections. Although the use of this resource holds certain limitations, such as the reliance on the audio features provided in the absence of proprietary knowledge (Maloney et al., 2021), there are affordances also. Spotify is the most widely accessed streaming service in the world (Götting, 2022) and, as a resource, the API has been applied elsewhere in extant research (e.g., Barone et al., 2017; Jin et al., 2018; Maloney et al., 2021). Moreover, the findings of Study 2 provide evidence that the API is a useful basis upon which data can be retrieved and reapplied through a platform widely accessed by many real-world listeners. Therefore, any tangible applications stemming from the use of an existing and popular system is a practical way of generating and distributing recommendations. With the relevant caveats in mind, using the Spotify API for the purpose of facilitating an exploratory recommendation procedure was deemed to be pragmatic due to the parity of applied measures between the present and previous study, and relative ease of implementation.

The recommendation endpoint of the Spotify API[6] accepts an array of arguments (a way of providing additional information or instruction to a given function) that includes minimum, maximum, and targeted values of audio features for returned tracks, as well as the input of *seeds* (a preliminary input to generate recommendations). The Spotify API allows genres, artists, or tracks to be used as seeds in any given combination (Jin et al., 2018), but it is not possible to utilise this feature without at least one form of seed data. Therefore, a means of selecting seed data from which to run recommendation requests, as well as a means of estimating target values for the audio features in given conditions, was needed.

The former of these two problems is simpler to resolve. The API accepts specified genres as seed data, and it makes substantive sense to utilise this option to enable a modest degree of taste-indication. There is evidence in the literature that individuals' music preference is at least partially represented by preferred genres (Rentfrow et al., 2011; 2012), and this is often leveraged as a general indicate of preference and taste (Ferrer et al., 2013). Accommodating this in some way enables a degree of taste-indication in the absence of prior information about users' preferences due to the prefiltering approach taken. For this purpose, 10 genres were identified as suitable seeds from which to run recommendations: classical, country, dance, folk, hip-hop (rap), indie, jazz, metal, pop, and rock. Bansal et al. (2020) describe these 10 genres as those most commonly applied in psychological research, and with the caveat that they are Western-centric and non-generalisable to other cultures, do provide a practical array of seed data from which to proceed given the Westernised context of this research. However, future research would do well to incorporate indicators of taste that are not culturally biased or overly

---

[6] https://developer.spotify.com/documentation/web-api/reference/#/operations/get-recommendations

summary of listeners' personal taste, as genre preference alone is something of an oversimplification in this regard.

Regarding the latter of the issues mentioned, estimates of appropriate values of the audio features to calibrate recommendations require either minimum, maximum, or target values, so a means of determining such values is essential. It was decided that targeting the raw values of audio features would be the most pragmatic approach to integrating content-based parameters in conjunction with the seed genres identified. Targeting audio features effectively does as one would expect. These are tuneable variables within the raw metric of a given audio feature, with tracks nearest to the specified target value preferred and returned over tracks further away from the specified value. For instance, if the target values for the features *Energy* and *Danceability* are set at 0.6 and 0.8 respectively, tracks closest to those values are selected. Moreover, any number of audio features can be assigned target values and are weighted equally in the corresponding results, making it relatively straightforward to implement whether one or multiple features are targeted. This was considered preferable to the alternative option of setting hard floor and ceiling values (minimum and maximum), that do not aim for a given value, but rather set hard cut-off points and use the intermediate range. A method of identifying the relevant target values from a path model is, therefore, described in the next section.

8.2.2 Determining target values of audio features

For the outlined approach to be implemented practically, estimated values of the audio features from the Spotify API need to be obtained in their raw metric. This is so meaningful target values can be used when implementing a request to the API's recommendations endpoint. Determining target values for audio features is not immediately straightforward given the nature of the analyses carried out in Study 2, in which the assessed mediation model simplified the interpretable structure of the audio features of interest by reducing them via PCA first and using factor scores as proxies for the latent variables in the structural path model. The component scores generated via PCA are standardised linear composites of observed variables weighted by eigenvectors (Suhr, 2005), and as such are not really interpretable or comparable with the original observed variables. Therefore, a way of estimating the level of each audio feature in its raw metric is necessary to provide interpretable values. Similarly, factor scores of the latent FML were computed in *lavaan* using the regression method (Andersson & Yang-Wallentin, 2021). Conceptually, the factor scores are hypothetical values that would have been observed if it were possible to measure each latent factor directly, but because latent variables have no natural unit of measurement, weighted standardised scores are used here also. As with the scores derived from PCA, the resulting factor scores are useful for the purpose of subsequent modelling, but uninterpretable on their own for the purposes of prediction, for example.

Since these latent factors serve as mediators in the structural path model, however, they are integral to the substantive conclusions drawn, and should thus be considered in the recommendation procedure given the nature of the relationship between constructs. This is because unweighted regression coefficients of indirect effects represent changes in the value of Y given one unit of change in X transmitted via M, whilst controlling for the direct effect of X on Y (Preacher & Kelley, 2011; Pearl, 2012). Therefore, interpretable, predicted values that can be directly embedded into the recommendation procedure (more accurately representing the path structure) are also necessary for the FML factors. To be clear, this is because it may be the case that the nature of a given effect is dependent on a given level of a mediator (e.g., in cases of complete/full mediation). It is not theoretically admissible to infer the presence of an effect onto an audio feature in the absence of a predicted function that mediates or otherwise influences the relationship. So, it is necessary to estimate values of the latent factors dynamically, so as to gauge the extent to which each factor is employed and provide recommendations based on whether the predicted *functionality* is present via some predicted value/level of the mediating variable, given a relative value of X.

Because of the reasons outlined above, neither factor nor component scores can be used to estimate values of the endogenous variables because they are weighted in the given data and cannot be used to predict future values in their raw metrics. To provide an explanatory model from which estimated values could be obtained therefore, it was considered that fitting a new model using the original audio features retained in Study 2 (rather than the use of component scores), and the use of a *coarse* factor scoring method would be a practical compromise in this instance to estimate predicted values for targeted audio features and thresholds for factors to infer indirect effects simultaneously.

Coarse factor scores are unweighted composites of the raw scores of each latent variables' indicators, calculated either by averaging or summing the indicators of each factor (Brown, 2006; DiStefano et al., 2009). Through such scores, an intrinsic unit in an interpretable metric can be estimated (e.g., a mean of a given set of indicators). It should be acknowledged, however, that coarse scores are unrefined (DiStefano et al., 2009) and (generally) represent latent factors less well than factor scoring methods. However, this trade-off is pragmatic as it provides an interpretable unit of measurement by which ratings for the latent factors can be estimated dynamically. By using a mean score for latent mediators, implied values for the factors can be integrated into the predictive modelling process. Mean scores were considered preferable for this purpose as opposed to the summing of items because the number of indicators per factor ranges from three to six, biasing factors with larger numbers of indicators.

Regarding the ways in which estimates obtained from a model such as this can be used, predictive modelling is the process of employing a statistical model or data mining algorithm to data to predict new or future observations (Shmeuli, 2010). MIR applications often leverage machine-learning techniques for this purpose (e.g., Lepa et al., 2020), however, recent research has been critical of the use of machine-learning techniques in fields like psychology as such methods often result in biased estimates and overfitting (Fokkema et al., 2022). Although machine-learning methods hold some bias-correction capabilities (James et al., 2017), they generally demonstrate little benefit over simpler regression methods in predictive modelling as they are only ever capable of improving on linear main-effects models by capturing additional nuanced non-linearities and model interactions (Fokkema et al., 2022). Because such effects are often small in size (as psychological studies generally have much smaller samples than big data applications) the price of capturing this increased nuance is the aforementioned tendency of machine-learning methods to overfit (that is, to become too dependent on the data to which the model was trained). Consequently, predictive modelling for smaller samples than those used in big data applications is desirable and an active research area for the complex behavioural models applied in fields like psychology (e.g., de Rooij et al., 2022; Fokkema et al., 2022).

Such criticism is seldom one-way, however. Overfitting remains an issue in regression-based explanatory models which tend to provide overly optimistic predictions as the linear interactions are also tied to the sample to which a model is fit (Yarkoni & Westfall, 2017). In addition, there is criticism from some that social scientists are generally too dependent on theoretical, explanatory models and that predictive modelling should be embraced more readily at the expense of explainability (Shmeuli, 2010). As such, predictive modelling is contentious both methodologically and practically given the kind and more limited volume of data psychological research handles, and the research questions it answers.

However, a relevant and motivating alternative to machine-learning approaches was recently published by de Rooij et al. (2022), in which predicted values are calculated for SEMs directly. Specifically, predicted values for observed variables are computed using a conditional distribution, in which predictions for endogenous variables given a predictor are based on the joint model-implied variance-covariance matrix (Σ) and mean vector (μ; see de Rooij et al., 2022; pp. 134-135). The applied method has uses for full SEMs (with measurement and structural components), measurement only models, and structural only models, but unfortunately cannot yet be applied to models with categorical variables. In conditions where the model is saturated, however, the predicted values of observed endogenous variables are identical to the predictions generated in ordinary least squares (OLS) regression. The *bias-variance* trade-off is an inherent consideration within predictive modelling (see Hastie et al., 2009), including for machine learning applications, and so is not something introduced by this different approach. As such, this approach is practical insofar as it supports estimation of future values of the observed variables of interest by using a theory-based structural model, rather than through machine-learning. It was therefore anticipated that by obtaining predicted values for each y variable in response to a given x, subsequently obtained predicted values ($\hat{y}$) could then be applied as target values for each audio feature.

Given these conceptual and practical requirements, a new saturated structural model was fit to the data gathered in the second study, in which instead of the reduced set of principal components specified as outcome variables, the five retained audio features (*Danceability*, *Energy*, *Acousticness*, *Instrumentalness*, and *Valence*) were directly added to the model as outcome (y) variables. The rest of the model remained structurally the same, with dummy-coded activity variables serving as predictors, but coarse (mean) factor scores operating as observed proxies for the latent factors. μ was added during model estimation, rendering model intercepts mean based in the absence of a predictor, rather than standardised as in estimation based solely on covariance structures (Ployhart & Oswald, 2004; Kline, 2016). This provides intercepts and subsequently unstandardised parameter estimates of observed variables in the original metric of the data. Once the saturated path model was fit, it was adjusted for non-*iid* as in Study 2 (Oberski, 2014).

In the saturated model, the paths are essentially a series of multiple regression estimates, providing values identical to OLS in the absence of model constraints as mentioned (de Rooij et al., 2022). In the OLS tradition, predicted values for outcome variables ($\hat{y}$) can be computed as the sum of the intercept (β_0) and regression coefficient (β_1) given a one unit increase in the predictor (x; Ployhart & Oswald, 2004; Hair et al., 2014; James et al., 2017), summarised in the following equation:

$$\hat{y} = \beta_0 + \beta_1 x, \tag{1}$$

Given that unstandardised regression coefficients from the model are analogous to those estimated in multiple regression, estimated values for specific effects can be calculated by following the same logic as above (Preacher & Kelley, 2011; Miles et al., 2015). In the present application, therefore, it is considered that the principle of (1) can be used to provide estimates for given values of audio features in model paths by plugging in respective path coefficients. In other words, this approach uses unstandardised regression coefficients from the fit explanatory model to estimate values for each audio feature where specific effects indicate there is an effect on y at the $p \leq .05$ level.

In extending this logic, conditional setups can be used to assess which predicted value should be used given the predicted value of a mediator (e.g., if the mean value of items relating to a given factor meets or exceeds the predicted level of the mediator, then indirect effects may be inferred and use the corresponding $\hat{y}$ based on that path coefficient). This is not solely the case for indirect effects, however, but also for direct and total effects since these conceptually all reflect changes in y given x. Therefore, estimating all of these values and using them to determine paths (e.g., *DE* only, *IE* only, *TE*) can then be used in a decision-tree style manner. Table 26 summarises these interactions and predicted values based on the fitted model. Specific effects were largely comparable with those observed in the prior model involving components.

Table 26 Summary $\hat{y}$ for Audio Features in Path Model

Activity (x)	Statistically significant effects in path model	Path coefficient (β_x)	Intercept (β_0)	Predicted Value ($\hat{y}$)
Work/study	→ Instrumentalness (*DE*)	0.090**	0.253	*Instrumentalness* = $\beta_0 + \beta_1 x$ = 0.343
	→ Instrumentalness (*TE*)	0.123**	0.253	*Instrumentalness* = $\beta_0 + \beta_1 x$ = 0.376
	→ Valence (*DE*)	-0.060*	0.476	*Valence* = $\beta_0 + \beta_1 x$ = 0.416
	→ *FaC* → Valence (*IE*)	-0.031**	0.476	*Valence* = $\beta_0 + \beta_1 x$ = 0.445
	→ Valence (*TE*)	-0.098***	0.476	*Valence* = $\beta_0 + \beta_1 x$ = 0.378
	→ *FaC* → Energy (*IE*)	-0.039**	0.611	*Energy* = $\beta_0 + \beta_1 x$ = 0.572
	→ Energy (*TE*)	-0.077**	0.611	*Energy* = $\beta_0 + \beta_1 x$ = 0.534
	→ *FaC* → Acousticness (*IE*)	0.036*	0.318	*Acousticness* = $\beta_0 + \beta_1 x$ = 0.354
	→ Acousticness (*TE*)	0.081*	0.318	*Acousticness* = $\beta_0 + \beta_1 x$ = 0.399
Travel	→ Valence (*DE*)	-0.061*	0.476	*Valence* = $\beta_0 + \beta_2 x$ = 0.415
	→ Valence (*TE*)	-0.055*	0.476	*Valence* = $\beta_0 + \beta_2 x$ = 0.421
Relaxation	→ *BaA* → Danceability (*IE*)	-0.010*	0.495	*Danceability* = $\beta_0 + \beta_3 x$ = 0.485
	→ Danceability (*TE*)	-0.039*	0.495	*Danceability* = $\beta_0 + \beta_3 x$ = 0.456
Exercise	→ *PA* → Danceability (*IE*)	0.047*	0.495	*Danceability* = $\beta_0 + \beta_4 x$ = 0.542
	→ *PA* → Energy (*IE*)	0.071*	0.611	*Energy* = $\beta_0 + \beta_4 x$ = 0.682

	→ *PA* → Acousticness (*IE*)	-0.102**	0.318	*Acousticness* = $\beta_0 + \beta_4 x = 0.216$
Socialising	→ Instrumentalness (*DE*)	-0.185*	0.253	*Instrumentalness* = $\beta_0 + \beta_5 x = 0.068$
	→ *BaA* → Danceability (*IE*)	-0.024*	0.495	*Danceability* = $\beta_0 + \beta_5 x = 0.471$

Note. *FaC* = Focus and Concentration. *BaA* = Background and Accompaniment. *PA* = Physiological Arousal. *DE* = Direct Effect. *IE* = Indirect Effect. *TE* = Total Effect. *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$. For ease of interpretation, equations are simplified in the right-hand column, and refer solely to the path coefficient of the reported effect.

Based on this, a series of eight possible outcomes in which audio features could be predicted were identified: (1) Work/Study with indirect effects and direct effects (plus relevant total effects), (2) Work/Study with direct effects only, (3) Travel with a direct effect only, (4) Relaxation with an indirect effect only, (5) Exercise with indirect effects only, (6) Socialising with an indirect effect and direct effect, (7) Socialising with a direct effect only, and (8) conditions in which no predicted values for the audio features depending on the categorical predictor could be estimated. Here, total effects are regarded as the overall effect of direct and indirect effects, and so are useful where possible. However, in most cases, the conditions are such that the total effect and an indirect effect is significant, but that the direct effect between *X* and *Y* is not. Therefore, the relationship is fully mediated and, as such, the estimated values of the audio features are hypothesised to be dependent on a given level of the mediator. As such, if the mediator's mean value does not align with the predicted value, then it is assumed that the influence of that mediator is absent. In such conditions, it is necessary to adjust the potential outcome by reverting to estimating any relevant direct effects only, or by not estimating any values at all if the interactions are based solely on indirect effects. Given the presence of indirect effects, this was repeated for the *a* paths in which the predictors influenced the mediating FML measures, shown in Table 27.

Table 27 Summary $\hat{y}$ for Coarse (Mean) Factor Scores in Path Model

Activity (x)	Statistically significant effects in path model	Path/effect coefficient (β_x)	Intercept (β_0)	Predicted Value ($\hat{y}$)
Work/Study	→ *FaC* (*a*31)	1.271***	2.047	$FaC = \beta_0 + \beta_1 x = 3.318$
Travel	→ *ER* (*a*22)	0.560***	1.866	$ER = \beta_0 + \beta_2 x = 2.426$
Relaxation	→ *ER* (*a*23)	0.401**	1.866	$ER = \beta_0 + \beta_3 x = 2.267$
	→ *BaA* (*a*43)	-0.432*	2.804	$BaA = \beta_0 + \beta_3 x = 2.372$
Exercise	→ *PA* (*a*54)	2.167***	1.705	$PA = \beta_0 + \beta_4 x = 3.872$
Socialising	→ *ISB* (*a*15)	1.174*	1.326	$ISB = \beta_0 + \beta_5 x = 2.500$
	→ *BaA* (*a*45)	-1.054*	2.804	$BaA = \beta_0 + \beta_5 x = 1.750$

Note. *ISB* = Identity and Social Bonding. *ER* = Emotion Regulation. *FaC* = Focus and Concentration. *BaA* = Background and Accompaniment. *PA* = Physiological Arousal. Path terms are in brackets. *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

8.2.3 Generating recommendation playlists using R

Once estimates for the raw audio features and coarse factor scores had been calculated and seed genres identified, playlists that target these values could be generated. This was done using *spotifyr* (Thompson et al., 2022), a wrapper for the Spotify API that allows users to call API functions in an R environment, in conjunction with the *httpuv* package (Cheng & Chang, 2023), a Hypertext Transfer Protocol (HTTP) and WebSocket server which handles requests between the server and local environment directly from within R.

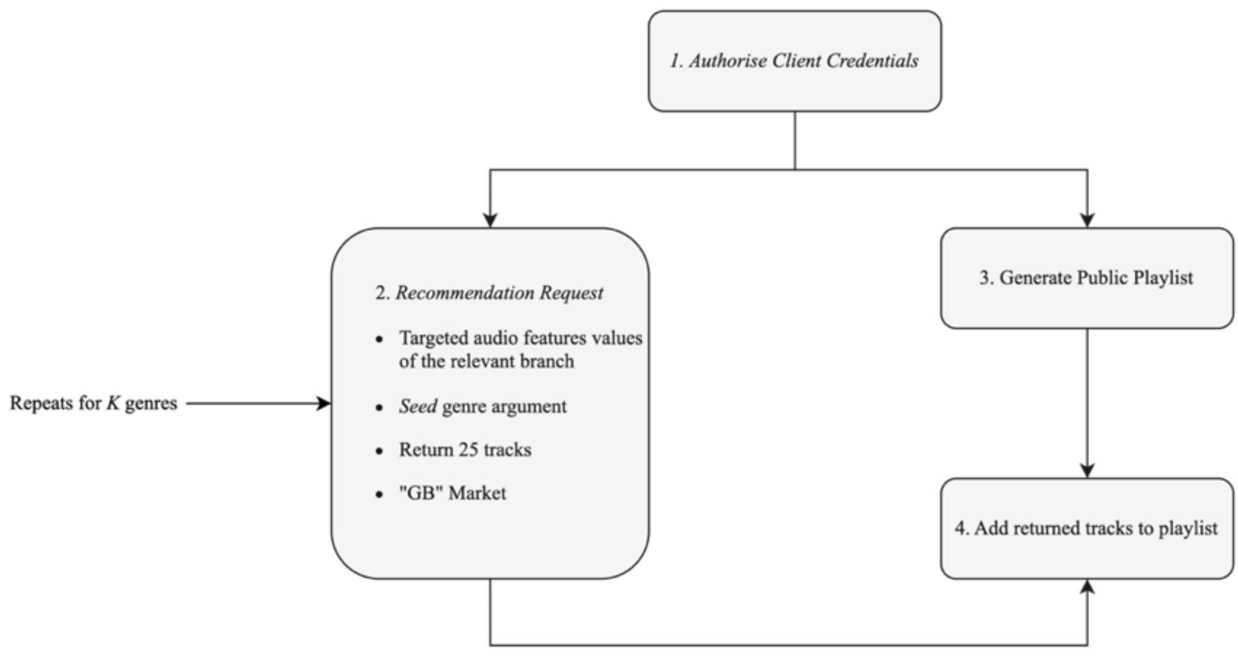
The recommendation procedure first operates by receiving server-side client permissions from the Spotify API. For this, a web app was generated in the Spotify Developer[7] page, providing a client ID, secret, and redirect URL (set to a localhost for these purposes). From there, API endpoints could be called and accessed through the researcher's Spotify account. First a seed genre was selected (one of *K* selected genres) and an argument added to retrieve 25 tracks from each request, and that all tracks be pulled from the relevant market ("GB" in this case)[8]. These

[7] https://developer.spotify.com/dashboard
[8] https://developer.spotify.com/documentation/web-api/reference/get-available-markets

were then accompanied by the target values for the audio features of each relevant category, as calculated in the steps outlined in section 8.2.2. This returned $n = 25$ tracks for each request that targeted audio features according to the given seed genre.

The final two steps were more concise. An empty (public) playlist was generated in the researcher's personal Spotify account (from within the R environment) and the IDs of the empty playlists were then parsed through another API endpoint[9], in which the recommended track URIs were appended, populating the empty playlist with the returned $n = 25$ tracks. This was repeated for K genres in K branches identified through the explanatory model (i.e., 80 playlists were generated in total, including a neutral set of playlists with no targeted audio features). This procedure is summarised in Figure 21, whilst the R syntax is provided in Appendix I.

Figure 21 Procedure for generating prefabricated playlists



[9] https://developer.spotify.com/documentation/web-api/reference/#/operations/add-tracks-to-playlist

8.3 Evaluation of generated playlists

In section 4.3.4, evaluation methods of recommender systems were discussed. It is therefore useful to situate system evaluation in the context of the present study and outline an approach to implementation and hypothesis testing using an identified framework. Of the three prominent approaches to evaluating recommendations (offline, online, and user-centric evaluation), it was argued that the user-centric approach was principally aligned with the motivations of psychology-informed approaches. This is because user-centric designs allow for the user-experience (UX) to be directly measured, which extends the principle of knowledge-integration that not only informs the recommendation process itself, but also the evaluation of UX thereafter (e.g., Knijnenburg et al., 2012; Lex et al., 2021). In addition, the use of user-centric evaluation enables inferences to be made regarding objective aspects of the methods under evaluation, extending the reach of claims made beyond circumstantial experiences of pleasant experience (Schedl et al., 2018). The integration of psychology-informed recommender systems and user-centric evaluation methods are thus conceptually aligned as a means of keeping the user as the central focus of the system in terms of orientation and (where possible) improvement.

To evaluate the proposed approach, the user-centric framework presented by Knijnenburg et al. (2012) was applied, in which personal characteristics of the user, as well as the user's response to situational characteristics of their experience with the system, are measured via a series of unidimensional latent constructs. This particular framework lends itself particularly well to user-centric evaluation as it holds a high degree of abstraction, providing a means of forming and testing hypotheses that are specific and relative to a given system (Lex et al., 2021). To the best of the researcher's knowledge, this would constitute the first application of Knijnenburg et al.'s (2012) framework to evaluate an MRS.

The framework integrates several important aspects of measurement to assess a system's performance, namely Objective System Aspects (OSAs), Subjective System Aspects (SSAs), and UX. Knijnenburg et al. also consider how personal characteristics (e.g., trust in technology or user expertise) influence users' perception of a system. To this end, a series of psychometric measures relevant to the evaluation of the proposed system were selected in the present study, which are hypothesised to operate as a system of variables. The simultaneous estimation of these measurement constructs as well as a hypothesised structural model lends itself well to the capabilities of SEM, which is often cited as a means of applying this kind of evaluative framework (e.g., Knijnenburg et al., 2012; Lex et al., 2021). Each of the latent constructs presented by Knijnenburg et al. are measured on either 5 or 7-point Likert scales, ranging from *Completely Disagree* to *Completely Agree*. For each of the selected constructs, a 7-point scale was used, with relevant adaptations made to item content to ensure they were relevant to the domain of interest (e.g., terms such as 'items' were replaced with 'tracks').

8.3.1 Evaluation Hypotheses

Regarding system aspects, OSA pertains to specific characteristics of the system, and so serve as predictors of system specific UX (e.g., perceived quality, perceived effectiveness). With regard to the present system, the method primarily looks to assess perceived difference in quality between playlists with targeted and non-targeted recommendations. Initially, this creates a binary predictor, whereby recommendations are targeted according to predicted audio features (T_1), and a reference category in which recommendations are non-targeted (T_0). To assess SSA, Knijnenburg et al. apply a measure of *perceived recommendation quality* (PRQ), which they in turn hypothesise to mediate relationships between OSAs and UX. Therefore, the latent construct *perceived system effectiveness* (PSE) is also applied as a measure of UX to assess this hypothesis with regard to the present system (i.e., $T_1 \rightarrow$ PRQ $\rightarrow$ PSE). Here, it is considered that targeted recommendations will result in greater PRQ and PSE than recommendations that are generated without any targeted audio features.

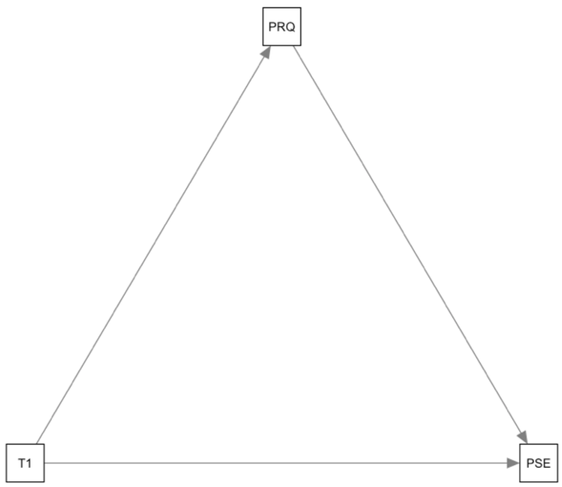
In addition to the above specific characteristics of the system's implementation, it is considered that, if evidence of a mediated relationship is found, the construct representing user *Expertise* (a personal characteristic conveying information about an individual's self-rated level of expertise) moderates the mediated relationship. Moderated-mediation occurs when the strength of an indirect effect is dependent on the level of a moderator, or moderators (Preacher et al., 2007). In this case it is hypothesised that *Expertise* moderates the strength of the simultaneously hypothesised indirect effect of T_1 on PSE, via PRQ. This is because users with greater expertise of a given domain perceive recommender systems as being less useful and effective than novice users, with users holding moderate levels of expertise considering systems most effective (Kamis & Davern, 2004; Hu & Pu, 2010; Knijnenburg et al., 2012).

To be clear however, the nature of this moderated interaction could theoretically take several forms. For instance, *Expertise* may hypothetically operate as a first-stage moderator (Kline, 2016), whereby it moderates the extent to which users perceive PRQ as a consequence of T_1 as users with higher levels of *Expertise* were hypothesised to be more critical of the recommendations and thus see reduced influence on PRQ. However, it may also be the case that *Expertise* is a second-stage moderator, whereby it rather moderates the *b* path of the model (PRQ $\rightarrow$ PSE), or moderates both of these paths. To this end, it was considered that from an analytic perspective, comparing these competing models would be a prudent way of assessing which best explains the observed variance in the sample. Specifically, it was reflected that comparing four models would be a practical way to assess the applicability of *Expertise* as a moderator in the given context, in which the first model is the simplest with no moderating factors (M_1), compared with the three models outlined above. These can be summarised as *Expertise* as a first-stage moderator only (M_2), a second stage moderator only (M_3), and finally as a first-and-second-stage moderator (M_4).

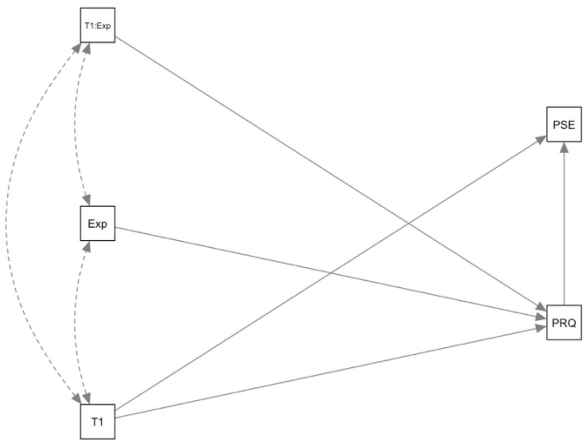
To examine the comparability between competing models, several options are available. In the SEM context, models are considered 'nested' when a set of free parameters estimated in the first model is a subset of those specified in a subsequent model, or vice versa (Chou & Huh, 2012; pp. 233-234). The χ^2 difference test can be used to assess the decrement in model fit as parameters are eliminated in model trimming, or as the improvement of fit as parameters are added (Kline, 2016), a process similar to the forward and backward approaches applied during stepwise regression (Chou & Huh, 2012). By comparison, in non-nested models, the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) can be used to compare competing models, where the model with the smallest AIC and BIC values is interpreted as having the best fit amongst the competing models (see Merkle et al., 2016; Lin et al., 2017). These measures can be used to compare the four proposed models, retaining that which best explains the variance in the observed data. SEM is preferred for this over the sequential regression approach as it incorporates both mediation and moderation hypotheses simultaneously and handles error arising from interaction terms more effectively (Sardeshmukh & Vandenburg, 2017). These four proposed models are illustrated in Figure 22 (to ease interpretation, all constructs are presented as manifest variables).

Figure 22 Hypothesised Moderated Mediation Models

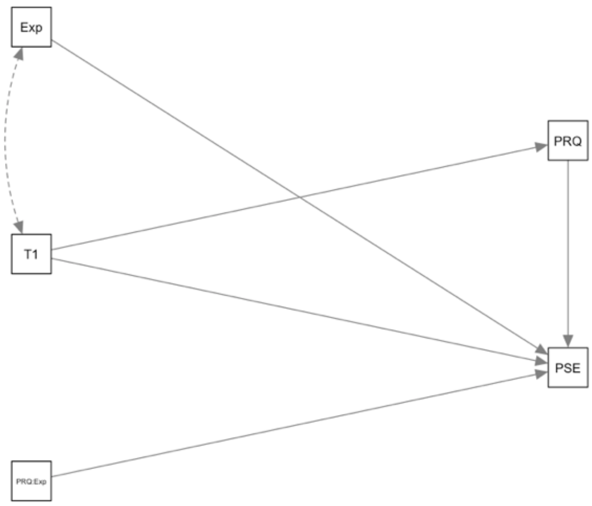
M_1 – *Mediation Only*



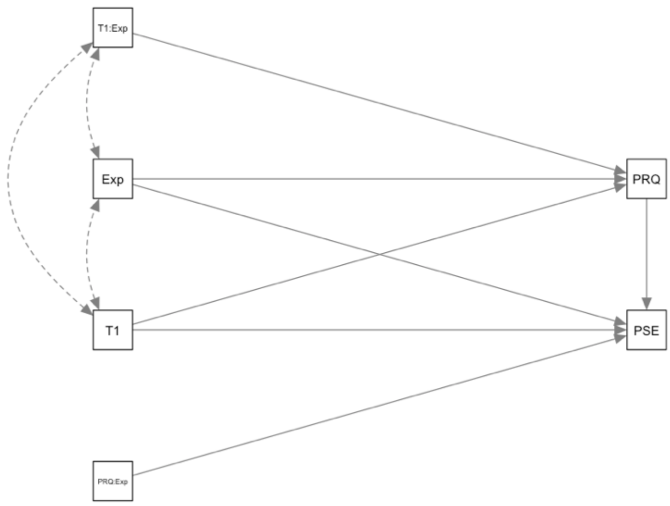
M_2 – *First-stage Moderation*



M_3 – *Second-stage Moderation*



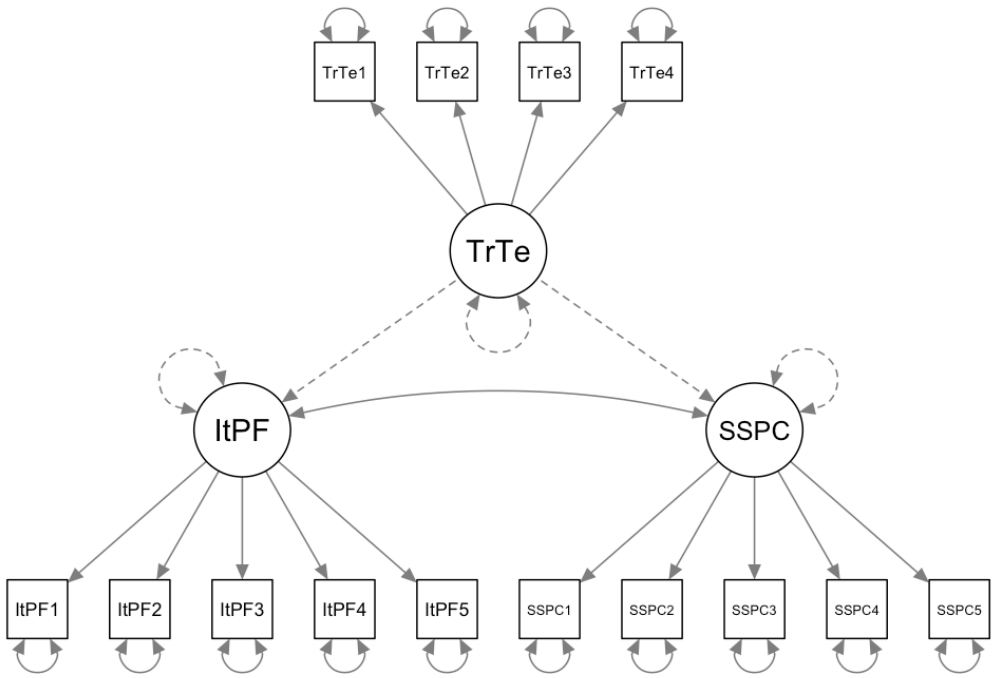
M_4 – *First-and-second stage Moderation*



Note. *Expertise* (Exp), *Perceived Recommendation Quality* (PRQ), and *Perceived System Effectiveness* (PSE) are all multi-indicator latent constructs with three, six, and six indicators respectively.

To be clear, however, it would only be theoretically permissible to estimate these models should the base mediation effect exist (M_1). As such, estimation of these relationships is dependent on the existence of mediation in the base model, for which if no evidence is found, then there would not be sufficient support to proceed with the estimation of subsequent models. Next, it was hypothesised that an understanding of users' trust in, and perception of, the system is of substantive value since this assesses the extent to which users consider the system trustworthy. Unidimensional constructs assessing this can be represented by a participant's general *trust in technology* (TrTe), *intention to provide feedback* (ItPF), and *system-specific privacy concern* (SSPC). TrTe is a personal characteristic, measured through four items, that is not specific to a system, but instead more general in the way it is conveyed. SSPC is, as the name suggests, system-specific, and is as such hypothesised to be dependent on participants general TrTe (i.e., the greater trust a user has in technology, will result in lower levels of SSPC). ItPF is characterised by Knijnenburg et al. (2012) as an 'interaction', conveying information about long-term intentions following use of a system. However, because measurement of this construct theoretically requires (1) long-term use of the system and (2) a greater degree of independence from an existing system than the current study provides, its application here is adapted to act as a more general indication of user intentions. Therefore, ItPF will also act as a personal characteristic, hypothesised to be dependent on TrTe as this conceptually conveys the notion that a user's general disposition to provide feedback with systems, they use is influenced by TrTe at the latent level. Consequently, it was hypothesised that a further SEM could be used to assess this whilst controlling for the covariance between ItPF and SSPC. Specifically, it was hypothesised that TrTe would have a negative effect on SSPC (i.e., result in lower levels of SSPC) and a positive effect on ItPF (i.e., lead to higher ratings of ItPF). This structure is illustrated in Figure 23.

Figure 23 Hypothesised Structural Equation Model of TrTe as a predictor of ItPF and SSPC



The models presented represent a priori hypotheses regarding the relations between constructs. These signify two key areas of substantive interest: (1) the efficacy and perceived quality of the recommendations targeted by audio features and (2) that users' TrTe is a negative predictor of SSPC and a positive predictor of ItPF. Given, however, that some of these measures can be considered traits (i.e., personal characteristics in the present study, such as *Expertise*, TrTe, and ItPF as conveyed) whilst others are more-state like (i.e., situational characteristics regarding the system such as PRQ, PSE, and SSPC), attention was needed as to the stage of measurement, so that the personal characteristics are measured prior to interaction with the system and not after. Alternatively, PRQ, PSE, and SSPC are obviously subject to engagement with the procedure and would thus need to be collated after system interaction. Regarding practical estimation of the models, the previously applied *lavaan* package (Rosseel, 2012) is able to fit both simpler SEMs as well as more complex models, such as those assessing moderated mediation (e.g., Miles et al., 2015; Balwant et al., 2019; Geng et al., 2021). It was therefore practical to reapply this software tool to test hypotheses.

8.4 Survey Design

To implement this user-study, an online survey was generated in Qualtrics. Participants were provided study information and an embedded consent form and would then be asked if they were willing and able to listen to music using Spotify. If participants indicated they were not in a situation where they were able to listen to music, then they could choose to defer participation, whereby they provided an email address to which the study could be sent for later use. When participants were able to take part, demographic information was collated (e.g., age, gender) and they would indicate which activity best described their current/ongoing task from a predefined list (the discrete categories generated during Study 2), and preferred genre from the 10 available options. Next, the 23-items from the FML measure were rated via a matrix table. Embedded code leveraged through the Qualtrics API[10] would then automatically compute the mean values of each factor's score for each participant, and where relevant would use the subsequent value in tandem with the selected activity and preferred genre to direct participants toward the appropriate playlist. Additional measures were included at this stage from Knijnenberg et al.'s (2012) framework, including *Expertise*, TrTe, and ItPF, as these measures are conceptually independent of any intervention arising from the recommendation procedure in the present context.

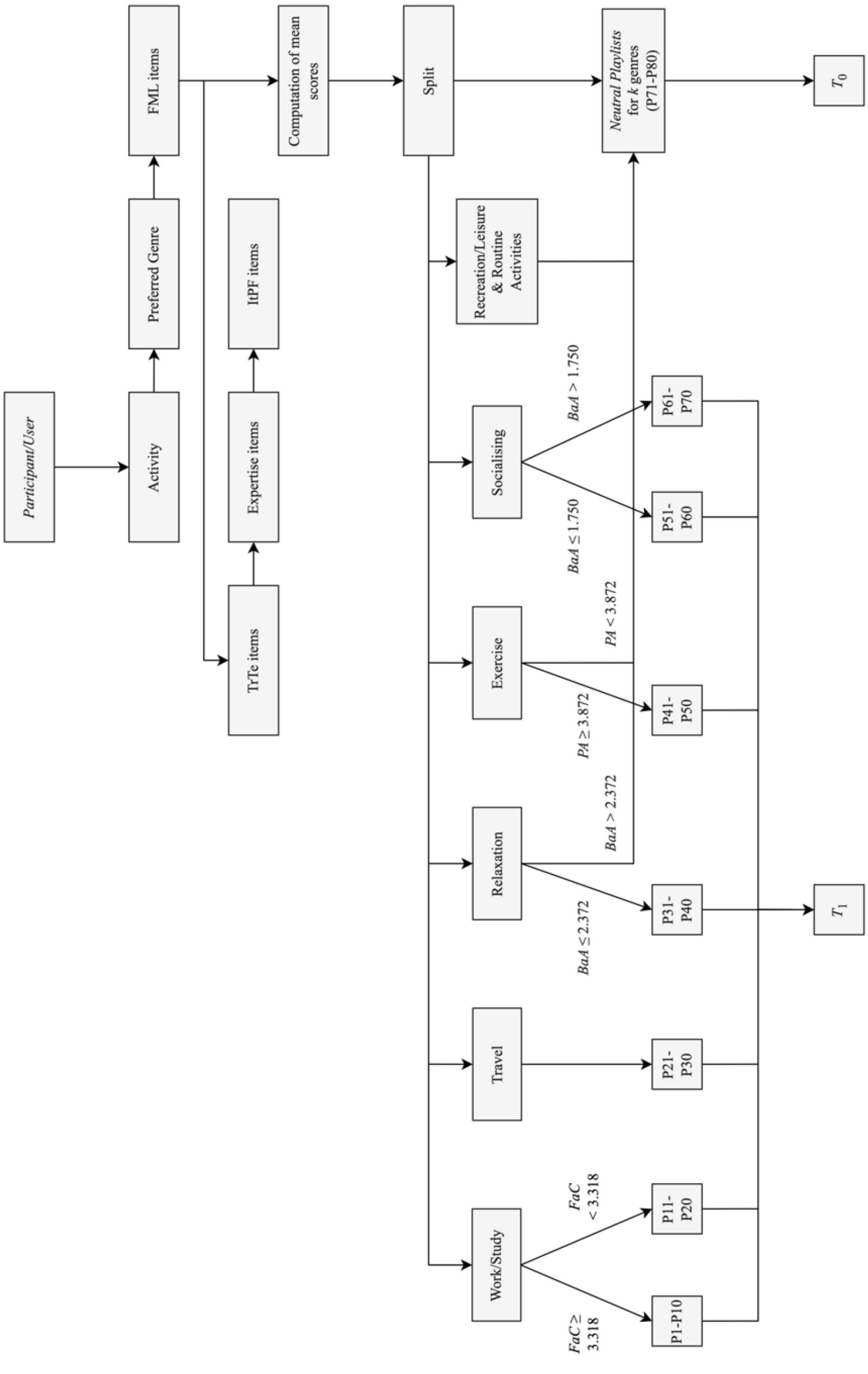
Participants' selected activity, genre preference, and ratings of the FML items were used to orientate participants to a suggested playlist with targeted audio features dynamically (i.e., following a decision-tree-style structure). If the conditions were such that no targeted values could be estimated, however, participants were directed to *neutral* playlists based on indicated genre preference only (i.e., with no targeted values for any audio feature). More importantly, however, in situations where predicted values could be estimated (and thus direction to a relevant playlist inferred), responses were randomised in such a way that participants would randomly be split between receiving targeted values and the *neutral* (i.e., non-targeted) playlists. This way, the sample received targeted and non-targeted features alternately, and thus

---

[10] https://api.qualtrics.com/

comparison between the efficacy of targeted (T_1) in comparison to non-targeted (T_0) playlists could be assessed in activity conditions whereby it was possible to estimate audio features. This decision-tree-style flow is summarised in Figure 24.

Figure 24 Summary decision-tree of study flow



Once this structure had been implemented, participants were presented with a hyperlink to the branched playlist resulting from the described process. Each hyperlink set a different value of an embedded variable which corresponded to respective playlist numbers (playlists were labelled with neutral tags so that participants could not see the intended outcome, function, or activity). This made it possible to see which playlist each participant was directed to. Similarly, an embedded dichotomous variable was included, which was set to 0 by default. Each time a playlist was selected which held *targeted* audio features, the value of this dichotomous variable changed from 0 to 1, hence automating the generation of the binary predictor (T_x) intended for use in the models shown in Figure 22.

It should be acknowledged that based on which values it was possible to estimate values for, some conditional branches had no targeted audio features estimated. Recreation/Leisure was not found to have any significant direct or indirect on any audio feature, and thus had no estimated features. Routine Activities served as the reference group in the model applied for estimation, hence informing model intercepts, but left to one side for the purpose of the present study estimation. Therefore, for both of these activities, participants were directed to the *neutral* playlists only. In addition, Relaxation and Exercise indicated indirect effects in the model only, and as such any effects on audio features were exclusively dependent on a given level of the relative mediating variable. As such, if the predicted values for the mediators were not met, this condition would also branch to the *neutral* playlists as the estimated values depended solely on the presence of the mediator.

Once completing this first phase of the study, participants would receive automated email follow-ups one hour after completion, containing a link to the evaluative section. There were no time-limits on the availability of this link (beyond the conclusion of data collection). Email addresses were first restated to enable alignment of study sections between participants. Next, participants were asked to describe their engagement with the playlist, which included an approximation of listening time in minutes and a choice of four engagement conditions. These included: (1) listening to the playlist in order and without skipping tracks, (2) listening to the playlist on shuffle and without skipping tracks, (3) listening to the playlist in order and skipping tracks, and (4) listening to the playlist on shuffle and skipping tracks. Those selecting the third and fourth descriptors were then asked to approximate the number of tracks they skipped through. Following these preliminary measures, the situation-specific unidimensional latent constructs provided by Knijnenburg et al. (2012) were presented, namely PRQ, PSE, and SSPC. Thus, when aligned, the respective sections of the study formed one dataset including direction to recommended playlists, and evaluation of those playlists. The order of items relating to all latent constructs (FML items and unidimensional latent variables from the evaluative framework) were randomised in both study stages. The contents and flow of these sections are shown in appendices J and K respectively. Based on the design described, ethical approval was granted by the School of Arts and Creative Technologies ethics committee at the University of York.

8.5 Results

This section outlines the results of the present study, including data-cleaning and analytic approaches. The study was distributed via an anonymous URL, primarily shared through social media, and emailing lists. All participants were provided with relevant study information and provided informed consent via checkboxes embedded in the Qualtrics survey. There were no inclusion/exclusion criteria beyond the requirement to be at least 18 years old. Participants completing both parts of the study were not compensated, however, a £50 raffle for an Amazon gift voucher was included for participants who opted in (this was distributed to a randomly selected participant following data collection).

8.5.1 Data Cleaning

Shortly after data collection began, some irregularities were noticed. A large volume of responses were obtained in a short time-period after the study was initially shared. These were accompanied by unreadable email addresses provided, as well as unusual or identical responses to a qualitative feedback question in the second part of the study led to suspicions that the online survey had been spammed by survey bots (Griffin et al., 2022). This led to a pause in data collection for the first wave of the study to prevent further impact. Following adjustments to the survey, such as obtaining a new study link and adding reCAPTCHA (Completely Automated Public Turing Test to tell Computers and Humans Apart) verification, the study was redistributed in a second wave. When reshared, no mentions of incentives were made as it was suspected that previously mentioning the raffle for a gift voucher is what had led to targeting by bots via social media.

<u>*Case removal*</u>

The bot-affected first wave of data collection obtained 335 responses to the initial survey, with 147 responses completing the second part of the study. Firstly, incomplete cases, in which participants indicated they could not listen to music (n = 9) were removed. To clean the data and identify genuine participants, a combination of qualitative steps outlined by Griffin et al. (2022) were followed. Cases were deemed suspicious through a combination of email addresses (particularly Gmail addresses) that were unreadable or ended in an excessive number of digits, responses that were impermissible (e.g., stating that the interaction was a repeat visit in the absence of any other cases), if qualitative feedback questions were repeated or unusual, and if these preceding features related to cases responded to in close time-proximity, indicating a survey farm (Wang et al., 2017; Griffin et al., 2022). Following this process, 33 cases (9.85%) were retained in which there was a high degree of confidence were human participants. Though disappointing and disruptive, this confidence is essential for the quality and validity of study findings. It may be the case that some of the removed cases were not survey bots, however, which constitutes an unforeseen limitation.

Following this, a further 60 complete responses were gathered in the second wave of data collection. This totalled for a sample of n = 93 initial observations in which there was confidence that each case was a human participant, from which n = 50 had completed both study stages. For these cases therefore, the retention rate between study phases was 53.76%.

8.5.2 Descriptive Statistics

In total, there were 93 complete cases for the first stage of the study. In this, n = 4 participants (identified through multiple email entries) took part more than once, collectively accounting for n = 10 cases (n = 3 participated twice, n = 1 participated four times). In the case of the latter, although the same email address had been used, demographic data differed in one case (i.e., age and gender), indicating two individuals using the same email address. As such, these were taken as two individuals from the response perspective, and hence the data were considered as

93 responses partially nested within 88 individuals (although there is still dependency analytically between the individuals sharing an email address as this is likely a case of two individuals in the same household). These 88 individuals ranged between the ages of 19 and 69 (M = 32.30, SD = 12.237). Of these, 43 (48.86%) identified as male, 44 (50%) as female, and 1 (1.14%) as non-binary/third gender. Of this sample, 32 (36.36%) identified as non-musicians, 26 (29.55%) as amateur musicians, 23 (26.14%) as higher-level musicians, and 7 (7.95%) as Professional musicians. Before proceeding further, participants were given anonymised IDs in replacement of email addresses, which were then removed from the dataset.

Regarding the 93 listening episodes overall, Table 28 shows the frequency of selected listener activities, and Table 29 shows frequencies of selected genres.

Table 28 Frequencies of Selected Activities

Activity	*N*	%	Cumulative %
Working/Studying	38	40.86	40.86
Relaxing	22	23.66	64.52
Routine Activity (e.g., chores)	14	15.05	79.57
Travelling	8	8.60	88.17
Recreational Activity	7	7.53	95.70
Socialising	3	3.23	98.93
Exercising	1	1.08	100
Total	93	100	

Table 29 Frequencies of Selected Genre

Preferred Genre	*N*	%	Cumulative %
Indie	26	27.96	27.96
Pop	17	18.28	46.24
Classical	16	17.20	63.44
Rock	11	11.83	75.27
Jazz	10	10.75	86.02
Dance	4	4.30	90.32
Folk	4	4.30	94.62
Hip-hop (Rap)	4	4.30	98.92
Metal	1	1.08	100
Country	0	0	100

Preferred Genre	N	%	Cumulative %
Total	93	100	

From these cases, 26 participants (27.96%) were directed to targeted playlists (T_1) whilst 67 (72.04%) were directed to non-targeted playlists (T_0). Regarding whether participants skipped tracks or not, of the 50 completing both study stages, 10 indicated they listened on shuffle without skipping, 20 in order without skipping, 3 on shuffle while skipping, and 17 in order and skipping. Participants from these latter two groups reported skipping M = 6.40 tracks on average (SD = 4.84). With this sample, it was possible to proceed to preliminary analyses, subject to additional filtering at each stage.

8.5.3 Model estimation

<u>*Cross-validation of FML Measure*</u>

Before proceeding to the primary analyses, a CFA was fit to the FML items to assess whether the factor structure held in the data. This was essential as it is the validity of this construct upon which the mediating role of predicted FML depends. For this, all available data was used as the measure was rated in the first survey section (i.e., before participant drop-off; n = 93). The model was fitted using *lavaan* in R (version 4.2.2). Due to modest dependency in the data (arriving from multiple interactions in n = 4 participants), *lavaan.survey* (Oberski, 2014) was again used to accommodate this dependency in the data. The factor structure indicated good fit ($\chi^2(220)$ = 280.233, p = .004, CFI = .950, TLI = .942, RMSEA = .059). Standardised factor loadings were all >.50, as was the AVE of each factor, again satisfying criterion provided by Hair et al. (2014) seen in studies 1 and 2. This provides confidence that the data was reflected by the factor structure, providing cross-validation across the three studies contained in this thesis. Moreover, this provides confidence that the factor structure is reflected in the characteristics of the observed data, affording confidence of validity in the predicted mediated relationships.

<u>*Model estimation under small n*</u>

In the final dataset, n was small for participants completing both study arms (n = 50). This presents problems insofar as traditional methods of model estimation in SEM typically require large samples (Kline, 2016), and traditional estimation methods (e.g., Maximum Likelihood) may produce inadmissible or inaccurate parameter estimates (e.g., Heywood cases and estimates that deviate significantly from the true population value; Nevitt & Hancock, 2004; Smid et al., 2020). Even then, such problems only occur if models successfully converge in the first place, with non-convergence (when the estimator is unable to find the maximum or minimum for the derivative of model parameters) often occurring when n is small (Smid & Rosseel, 2020; Smid & Winter, 2020). In addition, whilst regression-based factor scores work well as proxies for latent variables in sufficiently large samples, they are often biased when n is small (Andersson & Yang-Wallentin, 2021). As such, the FSR approach that segmented the measurement and structural models in Study 2 would also be unsuitable since this would likely yield biased estimates. In light of this, it was considered how internal and external validity could be optimised to maximise the extent to which meaningful conclusions could be drawn from the sample and provide a better understanding of the ways in which the variables of interest relate.

To address this problem, it was decided to follow the Structural-After-Measurement (SAM) approach provided by Rosseel and Loh (2022). SAM, like FSR, is a two-step approach to estimating SEMs, in which the parameters relating to measurement models are estimated first. Next, keeping these parameters fixed, the parameters of the structural model are estimated. Because the measurement models are estimated first, SAM disattenuates regression coefficients that may otherwise be biased under normal theory maximum likelihood or when using FSR with small n. Moreover, it utilises bias-corrections to accommodate a small sample size by adjusting standard errors to take the estimation uncertainty into account during the second (i.e., structural) stage, and thus operate at small n effectively. These features stabilise model estimation and maximise the ability of the model to reliably detect linear relationships amongst the measurement constructs (Rosseel & Loh, 2022).

Reliability and Validity of Unidimensional constructs

Although SAM is two-step in nature and allows researchers to better model regression coefficients in the structural part of the model when the sample is small, it is not designed to identify or handle misspecifications in the measurement models themselves, and rather assumes the measurement constructs are valid prior to the structural model (Rosseel & Loh, 2022). Therefore, it was first necessary to inspect the validity and reliability of each latent construct from the evaluative framework before applying them in SAM. Suitable unidimensional measurement ensures that each latent variable measures its target construct as theorised, and that poor indicators are identified before applying the model in a structural model (Anderson & Gerbing, 1988; Bollen, 1989; Kline, 2016). Dropping poorly associated items stabilises the construct as far as possible and has the added benefit of reducing the subsequent number of parameters fit in the final model, further mitigating measurement error.

To this end, CFAs were fit to all six evaluation factors to inspect AVE, standardised loadings, and reliability coefficients. Here, the aim was to maximise construct validity by dropping weak items, improving AVE, and construct reliability. In performing this step, the construct validity and reliability for each variable could be inspected before further use. In each case, the maximum amount of data available was used, regardless of whether the construct was measured in the first or second stage of the study. Dependency was accounted for using *lavaan.survey* in the estimation of each construct. Summary statistics for this process are reported in Table 30.

Table 30 Validity and Reliability Iterations of Unidimensional Evaluation Constructs

Factor	Label	β	AVE	ω	ω (if item dropped)	*n* observations
TrTe						93
Iteration 1			.475	.753		
Technology never works (reverse coded)	T1	.853***				
I'm less confident when I use technology (reverse coded)	T2	.671***				
The usefulness of technology is highly overrated (reverse coded)	T3	.843***				
Technology may cause harm to people (reverse coded)	T4	.171			.830	
*Iteration 2**			.622	.830		
Technology never works (reverse coded)	T1	.874***				
I'm less confident when I use technology (reverse coded)	T2	.670***				
The usefulness of technology is highly overrated (reverse coded)	T3	.822***				
ItPF						93
*Iteration 1**			.373	.744		
I like to give feedback on the content I engage with	I1	.629***				
Normally I wouldn't rate any tracks/songs (reverse coded)	I2	.664***				
I only sparingly give feedback (reverse coded)	I3	.547***				

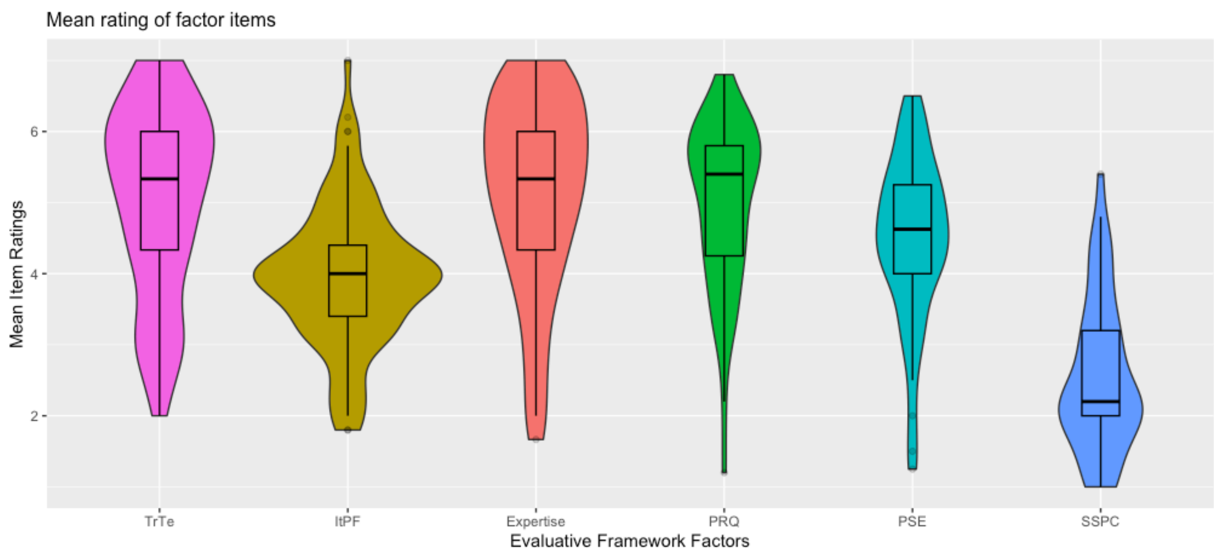
I don't mind rating tracks/songs	I4	.604***				
Overall, rating tracks/songs is not beneficial for me (reverse coded)	I5	.580***				
Expertise						93
*Iteration 1**			.642	.840		
Compared to my peers, I am an expert on music	Ex1	.753***				
Compared to my peers, I listen to a lot of music	Ex2	.843***				
I am a music lover	Ex3	.795***				
PRQ						50
Iteration 1			.611	.886		
I liked the tracks recommended by the system	PQ1	.882***				
The recommended tracks fitted my preference	PQ2	.935***				
The recommended tracks were well-chosen	PQ3	.893***				
The recommended tracks were relevant	PQ4	.749***				
The system recommended too many bad tracks (reverse coded)	PQ5	.732***				
I didn't like any of the recommended tracks (reverse coded)	PQ6	.392			.925	
*Iteration 2**			.699	.923		
I liked the tracks recommended by the system	PQ1	.877***				
The recommended tracks fitted my preference	PQ2	.943***				
The recommended tracks were well-chosen	PQ3	.890***				
The recommended tracks were relevant	PQ4	.751***				
The system recommended too many bad tracks (reverse coded)	PQ5	.726***				
PSE						50
Iteration 1			.454	.828		
I would recommend the playlist generation survey to others	PE1	.769***				
The playlist generation survey is useless (reverse coded)	PE2	.455**		.834		
The playlist generation survey makes me more aware of my choice options	PE3	.686***				
I make better choices with the playlist generation survey	PE4	.678***				
I can find better tracks using the playlist generation survey	PE5	.735***				
I can find better tracks without the help of the playlist generation survey (reverse coded)	PE6	.642***				
Iteration 2			.495	.816		
I would recommend the playlist generation survey to others	PE1	.672***				
The playlist generation survey makes me more aware of my choice options	PE3	.673***				
I make better choices with the playlist generation survey	PE4	.777***				
I can find better tracks using the playlist generation survey	PE5	.824***				
I can find better tracks without the help of the playlist generation survey (reverse coded)	PE6	.530**		.824		
*Iteration 3**			.530	.805		
I would recommend the playlist generation survey to others	PE1	.609***				
The playlist generation survey makes me more aware of my choice options	PE3	.654***				

I make better choices with the playlist generation survey	PE4	.807***			
I can find better tracks using the playlist generation survey	PE5	.851***			
SSPC					
<u>*Iteration 1**</u>			.566	.849	50
I'm afraid the playlist generation survey discloses private Information about me	S1	.860***			
The playlist generation survey invades my privacy	S2	.920***			
I feel confident that the playlist generation survey respects my privacy	S3	.683***			
I'm uncomfortable providing data to the playlist generation survey	S4	.634***			
I think the playlist generation survey respects the confidentiality of my data	S5	.551***			

Note. Iterations marked * represent those carried forward in subsequent analyses. β = standardised coefficient.

Each factor was inspected to gauge its validity and identify misspecifications and/or poorly associated items. Each factor held an adequate reliability coefficient, but improvements were made where it was possible to drop poorly associated items. Iterations marked with an asterisk denote the iterative set of items carried forward for subsequent modelling. To illustrate the ways these constructs were generally rated by participants, Figure 25 shows violin plots relating to the mean ratings of the items for each retained construct.

Figure 25 Violin Plots of Evaluation Factors' Mean Ratings



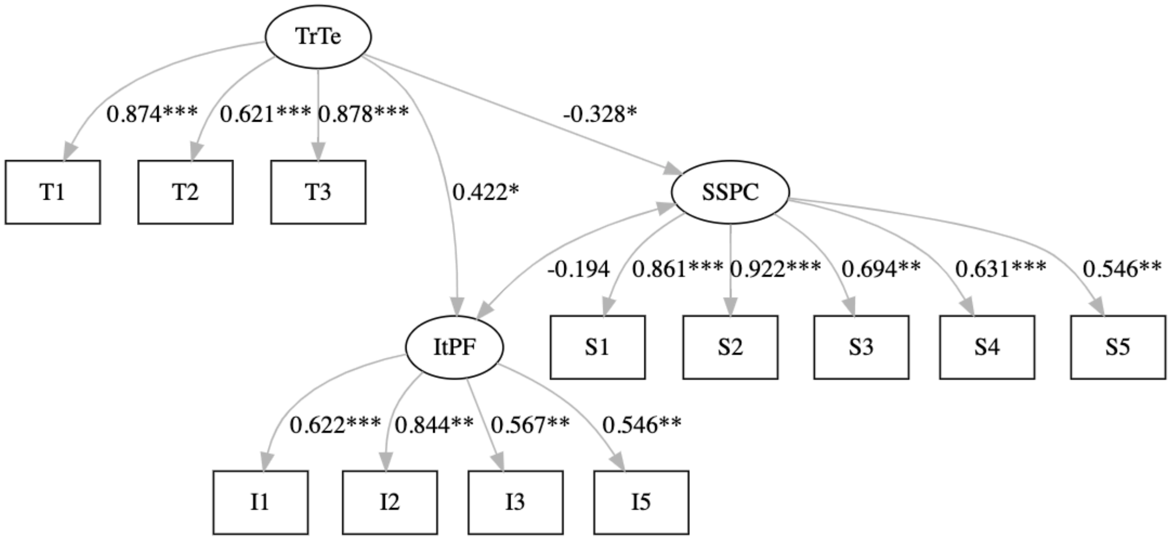
Note. TrTe = Trust in Technology, ItPF = Intention to Provide Feedback, PRQ = Perceived Recommendation Quality, PSE = Perceived System Effectiveness, SSPC = System-Specific Privacy Concern.

Following this reduction and optimisation, analyses proceeded to fitting the SAM models, with the retained iterations of each latent construct.

Trust Model

To fit the model estimating the effect from TrTe on ItPF and SSPC, twostep robust standard errors were computed using ML estimation and the 'local' SAM method, in which the mean vector and variance-covariance matrix of the latent variables is expressed as a function of the observed summary statistics and measurement model parameters (see Rosseel and Loh for a comparison of 'local' and 'global' SAM estimation). As before, the maximum amount of theoretical data permissible to estimate the model was used, which in this case refers to all participants completing both stages of the study. However, the SAM function does not operate with functions of *lavaan.survey*, and so any dependency in the data could not be modelled. As such, any duplicate cases of participant IDs were inspected before fitting the model, in which one participant had $n = 3$ cases. To avoid violating the assumption of independence, therefore, only the participant's first case was retained and the other two were dropped. Hence, the fitted data contained $n = 48$ observations with no dependency within participants. There were no hypotheses relating to T_1 with regard to this model, so there were no inclusion/exclusion criteria based on whether targeted recommendations could have been received. In an initial iteration, one item on the ItPF factor (I4, see Table 30) was found to have a low item loading when estimated in the model ($\beta = .456^*$). This item was therefore dropped to optimise reliability, and the model was re-fit. R code outlining model syntax can be seen in Appendix L, whilst the model results can be seen in Figure 26.

Figure 26 TrTe as a Predictor of SSPC and ItPF using SAM method



As can be seen in Figure 26, TrTe was observed to influence both ItPF and SSPC as hypothesised, whereby higher TrTe resulted in higher levels of ItPF and lower ratings of SSPC. Model-based reliability was high for each construct (TrTe $\omega = .937$; ItPF $\omega = .891$; SSPC $\omega = .956$). Table 31 provides summary results.

Table 31 Summary Statistics for TrTe SAM model

Regressions	Unstandardised Estimate	β	*SE*	*p*	CI Lower	CI Upper
TrTe → ItPF	0.330	.422	0.132	**.012***	0.072	0.589
TrTe → SSPC	-0.309	-.328	0.141	**.028***	-0.585	-0.033
Covariance						
ItPF ↔ SSPC	-0.168	-.194	0.229	.465	-0.617	0.282

Note. β = standardised estimate. *SE* = standard error. *p<.05.

This is consistent with relevant literature, indicating that greater trust in technology yields greater intentions to provide feedback and lower levels of privacy concern. There is, therefore, consistency in the influence of measures in the present study with other applications of the framework (i.e., Knijnenburg et al., 2012). In general, this reaffirms that trust is an important element in moderating privacy concerns, and that trust remains an important aspect to the efficacy of UX. Moreover, this demonstrates that parameters of user-centric evaluative frameworks are indeed applicable to MRS, as the validity of the constructs hold in the present application. Therefore, others generating recommendations may benefit from the application of such measures to better understand how and whether other systems, or characteristics of other systems, translate to SSPC which has implications for longer term use.

Mediation model

Next, the models reported in section 8.3 were fit, namely those assessing whether receiving Targeted versus Non-Targeted recommendations led to changes in PSE, directly and/or indirectly through PRQ (M_1). For this, only cases in which it was possible to estimate audio features were used, which meant filtering cases relating to Routine and Recreational Activities, and cases of Relaxation where $BaA > 2.372$. No cases of Exercise with $PA < 3.872$ were present in the remaining data. Next, any repeated uses were identified, of which $n = 2$ cases from one participant remained in the data, from which only the first case was retained. Hence, the data applied in this model contained $n = 33$ cases in which all conditions were satisfied (i.e., situations in which recommendations *could* have been targeted, were full and complete, and no dependency in the used cases). Of these cases, $n = 14$ were T_0 and $n = 19$ were T_1.

The resulting model (shown in Figure 27) indicated that although PRQ had a strong positive effect on PSE as hypothesised, no difference between the Targeted or Non-Targeted groups could be detected. As such, it does not appear to be the case, based on the available data, that receiving Targeted recommendations led to a positive increase in PRQ or PSE. This may be due to a lack of power, given that comparative group sizes were small, however, this is perhaps only a partial explanation, given that it is equally possible that there was simply no difference between the two groups of cases. Either way, it is difficult to gauge from the current data whether participants' experienced little to no difference between targeted and non-targeted recommendations, or whether there were simply not enough cases to reliably detect this difference. However, Table 32 reports the specific effects for this model, in which wide confidence intervals show a noted lack of precision around the parameter estimates for the Targeted recommendations, indicating a lack of power in the binary predictor (Fox, 2016).

Figure 27 Estimation of M1 Model

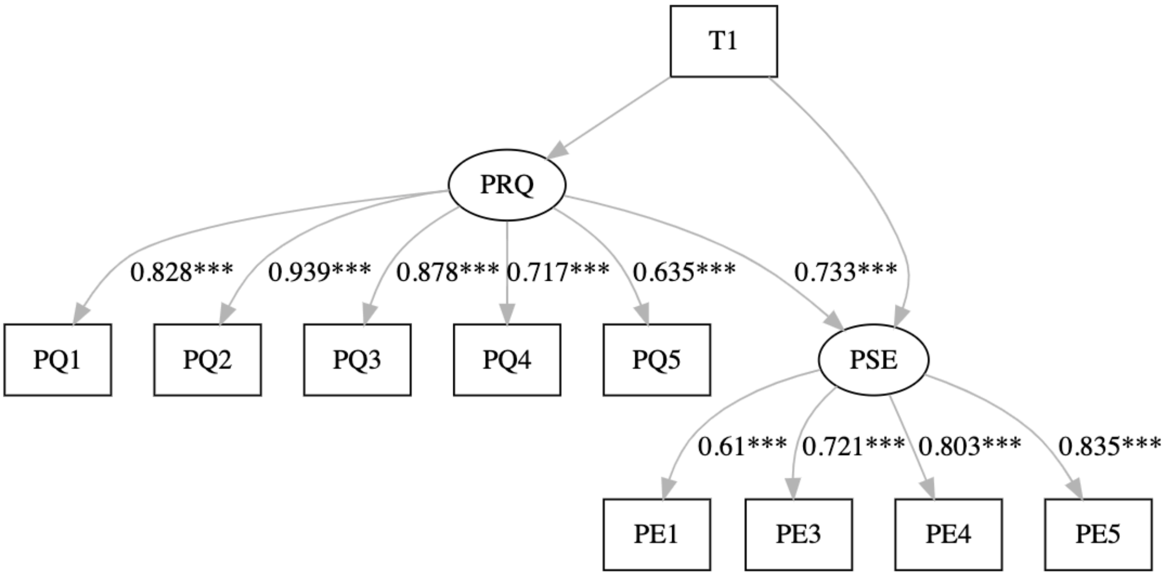


Table 32 Summary Statistics for M_1

Regressions	Unstandardised Estimate	β	*SE*	*p*	CI Lower	CI Upper
T_1 → PRQ	-0.213	-.103	0.369	.565	-0.936	0.511
T_1 → PSE	-0.068	-.035	0.277	.806	-0.612	0.476
PRQ → PSE	0.690	.733	0.218	**.002****	0.263	1.117
Indirect Effect						
T_1 → PRQ → *PSE*	-0.147	-.076	0.258	.570	-0.653	0.359
Total Effect						
T_1 → PSE	-0.215	-.111	0.367	.558	-0.934	0.504

Note. β = standardised estimate. *SE* = standard error. $^{**}p<.01$.

Since T_1 predicted neither PRQ nor PSE in M_1, it was not deemed appropriate to proceed with the assessment of the remaining models outlined in section 8.3.1. This is because the models integrated complex interaction terms amongst the variables, and since T_1 was not a predictor of PRQ or PSE in the simplest model, it is not theoretically permissible to estimate effects in more complex models containing the same paths. Therefore, the modelling process was re-evaluated to explore whether alternative hypotheses may still be derived. It was hypothesised that it may still be the case that listeners' ratings of *Expertise* may still influence PRQ and PSE directly and/or indirectly. Therefore, an alternative model was estimated in which *Expertise* operated as a predictor of PRQ, essentially as T_1 had in M_1. Theoretically, this model was subsumed by M_2, in which there was a hypothesised main effect of *Expertise* on PRQ regardless of the interaction between *Expertise* and T_1. Again, following SAM, the model was fit, which leveraged the maximum amount of data available whilst removing dependency (n = 48) and is shown in Figure 28, whilst Table 33 shows the summary statistics for this model.

Figure 28 Estimation of Alternative Expertise Model

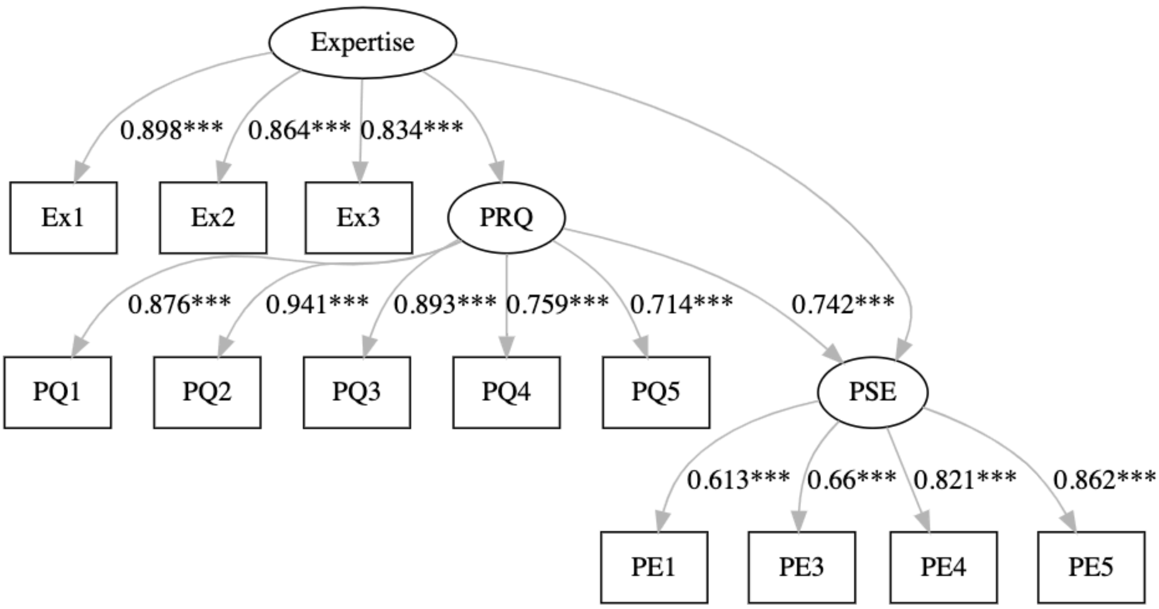


Table 33 Summary Statistics for Alternative Model

Regressions	Unstandardised Estimate	β	*SE*	*p*	CI Lower	CI Upper
Expertise → PRQ	0.060	.094	0.095	.532	-0.127	0.246
Expertise → PSE	-0.056	-.103	0.061	.356	-0.175	0.063
PRQ → PSE	0.633	.742	0.138	**<.001*****	0.362	0.903
Indirect Effect						
Expertise → PRQ → *PSE*	0.038	.069	0.061	.536	-0.082	0.157
Total Effect						

Expertise → PSE	-0.018	-.034	0.084	.828	-0.182	0.146

Note. β = standardised estimate. *SE* = standard error. ***p<.001.

The results indicated that like T_1, *Expertise* did not influence PRQ or PSE. This was surprising, as although n was small, it was hypothesised that in the case of bias-corrected latent variables with continuous outcomes, there should be sufficient ability to detect an effect, unlike the prior model in which comparatively small group sizes introduces further restrictions on test power. Therefore, if there were an effect, this is more likely to have been detected. As before, PRQ influenced PSE in a strong, positive direction. The following section expands on and discusses these findings.

8.6 Discussion

This exploratory study aimed to leverage inferences from an explanatory model to provide listeners with recommendations based on contextual indicators (namely listeners' activity and FML). Specifically, the approach taken incorporates interpretable values for relevant constructs (i.e., the raw metric of Spotify audio features and mean values of FML factors) by adding a mean vector during model estimation of a saturated structural, leveraging the resulting intercept and unstandardised coefficients to approximate values for audio features in given conditions as in the OLS tradition. As such, the values estimated via the model fit to the sample from Study 2 is taken as an estimate of the population values. Though there are limitations with this approach, it nevertheless enables approximations to be made regarding which features should be upweighted and downweighted in the resulting recommendation process. In this sense, this study addresses the fourth aim of this thesis at large (see section 1.2), by illustrating how a psychology-informed approach may be synthesised using relevant information from an explanatory model, rather than by relying solely on data-driven approaches. The underlying aim of this was essentially a proof of concept, namely that an explanatory model can be used in recommendation procedures, hence integrating knowledge about listening situations to approximate appropriate content given information about listeners' context and FML and to implement this. The evaluative framework put forward by Knijnenburg et al. (2012) was subsequently used to evaluate the resulting recommendations, providing a considered approach that incorporates users' experiences and perceptions of a system.

Regarding the findings of this study, it was first found that participants' TrTe had a positive effect on ItPF, as is consistent within the literature (Knijnenburg et al., 2012), though this effect was not system specific. The effect of TrTe on SSPC was also significant, illustrating a negative effect, providing evidence that trait TrTe will influence users' perceptions of privacy concern. This holds implications for researchers seeking to evaluate MRSs in general, as this contextualises and presents a framework which both researchers in academia and industry may be able to apply instead of *offline* and *online* evaluation methods respectively. The application of the framework therefore provides confidence that future work need not be limited to purely computational outcomes but will be able to incorporate users' experiences as an integral component of system evaluation also. Other measures or orientations may still occur alongside such implementations (e.g., A/B testing), but it is plausible to apply such measures in tandem with a user-centric approach to provide a principled, theoretically derived evaluation of system efficacy that balances UX with system outcomes, rather than be dependent on methods which diminish transparency and explainability.

At the system-specific level, however, the targeted recommendations were not observed to yield any meaningful differences in participants' experiences of PRQ or PSE. This may be due to a lack of power in comparative group sizes, or it may be the case that difference was limited between groups regardless. Though difficult to pinpoint, there was no support that the designated OSA influenced differences in UX based on the available data. Consequently, there was no theoretical support to proceed to measuring the interaction terms as hypothesised (see Figure 22). Moreover, isolating *Expertise* as the sole predictor of PRQ and PSE also yielded a null outcome, indicating that the extent to which a user perceived themselves as an expert did not influence their resulting evaluation of the recommendations they had received, contrary to previous findings (Knijnenburg et al., 2012). Yet, consistent with previous research, the effect of PRQ on PSE was strong and positive. This provides confidence in the applicability of the applied measures in the context of the present study, which was solely system specific. Therefore, whilst it cannot be argued the present approach yielded any difference in outcome between the relevant groups, PRQ and PSE were associated in UX measures, adding to confidence that the applied framework is amenable to MRS evaluation.

A useful contribution of this study, therefore, is that the applied framework appears a useful tool to apply in future work with different systems, applying different OSAs. Moreover, this has been drawn from a small sample setting, meaning that the applicability of the Knijnenburg et al.'s (2012) SEM method of evaluation remains practically implementable despite the challenges user-studies often face regarding sample size (Schedl et al., 2022). Therefore, it should be pointed out that future uses of frameworks like this remain up to date with relevant SEM literature to improve methods and extract as much meaningful information from available studies as feasibly possible, especially since small *n* is a common constraint of user studies.

Here, SAM (Rosseel & Loh, 2022) was applied to fix model parameters and segment measurement and structural components to avoid model bias and convergence problems. Alternatively, Bayesian approaches to SEM (Smid & Rosseel, 2020; Smid et al., 2020; Smid & Winter, 2020) may prove particularly useful in future work where *n* is small, but it should be noted that tight prior distributions are required in such settings (McNeish, 2016). However, frequentist uses of the framework in question, such as this study and others, may help inform researchers determining such distributions, should Bayesian approaches be sought. Informative Bayesian approaches may maximise the efficacy of the framework in small sample settings, mitigating a frequent limitation of user-centric studies in general.

8.6.1 Limitations

Limitations of this study include that, when initially distributed, the first wave of data collection experienced targeting by survey bots, automated algorithms scraping social media platforms to overload surveys and increase the likelihood of receiving rewards or monetary compensation (Griffin et al., 2022). This impacted study efficacy, distribution, and communication, plausibly contributing to small *n*. The small sample is limited in its ability to detect complex effects in the theoretical models of interest, for which data analyses were optimised to extract the most meaningful amount of inference from (e.g., using SAM).

Though an unforeseen circumstance prior to data collection, it may be reflected that online research involving compensation is susceptible to such problems, particularly when researchers use social media platforms, for example. As such, researchers would do well to take steps to mitigate the risk of impact from survey bots in online studies, for example by using reCAPTCHA verification and other measures (e.g., Qualtrics' in-built bot-detection function, if available). Though there was a relatively simple fix to prevent further impact in the case of the present study, by closing the original survey link and generating a new version with reCAPTCHA verification, this was nevertheless a circumstance not encountered or considered prior to the study, and so raising awareness of the risk of this negative impact may be of substantive use to researchers in general moving forward.

A further limitation with this study related to the ecological validity of an academic context. Note that the phrasing of SSPC items, for instance, was specific to the study, which was administered through a university associated (i.e., branded) Qualtrics survey, accompanied by procedural aspects not available in many real-world applications (e.g., participant consent, information sheet, researcher contact details). It is hypothesised that the self-apparent academic context of the present in the study might play some role in moderating privacy concerns that may plausibly be stronger with industry systems (e.g., societal trust in universities). Though hypothetical, this is something that future work may wish to consider in terms of design and user-interaction, particularly if the goal is to evaluate a system intended for long-term use. It is difficult to extrapolate this hypothesis further due to limited empirical research on trust between universities and the societies in which they operate (Law & Le, 2023), but other user-centric studies may wish to consider this aspect of UX and the implications this may have to best optimise ecological validity by providing user experiences as comparable to widely accessed systems as possible.

Finally, this study sought to exploratively target recommendations according to preferred genre and estimated values estimated in the OLS tradition via an explanatory path model (that is, estimating values for path conditions by summing the intercept and regression coefficient for a 1 unit increase in x). For the purposes of the present exploration, this illustrates just one way in which psychological data might be leveraged to approximate suitable content given relevant predictors and influencing factors. However, this is constrained by the requirement of participants to select one genre only, which may not plausibly always easily translate to targeted features based on FML. This is, however, a study-specific limitations due to the ways that seed data need to be collated in the applied tools. Future research may consider better ways of incorporating taste so as to not be constrained by such factors when possible. Moreover, recommendation procedures apply machine learning explicitly, in which a predictive model is generated, often at the expense of interpretability (Fokkema et al., 2022). Tree-based methods (e.g., regression and classification trees, random forests), neural networks, and *K*-Nearest Neighbours are just some of the available machine learning approaches available to researchers, with their application to psychology an area of ongoing research and debate (e.g., Shmeuli, 2010; Yarkoni & Westfall, 2017; Fokkema et al., 2022).

The problem with generating out-of-sample predictions via these methods, however, is subject to the criticism that they diminish validity and raise ethical questions stemming from their black-box nature, which typically require large samples to avoid overfitting. This feature of machine learning identifies some form of *what* but sacrifices understanding *why* (Shmeuli, 2010; Yarkoni & Westfall, 2017), and that seldom sits comfortably with social scientists who seek to elucidate the nature of phenomena of interest. On balance, and in being consistent with the motivations and endeavours of this thesis so far, it was opted to generate estimates based on an explanatory model that is likely overly optimistic. Here, explanation was retained over (potentially) more accurate prediction. This is a trade-off deemed pragmatic given the methods and motivations of this thesis at large, but nonetheless holds implications for the subsequent outcomes observed in the efficacy of OSAs in particular.

8.6.2 Conclusion

This study attempted to operationalise the cross-sectional determinants of music selection in everyday life, albeit with mixed success. A tangible approach was outlined, in which a structural explanatory model derived from the prior study was used to approximate audio features according to pathways in a structural model, which were then applied through a decision-tree style process that leveraged contextual prefiltering to direct participants toward resulting playlists. Broadly speaking, this addressed the fourth aim of the thesis, overall, by providing a conceptual approach to integrating knowledge into the recommendation process. In this targeting system, tracks with features closest to the estimated values are returned and thus hypothesised to be more effective overall than playlists that were non-targeted. However, the results did not support hypotheses that there would exist a meaningful difference between participants receiving Targeted recommendations in comparison to those receiving Non-Targeted recommendations. This may be due to limited differences between the suggested playlists for example, that did not lead to drastically different weightings between the playlists targeting estimated values and those left empty, or to limited sample size and test power with which to conduct group comparisons.

Regarding implementation, future work should seek to remain updated with advances in prediction-oriented explanatory modelling. For instance, the SEM-based out-of-sample prediction method proposed by de Rooij et al. (2022) provides a means of estimating effects using the explanatory model explicitly. At the time of writing, this is, however, constrained solely to models in which all variables are continuous, and so cannot yet be applied to models with discrete predictors, for example. A workaround was applied in the present context, by noting that in saturated models the estimates from SEM-based out-of-sample prediction are identical to those in OLS, and so the principle was applied manually in this study. Given that CAMRSs, however, as well as music listening research, often treats situational variables as discrete in nature, advances in such methodologies are worth researchers attention. Alternatively, a different approach could be taken in which situational predictors are measured through continuous means, such as the DIAMONDS model applied by Behbani and Steffens (2020). Using such constructs, in conjunction with continuous FML and audio feature measures, would be compatible with de Rooij et al.'s (2022) method, and thus provide a means for out-of-sample prediction in sufficiently large samples. This is one suggested avenue for future research on FML in everyday life, and also for those looking to further the development of psychology-informed recommender systems.

When it comes to evaluation, it was observed that whilst the OSA was not contributing to any meaningful conceptual difference in outcome, that those who reported higher PRQ generally reported higher PSE as a result. The same can be said regarding the influence of TrTe on SSPC, coupled with the hypothesised negative effect of TrTe on ItPF. Theoretically consistent observed effects within relevant measures of the framework as applied provides evidence that despite the limitations of the present study, the user-centric framework presented by

Knijnenburg et al. (2012) holds enough abstraction to provide insight into a system's performance, even in small sample settings. The observed effects demonstrate that the linear relationships amongst the relevant constructs hold relevance to UX under these limitations, and as such, corroborate the arguments in favour of the framework made by Lex et al. (2021) in particular. Future research on MRSs should, therefore, consider applying this framework, or at least others like it, to evaluate the effectiveness of a system under review. One caveat, however, is that researchers should inspect the validity and reliability of the unidimensional constructs before applying them where possible, as this process revealed poorly associated items in the present study. It may be the case that such results were specific to this sample, however, if other uses of the framework yield similar results, then there is an argument for updating the framework to reflect new insights into poor items specific to the MRSs.

In summary, the exploratory approach outlined in this study provides a method by which to implement a psychology-informed approach to music recommendation, supplemented with user-centric evaluation. These complementary aspects of curation and evaluation extend the findings of psychological research by applying such knowledge in place of black-box machine learning, whilst considering the UX as an integral part of understanding system efficacy. Other researchers may consider different approaches, and there is indeed scope to retain machine learning approaches (especially if they are supervised), but it is hoped the principles and methods outlined in this third and final study contribute to an acknowledgement that psychological research is able to contribute significantly to system generation and evaluation. Such principled approaches may help mitigate data-dependency in existing systems, whilst enhancing trust and explainability at a time when scrutiny of data usage and automation is increasing.

9.1 Discussion and Concluding Remarks

This thesis has sought to make a novel contribution to knowledge on several fronts, including to the music psychology domain on the subject of FML, the role of psychology in and approaches to music recommendations, and the cross-over between these two areas. This has been to address the underlying research question of interest, as outlined in the introduction to this thesis: By what means might it be possible to curate everyday listening through a psychology-informed approach to music recommendations?

This final chapter therefore aims to summarise the key findings of this thesis, as well as provide context and direction future work may consider building on. This is partitioned into brief discussions of the implications for the relevant domains this thesis has been concerned with and supplemented with reference to the overall aims of this thesis as a whole. For reference, these aims (intended to address the primary research question) are restated here before proceeding to the subsequent discussion:

1. To be able to validate a measure of FML from the utilitarian perspective
2. To be able to associate a validated measure of FML with listening contexts and music content
3. To be able to associate musical content with listening contexts
4. To be able to propose an actionable method of integrating knowledge generated in steps 1-3 into a recommendation procedure

The first of these aims has been realised through the study presented in Chapter 6, in which a utilitarian measure of FML was generated based on an existing theoretical framework aligned with this perspective. The second and third aims were achieved by triangulating this FML construct, alongside listening activities and audio features. This provided relationships into the direct and indirect relationships amongst the three focal constructs. Finally, the fourth aim essentially sought to bring these preceding components together to formulate an approach to estimating appropriate music content given cross-sectional information about listeners' activities and FML. In this, a procedure was outlined as to how this information may be used to approximate and implement an approach to providing listeners with music recommendations, ultimately linking each aim, and addressing the underlying research question. What follows are more detailed discussions relating to the relevance and utility of the implications of this research for the areas of interest discussed.

Implications for functions research

Throughout this thesis, *functionality* has been understood as stemming from goal-orientated constructs. Consistent with *uses and gratifications* approaches to understanding utility, the first study presented in Chapter 6 sought to extend Maloney's (2019) contribution by developing a quantitative measure of FML from the utilitarian perspective, formed from an item pool derived through qualitative assessment of the CFF. The construct generated is formed of five factors representing FML: *Identity and Social Bonding*, *Emotion Regulation*, *Focus and Concentration*, *Background and Accompaniment*, and *Physiological Arousal*. This was uncovered through factor analyses, in which the latent structure was identified in unrestricted conditions (EFA) and inspected in restrictive (CFA) conditions. The specification of the more concise 23-item model was deemed prudent in the interest of mitigating length and subsequent burden and was cross-validated in the two later studies. This cross-validation provides confidence in the validity and reliability of the construct, and thus the first tangible contribution of this thesis is a utilitarian measure of FML. However, caveats remain, for example that the samples from which data were gathered in this thesis have been presumably based in the global north and may not be representative of those in other cultural contexts. As such, this model may not be exhaustive of all key underlying dimensions of *functionality* given the role that culture may plausibly also have. As such, a need for cross-cultural validation and/or exploration remains. This may be an area that future research could look to establish in the case of this specific model.

The factors of the retained construct were later observed in Study 2 to be influenced by listeners' concurrent activities in ecologically valid settings. Specifically, there were theoretically consistent indicators that activities such as Work/Study and Exercise influenced the extent to which participants reported using *Focus and Concentration* and (in the case of the latter) also *Physiological Arousal*. Such results provide evidence on two notable points of interest: (1) that factors relating to the latent constructs are rated according to theoretical consistencies, adding to assuredness in the validity of the model, and (2) doing so provides further evidence in general of the link between context and FML. In addition to the above, not only was it found that overall, *functionality* was seemingly affected by activity, but also that activities led to changes in the audio content of participants' music selection.

Following dimension reduction, three components (*Arousal*, *Valence*, *Instrumentalness*) were extracted from a subset of audio features from the Spotify API, and these were triangulated in a path model as outcome variables, in which they were predicted to be affected by activity directly and/or indirectly through FML (see section 7.4.6). This path model yielded a small but theoretically consistent number of meaningful effects, outlining how activities influenced not only *functionality* but also music selection in everyday listening episodes. This contribution holds relevant applications as it first contextualises and applies MIR-generated audio features into data collated in a psychological study. Studies typically rely on self-reports to describe the affective content of music selected in everyday life, but this study demonstrates it is possible to obtain audio features from an MIR application and to apply these directly in psychological research, which may hold further uses in related music research. The finding that there are, overall, theoretically consistent correlations between the MIR-generated audio features and self-reports also provides further confidence that there is comparative consistency in the features attributed by the application and participants' perceptions, thus extending the validity of such an approach. This extends options and possibilities for researchers in this area, as it provides a tangible example of how MIR-generated audio features can be accessed and used in psychological research. However, it should also be acknowledged that this seldom removes response biases during such studies altogether, as FML measures remain dependent on self-reports, for example. As such, considerations as to the noise that may be present in the measurement of other constructs remain relevant. Overall, this process broadly achieved the second and third goals that were initially outlined, which emphasised the need to associate the three focal constructs relevant to this thesis (contextual factors, *functionality*, and content/characteristics of music).

Implications for psychology-informed recommender systems

Regarding the second strand of substantive interest to this thesis, the third and final study sought to essentially operationalise the findings of the second study (or, at least, the data from the second study), and reverse engineer the identified paths relating to music selection. In this, prefiltering was used to build a decision-tree-style process in which recommendations are provided based on node outcomes, coupled with indications of genre preference. Raw values for audio features obtained from the Spotify API were used in conjunction with mean scores of the five latent factors which, when fit with the mean vector in model estimation, provided model intercepts and unstandardised regression coefficients that could be interpreted in the OLS tradition (de Rooij et al., 2022). In extending the logic of estimating values based on such units (by plugging in an unstandardised regression coefficient to model intercepts for given predictors; Preacher & Kelley, 2011), a series of estimated values for specific audio features in specific paths were generated. Recommendations targeting these values were then generated using an API wrapper in R (Thompson et al., 2022) to call functions and gather recommendations based on such values, running through 10 genres for each conditional branch. Hence, this study first synthesised an approach to generating recommendations for specific situations, based on the saturated structure.

Though this study was of mixed success, the retained data is a generative proof of concept that the general outcome of this approach is consistent with prior theoretical frameworks insofar as listeners' perceptions of quality influence perceptions of effectiveness. However, OSAs of the present approach were not able to detect differences between received recommendations that were targeted to estimated values and those that were not. This may be due to a limited sample size, but also it may simply be the case that the playlists providing targeted recommendations did not yield particularly strong differences in user perceptions of PRQ and PSE. However, the fact that PRQ influenced PSE and TrTe influenced ItPF and SSPC, does convey information about the applicability of the evaluative framework to the overall approach. Insofar as such relationships are consistent with the framework as applied in other settings, the ability to extend this into a psychology-informed setting is constructive.

Though limited, future work may wish to build on and reapply the applied framework which was, overall, effective in evaluating recommendations. In other words, though the system/output presented in this thesis cannot be argued to be in anyway more effective than other approaches, the combination of factors that led to study design (e.g., selection of OSAs, applied theoretical frameworks with unidimensional latent constructs) provides a tangible example of the ways in which psychology-informed recommender systems may be generated and evaluated (using a user-centric approach). Future research may wish to carry forward and further improve these motivations to reduce data-dependency in systems and curate experiences by incorporating relevant knowledge of given domains. Nonetheless, this final study implemented the fourth of this thesis in general terms, in which it was intended to employ knowledge about FML, context, and content to implement a recommender procedure (though there remains room for refinement to generate higher-resolution recommendations).

9.2 Future Directions and Recommendations

An important aspect relating to limited take-up in generating psychology-informed recommender systems is the perceived trade-off between explanatory and predictive modelling. Recommender systems typically rely on machine learning techniques that are black-box in nature. Though such models may yield reasonably high-predictive accuracy, they (1) require (very) large amounts of data to do so, and (2) are subject to a series of biases (e.g., biases in training data leads to biases in the trained model, which leads to biases in outputs). Moreover, the general black-box nature of machine learning models means that the decisions behind subsequent predictions are unexplainable. The lack of explainability further reduces transparency and diminishes trust. Explanatory models, on the other hand, are theoretically derived and knowledge driven. Because they are typically applied in psychological studies, for instance, knowledge and understanding of the data characteristics can be more closely considered and corrected for, depending on the area of substantive interest. Such models detect linear relationships amongst variables of interest (e.g., via regression techniques), but tend to be overly optimistic as they are intrinsically tied to the data on which they are fit.

This makes explanatory models unappealing to predictive endeavours, as they tend to be overly optimistic and hold poorer predictive accuracy than machine-learning outputs that detect lower-level linearities. This broad methodological problem has not yet been solved, and in this thesis the latter of these two approaches was taken to approximate optimal audio-content values for new cases of music listening where possible (the frank limitations of which have been acknowledged). However, given that recommender systems are generally effective at predicting long-term taste, they do not provide recommendations based on short-term dynamic needs of listeners. To provide recommendations for short-term, cross-sectional needs, MIR-research has sought to provide CAMRSs (Wang et al., 2012). These systems hold particular limitations as the training data often includes mobile phone information, such as microphones and accelerometer data, plausibly reducing transparency further and exacerbating privacy concerns.

However, there are promising signs in the methodological literature that out-of-sample predictions generated through explanatory models may soon have simpler means of estimation with reduced trade-offs. De Rooij et al.'s (2022) method of estimating predicted values in SEMs is one such example, in which the model-implied variance-covariance matrix (Σ) and mean vector (μ) are used to produce estimates for y given some value of x. At the time of writing, however, this requires all variables in the model to be continuous, as in models with categorical predictors there is no joint distribution of x and y through which Σ and μ can be leveraged for prediction (hence this method could not be applied during the third study). Methods will continue to evolve, however, that allow researchers to accommodate categorical variables in such models, though this is not yet available. Future research would do well to remain well-informed on such developments and explore and apply such methods as they become available to generate more accurate predictions of target variables via explanatory models. These may reduce the need to compromise on overfitting via the traditional regression methods or lose interpretability through regression-based machine learning models. This specific method holds future implications for the practical capabilities of psychology-informed recommender systems in particular.

We may hypothesise that the future development of predictive modelling in the social sciences will provide researchers developing MRSs with a meaningful opportunity to synthesise content-based recommendations through an understanding of psychology-informed behavioural modelling and principled methodological applications as they develop. The use of such methods may help reduce the amount of data required for predictive modelling, enabling researchers to operate at smaller scales and reduce dependency on user transaction data. This has been argued here to hold particular relevance for CAMRSs, for which data-dependency could prove a real problem should systems not move away from triangulation approaches to identify users' contexts. Moreover, consider that one of this thesis' central criticisms of recommender systems is that they implement a *what* without understanding *why*. Maximising the effectiveness of our shared knowledge of *why* may be invaluable, therefore, in increasing transparency and curtailing distrust. When such approaches become available in practice, it would be interesting to consider applying them in a similar design to the third study, with user-centric evaluation. Such approaches are complementary and, as the study showed, actionable if complex.

9.3 Concluding remarks

In summary, the crux of this endeavour has been to outline and explore an approach by which recommendations can be targeted according to information derived from psychological data. Novel contributions to the field include the development of a psychometric instrument measuring FML from the utilitarian perspective, the integration of audio features via programmatic use of online MIR tools, and a conceptual process that estimates audio content (applied and evaluated via a user-centric study design). Pros and cons of the approaches used have been discussed, and there remains room for further refinement of the applied approaches. Since the focus has largely been on methods throughout this thesis, four methodological suggestions are made from which future research may benefit given the reflections contained within this thesis:

1. To make use (where possible) of weighted effects coding to remove the need of selecting an arbitrary reference variable in structural path models
2. To re-introduce person-level variables to explore additional effects at multiple levels and explore their relevant interactions

3. To apply variable selection methods (e.g., Ridge regression, LASSO) to help formulate tight path models that increase *df* when applying SEM (to enable the assessment of model fit)
4. To consider and keep updated on methodological progressions with regard to SEM-based on out-of-sample predictions in general

Overall, this thesis has aimed to explore an approach of generating psychology-informed music recommendations by iteratively modelling and characterising relationships between key constructs. Inferences have been drawn regarding the ways in which FML differ according to concurrent activities, and how these translate to changes in the audio content of music selection. By applying MIR-generated measures of musical content as indicators of music selection, it was outlined how such measures could be reapplied to target recommended tracks by targeting values of such indicators. Although this was of mixed success, this nonetheless threads through the original motivations of this work, with an applied example, demonstrating that such approaches are actionable. Scalability of this work is limited, however, and future endeavours may not only seek to integrate some of the reflections mentioned, but also to apply this at greater scale if resources allow. Music listening is an extremely complex yet ubiquitous phenomena, and difficult to measure with complete ecological validity. Continued use of ESM may also be particularly helpful, especially if/when integrated as part of a study on curating music listening, for example.

One broader benefit of this, however, is that this highlights some of the complementary aspects of music psychology and MIR research. This thesis has sought to contextualise and provide some meaningful contributions to areas of both substantive and methodological interest to researchers in both fields and highlights that potential collaborations may be of benefit to those working in these areas. Collaboration between researchers of this respective fields may yield stronger outputs to the benefit of listeners also, therefore furthering our principled understanding of given phenomena, the explainability and accessibility of technological applications (e.g., that underpin curation and listening), and efficacy in the listening experience itself and the benefits this may provide people with. Through continued collaboration and interdisciplinarity, the efficacy and beneficence of listening technologies may be maximised, providing meaningful outcomes across a broad array of outputs, not exclusively that in relation to recommender systems.

It is therefore hoped that this thesis provides an illustration of the benefits of this crossover and would hope that as the pool of expertise in relation to the areas of interest become more closely integrated, then continued growth and development of issues such as those discussed here may be broadly beneficial to all stakeholders. Beneficiaries of such collaborations would ultimately include end-users, who may be provided with a greater understanding of why they receive the suggestions they ultimately do when engaging with online systems or streaming services. This need not be limited to CAMRSs as this thesis has explored but may span a myriad of other applications. For example, it seems plausible to consider how collaborations between MIR and music psychology researchers may be able to provide targeted recommendations relating to listeners' situational mood, plausibly providing means of curated access beneficial to individuals' mental health and wellbeing. Though just one example, this highlights that potential benefits of the approaches and motivations outlined within this thesis may provide meaningful outcomes that maximise the beneficence of applied research.

To conclude, listeners engage with music for a myriad of complex reasons. The situations in which listening occurs influences *functionality*, which in turn affects music selection. Better understanding of these behaviours may serve to increase knowledge of not only music's utility in everyday life, but also stimulate novel approaches to curation. In an increasingly digitised world where music may be listened to on demand, it is more important than ever to appreciate what drives this diversified and seemingly universal behaviour. Coupling this understanding with appropriate methodological approaches may, however, situate knowledge at the heart of curation. Data-dependency constitutes an ethical and increasingly legislative problem, and as such, methods of curation that reduce such dependencies are amenable to both listeners and industry moving forwards. Therefore, given increased scrutiny on the kinds of data that systems access and the ways in which these are applied, investing, and engaging with principled behavioural research that retains interpretability stands to provide clarity, transparency, and explainability in light of such challenges. It is therefore mutually beneficial to engage with psychological research relating to the use of media like music, and to uncover ways of applying this knowledge proactively in systems, rather than relying on systems to uncover low-level linearities as a substitute for applying this knowledge. To this end, it is hoped this thesis is a pebble in the pile of getting this balance right, and that research may build upon some of the approaches that have been applied here and elsewhere to the betterment of all stakeholders.

Appendices

Appendix A: Definitions of the 53 functions of music listening included in the CFF (adapted from Maloney, 2019)

Cognitive Functions

i. Aestheticisation & filmic listening

Music is applied to allow visual stimuli to become filmic. Listeners become passive observers as the world around them plays out a personalised soundtrack. This mostly occurs through the use of portable listening devices which allows listeners to observe any synchronisation between the listening, environment, and perceived moods or emotions (even if these emotional factors are not necessarily felt by the listener).

ii. Creativity

Music is applied to enhance or maintain the listener's creativity. Music may function as a source of inspiration and thus, may become a tool that allows the listener's imaginations to “flourish” and act as a means through which to fantasise. However, this does not necessarily predicate occurrences of flow states or cognitive visual imagery.

iii. Distraction

Music is applied in order to act as a stimulus to distract from factors such as current activities or thoughts. Distraction may relieve boredom and occupy unused attention during tasks; or become enacted when an individual's full attention is not required in a given task.

iv. Earworm fulfilment

Music is applied to satisfy and/or clear a musical phrase that is ‘stuck’ in the listener's minds. This may involve utilising music in order to fulfil a current earworm and allows listeners to remove focus on a piece and thus remove the distraction by satisfying the listener.

v. Focus & concentration

Music is applied as a means to facilitate greater concentration on certain tasks, or to stimulate focus within listeners. Generally, music can narrow the parameters of attention, or allow listeners to attain desired levels of concentration by preventing other, external stimuli from distracting the listener. Within this function, individuals may also achieve some form of flow state during accompanying activities.

vi. Habitual listening

Music is applied based on its being habit. This function generally manifests itself once listeners have developed, or learnt, expectant behaviours in certain situations or activity. In this regard, music may function to satisfy learned responses.

vii. Mental state

Music is applied to enable listeners to attain particular mental states. This allows the listener to select pieces of music that are congruent towards the attainment of specific mental states desired by the listener. Maloney (2019) writes that "the exact nature of these mental states may not be easily explained but listeners are aware of the specific state required" (p. 250).

viii. Motivation

Music is applied to apply psychological motivation. This function facilitates listeners the ability to select music that is congruent towards the achievement of particular psychological goals. There may be consequences for emotional or physical modes of action by enacting psychological drivers that lead to further action.

ix. Reflection

Music is applied to sustain and enhance healthy psychological states. Music may serve as a reflective tool to afford individuals the ability to analyse their perceptions, experiences, and behaviours from alternate perspectives. This can allow an individual to perceive thoughts and feelings within music that are expressive in terms to understand their own world and themselves.

x. Visual imagery

Music is applied in order to generate mental visual imagery that is spontaneous. This expands creativity functions insofar as it allows listeners to build mental images in their heads, that concern the music or extra-musical features. Such images may be abstract, programmatic in nature or simply offer listeners the means to psychologically play our personal scenarios.

Emotional Functions

i. Entertainment & hedonic motive

Music is applied for the explicit purpose of enjoyment or entertainment. This function directly involves the induction of positive emotional states through musical engagement, rather than engaging with music for the purposes of attaining cognitive goals.

ii. Escapism & venting

Music is applied to distract from stressors and attain catharsis. Music may therefore serve as an alternative stimulus on which listeners are able to infer attention and focus, facilitating temporary escape from stressful events or situations. This may also serve an ability to alleviate stress and negative emotions.

iii. Solace

Music is applied in order to generate feelings of comfort and solace within listeners. This allows music to be selected by an individual to provide feelings of comfort and kinship, or to facilitate particular emotional states. Whilst the music itself is not a presence it provides emotional support to listener, although this may be somewhat detached.

iv. Therapy

Music is applied as a therapeutic tool to mitigate negative emotions. This may refer to both clinical scenarios of music therapy (within psychotherapy) or self-administered listening therapy. Listeners can find meaning within musical stimuli that allows them to mitigate negative emotions or moods.

Specific Regulatory Strategies

i. Accentuate emotion/mood

Music is applied in order to enhance or accentuate a listener's current emotional state. This allows the listener to intensify their current states via emotional induction or through concurrent activities that lead to increased or deeper levels of emotion or mood.

ii. Change or shift emotion/mood

Music is applied in order to alter listeners' current emotional states. It allows listeners to alter their current mood through emotional indication through concurrent activities leading to altered emotional states. This typically relieves negative emotions; however, it may also be used to attenuate positive emotions deemed too intense.

iii. Express or convey emotion/mood

Music is applied to make the listener's emotions apparent to themselves and/or others. Listeners may use music in order to reflect and reinforce their own mood states. Further to this, it may allow individuals to express their emotional states through physical action, such as crying or smiling, or via third-party uses (mixtapes or music presentation therapy).

iv. Regulate & maintain emotion/mood

Music is applied to maintain listeners' current mood states. Listeners use music in this regard to sustain or reinforce current states without accentuating or attenuating their emotional states. Thus, this function facilitates listeners' ability to manage current emotions, positive or negative, and preserve them in spite of external influences.

v. Trigger or elicit emotion/mood

Music is applied in order to engender a specific emotional state. Music may be employed to initiate a specific mood state, often through rehearsed or understood musical stimuli. However, this may only occur when listeners' moods can be considered to be somewhat "neutral", as otherwise listeners are initiating *Change or shift* functions.

Physiological functions

i. Accompaniment

Music is applied to accompany, or soundtrack everyday events. Unlike closely related cognitive functions, music is employed in this regard as a passive activity to accompany physical tasks but does not provide perceptions of pace. In this instance, music improves the current activity in some regard. This function is often reported during travelling.

ii. Activation, Arousal & Response

Music is applied to alter levels of physiological arousal or facilitate physical responses from listeners. This may occur before, during or after a particular task that allows individuals to prepare, maintain or readjust arousal levels as are required.

iii. Dancing

Music is applied in order to act as a stimulus for physical movement (dance). Dancing can occur incidentally, but functional employment and music selection can also be a directive action on the listener's part.

iv. Enhance Activity or Ability

Music is applied in order to improve task performance or otherwise reduce levels of perceived effort. In this regard, music may be employed as a cognitive tool to alter physiological performance which may subsequently improve performance.

v. Environmental Control & Aural Filtering

Music is applied as a blocking tool against external stimuli from the outside world. This function is primarily reported during the use of portable listening devices. Musical stimuli are utilised in order to prevent undesirable external auditory sources from distracting the listener. This may also be used in seeking to avoid uncomfortable silences and distract the listener from the location in which they find themselves.

vi. Pacing & Movement

Music is applied to track time or motion to tasks. For instance, music may be applied as a mechanism for maintaining rhythm for specific physical activities, or by acting as a pacer for specific movements. This allows for a level of entertainment to manifest itself with music being used to enhance motor control, fine muscle movement and perceptions of increased stamina levels.

vii. Physical Discomfort

Music is applied to induce physical discomfort to oneself or to others. Music may be used in order to cause physical discomfort via uncomfortable high frequencies or high volume. This is mostly reported in military situations or in the pursuit of control in public areas. It also causes discomfort to oneself to induce emotional responses in order to enhance physical aggression leading to perceived increases in stamina and strength.

viii. Structuring Time

Music is applied as a means by which listeners keep time. Music may allow listeners to track time's passing or compartmentalise physical tasks into chunks of time. This function may also serve to reduce the perceived length of journeys as well as other tasks, likely by means of distraction.

Social functions

i. Approval & Cultural Capital

Music is applied to obtain approval from the wider social group on the part of the individual. This function allows the individual to gather cultural or social capital from a group by engaging in listening behaviours that are socially acceptable. In general, this function pertains to Western art music, however, may also be present in subcultural settings to achieve similar results.

ii. Boundary Demarcation

Music is applied to serve as an external signifier of disengagement from social situations. This may typically be employed by listeners through portable/personal devices to create a perceived barrier between the listener and the social setting. This serves a defensive utility insofar as it allows the individual to detach themselves from unwanted interactions or social engagement.

iii. Communication

Music is applied as a means to communicate between individuals and groups. This gives listeners a tool through which they express emotions, thoughts and meaning with others (in particular ones that cannot be easily verbalised). Music may subsequently serve as a topic of conversation which may encourage communication and interaction with others.

iv. Control & Conformity

Music is applied to modify or control intra-group behaviour. This allows culturally coded extra-musical messages to be expressed to an individual from the group, thus communicating acceptable behaviour and compliance within the group. Musical stimuli may therefore be applied as a means for behaviour modification.

v. Group Identity

Music is applied for the purposes of constructing group identity. Music may be used to align individuals together to form a group. This subsequently allows groups to demonstrate their values and culture to one another via music. In turn, this may serve to identify and reinforce intra-group points of connection. Music may therefore serve as a source of collective identity.

vi. Interaction & Bonding

Music is applied to allow individuals to bond and interact. Music subsequently allows socialisation and feelings of belonging to occur. This allows interpersonal bonds to manifest and intensify and may be applied in relationships that are familial, fraternal, or romantic; subsequently providing interpersonal cohesion.

vii. Maintain & Express Cultural Values

Music is applied to outwardly express the values of a group or culture. This allows a group to bolster and maintain their cultural values by referencing musical stimuli. The same stimuli may serve as external demonstrations of these values to extra-group individuals.

viii. Surveillance

Music is applied to monitor the listening behaviours of other groups or individuals. This allows others to monitor and qualify the listening behaviours of other individuals. This is particularly relevant to modern settings through the digital sphere thanks to the "celestial jukebox". Subsequently, commercial bodies may monetise the listening habits of individuals.

ix. Symbolic Representation & Difference

Music is applied to represent and differentiate groups and individuals. This function allows groups to symbolise themselves through music, and exclude others based on their musical preferences or group. Music may serve as a 'totem' to allow a group or an individual to differentiate or separate themselves from a wider group.

Identity Functions

i. Create & Maintain Identity

Music is applied as a tool to develop an individual identity. Music may serve to aid individuals in the exploration and definition of their personal identities. This may also assist in the fluidity of identity.

ii. Express Identity & Values

Music is applied as an external expression of individual values or identities. This allows individuals to enhance or maintain their identity by referencing music that is reflective of that identity. Those same musical stimuli may function as an external demonstration of identity and values to other individuals or groups.

iii. Personal Development & Understanding

Music is applied to help personal growth and development. Individuals may use music to facilitate personal development and change. This may also help in reaching maturity and increase social, personal, and emotional understanding within individuals.

iv. Promote Autonomy & Agency

Music is applied to offer individuals feelings of control and agency. This allows the listener to enhance their feelings of control and facilitate the occurrence of perceived mental or social emancipation. This is particularly present during adolescence but is not exclusive to that life period.

v. Symbolic Representation

Music is applied to help individuals represent themselves. Music may be used to symbolically represent or 'stand-in' for an individual within wider social contexts. This subsequently allows an individual to differentiate themself from the wider group by using music that is viewed as inappropriate by the group as it deviates from the group identity.

<u>Meta-domain Functions</u>

i. Background

Music is applied to provide a sonic background or stimuli within a given space. Whilst this does not necessarily require a specific mood or atmosphere to be present, the music nevertheless adds stimuli within the listener's environment.

- *ia. Create & Maintain Atmosphere*

Music is applied in order to construct and sustain an atmosphere within a social setting, either as part as a group or whilst alone. The music may enhance feelings of ambience within the environment and is primarily used to ease social interaction. This may also stimulate cues concerning appropriate or desired modes of action within listeners.

ii. Company & Music as Proxy

Music is applied to provide feelings of company for the listener. Listening devices act as physical companions to listeners and/or the music listening that they facilitate can act in the stead of social interaction and mitigate feelings of loneliness.

- *iia. Silence Avoidance*

Music is applied to mitigate silence when alone. Whilst the music may not act as a proxy or substitute for company, it removes silence within uninhabited spaces and thus allows individuals to mitigate feelings of loneliness or of being physically alone.

iii. Memories

Music is applied to reference or recall memories. Music may serve to aid memories but more often serves as an autobiographical referencing tool. A listener can use music to elicit memories or relive their own pasts, via previously established music associations. This is particularly relevant when recalling previous phases of personal identities.

iv. Mimesis & Matching

Music is applied in order to match or mimic a space, place, time, or external feature. This allows the listener to match their music listening to an external feature, such as location, weather or time, perhaps through auditory mimesis. This may also allow listeners to feel active or invested within a situation or to "feel right" in that situation.

v. Musicking

Music is applied for the purpose of music making. The act of music making, such as playing, singing, conducting, or learning, can be a function in and of itself. Whilst a substantial portion of music making may take place through a cognitive process the subsequent expression is physical and manifest. This can occur in scenarios more formal than most, such within orchestras but also within amateur environments, such as at home.

- *va. Aesthetic Appreciation*

Music is applied to serve as a higher art medium for the purpose of appreciation and scrutiny. As opposed to functioning for the purposes of pure entertainment, music in this regard is often engaged with in focused manners. The aesthetic appreciation manifests without emotional engagement on the listener's part by rather functioning as a purely intellectual process.

- *vb. Listening Behaviours*

Music is applied by listeners to enhance personal musical understanding and taste. This allows the listener to engage with music for the purposes of exploration which may subsequently initiate further functions of music listening if reused once the music is known to the listener.

- *vc. Musical & Lyric Analysis*

Music is applied as the primary focus of analytical listening, thinking or behaviour. This provides listeners with cognitive stimuli which facilitates analytical thinking to take place. This may allow listeners to specifically engage with the lyrical content of music, thus aiding in a process of identity creation. However, that is not consistent within this function.

vi. Relaxation & Stress Relief

Music is applied to mitigate stress and/or facilitate relaxation. This may serve to act on cognition, emotions, and physiology either simultaneously or exclusively. Music may serve as a coping mechanism and allows stressors to be temporarily or permanently removed. This may subsequently provide listeners with the means to alter their perceptions of the stressors.

vii. Situational Relevance

Music is applied within specific scenarios to provide clues and/or cues as to modes of action deemed appropriate. Music may have meanings that are culturally encoded within these situations which are likely to vary between different cultures. Nevertheless, this function may act as a lens through which listeners are able to decipher extra-musical semantic prompts that remind listeners of appropriate modes of behaviour within the situation.

Appendix B: Study 1 questionnaire (initial 114 item pool)

Q1, What is your age?

- 18 - 24 (1)
- 25 - 34 (2)
- 35 - 44 (3)
- 45 - 54 (4)
- 55 - 64 (5)
- 65 - 74 (6)
- 75 - 84 (7)
- 85 or older (8)

Q2, What is your gender?

- Male (1)
- Female (2)
- Non-binary / third gender (3)
- Prefer not to say (4)

Q3, On a normal day, how much time do you spend listening to music?

- 0-1 hour (1)
- 1-2 hours (2)
- 2-3 hours (3)
- 3-4 hours (4)
- 4+ hours (5)

Q4, On a scale from Never (being you do not recall ever using music for that purpose) to Very often (being you use music for that purpose very frequently), to what extent do you use music to...

Cognitive Functions

i. Aestheticisation & Filmic Listening

Item 1: To make your environment seem more cinematic or 'film like'
Item 2: To synchronise, or 'line-up', with events in your daily life as though it were a soundtrack

ii. Creativity

Item 3: To help you become creative, or to maintain creativity
Item 4: To act as a source of inspiration to you
Item 5: To allow you to fantasise in order to become inspired

iii. Distraction

Item 6: To distract yourself from events or activities going on around you
Item 7: To relieve boredom during mundane tasks

iv. Earworm Fulfilment

Item 8: To satisfy or clear songs that are 'stuck' in your head
Item 9: To remove songs that are 'stuck' in your head to prevent distraction

v. Focus & Concentration

Item 10: To help you focus or concentrate on tasks
Item 11: To stop external factors from distracting you when trying to concentrate on a task
Item 12: To help you 'flow' when trying to concentrate on something

vi. Habitual Listening

Item 13: To satisfy listening habits, based on what you expect from your past experiences with music
Item 14: Alongside your daily routine (e.g., such as commuting)

vii. Mental State

Item 15: To help you attain the necessary mindset to working on certain tasks
Item 16: To help you attain the necessary attitude to working on certain tasks

viii. Motivation

Item 17: To motivate yourself to achieve a particular goal (for example, emotional goals such as feeling happy, or physical goals such as exercise or relaxation)
Item 18: To help you achieve goals by motivating you to further action (such as increased effort during exercise)

ix. Reflection

Item 19: To help you reflect on your experiences and learn from them
Item 20: To help you to think on your experiences and behaviours from different perspectives

Item 21: To perceive thoughts and feelings within music that express your experiences

x. Visual Imagery

Item 22: To generate mental visual images that are spontaneous
Item 23: To help you to build mental images in your mind that play out psychological scenarios, such as personal fantasies

Emotional Functions

i. Entertainment & Hedonic Motive

Item 24: To enjoy yourself and/or be entertained
Item 25: To make yourself feel happy or positive

ii. Escapism & Venting

Item 26: To distract yourself from negative or stressful situations
Item 27: To help you escape stressful events or situations
Item 28: To distract yourself from unwanted thoughts and/or feelings

iii. Solace

Item 29: To generate feelings of comfort or solace
Item 30: To feel like you are being comforted by another person

iv. Therapy

Item 31: To act as a therapeutic tool to help you reduce negative emotions
Item 32: To find meaning within music that allows you to reduce negative emotions or moods

Specific Regulatory Functions

i. Accentuate Emotion/Mood

Item 33: To enhance particular moods or feelings that are a consequence of an activity that you are engaging with
Item 34: To accentuate particular emotions or moods alongside an activity that I am engaging with (for example, listening during exercise to enhance performance)

ii. Change or Shift Emotion/Mood

Item 35: To help you reverse your emotions or moods
Item 36: To help you mitigate positive emotions that may feel too intense

iii. Express or Convey Emotion/Mood

Item 37: To help you express feelings outwardly
Item 38: To help you express your feelings physically, such as helping you to cry or smile

iv. Regulate & Maintain Emotion/Mood

Item 39: To help sustain certain moods or emotions you may be experiencing
Item 40: To manage emotions that you may be experiencing despite external influences, whether they are positive or negative

v. Trigger or Elicit Emotion/Mood

Item 41: To feel certain specific emotions, such as joy or sadness
Item 42: To help you feel certain specific emotions when feeling 'neutral' (e.g., neither happy nor sad)

<u>Physiological Functions</u>

i. Accompaniment

Item 43: To accompany, or soundtrack everyday events (such as commuting to work or whilst walking)
Item 44: To accompany mundane daily tasks and make them more enjoyable

ii. Activation, Arousal & Response

Item 45: To help physically stimulate you to carry out physical tasks, such as exercise or sports
Item 46: To prepare, maintain or adjust levels of the stimulation appropriate for tasks before, during or after tasks

iii. Dancing

Item 47: To have something to dance to
Item 48: To help you to move in response to music

iv. Enhance Activity or Ability

Item 49: To improve your effectiveness during certain tasks (e.g., during cleaning)
Item 50: To reduce feelings of effort during tasks that you feel require a lot of effort (e.g., during cooking)

v. Environmental Control & Aural filtering

Item 51: To stop unpleasant or uncontrolled sounds from distracting or affecting you
Item 52: To avoid uncomfortable silences and/or distract you from the location in which you are listening

vi. Pacing & Movement

Item 53: To help you track time of physical motions during tasks
Item 54: To help you maintain pacing during physical activities, such as yoga, walking or whilst in the gym
Item 55: To perform tasks for longer or to a greater extent than you would be able to without music

vii. Physical Discomfort

Item 56: To intentionally cause discomfort to yourself or others (such as through high volume and/or frequencies)
Item 57: To cause yourself discomfort (such as through high volume and/or frequencies) to feel enhanced aggression leading to increased feelings of strength or stamina

viii. Structuring Time

Item 58: To help you keep track of time

Item 59: To reduce the perceived length of time of journeys, such as when in a car or on public transport
Item 60: To cluster together chunks of time to compartmentalise different tasks

Social Functions

i. Approval & Cultural Capital
Item 61: To feel validation or approval as a part of a group
Item 62: To match a group's dynamic so you are able to bond with group members when listening with others

ii. Boundary Demarcation
Item 63: To disengage from others in social settings
Item 64: To create a perceptible barrier between myself and others in social settings by using my device (e.g., smartphone or tablet) and/or headphones

iii. Communication
Item 65: To help you express your emotions and thoughts to others
Item 66: To act as a topic of discussion with others and ease communication or interaction
Item 67: To share content with others by sharing (i.e., sharing playlists or mixtapes)

iv. Control & Conformity
Item 68: To communicate appropriate behaviour within group dynamics (e.g., being appropriate to dance or to let it take a background role)
Item 69: To help you understand appropriate behaviour in social or group settings

v. Group Identity
Item 70: To identify with others through your shared values and/or culture
Item 71: To identify or feel connection with others who share your taste in music

vi. Interaction & Bonding
Item 72: To help to bond and/or interact with others
Item 73: To help you bond with others, and to subsequently feel a sense of belonging with those individuals
Item 74: To foster and develop new personal relationships

vii. Maintain & Express Cultural Values
Item 75: To help you and your social group to express your culture or values
Item 76: To act as a reference point for your social groups to maintain your shared culture (e.g., feelings of nostalgia with others)

viii. Surveillance
Item 77: To help you to monitor the behaviour of others, and gather information about their character
Item 78: To act as a tool through which you can assess the behaviour of other groups, based on their listening behaviour
Item 79: To allow others to survey your music taste and gather information about you

ix. Symbolic Representation & Difference

Item 80: To feel that certain artists, pieces, or genres of music are central to your social group's culture and sets you apart from others

Item 81: To feel that you may wish to exclude others from sharing in your social settings, if their music culture deviates from you or your group's

Identity Functions

i. Create & Maintain Identity

Item 82: To establish and maintain a part of your personal identity

Item 83: To explore different identities or music cultures you may wish to share in

ii. Express Identity & Values

Item 84: To help express your identities and values to others

Item 85: To help you maintain your identity as it reflects who you are as a person

iii. Personal Development & Understanding

Item 86: To help you grow and develop as an individual

Item 87: To help you understand your social, personal, and emotional experiences and come to terms with them

Item 88: To allow you to reflect on your previous identities

iv. Promote Autonomy & Agency

Item 89: To help you to feel control and agency within your daily life

Item 90: To enhance feelings of control and empowerment in your environment

v. Symbolic Representation

Item 91: To act as a point of symbolic representation of who you are

Item 92: To differentiate yourself from others in order to stand out

Meta-Domain Functions

i. Background

Item 93: To provide background noise and remove silence

- *ia. Create and Maintain Atmosphere*

Item 94: To generate a certain atmosphere or feeling within a given space, whether by yourself or with others

Item 95: To provide an ambience to make social interaction easier

ii. Company & Music as Proxy

Item 96: To feel a sense of company in the absence of others (e.g., playing the radio when home alone)

Item 97: To reduce feelings of being lonely when social interaction is not possible

- *iia. Silence Avoidance*

Item 98: To avoid silence when you’re alone (e.g., playing music when nobody else is home)

Item 99: To reduce feelings of loneliness when you are alone

iii. Memories

Item 100: To trigger or elicit certain memories
Item 101: To relive your own past and remember previous phases of your life

iv. Mimesis and Matching

Item 102: To replicate places, times, or environments
Item 103: To match external features (such as weather or time of day) so that you feel right within that environment

v. Musicking

Item 104: To experience music whilst you are making it yourself (e.g., singing, playing an instrument, or conducting)
Item 105: To perform or generate music

- *va. Aesthetic Appreciation*

Item 106: To focus on it in an intellectual manner so you can understand it technically, rather than listening to music for pure entertainment

- *vb. Listening Behaviours*

Item 107: To explore and listen to repertoire that is new to you
Item 108: To listen to music that is new so that you may find new potential purposes of music listening (e.g., finding music that is appropriate when accompanying a new hobby such as yoga or meditation)

- *vc. Musical and Lyric Analysis*

Item 109: To analyse music through its musical content
Item 110: To analyse music through its lyrical content

vi. Relaxation & Stress Relief

Item 111: To relieve stress and/or to help you to relax
Item 112: To relieve stress and negative emotions associated with negative events or situations

vii. Situational Relevance

Item 113: To inform appropriate behaviour in different situations (e.g., initiating dancing or social interaction at dinner parties)
Item 114: To take cues as to inform appropriate behaviour in different situations (e.g., initiating dancing or social interaction at dinner parties)

Note: Items were randomised across 11 matrix tables in study. For ease and clarity, they are listed in thematic order.

Appendix C: Study 1: Descriptive statistics of the initial pool of 114 items

Items	*N*	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
1. To make your environment seem more cinematic or 'film like'	327	0	4	1.49	1.320	0.453	-0.945
2. To synchronise, or ‘line-up’, with events in your daily life as though it were a soundtrack	327	0	4	1.47	1.230	0.413	-0.835
3. To help you become creative, or to maintain creativity	327	0	4	2.11	1.312	-0.169	-1.037
4. To act as a source of inspiration to you	327	0	4	2.39	1.116	-0.306	-0.493
5. To allow you to fantasise in order to become inspired	327	0	4	1.86	1.290	0.098	-1.058
6. To distract yourself from events or activities going on around you	327	0	4	2.35	1.103	-0.172	-0.676
7. To relieve boredom during mundane tasks	327	0	4	3.09	0.917	-0.862	0.508
8. To satisfy or clear songs that are 'stuck' in your head	327	0	4	1.84	1.271	0.160	-0.952
9. To remove songs that are 'stuck' in your head to prevent distraction	327	0	4	1.71	1.235	0.233	-0.832
10. To help you focus or concentrate on tasks	327	0	4	2.57	1.165	-0.471	-0.555
11. To stop external factors from distracting you when trying to	327	0	4	2.28	1.173	-0.259	-0.709

concentrate on a task							
12. To help you 'flow' when trying to concentrate on something	327	0	4	2.40	1.234	-0.339	-0.842
13. To satisfy listening habits, based on what you expect from your past experiences with music	327	0	4	2.35	1.121	-0.292	-0.516
14. Alongside your daily routine (e.g., such as commuting)	327	0	4	3.13	1.070	-1.292	1.170
15. To help you attain the necessary mindset to working on certain tasks	327	0	4	2.42	1.085	-0.327	-0.425
16. To help you attain the necessary attitude to working on certain tasks	327	0	4	2.37	1.156	-0.367	-0.582
17. To motivate yourself to achieve a particular goal (for example, emotional goals such as feeling happy, or physical goals such as exercise or relaxation)	327	0	4	2.67	1.086	-0.506	-0.361
18. To help you achieve goals by motivating you to further action (such as increased effort during exercise)	327	0	4	2.61	1.129	-0.613	-0.263
19. To help you reflect on your	327	0	4	1.55	1.139	0.245	-0.790

experiences and learn from them							
20. To help you to think on your experiences and behaviours from different perspectives	327	0	4	1.48	1.164	0.342	-0.744
21. To perceive thoughts and feelings within music that express your experiences	327	0	4	2.14	1.179	-0.016	-0.796
22. To generate mental visual images that are spontaneous	327	0	4	1.42	1.218	0.541	-0.633
23. To help you to build mental images in your mind that play out psychological scenarios, such as personal fantasies	327	0	4	1.53	1.298	0.437	-0.895
24. To enjoy yourself and/or be entertained	327	0	4	3.37	0.78	-1.182	1.390
25. To make yourself feel happy or positive	327	0	4	2.96	0.957	-0.745	0.411
26. To distract yourself from negative or stressful situations	327	0	4	2.41	1.177	-0.313	-0.744
27. To help you escape stressful events or situations	327	0	4	2.43	1.181	-0.32	-0.694
28. To distract yourself from unwanted thoughts and/or feelings	327	0	4	2.29	1.192	-0.144	-0.870
29. To generate feelings of comfort or solace	327	0	4	2.60	1.094	-0.589	-0.120

30. To feel like you are being comforted by another person	327	0	4	1.39	1.206	0.482	-0.667
31. To act as a therapeutic tool to help you reduce negative emotions	327	0	4	2.41	1.232	-0.305	-0.869
32. To find meaning within music that allows you to reduce negative emotions or moods	327	0	4	2.15	1.257	-0.093	-0.911
33. To enhance particular moods or feelings that are a consequence of an activity that I am engaging with	327	0	4	2.36	1.132	-0.291	-0.587
34. To accentuate particular emotions or moods alongside an activity that you are engaging with (for example, listening during exercise to enhance performance)	327	0	4	2.63	1.143	-0.589	-0.410
35. To help you reverse your emotions or moods	327	0	4	2.19	1.160	-0.121	-0.642
36. To help you mitigate positive emotions that may feel too intense	327	0	4	1.24	1.261	0.694	-0.619
37. To help you express feelings outwardly	327	0	4	1.72	1.189	0.293	-0.733
38. To help you express your feelings physically, such	327	0	4	1.97	1.227	0.043	-0.844

as helping you to cry or smile							
39. To help sustain certain moods or emotions you may be experiencing	327	0	4	2.44	1.106	-0.334	-0.461
40. To manage emotions that you may be experiencing despite external influences, whether they are positive or negative	327	0	4	2.09	1.255	-0.118	-0.923
41. To feel certain specific emotions, such as joy or sadness	327	0	4	2.51	1.079	-0.367	-0.362
42. To help you feel certain specific emotions when feeling 'neutral' (e.g., neither happy nor sad)	327	0	4	2.06	1.208	-0.103	-0.773
43. To accompany, or soundtrack everyday events (such as commuting to work or whilst walking)	327	0	4	2.82	1.220	-0.818	-0.237
44. To accompany mundane daily tasks and make them more enjoyable	327	0	4	3.15	0.892	-0.742	-0.259
45. To help physically stimulate you to carry out physical tasks, such as exercise or sports	327	0	4	2.79	1.119	-0.712	-0.225
46. To prepare, maintain or adjust levels of the stimulation appropriate for	327	0	4	2.13	1.151	-0.205	-0.592

tasks before, during or after tasks							
47. To have something to dance to	327	0	4	2.00	1.217	-0.057	-0.888
48. To help you to move in response to music	327	0	4	1.77	1.142	0.142	-0.691
49. To improve your effectiveness during certain tasks (e.g., during cleaning)	327	0	4	2.83	1.102	-0.673	-0.311
50. To reduce feelings of effort during tasks that you feel require a lot of effort (e.g., during cooking)	327	0	4	2.78	1.055	-0.585	-0.243
51. To stop unpleasant or uncontrolled sounds from distracting or affecting you	327	0	4	2.17	1.216	-0.151	-0.796
52. To avoid uncomfortable silences and/or distract you from the location in which you are listening	327	0	4	2.11	1.219	-0.116	-0.854
53. To help you track time of physical motions during tasks	327	0	4	1.57	1.318	0.365	-0.976
54. To help you maintain pacing during physical activities, such as yoga, walking or whilst in the gym	327	0	4	2.37	1.299	-0.383	-0.948
55. To perform tasks for longer or to a greater extent than you would be able to without music	327	0	4	2.59	1.081	-0.439	-0.386

56. To intentionally cause discomfort to yourself or others (such as through high volume and/or frequencies)	327	0	4	0.41	0.852	2.311	4.925
57. To cause yourself discomfort (such as through high volume and/or frequencies) to feel enhanced aggression leading to increased feelings of strength or stamina	327	0	4	0.61	1.018	1.728	2.172
58. To help you keep track of time	327	0	4	1.24	1.177	0.653	-0.489
59. To reduce the perceived length of time of journeys, such as when in a car or on public transport	327	0	4	3.01	1.077	-1.078	0.717
60. To cluster together chunks of time to compartmentalise different tasks	327	0	4	1.36	1.177	0.454	-0.670
61. To feel validation or approval as a part of a group	327	0	4	0.93	0.972	0.933	0.361
62. To match a group's dynamic so you are able to bond with group members when listening with others	327	0	4	1.38	1.123	0.35	-0.745
63. To disengage from others in social settings	327	0	4	1.43	1.154	0.443	-0.621
64. To create a perceptible barrier between myself and others in	327	0	4	1.61	1.216	0.329	-0.792

social settings by using my device (e.g., smartphone or tablet) and/or headphones							
65. To help you express your emotions and thoughts to others	327	0	4	1.56	1.229	0.41	-0.798
66. To act as a topic of discussion with others and ease communication or interaction	327	0	4	1.71	1.123	0.199	-0.582
67. To share content with others by sharing (i.e., sharing playlists or mixtapes)	327	0	4	1.59	1.242	0.321	-0.914
68. To communicate appropriate behaviour within group dynamics (e.g., being appropriate to dance or to let it take a background role)	327	0	4	1.36	1.159	0.435	-0.802
69. To help you understand appropriate behaviour in social or group settings	327	0	4	0.91	1.029	1.068	0.619
70. To identify with others through your shared values and/or culture	327	0	4	1.49	1.143	0.354	-0.702
71. To identify or feel connection with others who share your taste in music	327	0	4	1.89	1.128	0.005	-0.681
72. To help to bond and/or interact with others	327	0	4	1.79	1.085	0.044	-0.483
73. To help you bond with others, and	327	0	4	1.62	1.070	0.114	-0.596

to subsequently feel a sense of belonging with those individuals							
74. To foster and develop new personal relationships	327	0	4	1.25	1.076	0.617	-0.220
75. To help you and your social group to express your culture or values	327	0	4	1.27	1.119	0.44	-0.736
76. To act as a reference point for your social groups to maintain your shared culture (e.g., feelings of nostalgia with others)	327	0	4	1.57	1.162	0.264	-0.751
77. To help you to monitor the behaviour of others, and gather information about their character	327	0	4	0.82	0.996	1.111	0.462
78. To act as a tool through which you can assess the behaviour of other groups, based on their listening behaviour	327	0	4	0.93	1.043	0.952	0.102
79. To allow others to survey your music taste and gather information about you	327	0	4	1.22	1.103	0.68	-0.282
80. To feel that certain artists, pieces, or genres of music are central to your social group's culture and sets	327	0	4	1.35	1.216	0.603	-0.545

you apart from others							
81. To feel that you may wish to exclude others from sharing in your social settings, if their music culture deviates from you or your group's	327	0	4	0.57	0.920	1.743	2.602
82. To establish and maintain a part of your personal identity	327	0	4	1.90	1.243	0.000	-0.960
83. To explore different identities or music cultures you may wish to share in	327	0	4	1.79	1.186	0.115	-0.780
84. To help express your identities and values to others	327	0	4	1.44	1.165	0.428	-0.731
85. To help you maintain your identity as it reflects who you are as a person	327	0	4	1.75	1.289	0.200	-1.026
86. To help you grow and develop as an individual	327	0	4	1.65	1.206	0.267	-0.836
87. To help you understand your social, personal, and emotional experiences and come to terms with them	327	0	4	1.59	1.202	0.387	-0.718
88. To allow you to reflect on your previous identities	327	0	4	1.50	1.211	0.334	-0.819
89. To help you to feel control and agency within your daily life	327	0	4	1.66	1.24	0.279	-0.885

90. To enhance feelings of control and empowerment in your environment	327	0	4	1.70	1.234	0.196	-0.981
91. To act as a point of symbolic representation of who you are	327	0	4	1.56	1.191	0.319	-0.808
92. To differentiate yourself from others in order to stand out	327	0	4	1.02	1.084	0.758	-0.402
93. To provide background noise and remove silence	327	0	4	2.76	1.102	-0.658	-0.106
94. To generate a certain atmosphere or feeling within a given space, whether by yourself or with others	327	0	4	2.52	1.068	-0.311	-0.460
95. To provide an ambience to make social interaction easier	327	0	4	2.16	1.092	-0.226	-0.436
96. To feel a sense of company in the absence of others (e.g., playing the radio when home alone)	327	0	4	2.49	1.282	-0.484	-0.796
97. To reduce feelings of being lonely when social interaction is not possible	327	0	4	2.33	1.297	-0.270	-0.988
98. To avoid silence when you're alone (e.g., playing music when nobody else is home)	327	0	4	2.61	1.195	-0.534	-0.586
99. To reduce feelings of	327	0	4	2.34	1.279	-0.353	-0.896

loneliness when you are alone							
100. To trigger or elicit certain memories	327	0	4	2.11	1.126	0.060	-0.639
101. To relive your own past and remember previous phases of your life	327	0	4	2.23	1.109	-0.145	-0.594
102. To replicate places, times, or environments	327	0	4	1.91	1.137	0.181	-0.643
103. To match external features (such as weather or time of day) so that you feel right within that environment	327	0	4	1.53	1.115	0.314	-0.709
104. To experience music whilst you are making it yourself (e.g., singing, playing an instrument, or conducting)	327	0	4	1.34	1.483	0.667	-1.042
105. To perform or generate music	327	0	4	1.17	1.380	0.860	-0.594
106. To focus on it in an intellectual manner so you can understand it technically, rather than listening to music for pure entertainment	327	0	4	1.32	1.254	0.610	-0.712
107. To explore and listen to repertoire that is new to you	327	0	4	2.12	1.115	-0.136	-0.609
108. To listen to music that is new so that you may find new potential purposes of music listening (e.g., finding music that is appropriate when accompanying a new hobby such	327	0	4	1.99	1.192	0.040	-0.785

as yoga or meditation)							
109. To analyse music through its musical content	327	0	4	1.48	1.240	0.463	-0.732
110. To analyse music through its lyrical content	327	0	4	1.70	1.236	0.188	-0.955
111. To relieve stress and/or to help you to relax	327	0	4	2.94	0.962	-0.611	-0.128
112. To relieve stress and negative emotions associated with negative events or situations	327	0	4	2.54	1.153	-0.420	-0.579
113. To inform appropriate behaviour in different situations (e.g., initiating dancing or social interaction at dinner parties)	327	0	4	1.54	1.177	0.251	-0.897
114. To take cues as to inform appropriate behaviour in different situations (e.g., initiating dancing or social interaction at dinner parties)	327	0	4	1.46	1.117	0.304	-0.690
Valid *N* (listwise)	327						

Note. Items have been abbreviated. For full items please refer to Appendix A. * denotes item removed from analyses according to section 6.4.2.

Appendix D: Study 2: Initial Survey questionnaire

Q1) What is your age in years?

____

Q2) What is your gender?

○ Male (1)

○ Female (2)

○ Non-binary / third gender (3)

○ Prefer not to say (4)

The following questions are going to ask you about the most recent experience in which **you chose to listen to music**.

Specifically, we are thinking about **a situation in which you were in control of the music**.

This means that we are ***not*** including live music performances or situations in which you were ***not*** in control of the music.

○ Cool! 👍 (1)

Q3) When did you last decide to listen to music?

- ◯ I am listening right now! (1)
- ◯ in the last hour (2)
- ◯ in the last 2 hours (3)
- ◯ in the last 2-3 hours (4)
- ◯ in the last 3-4 hours (5)
- ◯ in the last 4-12 hours (6)
- ◯ in the last 12-24 hours (7)
- ◯ Other (please state) (8) ____________________
- ◯ I do not remember (9)

Q4) Where were/are you listening to music?

- ○ Work (1)
- ○ Home (2)
- ○ Friend's Home (3)
- ○ Gym (4)
- ○ Transitory space (e.g., walking, driving or being on public transport) (5)
- ○ Urban location (i.e., in town or the city) (6)
- ○ Restaurant/bar (7)
- ○ Cultural location (e.g., place of worship) (8)
- ○ Musicking location (e.g., rehearsal or recording studio) (9)
- ○ Other (10) ____

Q5) What were/are you doing whilst listening to music? (e.g., working, exercising, relaxing)

____

Q6) What format of music did you listen to (excluding live music)?

- ○ Streaming service (e.g., Spotify, Apple Music) (1)
- ○ Radio (2)
- ○ Physical format (e.g., CD, Vinyl) (3)
- ○ Digital file (e.g., mp3, iTunes) (4)
- ○ Audio-visual content (e.g., YouTube) (5)
- ○ Other (6)

Q7) What form of music did you decide to listen to?

- ○ A specific song/track (1)
- ○ A playlist (2)
- ○ An album (3)
- ○ Other (4) ____

Display This Question:

If What form of music did you decide to listen to? = A specific song/track

Q8) Please name a song and artist that you listened to:

- ○ Track/Song title (1) ____
- ○ Artist/Musician (2) ____

Display This Question:

If What form of music did you decide to listen to? = A playlist

Q9) Was the playlist private or public?

- ○ Private (1)
- ○ Public (2)

Display This Question:

If What form of music did you decide to listen to? = A playlist

Q10) Can you name a song that you heard on this playlist? (e.g., She Loves You by The Beatles)

◯ Track/Song title (1) ____

◯ Artist/Musician (2) ____

Display This Question:

If What form of music did you decide to listen to? = An album

Q11) Can you name the artist and a track from the album you decided to listen to?

◯ Track/Song title (3) ____

◯ Album (1) ____

◯ Artist (2) ____

Display This Question:

If What form of music did you decide to listen to? = Other

Q12) Can you name a track and/or artist that you heard?

◯ Track (1) ____

◯ Artist (2) ____

Q13) How would you characterise the music?

	1 (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	7 (7)	
Calming	○	○	○	○	○	○	○	Exciting
Slow	○	○	○	○	○	○	○	Fast
Sad	○	○	○	○	○	○	○	Happy
Unfamiliar	○	○	○	○	○	○	○	Familiar
Less melodic	○	○	○	○	○	○	○	Very melodic
Less rhythmic	○	○	○	○	○	○	○	Very rhythmic
Simple	○	○	○	○	○	○	○	Complex
Peaceful	○	○	○	○	○	○	○	Aggressive
Less intense	○	○	○	○	○	○	○	Very intense

Q14) How would you characterise the music?

○ Instrumental (1)

○ Vocal (2)

Q15) In your most recent listening experience, to what extent did you use music for the following reasons:

	Not very important (0)	Slightly important (1)	Somewhat important (2)	Quite important (3)	Extremely important! (4)
To differentiate yourself from others in order to stand out (Item1)	○	○	○	○	○
To help you maintain your identity as it reflects who you are as a person (Item2)	○	○	○	○	○
To help express your identities and values to others (Item3)	○	○	○	○	○
To establish and maintain a part of your personal identity (Item4)	○	○	○	○	○

To feel that certain artists, pieces, or genres of music are central to your social group's culture and sets you apart from others (Item5)	○	○	○	○	○
To allow others to survey your music taste and gather information about you (Item6)	○	○	○	○	○
To act as a reference point for your social groups to maintain your shared culture (e.g., feelings of nostalgia with others) (Item7)	○	○	○	○	○

To help you and your social group to express your culture or values (Item8)	○	○	○	○	○
To foster and develop new personal relationships (Item9)	○	○	○	○	○
To help you bond with others, and to subsequently feel a sense of belonging with those individuals (Item10)	○	○	○	○	○
To help to bond and/or interact with others (Item11)	○	○	○	○	○
To identify or feel connection with others who share your taste in music (Item12)	○	○	○	○	○

To identify with others through your shared values and/or culture (Item13)	○	○	○	○	○
To share content with others by sharing (i.e., sharing playlists or mixtapes) (Item14)	○	○	○	○	○
To act as a topic of discussion with others and ease communication or interaction (Item15)	○	○	○	○	○
To match a group's dynamic so you are able to bond with group members when listening with others (Item16)	○	○	○	○	○

To act as a point of symbolic representation of who you are (Item17)	○	○	○	○	○
To distract yourself from negative or stressful situations (Item18)	○	○	○	○	○
To help you escape stressful events or situations (Item19)	○	○	○	○	○
To distract yourself from unwanted thoughts and/or feelings (Item20)	○	○	○	○	○
To act as a therapeutic tool to help you reduce negative emotions (Item21)	○	○	○	○	○

To find meaning within music that allows you to reduce negative emotions or moods (Item22)	○	○	○	○	○
To help you reverse your emotions or moods (Item23)	○	○	○	○	○
To manage emotions that you may be experiencing despite external influences, whether they are positive or negative (Item24)	○	○	○	○	○
To feel certain specific emotions, such as joy or sadness (Item25)	○	○	○	○	○

To help you feel certain specific emotions when feeling 'neutral' (e.g., neither happy nor sad) (Item26)					
To relieve stress and negative emotions associated with negative events or situations (Item27)					
To help you focus or concentrate on tasks (Item28)					
To stop external factors from distracting you when trying to concentrate on a task (Item29)					

To help you 'flow' when trying to concentrate on something (Item30)	○	○	○	○	○
To help you attain the necessary mindset to working on certain tasks (Item31)	○	○	○	○	○
To provide background noise and remove silence (Item32)	○	○	○	○	○
To feel a sense of company in the absence of others (e.g., playing the radio when home alone) (Item33)	○	○	○	○	○

To reduce feelings of being lonely when social interaction is not possible (Item34)	○	○	○	○	○
To avoid silence when you're alone (e.g., playing music when nobody else is home) (Item35)	○	○	○	○	○
To help you achieve goals by motivating you to further action (such as increased effort during exercise) (Item36)	○	○	○	○	○
To help physically stimulate you to carry out physical tasks, such as exercise or sports (Item37)	○	○	○	○	○

To help you maintain pacing during physical activities, such as yoga, walking or whilst in the gym (Item38)	○	○	○	○	○

Note. Additional items from larger structure identified in Study 1 were retained for exploratory analyses in case of poor fit.

Q16) Were there other reasons you had for listening to music that were not covered on the previous pages?

○ Yes (1)

○ No (2)

Q17) Please describe these other reasons for listening:

____

Q18) Please read the following statement carefully and move the scale according to the extent to which you agree:

The music fit the situation in which I was listening

What we mean by '*fit*' - By fit, we are referring to how appropriate you feel the music was according to this particular listening experience.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
	1	2	3	4	5	6	7
On a scale from 1 (Strongly Disagree) to 7 (Strongly Agree), please move the slider according to the above statement ()							

Q19) Please read the following statement carefully and move the scale according to the extent to which you agree:

The music was effective in helping me achieve my reasons for listening

What we mean by '*effective*': By effective, we mean how successful was the music in achieving your desired listening outcome

	Had no effect		Moderately Effective			Highly Effective	
	1	2	3	4	5	6	7
On a scale from 1 (Strongly Disagree) to 7 (Strongly Agree), please move the slider according to the above statement ()							

Appendix E: Study 2: ESM Study questionnaire

Q1) Have you listened to music since the last notification?

1. Yes
2. No – *Survey Terminates*

Q2) When did you last listen to music?

1. I am listening right now!
2. In the last hour
3. In the last 2 hours
4. In the last 2-3 hours
5. In the last 3-4 hours
6. In the last 4+ hours

Q3) Where were/are you listening to music?

1. Work
2. Home
3. Friend's Home
4. Gym
5. Transitory space
6. Urban location
7. Restaurant/bar
8. Cultural location (e.g., place of worship)
9. Musicking location (e.g., rehearsal or recording studio)
10. Other

- *Open text field –*

Q4) What were/are you doing whilst listening to music?

- *Open text field –*

Q5) What format of music did you listen to (*excluding* live music)?

1. Streaming service (e.g., Spotify, Apple Music)
2. Radio
3. Physical format
4. Digital file
5. Audio-visual content
6. Other

Q6/7) What medium of music did you decide to listen to?

1. A specific song/track

   → Can you name the song and artist that you listened to?

- *Open text field –*

2. A playlist

   → Was this playlist private or public?

   1. Private
   2. Public

   → Can you name a song you heard on this playlist?

- *Open text field –*

3. An album

   → Can you name the album and artist that you listened to?

- *Open text field –*

   → Can you name a song that you heard?

- *Open text field –*

4. Other

   → What other medium did you listen to?

- *Open text field –*

   → Can you name the song and artist that you listened to?

5. I was not in control of the music

Q8) How would you characterise the music?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Calming	○	○	○	○	○	○	○	Exciting
Slow	○	○	○	○	○	○	○	Fast
Sad	○	○	○	○	○	○	○	Happy
Unfamiliar	○	○	○	○	○	○	○	Familiar
Less melodic	○	○	○	○	○	○	○	Very melodic
Less rhythmic	○	○	○	○	○	○	○	Very rhythmic
Simple	○	○	○	○	○	○	○	Complex
Peaceful	○	○	○	○	○	○	○	Aggressive
Less intense	○	○	○	○	○	○	○	Very intense

Note. Items presented in table for ease but individually during ESM.

Q9) Why were/are you listening to music?

Identity and Social Bonding

1. To help express your identities and values to others
2. To feel that certain artists, pieces, or genres of music are central to your social group's culture and sets you apart from others
3. To help you bond with others, and to subsequently feel a sense of belonging with those individuals

4. To identify with others through your shared values and/or culture
5. To act as a topic of discussion with others and ease communication or interaction
6. To match a group's dynamic so you are able to bond with group members when listening with others

Emotion Regulation

7. To distract yourself from negative or stressful situations
8. To find meaning within music that allows you to reduce negative emotions or moods
9. To manage emotions that you may be experiencing despite external influences, whether they are positive or negative
10. To feel certain specific emotions, such as joy or sadness
11. To relieve stress and negative emotions associated with negative events or situations
12. To help you reverse your emotions or moods

Focus and Concentration

13. To help you focus or concentrate on tasks
14. To stop external factors from distracting you when trying to concentrate on a task
15. To help you 'flow' when trying to concentrate on something
16. To help you attain the necessary mindset to working on certain tasks

Background and Accompaniment

17. To provide background noise and remove silence
18. To feel a sense of company in the absence of others (e.g., playing the radio when home alone)
19. To reduce feelings of being lonely when social interaction is not possible (8)
20. To avoid silence when you're alone (e.g., playing music when nobody else is home)

Physiological Arousal

21. To help you achieve goals by motivating you to further action (such as increased effort during exercise)
22. To help physically stimulate you to carry out physical tasks, such as exercise or sports
23. To help you maintain pacing during physical activities, such as yoga, walking or whilst in the gym

Q10) Were there other reasons you had for listening to music that was not covered on the previous page?

(1) Yes
→ Please describe you reasons for listening

- *Open text field –*

(2) No

Q11) Please read the following statement carefully and move the scale according to the extent to which you agree:

The music fit the situation in which I was listening

What we mean by '*fit*' - By fit, we are referring to how appropriate you feel the music was according to this particular listening experience.

Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
1	2	3	4	5	6	7

On a scale from 1 (Strongly Disagree) to 7 (Strongly Agree), please move the slider according to the above statement ()	

Q12) Please read the following statement carefully and move the scale according to the extent to which you agree:

The music was effective in helping me achieve my reasons for listening

What we mean by '*effective*': By effective, we mean how successful was the music in achieving your desired listening outcome

Had no effect		Moderately Effective			Highly Effective	
1	2	3	4	5	6	7

On a scale from 1 (Strongly Disagree) to 7 (Strongly Agree), please move the slider according to the above statement ()	

Note. Participants were responded to items they deemed applicable in each case, rating them on a reduced 1-4 scale. Cases where items were not selected assumed the lowest value (0) by default.

Appendix F: Bash script syntax for gathering Spotify audio features

```
#! /bin/bash

array=( TRACK IDs GO HERE )
for i in "${array[@]}"
do
curl -X "GET" "https://api.spotify.com/v1/audio-features/$i" -H "Accept:
application/json" -H "Content-Type: application/json" -H "Authorization:
Bearer BEARER_TOKEN_HERE"
done
```

Appendix G: R syntax for Study 2 (pooled CFA and Mediation)

```
# Load data
library(haven)
MLSData <- # load data
View(MLSData)

# Load packages
library(lavaan) # install.packages("lavaan")
library(lavaan.survey) # install.packages("lavaan.survey")

# Specify CFA model
fml.model <-
'social_interaction =~  SIB_ExpressIdentityToOthers + SIB_BondFB +
SIB_FeelArtistsCentralToCulture + SIB_TopicOfDiscussion +
SIB_IdentifyWithOthers + SIB_MatchGroupDynamic
emotion_regulation =~ ER_DistractFromNegativeSituation +
ER_FeelSpecificEmotions + ER_ReduceNegativeEmotions + ER_ReverseEmotions
+ ER_ManageEmotions + ER_RelieveStress
focus_concentration =~ FAC_HelpFlow + FAC_FocusOnTask +
FAC_AttainNecessaryMindset + FAC_StopExternalFactorsDistracting
background_and_accompaniment =~ CAMP_BackgroundNoise +
CAMP_ReduceLoneliness + CAMP_SenseOfCompany + CAMP_AvoidSilence
physiological_arousal =~ PA_MotivationToAction + PA_PhysicallyStimulate +
PA_MaintainPacing'

# Initial fit of CFA (clustering ignored)
cfa.fit <- cfa(model = fml.model, data = MLSData, std.lv = TRUE,
estimator = 'MLM')
fitmeasures(cfa.fit, c('chisq.scaled', 'df', 'pvalue.scaled',
'cfi.robust', 'tli.robust', 'rmsea.robust'))

# Adjusted SEs and model fit according to clustered data
```

```
survey.design <- svydesign(ids =~ PID, nest = TRUE, prob =~ NULL, strata
= NULL, data = MLSData)
survey.fitCFA <- lavaan.survey(lavaan.fit = cfa.fit, survey.design =
survey.design)
fitmeasures(survey.fitCFA, c('chisq.scaled', 'df', ‘pvalue.scaled’,
'cfi.robust', 'tli.robust', 'rmsea.robust'))

# Generate factor scores and add to dataset
fscores <- lavPredict(survey.fitCFA, newdata = MLSData)
idx <- lavInspect(cfa.fit, 'case.idx')
for (fs in colnames(fscores)) {
  MLSData[idx, fs] <- fscores[ , fs]
}
write_sav(MLSData, 'MLSDataFScores.sav') # optional (creates new dataset)
getwd()

# Specify Mediation model
mediation.model <-
'# outcome variables regression – paths b and c
PCArousal ~ b11*social_interaction + b12*emotion_regulation +
b13*focus_concentration + b14*background_and_accompaniment +
b15*physiological_arousal + c11*Work + c12*Travel + c13*Relaxation +
c14*Exercise + c15*Socialising + c16*RecreationalActivity

PCValence ~ b21*social_interaction + b22*emotion_regulation +
b23*focus_concentration + b24*background_and_accompaniment +
b25*physiological_arousal + c21*Work + c22*Travel + c23*Relaxation +
c24*Exercise + c25*Socialising + c26*RecreationalActivity

PCInstrumental ~ b31*social_interaction + b32*emotion_regulation +
b33*focus_concentration + b34*background_and_accompaniment +
b35*physiological_arousal + c31*Work + c32*Travel + c33*Relaxation +
c34*Exercise + c35*Socialising + c36*RecreationalActivity

# Exogenous to mediator regressions (path a)
social_interaction ~ a11*Work + a12*Travel + a13*Relaxation +
a14*Exercise + a15*Socialising + a16*RecreationalActivity
emotion_regulation ~ a21*Work + a22*Travel + a23*Relaxation +
a24*Exercise + a25*Socialising + a26*RecreationalActivity
focus_concentration ~ a31*Work + a32*Travel + a33*Relaxation +
a34*Exercise + a35*Socialising + a36*RecreationalActivity
background_and_accompaniment ~ a41*Work + a42*Travel + a43*Relaxation +
a44*Exercise + a45*Socialising + a46*RecreationalActivity
```

```
physiological_arousal ~ a51*Work + a52*Travel + a53*Relaxation +
a54*Exercise + a55*Socialising + a56*RecreationalActivity

# M variable residual covariances
social_interaction ~~ emotion_regulation
social_interaction ~~ focus_concentration
emotion_regulation ~~ focus_concentration
social_interaction ~~ background_and_accompaniment
emotion_regulation ~~ background_and_accompaniment
focus_concentration ~~ background_and_accompaniment
social_interaction ~~ physiological_arousal
emotion_regulation ~~ physiological_arousal
focus_concentration ~~ physiological_arousal
background_and_accompaniment ~~ physiological_arousal

# Y variable residual covariances
PCArousal ~~ PCValence
PCArousal ~~ PCInstrumental
PCValence ~~ PCInstrumental

# Effect decomposition – specifies all IEs and TEs
# y1 ~ x1
ind_x1_m1_y1 := a11*b11
ind_x1_m2_y1 := a21*b12
ind_x1_m3_y1 := a31*b13
ind_x1_m4_y1 := a41*b14
ind_x1_m5_y1 := a51*b15
ind_x1_y1    := ind_x1_m1_y1 + ind_x1_m2_y1 + ind_x1_m3_y1 + ind_x1_m4_y1
+ ind_x1_m5_y1
tot_x1_y1    := ind_x1_y1 + c11

# y1 ~ x2
ind_x2_m1_y1 := a12*b11
ind_x2_m2_y1 := a22*b12
ind_x2_m3_y1 := a32*b13
ind_x2_m4_y1 := a42*b14
ind_x2_m5_y1 := a52*b15
ind_x2_y1    := ind_x2_m1_y1 + ind_x2_m2_y1 + ind_x2_m3_y1 + ind_x2_m4_y1
+ ind_x2_m5_y1
tot_x2_y1    := ind_x2_y1 + c12

# y1 ~ x3
ind_x3_m1_y1 := a13*b11
```

```
ind_x3_m2_y1 := a23*b12
ind_x3_m3_y1 := a33*b13
ind_x3_m4_y1 := a43*b14
ind_x3_m5_y1 := a53*b15
ind_x3_y1    := ind_x3_m1_y1 + ind_x3_m2_y1 + ind_x3_m3_y1 + ind_x3_m4_y1
+ ind_x3_m5_y1
tot_x3_y1    := ind_x3_y1 + c13

# y1 ~ x4
ind_x4_m1_y1 := a14*b11
ind_x4_m2_y1 := a24*b12
ind_x4_m3_y1 := a34*b13
ind_x4_m4_y1 := a44*b14
ind_x4_m5_y1 := a54*b15
ind_x4_y1    := ind_x4_m1_y1 + ind_x4_m2_y1 + ind_x4_m3_y1 + ind_x4_m4_y1
+ ind_x4_m5_y1
tot_x4_y1    := ind_x4_y1 + c14

# y1 ~ x5
ind_x5_m1_y1 := a15*b11
ind_x5_m2_y1 := a25*b12
ind_x5_m3_y1 := a35*b13
ind_x5_m4_y1 := a45*b14
ind_x5_m5_y1 := a55*b15
ind_x5_y1    := ind_x5_m1_y1 + ind_x5_m2_y1 + ind_x5_m3_y1 + ind_x5_m4_y1
+ ind_x5_m5_y1
tot_x5_y1    := ind_x5_y1 + c15

# y1 ~ x6
ind_x6_m1_y1 := a16*b11
ind_x6_m2_y1 := a26*b12
ind_x6_m3_y1 := a36*b13
ind_x6_m4_y1 := a46*b14
ind_x6_m5_y1 := a56*b15
ind_x6_y1    := ind_x6_m1_y1 + ind_x6_m2_y1 + ind_x6_m3_y1 + ind_x6_m4_y1
+ ind_x6_m5_y1
tot_x6_y1    := ind_x6_y1 + c16

# y2 ~ x1
ind_x1_m1_y2 := a11*b21
ind_x1_m2_y2 := a21*b22
ind_x1_m3_y2 := a31*b23
ind_x1_m4_y2 := a41*b24
```

```
ind_x1_m5_y2 := a51*b25
ind_x1_y2    := ind_x1_m1_y2 + ind_x1_m2_y2 + ind_x1_m3_y2 + ind_x1_m4_y2
+ ind_x1_m5_y2
tot_x1_y2    := ind_x1_y2 + c21

# y2 ~ x2
ind_x2_m1_y2 := a12*b21
ind_x2_m2_y2 := a22*b22
ind_x2_m3_y2 := a32*b23
ind_x2_m4_y2 := a42*b24
ind_x2_m5_y2 := a52*b25
ind_x2_y2    := ind_x2_m1_y2 + ind_x2_m2_y2 + ind_x2_m3_y2 + ind_x2_m4_y2
+ ind_x2_m5_y2
tot_x2_y2    := ind_x2_y2 + c22

# y2 ~ x3
ind_x3_m1_y2 := a13*b21
ind_x3_m2_y2 := a23*b22
ind_x3_m3_y2 := a33*b23
ind_x3_m4_y2 := a43*b24
ind_x3_m5_y2 := a53*b25
ind_x3_y2    := ind_x3_m1_y2 + ind_x3_m2_y2 + ind_x3_m3_y2 + ind_x3_m4_y2
+ ind_x3_m5_y2
tot_x3_y2    := ind_x3_y2 + c23

# y2 ~ x4
ind_x4_m1_y2 := a14*b21
ind_x4_m2_y2 := a24*b22
ind_x4_m3_y2 := a34*b23
ind_x4_m4_y2 := a44*b24
ind_x4_m5_y2 := a54*b25
ind_x4_y2    := ind_x4_m1_y2 + ind_x4_m2_y2 + ind_x4_m3_y2 + ind_x4_m4_y2
+ ind_x4_m5_y2
tot_x4_y2    := ind_x4_y2 + c24

# y2 ~ x5
ind_x5_m1_y2 := a15*b21
ind_x5_m2_y2 := a25*b22
ind_x5_m3_y2 := a35*b23
ind_x5_m4_y2 := a45*b24
ind_x5_m5_y2 := a55*b25
ind_x5_y2    := ind_x5_m1_y2 + ind_x5_m2_y2 + ind_x5_m3_y2 + ind_x5_m4_y2
+ ind_x5_m5_y2
```

```
tot_x5_y2    := ind_x5_y2 + c25

# y2 ~ x6
ind_x6_m1_y2 := a16*b21
ind_x6_m2_y2 := a26*b22
ind_x6_m3_y2 := a36*b23
ind_x6_m4_y2 := a46*b24
ind_x6_m5_y2 := a56*b25
ind_x6_y2    := ind_x6_m1_y2 + ind_x6_m2_y2 + ind_x6_m3_y2 + ind_x6_m4_y2
+ ind_x6_m5_y2
tot_x6_y2    := ind_x6_y2 + c26
# y3 ~ x1
ind_x1_m1_y3 := a11*b31
ind_x1_m2_y3 := a21*b32
ind_x1_m3_y3 := a31*b33
ind_x1_m4_y3 := a41*b34
ind_x1_m5_y3 := a51*b35
ind_x1_y3    := ind_x1_m1_y3 + ind_x1_m2_y3 + ind_x1_m3_y3 + ind_x1_m4_y3
+ ind_x1_m5_y3
tot_x1_y3    := ind_x1_y3 + c31

# y3 ~ x2
ind_x2_m1_y3 := a12*b31
ind_x2_m2_y3 := a22*b32
ind_x2_m3_y3 := a32*b33
ind_x2_m4_y3 := a42*b34
ind_x2_m5_y3 := a52*b35
ind_x2_y3    := ind_x2_m1_y3 + ind_x2_m2_y3 + ind_x2_m3_y3 + ind_x2_m4_y3
+ ind_x2_m5_y3
tot_x2_y3    := ind_x2_y3 + c32

# y3 ~ x3
ind_x3_m1_y3 := a13*b31
ind_x3_m2_y3 := a23*b32
ind_x3_m3_y3 := a33*b33
ind_x3_m4_y3 := a43*b34
ind_x3_m5_y3 := a53*b35
ind_x3_y3    := ind_x3_m1_y3 + ind_x3_m2_y3 + ind_x3_m3_y3 + ind_x3_m4_y3
+ ind_x3_m5_y3
tot_x3_y3    := ind_x3_y3 + c33

# y3 ~ x4
ind_x4_m1_y3 := a14*b31
```

```
ind_x4_m2_y3 := a24*b32
ind_x4_m3_y3 := a34*b33
ind_x4_m4_y3 := a44*b34
ind_x4_m5_y3 := a54*b35
ind_x4_y3    := ind_x4_m1_y3 + ind_x4_m2_y3 + ind_x4_m3_y3 + ind_x4_m4_y3
+ ind_x4_m5_y3
tot_x4_y3    := ind_x4_y3 + c34

# y3 ~ x5
ind_x5_m1_y3 := a15*b31
ind_x5_m2_y3 := a25*b32
ind_x5_m3_y3 := a35*b33
ind_x5_m4_y3 := a45*b34
ind_x5_m5_y3 := a55*b35
ind_x5_y3    := ind_x5_m1_y3 + ind_x5_m2_y3 + ind_x5_m3_y3 + ind_x5_m4_y3
+ ind_x5_m5_y3
tot_x5_y3    := ind_x5_y3 + c35

# y3 ~ x6
ind_x6_m1_y3 := a16*b31
ind_x6_m2_y3 := a26*b32
ind_x6_m3_y3 := a36*b33
ind_x6_m4_y3 := a46*b34
ind_x6_m5_y3 := a56*b35
ind_x6_y3    := ind_x6_m1_y3 + ind_x6_m2_y3 + ind_x6_m3_y3 + ind_x6_m4_y3
+ ind_x6_m5_y3
tot_x6_y3    := ind_x6_y3 + c36'

# Fit mediation model (no clustering/adjusted SEs)
mediation.fit <- sem(mediation.model, data = MLSData)

# Re-specify complex survey model (optional)
survey.design <- svydesign(ids =~ PID, nest = TRUE, prob =~ NULL, strata
= NULL, data = MLSData)

# Adjust SEs for clustered data
adjusted.mediation.fit <- lavaan.survey(lavaan.fit = mediation.fit,
survey.design = survey.design)

# Inspect adjusted model
options(max.print = 100000)
summary(adjusted.mediation.fit, standardized = TRUE, fit.measures = TRUE,
rsquare = TRUE)
```

```
# To see confidence intervals
parameterestimates(adjusted.mediation.fit)
```

Note. This syntax does *not* conduct CFAs for separate online survey and ESM responses (see section 7.4.3). To conduct these, segment `MLSData` by the variable `Sample` (1 = ESM responses, 2 = Survey responses) and follow steps 1-5 when `Sample` = 1 and steps 1-4 when `Sample` = 2.

Appendix H: Full table of study 2 mediation results

Exogenous variable	*a*	Mediator variable	*b*	Component/Outcome variable	Parameter Estimate	*SE*	*p*
Work/Study	→	Social Interaction	→	Arousal	-0.010	0.01	.330
Work/Study	→	Emotion Regulation	→	Arousal	-0.005	0.01	.620
Work/Study	→	Focus and Concentration	→	Arousal	-0.131	0.052	**.012**
Work/Study	→	Background and Accompaniment	→	Arousal	-0.004	0.009	.662
Work/Study	→	Physiological Arousal	→	Arousal	0.016	0.019	.417
Work/Study		*TIE* →		Arousal	-0.134	0.051	**.009**
Work/Study		*TE* →		Arousal	-0.226	0.114	**.047**
Travelling	→	Social Interaction	→	Arousal	-0.007	0.009	.407
Travelling	→	Emotion Regulation	→	Arousal	-0.013	0.026	.600
Travelling	→	Focus and Concentration	→	Arousal	0.005	0.02	.803
Travelling	→	Background and Accompaniment	→	Arousal	0.003	0.007	.669
Travelling	→	Physiological Arousal	→	Arousal	0.036	0.025	.155
Travelling		*TIE* →		Arousal	0.024	0.03	.429
Travelling		*TE* →		Arousal	0.064	0.114	.571
Relaxation	→	Social Interaction	→	Arousal	-0.01	0.01	.350
Relaxation	→	Emotion Regulation	→	Arousal	-0.008	0.016	.604
Relaxation	→	Focus and Concentration	→	Arousal	0.039	0.028	.164
Relaxation	→	Background and Accompaniment	→	Arousal	-0.006	0.013	.624
Relaxation	→	Physiological Arousal	→	Arousal	-0.024	0.023	.286
Relaxation		*TIE* →		Arousal	-0.01	0.03	.735
Relaxation		*TE* →		Arousal	0.021	0.13	.872
Exercise	→	Social Interaction	→	Arousal	-0.023	0.023	.304
Exercise	→	Emotion Regulation	→	Arousal	-0.017	0.033	.609
Exercise	→	Focus and Concentration	→	Arousal	-0.075	0.04	.064†
Exercise	→	Background and Accompaniment	→	Arousal	0.003	0.009	.720
Exercise	→	Physiological Arousal	→	Arousal	0.281	0.113	**.013**
Exercise		*TIE* →		Arousal	0.169	0.094	.073†
Exercise		*TE* →		Arousal	0.293	0.212	.167
Socialising	→	Social Interaction	→	Arousal	-0.094	0.088	.286
Socialising	→	Emotion Regulation	→	Arousal	0.001	0.011	.953

Socialising	→	Focus and Concentration	→	Arousal	0.073	0.059	.218
Socialising	→	Background and Accompaniment	→	Arousal	-0.021	0.044	.629
Socialising	→	Physiological Arousal	→	Arousal	-0.007	0.065	.911
Socialising		*TIE* →		Arousal	-0.048	0.087	.577
Socialising		*TE* →		Arousal	0.108	0.303	.721
Recreational Activity	→	Social Interaction	→	Arousal	-0.009	0.012	.442
Recreational Activity	→	Emotion Regulation	→	Arousal	-0.004	0.008	.648
Recreational Activity	→	Focus and Concentration	→	Arousal	0.004	0.023	.862
Recreational Activity	→	Background and Accompaniment	→	Arousal	0	0.007	.979
Recreational Activity	→	Physiological Arousal	→	Arousal	-0.004	0.029	.177
Recreational Activity		*TIE* →		Arousal	-0.048	0.025	**.049**
Recreational Activity		*TE* →		Arousal	0.072	0.134	.591
Work/Study	→	Social Interaction	→	Valence	-0.002	0.009	.831
Work/Study	→	Emotion Regulation	→	Valence	-0.013	0.012	.259
Work/Study	→	Focus and Concentration	→	Valence	-0.099	0.047	**.034**
Work/Study	→	Background and Accompaniment	→	Valence	-0.018	0.019	.335
Work/Study	→	Physiological Arousal	→	Valence	0.015	0.02	.471
Work/Study		*TIE* →		Valence	-0.118	0.053	**.026**
Work/Study		*TE* →		Valence	-0.249	0.101	**.014**
Travelling	→	Social Interaction	→	Valence	-0.001	0.007	.836
Travelling	→	Emotion Regulation	→	Valence	-0.036	0.026	.164
Travelling	→	Focus and Concentration	→	Valence	0.004	0.015	.799
Travelling	→	Background and Accompaniment	→	Valence	0.015	0.02	.444
Travelling	→	Physiological Arousal	→	Valence	0.033	0.029	.252
Travelling		*TIE* →		Valence	0.015	0.028	.583
Travelling		*TE* →		Valence	-0.224	0.113	**.047**
Relaxation	→	Social Interaction	→	Valence	-0.002	0.009	.830
Relaxation	→	Emotion Regulation	→	Valence	-0.022	0.017	.192
Relaxation	→	Focus and Concentration	→	Valence	0.029	0.021	.155

Relaxation	→	Background and Accompaniment	→	Valence	-0.031	0.022	.154
Relaxation	→	Physiological Arousal	→	Valence	-0.023	0.02	.264
Relaxation		*TIE* →		Valence	-0.049	0.03	.104
Relaxation		*TE* →		Valence	-0.256	0.123	**.037**
Exercise	→	Social Interaction	→	Valence	-0.005	0.021	.827
Exercise	→	Emotion Regulation	→	Valence	-0.046	0.035	.194
Exercise	→	Focus and Concentration	→	Valence	-0.057	0.039	.152
Exercise	→	Background and Accompaniment	→	Valence	0.015	0.033	.635
Exercise	→	Physiological Arousal	→	Valence	0.262	0.139	.061†
Exercise		*TIE* →		Valence	0.17	0.108	.114
Exercise		*TE* →		Valence	-0.091	0.218	.677
Socialising	→	Social Interaction	→	Valence	-0.018	0.082	.825
Socialising	→	Emotion Regulation	→	Valence	0.002	0.03	.953
Socialising	→	Focus and Concentration	→	Valence	0.055	0.046	.227
Socialising	→	Background and Accompaniment	→	Valence	-0.103	0.058	.075†
Socialising	→	Physiological Arousal	→	Valence	-0.007	0.06	.911
Socialising		*TIE* →		Valence	-0.071	0.097	.465
Socialising		*TE* →		Valence	0.015	0.333	.965
Recreational Activity	→	Social Interaction	→	Valence	-0.002	0.008	.831
Recreational Activity	→	Emotion Regulation	→	Valence	-0.01	0.015	.512
Recreational Activity	→	Focus and Concentration	→	Valence	0.003	0.017	.859
Recreational Activity	→	Background and Accompaniment	→	Valence	0.001	0.035	.979
Recreational Activity	→	Physiological Arousal	→	Valence	-0.037	0.027	.166
Recreational Activity		*TIE* →		Valence	-0.045	0.032	.165
Recreational Activity		*TE* →		Valence	0.015	0.154	.922
Work/Study	→	Social Interaction	→	Instrumentalness	0.007	0.01	.473
Work/Study	→	Emotion Regulation	→	Instrumentalness	-0.003	0.008	.716
Work/Study	→	Focus and Concentration	→	Instrumentalness	0.049	0.052	.354
Work/Study	→	Background and Accompaniment	→	Instrumentalness	0.020	0.024	.398

Work/Study	→	Physiological Arousal	→	Instrumentalness	-0.008	0.01	.458
Work/Study		*TIE* →		Instrumentalness	0.066	0.062	.292
Work/Study		*TE* →		Instrumentalness	0.319	0.123	**.010**
Travelling	→	Social Interaction	→	Instrumentalness	0.005	0.008	.494
Travelling	→	Emotion Regulation	→	Instrumentalness	-0.008	0.023	.720
Travelling	→	Focus and Concentration	→	Instrumentalness	-0.002	0.007	.801
Travelling	→	Background and Accompaniment	→	Instrumentalness	-0.017	0.022	.441
Travelling	→	Physiological Arousal	→	Instrumentalness	-0.017	0.017	.300
Travelling		*TIE* →		Instrumentalness	-0.039	0.029	.183
Travelling		*TE* →		Instrumentalness	-0.072	0.104	.488
Relaxation	→	Social Interaction	→	Instrumentalness	0.007	0.01	.449
Relaxation	→	Emotion Regulation	→	Instrumentalness	-0.005	0.014	.723
Relaxation	→	Focus and Concentration	→	Instrumentalness	-0.014	0.018	.430
Relaxation	→	Background and Accompaniment	→	Instrumentalness	0.034	0.025	.177
Relaxation	→	Physiological Arousal	→	Instrumentalness	0.012	0.013	.389
Relaxation		*TIE* →		Instrumentalness	0.034	0.033	.302
Relaxation		*TE* →		Instrumentalness	0.127	0.106	.229
Exercise	→	Social Interaction	→	Instrumentalness	0.018	0.023	.440
Exercise	→	Emotion Regulation	→	Instrumentalness	-0.01	0.029	.722
Exercise	→	Focus and Concentration	→	Instrumentalness	0.028	0.032	.386
Exercise	→	Background and Accompaniment	→	Instrumentalness	-0.017	0.035	.631
Exercise	→	Physiological Arousal	→	Instrumentalness	-0.135	0.102	.188
Exercise		*TIE* →		Instrumentalness	-0.116	0.084	.166
Exercise		*TE* →		Instrumentalness	0.158	0.223	.478
Socialising	→	Social Interaction	→	Instrumentalness	0.072	0.093	.440
Socialising	→	Emotion Regulation	→	Instrumentalness	0	0.007	.953
Socialising	→	Focus and Concentration	→	Instrumentalness	-0.027	0.035	.444
Socialising	→	Background and Accompaniment	→	Instrumentalness	0.113	0.065	.079†
Socialising	→	Physiological Arousal	→	Instrumentalness	0.003	0.031	.911
Socialising		*TIE* →		Instrumentalness	0.162	0.111	.143
Socialising		*TE* →		Instrumentalness	-0.394	0.211	.062†
Recreational Activity	→	Social Interaction	→	Instrumentalness	0.007	0.011	.540

Recreational Activity	→	Emotion Regulation	→	Instrumentalness	-0.002	0.007	.748
Recreational Activity	→	Focus and Concentration	→	Instrumentalness	-0.001	0.008	.860
Recreational Activity	→	Background and Accompaniment	→	Instrumentalness	-0.001	0.038	.979
Recreational Activity	→	Physiological Arousal	→	Instrumentalness	0.019	0.018	.303
Recreational Activity		*TIE* →		Instrumentalness	0.021	0.047	.649
Recreational Activity		*TE* →		Instrumentalness	0.081	0.147	.580

Note. *TIE* = Total Indirect Effect. *TE* = Total Effect. For individual *a* and *b* paths, refer to Tables 25 and 26 in section 8.5.6.2.

Appendix I: R syntax for Study 3 recommendations

```
# load packages
library(spotifyr)
library(lubridate)
library(tidyverse)
library(knitr)
library(httpuv)

# create an app and set credentials
(https://developer.spotify.com/dashboard/applications)
Sys.setenv(SPOTIFY_CLIENT_ID = "CLIENT ID HERE")
Sys.setenv(SPOTIFY_CLIENT_SECRET = "CLIENT SECRET HERE")
access_token <- get_spotify_access_token()

# K genre strings (Bansal et al., 2020): "classical", "country", "dance", "folk",
"hip-hop", "indie", "jazz", "metal", "pop", "rock"

# 1. Work & Study - as mediated by FaC
recommended_tracks.work_study <- as.data.frame(get_recommendations(limit = 50,
market = "GB", seed_artists = NULL, seed_genres = "", seed_tracks = NULL,
target_acousticness = 0.399, target_danceability = NULL, target_duration_ms = NULL,
target_energy = 0.534, target_instrumentalness = 0.376, target_popularity = NULL,
target_speechiness = NULL, target_tempo = NULL, target_valence = 0.378,
authorization = get_spotify_access_token(), include_seeds_in_response = FALSE))

# create playlist
work_study_playlist <- create_playlist("Username", name = "", public = TRUE,
collaborative = FALSE, description = "", authorization =
get_spotify_authorization_code())

# add recommended tracks to playlist
add_tracks_to_playlist(work_study_playlist$id, uris =
recommended_tracks.work_study$uri, authorization =
get_spotify_authorization_code())
```

```
# 2. Work & Study - absence of mediator (only the DE on valence and TE on Inst.)
recommended_tracks.work_study <- get_recommendations(limit = 50, market = "GB",
seed_artists = NULL, seed_genres = "", seed_tracks = NULL, target_acousticness =
NULL, target_danceability = NULL, target_duration_ms = NULL, target_energy = NULL,
target_instrumentalness = 0.376, target_popularity = NULL, target_speechiness =
NULL, target_tempo = NULL, target_valence = 0.416, authorization =
get_spotify_access_token(), include_seeds_in_response = FALSE)

# create playlist
work_study_playlist <- create_playlist("Username", name = "", public = TRUE,
collaborative = FALSE, description = "", authorization =
get_spotify_authorization_code())

add_tracks_to_playlist(work_study_playlist$id, uris =
recommended_tracks.work_study$uri, authorization =
get_spotify_authorization_code())

# 3. Travel (TE value on Valence - significant DE)
recommended_tracks.travel <- get_recommendations(limit = 50, market = "GB",
seed_artists = NULL,seed_genres = "", seed_tracks = NULL,target_acousticness =
NULL, target_danceability = NULL, target_duration_ms = NULL, target_energy = NULL,
target_instrumentalness = NULL, target_popularity = NULL, target_speechiness =
NULL, target_tempo = NULL, target_valence = 0.421, authorization =
get_spotify_access_token(), include_seeds_in_response = FALSE)

travel_playlist <- create_playlist("Username", name = "", public = TRUE,
collaborative = FALSE, description = "", authorization =
get_spotify_authorization_code())

add_tracks_to_playlist(travel_playlist$id, uris = recommended_tracks.travel$uri,
authorization = get_spotify_authorization_code())

# 4. Relaxation (TE mediated by BaA)
recommended_tracks.relaxation <- get_recommendations(limit = 50, market = "GB",
seed_artists = NULL,
                                                     seed_genres = "", seed_tracks
= NULL,
                                                     target_acousticness = NULL,
target_danceability = 0.456,
                                                     target_duration_ms = NULL,
target_energy = NULL,
                                                     target_instrumentalness =
NULL, target_popularity = NULL,
                                                     target_speechiness = NULL,
target_tempo = NULL, target_valence = NULL,
                                                     authorization =
get_spotify_access_token(), include_seeds_in_response = FALSE)

relaxing_playlist <- create_playlist("Username", name = "", public = TRUE,
collaborative = FALSE,
                                     description = "", authorization =
get_spotify_authorization_code())

add_tracks_to_playlist(relaxing_playlist$id, uris =
recommended_tracks.relaxation$uri, authorization =
get_spotify_authorization_code())

# 5. Exercise (mediated by PA - IEs only!)
recommended_tracks.exercise <- get_recommendations(limit = 50, market = "GB",
seed_artists = NULL,
                                                   seed_genres = "", seed_tracks =
NULL,
```

```
                                                            target_acousticness = 0.216,
target_danceability = 0.542,
                                                            target_duration_ms = NULL,
target_energy = 0.682,
                                                            target_instrumentalness = NULL,
target_popularity = NULL,
                                                            target_speechiness = NULL,
target_tempo = NULL, target_valence = 0.536,
                                                            authorization =
get_spotify_access_token(), include_seeds_in_response = FALSE)

exercise_playlist <- create_playlist("Username", name = "", public = TRUE,
collaborative = FALSE,
                                         description = "", authorization =
get_spotify_authorization_code())

add_tracks_to_playlist(exercise_playlist$id, uris =
recommended_tracks.exercise$uri, authorization = get_spotify_authorization_code())

# 6. Socialising (DE only)
recommended_tracks.socialising <- get_recommendations(limit = 50, market = "GB",
seed_artists = NULL,
                                                               seed_genres = "", seed_tracks
= NULL,
                                                               target_acousticness = NULL,
target_danceability = NULL,
                                                               target_duration_ms = NULL,
target_energy = NULL,
                                                               target_instrumentalness =
0.068, target_popularity = NULL,
                                                               target_speechiness = NULL,
target_tempo = NULL, target_valence = NULL,
                                                               authorization =
get_spotify_access_token(), include_seeds_in_response = FALSE)

socialising_playlist <- create_playlist("Username", name = "", public = TRUE,
collaborative = FALSE,
                                            description = "", authorization =
get_spotify_authorization_code())

add_tracks_to_playlist(socialising_playlist$id, uris =
recommended_tracks.socialising$uri, authorization =
get_spotify_authorization_code())

# 7. neutral
recommended_tracks.neutral <- get_recommendations(limit = 50, market = "GB",
seed_artists = NULL,
                                                           seed_genres = "", seed_tracks =
NULL,
                                                           target_acousticness = NULL,
target_danceability = NULL,
                                                           target_duration_ms = NULL,
target_energy = NULL,
                                                           target_instrumentalness = NULL,
target_popularity = NULL,
                                                           target_speechiness = NULL,
target_tempo = NULL, target_valence = NULL,
                                                           authorization =
get_spotify_access_token(), include_seeds_in_response = FALSE)

neutral_playlist <- create_playlist("Username", name = "", public = TRUE,
collaborative = FALSE,
```

```
                                             description = "", authorization =
get_spotify_authorization_code())

add_tracks_to_playlist(neutral_playlist$id, uris = recommended_tracks.neutral$uri,
authorization = get_spotify_authorization_code())
```

Appendix J: Study 3: Qualtrics form for part 1 (direction to a given playlist)

Q1) This study provides recommendations that are intended to fit your current activity. Are you currently in a situation where you can listen to music?

○ Yes, I can listen to music right now (1)

○ No, I cannot listen to music right now (if you wish to defer participation, provide an email address to receive a link and take part when convenient) (2) ____

Q2) What is your age in years?

____

Q3) What is your gender?

○ Male (1)

○ Female (2)

○ Non-binary/third gender (3)

○ Prefer not to say (4)

Q4) Which of the following best describes your musical background?

○ Non-musician (1)

○ Amateur musician (2)

○ Higher level musician (3)

○ Professional musician (4)

Q5) Have you taken part in this study before?

○ Yes, I have taken part before (1)

○ No, this is the first time I've taken part (2)

Please provide a valid email address to which we can send our follow-up survey:

____

Opt-in to be in with a chance of winning a £50 Amazon gift voucher! (upon completing both parts 1 and 2 of our study):

Please note: This will be transferred to the winner via email, so please ensure your address is correct. The winner will be selected in a raffle upon completion of data collection. This is offered in GBP only so is subject to relevant exchange rates if you are outside the UK.

- ◯ Opt-in (1)
- ◯ Opt-out (2)

Q6) Which of the following activities best describes what you are currently doing/will be doing for the next hour or so?

*Besides the current study that is!

- ◯ Working/Studying (1)
- ◯ Travelling (e.g., walking, public transport) (2)
- ◯ Relaxing (3)
- ◯ Routine Activity (e.g., chores) (4)
- ◯ Exercising (5)
- ◯ Socialising (6)

Q7) Which of the following 10 genres is most closely aligned with your personal taste:

- ◯ Classical (1)
- ◯ Country (2)
- ◯ Dance (3)
- ◯ Folk (4)
- ◯ Hip-hop (5)
- ◯ Indie (6)
- ◯ Jazz (7)

○ Metal (8)

○ Pop (9)

○ Rock (10)

Q8) **In your current situation why are/do you want to listen to music?**

Please read each of the following statements carefully and respond according to the extent to which the described reason for music listening by degree of importance.

	Not very important (1)	Slightly important (2)	Somewhat important (3)	Quite Important (4)	Extremely important (5)
To identify with others through your shared values and/or culture (S1)	○	○	○	○	○
To help express your identities and values to others (S2)	○	○	○	○	○
To feel that certain artists, pieces, or genres of music are central to your social group's culture and sets you apart from others (S3)	○	○	○	○	○
To help you bond with others, and to subsequently feel a sense of belonging with those individuals (S4)	○	○	○	○	○

To act as a topic of discussion with others and ease communication or interaction (S5)	○	○	○	○	○
To match a group's dynamic so you are able to bond with group members when listening with others (S6)	○	○	○	○	○
To distract yourself from negative or stressful situations (E1)	○	○	○	○	○
To relieve stress and negative emotions associated with negative events or situations (E2)	○	○	○	○	○

To manage emotions that you may be experiencing despite external influences, whether they are positive or negative (E3)	○	○	○	○	○
To find meaning within music that allows you to reduce negative emotions or moods (E4)	○	○	○	○	○
To feel certain specific emotions, such as joy or sadness (E5)	○	○	○	○	○
To help you reverse your emotions or moods (E6)	○	○	○	○	○
To help you focus or concentrate on tasks (F1)	○	○	○	○	○

To help you 'flow' when trying to concentrate on something (F2)	○	○	○	○	○
To stop external factors from distracting you when trying to concentrate on a task (F3)	○	○	○	○	○
To help you attain the necessary mindset to working on certain tasks (F4)	○	○	○	○	○
To avoid silence when you're alone (e.g., playing music when nobody else is home) (B1)	○	○	○	○	○

To feel a sense of company in the absence of others (e.g., playing the radio when home alone) (B2)	○	○	○	○	○
To reduce feelings of being lonely when social interaction is not possible (B3)	○	○	○	○	○
To provide background noise and remove silence (B4)	○	○	○	○	○
To help you maintain pacing during physical activities, such as yoga, walking or whilst in the gym (P1)	○	○	○	○	○

To help physically stimulate you to carry out physical tasks, such as exercise or sports (P2)	○	○	○	○	○
To help you achieve goals by motivating you to further action (such as increased effort during exercise) (P3)	○	○	○	○	○

Q9) Please read the following statements carefully and indicate the extent to which you agree with each one:

	Completely disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Completely agree (7)
I am a music lover (Expertise_MusicLover)	○	○	○	○	○	○	○
Compared to my peers, I listen to a lot of music (Expertise_ListensALot)	○	○	○	○	○	○	○
Compared to my peers, I am an expert on music (Expertise_ExpertOnMusic)	○	○	○	○	○	○	○

Q10) Please read the following statements carefully and indicate the extent to which you agree with each one:

	Completely disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Completely agree (7)
Technology never works (TrTe_TechnologyNeverWorks)							
I'm less confident when I use technology (TrTe_LessConfidentWhenUsingTechnology)							
The usefulness of technology is highly overrated (TrTe_UsefulnessOfTechnologyIsOverrated)							
Technology may cause harm to people (TrTe_TechnologyMayCauseHarm)							

Q11) Please read the following statements carefully and indicate the extent to which you agree with each one:

	Completely disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Completely agree (7)
I like to give feedback on the content I engage with (ItPF_LikesToGiveFeedback)	○	○	○	○	○	○	○
Normally I wouldn't rate any tracks/songs (ItPF_NormallyWouldntRate)	○	○	○	○	○	○	○
I only sparingly give feedback (ItPF_GivesFeedbackSparingly)	○	○	○	○	○	○	○
I don't mind rating tracks/songs (ItPF_DontMindRatingItems)	○	○	○	○	○	○	○
Overall, rating tracks/songs is not beneficial for me (ItPF_RatingItemsNotBeneficial)	○	○	○	○	○	○	○

Branches to given Spotify playlists embedded as hyperlinks.

Playlists are available upon reasonable request.

Appendix K: Study 3: Qualtrics form for part 2 (evaluation of given playlist)

Q1) The following survey aims to assess the effectiveness of the recommendations we provided you with in the previous part of the study. As such, please respond from the perspective of the recommendations we provided you with, rather than with Spotify more generally.

○ Yep, I understand that I should respond from the perspective of the playlist you gave me! (1)

Q2) For approximately how many minutes did you listen to the recommended playlist:

____

Q3) Please select the option that best describes the way in which you interacted with the playlist:

○ I listened to the playlist on shuffle and without skipping tracks (1)

○ I listened to the playlist in order and without skipping tracks (2)

○ I listened to the playlist on shuffle and I skipped tracks (3)

○ I listened to the playlist in order and I skipped tracks (4)

Display This Question:

If Please select the option that best describes the way in which you interacted with the playlist: = 3

Or Please select the option that best describes the way in which you interacted with the playlist: = 4

Q4) Approximately how many tracks did you skip through?

____

Perceived Quality (PRQ)

Q5) Please read each statement carefully and indicate the extent to which you agree:

	Completely disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Completely agree (7)
I liked the tracks recommended by the system (PRQ_LikedTheItems)	○	○	○	○	○	○	○
The recommended tracks fitted my preference (PRQ_FitPreference)	○	○	○	○	○	○	○
The recommended tracks were well-chosen (PRQ_ItemsWellChosen)	○	○	○	○	○	○	○
The recommended tracks were relevant (PRQ_ItemsWereRelevant)	○	○	○	○	○	○	○
The system recommended too many bad tracks (PRQ_RecommendedTooManyBadItems)	○	○	○	○	○	○	○
I didn't like any of the recommended tracks (PRQ_DidntLikeAnyRecommendedItems)	○	○	○	○	○	○	○

Perceived Efficacy

Q6) Please read each statement carefully and indicate the extent to which you agree:

	Completely disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Completely agree (7)
I would recommend the system to others (PSE_WouldRecommendToOthers)	○	○	○	○	○	○	○
The system is useless (PSE_SystemIsUseless)	○	○	○	○	○	○	○
The system makes me more aware of my choice options (PSE_MakesMeMoreAwareOfChoiceOptions)	○	○	○	○	○	○	○
I make better choices with the system (PSE_MakeBetterChoicesWithSystems)	○	○	○	○	○	○	○
I can find better items using the recommender system (PSE_CanFindBetterItemsWithSystem)	○	○	○	○	○	○	○

Q8) SSPC

Please read each statement carefully and indicate the extent to which you agree:

	Completely disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Completely agree (7)
I'm afraid the system discloses private Information about me (SSPC_AfraidTheSystemDisclosesPrivateInfo)	○	○	○	○	○	○	○
The system invades my privacy (SSPC_SystemInvadesPrivacy)	○	○	○	○	○	○	○
I feel confident that the system respects my privacy (SSPC_SystemRespectsPrivacy)	○	○	○	○	○	○	○
I'm uncomfortable providing data to the system (SSPC_UncomfortableProvidingDataToSystem)	○	○	○	○	○	○	○
I think the system respects the confidentiality of my data (SSPC_SystemRespectsConfidentiality)	○	○	○	○	○	○	○

Q9) Would you like to be updated on developments and/or outputs of this research (e.g., publications, conference presentations)?

Note that if you wish to be updated on research outputs, we will send these directly to the email address you have provided.

- ○ Yes, keep me up to date with developments/outputs of this research (1)
- ○ No thanks! (2)

Appendix L: R Syntax for Study 3 SAM analysis

```
# load required packages
library(readr)
library(tidyr)
library(dplyr)
library(psych)
library(lavaan)
library(lavaan.survey)
library(lavaanPlot)
library(semTools)
library(semPlot)
options(max.print = 100000)

# load data
data_all <- # load data

## reverse code items
# TrTe key
keys <- c(-1,-1,-1,-1)
RecodedTrTe <-as.data.frame(reverse.code(keys,data_all[,
c('TrTe_TechnologyNeverWorks', 'TrTe_LessConfidentWhenUsingTechnology',

'TrTe_UsefulnessOfTechnologyIsOverrated', 'TrTe_TechnologyMayCauseHarm')],
mini = 1, maxi = 7))
# ItPF key
keys <- c(1,-1,-1,-1,1)
RecodedItPF <- as.data.frame(reverse.code(keys, data_all[,
c('ItPF_LikesToGiveFeedback', 'ItPF_NormallyWouldntRate',
'ItPF_GivesFeedbackSparingly', 'ItPF_RatingItemsNotBeneficial',
'ItPF_DontMindRatingItems')], mini = 1, maxi = 7))

# SSPC key
keys <- c(1,1,-1,1,-1)
RecodedSSPC <- as.data.frame(reverse.code(keys, data_all[,
c('SSPC_AfraidTheSystemDisclosesPrivateInfo','SSPC_SystemInvadesPrivacy',

'SSPC_SystemRespectsPrivacy', 'SSPC_UncomfortableProvidingDataToSystem',

'SSPC_SystemRespectsConfidentiality')], mini = 1, maxi = 7))
# PRQ key
keys <- c(1,1,1,1,-1,-1)
RecodedPRQ <- as.data.frame(reverse.code(keys, data_all[,
c('PRQ_LikedTheItems', 'PRQ_FitPreference', 'PRQ_ItemsWellChosen',

'PRQ_ItemsWereRelevant','PRQ_RecommendedTooManyBadItems',

'PRQ_DidntLikeAnyRecommendedItems')], mini = 1, maxi = 7))
# PSE key
keys <- c(1,-1,1,1,1,-1)
RecodedPSE <- as.data.frame(reverse.code(keys, data_all[,
c('PSE_WouldRecommendToOthers', 'PSE_SystemIsUseless',
'PSE_MakesMeMoreAwareOfChoiceOptions','PSE_MakeBetterChoicesWithSystems',
'PSE_CanFindBetterItemsWithSystem','PSE_CanFindBetterTracksWithoutSystem')]
, mini = 1, maxi = 7))


# bind reversed items together
bound.items <- cbind(RecodedTrTe, RecodedItPF, RecodedSSPC, RecodedPRQ,
RecodedPSE)
```

```
# list (non-reversed) items
non_reversed_items <-
c('ItPF_LikesToGiveFeedback','ItPF_DontMindRatingItems','SSPC_AfraidTheSyst
emDisclosesPrivateInfo',

'SSPC_SystemInvadesPrivacy','SSPC_UncomfortableProvidingDataToSystem','PRQ_
LikedTheItems',
                        'PRQ_FitPreference', 'PRQ_ItemsWellChosen',
'PRQ_ItemsWereRelevant','PSE_WouldRecommendToOthers',

'PSE_MakesMeMoreAwareOfChoiceOptions','PSE_MakeBetterChoicesWithSystems',
'PSE_CanFindBetterItemsWithSystem')

# update bound.items by dropping non_reversed_items
bound.items <- bound.items[, !(names(bound.items) %in% non_reversed_items)]

# relabel columns to make reverse codes clear
colnames(bound.items) = c('TrTe_TechnologyNeverWorks_R',
'TrTe_LessConfidentWhenUsingTechnology_R',
                          'TrTe_UsefulnessOfTechnologyIsOverrated_R',
'TrTe_TechnologyMayCauseHarm_R',

"ItPF_NormallyWouldntRate_R",'ItPF_GivesFeedbackSparingly_R',
                          'ItPF_RatingItemsNotBeneficial_R',
"SSPC_SystemRespectsPrivacy_R",
                          "SSPC_SystemRespectsConfidentiality_R",
"PRQ_RecommendedTooManyBadItems_R",
                          "PRQ_DidntLikeAnyRecommendedItems_R",
                          "PSE_SystemIsUseless_R",
"PSE_CanFindBetterTracksWithoutSystem_R")

# add recoded items to main dataset and (optionally) remove redundant dfs
data_all <- cbind(data_all, bound.items)
rm(keys, bound.items, non_reversed_items, RecodedTrTe, RecodedItPF,
RecodedSSPC, RecodedPRQ, RecodedPSE)

# only retain cases completing both study sections
data_twostagecomplete <- data_all %>% drop_na(ResponseId...70)

# filter out cases of Routine/Recreational activities
data_filtered <- filter(data_twostagecomplete, Activity %in%
c('1','2','3','5','6'))

# if relaxation + BaA >2.372, Exercie + PA < 3.872
data_doublefilteredRelaxation <- filter(data_filtered, Activity == '3' &
Background_and_Accompaniment_Value > 2.372) # n = 7
data_doublefilteredExercise <- filter(data_filtered, Activity == '5' &
Physiological_Arousal_Value < 3.872) # n = 0 (nothing else needs doing)
IDs <- as.character(data_doublefilteredRelaxation$ResponseId...9) # compile
response IDs

# filter out rows containing those IDs in list
data_filtered <-
data_filtered[!grepl("R_1mXgbgBXSlKuCpw|R_9KKXJQi59OkddKt|R_3QQyg6JMCPSCHYv
|R_2QYCNbSdNTnGTHz|R_1Ho7iL5iNW1CGOQ|R_10wQWTePxOkh6HL|R_1hzhHTAODnlAsYL",
                          data_filtered$ResponseId...9),]
# clean environment (again!)
rm(data_doublefilteredRelaxation,data_doublefilteredExercise,IDs)

# cross-validate FML model
FML.Model <-
'ISB =~ S1 + S2 + S3 + S4 + S5 + S6
```

```
ER =~ E1 + E2 + E3 + E4 + E5 + E6
FaC =~ F1 + F2 + F3 + F4
BaA =~ B1 + B2 + B3 + B4
PA =~ P1 + P2 + P3'

# initial fit
cfa.fit <- cfa(model = FML.Model, data = data_all, std.lv = TRUE, estimator
= 'MLM')
# model dependency in data
survey.design <- svydesign(ids =~ Participant_ID, nest = TRUE, prob =~
NULL, strata = NULL, data = data_all)
# adjust initial fit
survey.fitCFA <- lavaan.survey(lavaan.fit = cfa.fit, survey.design =
survey.design)
# model fit
fitmeasures(survey.fitCFA, c('chisq.scaled', 'df', 'pvalue.scaled',
'cfi.robust', 'tli.robust', 'rmsea.robust'))
# summary statistics
summary(survey.fitCFA, standardized = TRUE, rsquare = TRUE)
# AVE
AVE(survey.fitCFA, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)
# reliability coeff
compRelSEM(survey.fitCFA, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)
# plot model
semPaths(survey.fitCFA, what = "std", residuals = TRUE, reorder = TRUE,
cardinal = TRUE, weighted = FALSE, layout = "tree2", rotation = 2, curve =
TRUE, curvature = 2, style = "OpenMX", intercepts = FALSE, sizeMan = 3,
sizeLat = 5, nDigits = 3, theme = "colorblind")
rm(cfa.fit,survey.fitCFA,FML.Model)

### Individual fit and inspection of EF factors
# 1. TrTe model + scores
TrTe.model <- 'TrTe_Fscore =~ TrTe_TechnologyNeverWorks_R +
TrTe_LessConfidentWhenUsingTechnology_R +
TrTe_UsefulnessOfTechnologyIsOverrated_R + TrTe_TechnologyMayCauseHarm_R'

# fit
TrTe.fit <- cfa(model = TrTe.model, data = data_all, std.lv = TRUE,
estimator = 'MLM')

# model dependency in data
survey.design <- svydesign(ids =~ Participant_ID, nest = TRUE, prob =~
NULL, strata = NULL, data = data_all)

# adjust initial fit
TrTe.fit <- lavaan.survey(lavaan.fit = TrTe.fit, survey.design =
survey.design)

# model fit
fitmeasures(TrTe.fit, c('chisq.scaled', 'df', 'pvalue.scaled',
'cfi.robust', 'tli.robust', 'rmsea.robust'))

# summary statistics
summary(TrTe.fit, standardized = TRUE, rsquare = TRUE)
```

```
# AVE
AVE(TrTe.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff: if tau.eq = FALSE, coeff is omega. if TRUE, coeff is
alpha.
compRelSEM(TrTe.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

# reliability coeff if item removed
compRelSEM(TrTe.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = 'TrTe_TechnologyMayCauseHarm_R', omit.imps =
c("no.conv", "no.se"),
           return.df = TRUE)

# respecify model (dropping item)
TrTe.model <- 'TrTe_Fscore =~ TrTe_TechnologyNeverWorks_R +
TrTe_LessConfidentWhenUsingTechnology_R +
TrTe_UsefulnessOfTechnologyIsOverrated_R'

# fit
TrTe.fit <- cfa(model = TrTe.model, data = data_all, std.lv = TRUE,
estimator = 'MLM')

# adjust initial fit
TrTe.fit <- lavaan.survey(lavaan.fit = TrTe.fit, survey.design =
survey.design)

# summary statistics
summary(TrTe.fit, standardized = TRUE, rsquare = TRUE)

# AVE
AVE(TrTe.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff: if tau.eq = FALSE, coeff is omega. if TRUE, coeff is
alpha.
compRelSEM(TrTe.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

# 2. ItPF model + scores
ItPF.model <- 'ItPF_Fscore =~ ItPF_LikesToGiveFeedback +
ItPF_NormallyWouldntRate_R + ItPF_GivesFeedbackSparingly_R +
ItPF_DontMindRatingItems + ItPF_RatingItemsNotBeneficial_R'
```

```
# fit
ItPF.fit <- cfa(model = ItPF.model, data = data_all, std.lv = TRUE,
estimator = 'MLM')

# adjust initial fit
ItPF.fit <- lavaan.survey(lavaan.fit = ItPF.fit, survey.design =
survey.design)

# summary statistics
summary(ItPF.fit, standardized = TRUE, rsquare = TRUE)

# AVE
AVE(ItPF.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff
compRelSEM(ItPF.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

# 3. Expertise model + scores
Expertise.model <- 'Expertise_Fscore =~ Expertise_ExpertOnMusic +
Expertise_ListensALot + Expertise_MusicLover'

# fit
Expertise.fit <- cfa(model = Expertise.model, data = data_all, std.lv =
TRUE, estimator = 'MLM')

# adjust initial fit
Expertise.fit <- lavaan.survey(lavaan.fit = Expertise.fit, survey.design =
survey.design)

# summary statistics
summary(Expertise.fit, standardized = TRUE, rsquare = TRUE)

# AVE
AVE(Expertise.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff
compRelSEM(Expertise.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

# 4. PRQ model + scores
PRQ.model <- 'PRQ_Fscore =~ PRQ_LikedTheItems + PRQ_FitPreference +
PRQ_ItemsWellChosen + PRQ_ItemsWereRelevant +
PRQ_RecommendedTooManyBadItems_R + PRQ_DidntLikeAnyRecommendedItems_R'

# fit
PRQ.fit <- cfa(model = PRQ.model, data = data_twostagecomplete, std.lv =
TRUE, estimator = 'MLM')
```

```
survey.design <- svydesign(ids =~ Participant_ID, nest = TRUE, prob =~
NULL, strata = NULL, data = data_all)

# adjust initial fit
PRQ.fit <- lavaan.survey(lavaan.fit = PRQ.fit, survey.design =
survey.design)

# summary statistics
summary(PRQ.fit, standardized = TRUE, rsquare = TRUE)

# AVE
AVE(PRQ.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff (tau.eq = FALSE = Omega, tau.eq = TRUE = Alpha)
compRelSEM(PRQ.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

# reliability coeff (tau.eq = FALSE = Omega, tau.eq = TRUE = Alpha)
compRelSEM(PRQ.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = 'PRQ_DidntLikeAnyRecommendedItems_R',
omit.imps = c("no.conv", "no.se"),
           return.df = TRUE)

PRQ.model <- 'PRQ_Fscore =~ PRQ_LikedTheItems + PRQ_FitPreference +
PRQ_ItemsWellChosen + PRQ_ItemsWereRelevant +
PRQ_RecommendedTooManyBadItems_R'

PRQ.fit <- cfa(model = PRQ.model, data = data_twostagecomplete, std.lv =
TRUE, estimator = 'MLM')

# adjust initial fit
PRQ.fit <- lavaan.survey(lavaan.fit = PRQ.fit, survey.design =
survey.design)

# summary statistics
summary(PRQ.fit, standardized = TRUE, rsquare = TRUE)

# AVE
AVE(PRQ.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff (tau.eq = FALSE = Omega, tau.eq = TRUE = Alpha)
compRelSEM(PRQ.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)
```

```
# 5. PSE model + scores
PSE.model <- 'PSE_Fscore =~ PSE_WouldRecommendToOthers +
PSE_SystemIsUseless_R + PSE_MakesMeMoreAwareOfChoiceOptions +
PSE_MakeBetterChoicesWithSystems + PSE_CanFindBetterItemsWithSystem +
PSE_CanFindBetterTracksWithoutSystem_R'

PSE.fit <- cfa(model = PSE.model, data = data_twostagecomplete, std.lv =
TRUE, estimator = 'MLM')

# adjust initial fit
PSE.fit <- lavaan.survey(lavaan.fit = PSE.fit, survey.design =
survey.design)

# summary statistics
summary(PSE.fit, standardized = TRUE, rsquare = TRUE)

# AVE
AVE(PSE.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff
compRelSEM(PSE.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

# reliability coeff
compRelSEM(PSE.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = 'PSE_SystemIsUseless_R', omit.imps =
c("no.conv", "no.se"),
           return.df = TRUE)

PSE.model <- 'PSE_Fscore =~ PSE_WouldRecommendToOthers +
PSE_MakesMeMoreAwareOfChoiceOptions + PSE_MakeBetterChoicesWithSystems +
PSE_CanFindBetterItemsWithSystem + PSE_CanFindBetterTracksWithoutSystem_R'

PSE.fit <- cfa(model = PSE.model, data = data_twostagecomplete, std.lv =
TRUE, estimator = 'MLM')

# adjust initial fit
PSE.fit <- lavaan.survey(lavaan.fit = PSE.fit, survey.design =
survey.design)

# summary statistics
summary(PSE.fit, standardized = TRUE, rsquare = TRUE)

# AVE
AVE(PSE.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff
compRelSEM(PSE.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
```

```
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

compRelSEM(PSE.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = 'PSE_CanFindBetterTracksWithoutSystem_R',
omit.imps = c("no.conv", "no.se"),
           return.df = TRUE)

PSE.model <- 'PSE_Fscore =~ PSE_WouldRecommendToOthers +
PSE_MakesMeMoreAwareOfChoiceOptions + PSE_MakeBetterChoicesWithSystems +
PSE_CanFindBetterItemsWithSystem'

PSE.fit <- cfa(model = PSE.model, data = data_twostagecomplete, std.lv =
TRUE, estimator = 'MLM')

# adjust initial fit
PSE.fit <- lavaan.survey(lavaan.fit = PSE.fit, survey.design =
survey.design)

# summary statistics
summary(PSE.fit, standardized = TRUE, rsquare = TRUE)

# AVE
AVE(PSE.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff
compRelSEM(PSE.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

# 6. SSPC model + scores
SSPC.model <- 'SSPC_Fscore =~ SSPC_AfraidTheSystemDisclosesPrivateInfo +
SSPC_SystemInvadesPrivacy + SSPC_SystemRespectsPrivacy_R +
SSPC_UncomfortableProvidingDataToSystem +
SSPC_SystemRespectsConfidentiality_R'

SSPC.fit <- cfa(model = SSPC.model, data = data_twostagecomplete, std.lv =
TRUE, estimator = 'MLM')

# adjust initial fit
SSPC.fit <- lavaan.survey(lavaan.fit = SSPC.fit, survey.design =
survey.design)

# summary statistics
summary(SSPC.fit, standardized = TRUE, rsquare = TRUE)

# AVE
```

```
AVE(SSPC.fit, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff
compRelSEM(SSPC.fit, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

# remove all models as values (just to keep things neat)
rm(Expertise.model,ItPF.model,PRQ.model,PSE.model,SSPC.model,TrTe.model)

# sam trust
trust.model <-
'# latent variables
TrTe =~ TrTe_TechnologyNeverWorks_R +
TrTe_LessConfidentWhenUsingTechnology_R +
TrTe_UsefulnessOfTechnologyIsOverrated_R
ItPF =~ ItPF_LikesToGiveFeedback + ItPF_NormallyWouldntRate_R +
ItPF_GivesFeedbackSparingly_R + ItPF_DontMindRatingItems +
ItPF_RatingItemsNotBeneficial_R
SSPC =~ SSPC_AfraidTheSystemDisclosesPrivateInfo +
SSPC_SystemInvadesPrivacy + SSPC_SystemRespectsPrivacy_R +
SSPC_UncomfortableProvidingDataToSystem +
SSPC_SystemRespectsConfidentiality_R

# regressions
ItPF ~ TrTe
SSPC ~ TrTe

# residual covariance
ItPF ~~ SSPC'

# compute alpha correction integer (N-1)/2
correction.integer <- round((nrow(data_twostagecomplete)-1)/2)

# SAM model: TrTe -> SSPC + ItPF
trust.sam <- sam(trust.model, data = data_twostagecomplete, cmd = 'sem', se
= 'twostep',
                 mm.list = NULL, mm.args = list(bounds = 'standard', se =
'robust.sem'),
                 struc.args = list(estimator = 'ML', se = 'standard'),
sam.method = 'local', alpha.correction = correction.integer)

# summary results
summary(trust.sam, standardized = TRUE, rsquare = TRUE)

# plot labels
labels <- list(TrTe_TechnologyNeverWorks_R = 'T1',
               TrTe_LessConfidentWhenUsingTechnology_R = 'T2',
               TrTe_UsefulnessOfTechnologyIsOverrated_R = 'T3',
               ItPF_LikesToGiveFeedback = 'I1',
               ItPF_NormallyWouldntRate_R = 'I2',
               ItPF_GivesFeedbackSparingly_R = 'I3',
               ItPF_DontMindRatingItems = 'I4',
               ItPF_RatingItemsNotBeneficial_R = 'I5',
               SSPC_AfraidTheSystemDisclosesPrivateInfo = 'S1',
               SSPC_SystemInvadesPrivacy = 'S2',
```

```
               SSPC_SystemRespectsPrivacy_R = 'S3',
               SSPC_UncomfortableProvidingDataToSystem = 'S4',
               SSPC_SystemRespectsConfidentiality_R = 'S5')

# plot
lavaanPlot(model = trust.sam, labels = labels, node_options = list(shape =
"box", fontname = "Times"),
           edge_options = list(color = "grey"), coefs = TRUE, stand = TRUE,
covs = TRUE, stars = c("regress","covs",'latent'), digits = 3)

# AVE
AVE(trust.sam, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff
compRelSEM(trust.sam, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

trust.model <-
  '# latent variables
TrTe =~ TrTe_TechnologyNeverWorks_R +
TrTe_LessConfidentWhenUsingTechnology_R +
TrTe_UsefulnessOfTechnologyIsOverrated_R
ItPF =~ ItPF_LikesToGiveFeedback + ItPF_NormallyWouldntRate_R +
ItPF_GivesFeedbackSparingly_R + ItPF_RatingItemsNotBeneficial_R
SSPC =~ SSPC_AfraidTheSystemDisclosesPrivateInfo +
SSPC_SystemInvadesPrivacy + SSPC_SystemRespectsPrivacy_R +
SSPC_UncomfortableProvidingDataToSystem +
SSPC_SystemRespectsConfidentiality_R

# regressions
ItPF ~ TrTe
SSPC ~ TrTe


# residual covariance
ItPF ~~ SSPC'

# compute alpha correction integer (N-1)/2
correction.integer <- round((nrow(data_twostagecomplete)-1)/2)

# SAM model: TrTe -> SSPC + ItPF
trust.sam <- sam(trust.model, data = data_twostagecomplete, cmd = 'sem', se
= 'twostep',
                 mm.list = NULL, mm.args = list(bounds = 'standard', se =
'robust.sem'),
                 struc.args = list(estimator = 'ML', se = 'standard'),
sam.method = 'local', alpha.correction = correction.integer)

# summary results
summary(trust.sam, standardized = TRUE, rsquare = TRUE)

# plot labels
labels <- list(TrTe_TechnologyNeverWorks_R = 'T1',
               TrTe_LessConfidentWhenUsingTechnology_R = 'T2',
               TrTe_UsefulnessOfTechnologyIsOverrated_R = 'T3',
```

```
              ItPF_LikesToGiveFeedback = 'I1',
              ItPF_NormallyWouldntRate_R = 'I2',
              ItPF_GivesFeedbackSparingly_R = 'I3',
              ItPF_RatingItemsNotBeneficial_R = 'I5',
              SSPC_AfraidTheSystemDisclosesPrivateInfo = 'S1',
              SSPC_SystemInvadesPrivacy = 'S2',
              SSPC_SystemRespectsPrivacy_R = 'S3',
              SSPC_UncomfortableProvidingDataToSystem = 'S4',
              SSPC_SystemRespectsConfidentiality_R = 'S5')

# plot
lavaanPlot(model = trust.sam, labels = labels, node_options = list(shape =
"box", fontname = "Times"),
           edge_options = list(color = "grey"), coefs = TRUE, stand = TRUE,
covs = TRUE, stars = c("regress","covs",'latent'), digits = 3)

# AVE
AVE(trust.sam, obs.var = TRUE, omit.imps = c("no.conv", "no.se"),
    omit.factors = character(0), dropSingle = TRUE, return.df = TRUE)

# reliability coeff
compRelSEM(trust.sam, obs.var = TRUE, tau.eq = FALSE, ord.scale = TRUE,
           config = character(0), shared = character(0), higher =
character(0),
           return.total = FALSE, dropSingle = TRUE, omit.factors =
character(0),
           omit.indicators = character(0), omit.imps = c("no.conv",
"no.se"),
           return.df = TRUE)

# M1: mediation-only model
M1 <-
'# lvs
PRQ_lv =~ PRQ_LikedTheItems + PRQ_FitPreference + PRQ_ItemsWellChosen +
PRQ_ItemsWereRelevant + PRQ_RecommendedTooManyBadItems_R

PSE_lv =~ PSE_WouldRecommendToOthers + PSE_MakesMeMoreAwareOfChoiceOptions
+ PSE_MakeBetterChoicesWithSystems + PSE_CanFindBetterItemsWithSystem

# regressions
PRQ_lv ~ a*Targeted
PSE_lv ~ b*PRQ_lv + c*Targeted

# indirect
T_PRQ_PSE := a*b
Total_Effect := (a*b) + c'

# compute alpha correction integer (N-1)/2
correction.integer <- round((nrow(df)-1)/2,0)

# SAM M1: T1 -> PRQ -> PSE
M1.sam <- sam(M1, data = df, cmd = 'sem', se = 'twostep',
                 mm.list = NULL, mm.args = list(bounds = 'standard', se =
'robust.sem'),
                 struc.args = list(estimator = 'ML', se = 'standard'),
sam.method = 'global', alpha.correction = correction.integer)

summary(M1.sam, standardized = TRUE, fit.measures = FALSE, rsquare = TRUE)

labels <- list(PRQ_lv = 'PRQ',
              PSE_lv = 'PSE',
              Targeted = 'T1',
```

```
               PRQ_LikedTheItems  = 'PQ1',
               PRQ_FitPreference  = 'PQ2' ,
               PRQ_ItemsWellChosen  = 'PQ3',
               PRQ_ItemsWereRelevant  = 'PQ4',
               PRQ_RecommendedTooManyBadItems_R  = 'PQ5',
               PSE_WouldRecommendToOthers = 'PE1',
               PSE_MakesMeMoreAwareOfChoiceOptions = 'PE3',
               PSE_MakeBetterChoicesWithSystems = 'PE4',
               PSE_CanFindBetterItemsWithSystem = 'PE5')


lavaanPlot(model = M1.sam, labels = labels, node_options = list(shape =
"box", fontname = "Times"),
           edge_options = list(color = "grey"), coefs = TRUE, stand = TRUE,
covs = TRUE, stars = c("regress","covs",'latent'), digits = 3, sig = .05)

# exploratory model looking at main-effect of Expertise only
exp.model <-
'# lvs
PRQ_lv =~ PRQ_LikedTheItems + PRQ_FitPreference + PRQ_ItemsWellChosen +
PRQ_ItemsWereRelevant + PRQ_RecommendedTooManyBadItems_R

PSE_lv =~ PSE_WouldRecommendToOthers + PSE_MakesMeMoreAwareOfChoiceOptions
+ PSE_MakeBetterChoicesWithSystems + PSE_CanFindBetterItemsWithSystem

Expertise_lv =~ Expertise_ExpertOnMusic + Expertise_ListensALot +
Expertise_MusicLover

# regressions
PRQ_lv ~ a*Expertise_lv
PSE_lv ~ b*PRQ_lv + c*Expertise_lv

# indirect
Exp_PRQ_PSE := a*b
Total_Effect := (a*b) + c'

# compute alpha correction integer (N-1)/2
correction.integer <- round((nrow(data_twostagecomplete)-1)/2,0)

# SAM M1: Expertise -> PRQ -> PSE
exp.sam <- sam(exp.model, data = data_twostagecomplete, cmd = 'sem', se =
'twostep',
                 mm.list = NULL, mm.args = list(bounds = 'standard', se =
'robust.sem'),
                 struc.args = list(estimator = 'ML', se = 'standard'),
sam.method = 'local', alpha.correction = correction.integer)

summary(exp.sam, standardized = TRUE, rsquare = TRUE)

labels <- list(PRQ_lv = 'PRQ',
               PSE_lv = 'PSE',
               Expertise_lv = 'Expertise',
               PRQ_LikedTheItems  = 'PQ1',
               PRQ_FitPreference  = 'PQ2' ,
               PRQ_ItemsWellChosen  = 'PQ3',
               PRQ_ItemsWereRelevant  = 'PQ4',
               PRQ_RecommendedTooManyBadItems_R  = 'PQ5',
               PSE_WouldRecommendToOthers = 'PE1',
               PSE_MakesMeMoreAwareOfChoiceOptions = 'PE3',
               PSE_MakeBetterChoicesWithSystems = 'PE4',
               PSE_CanFindBetterItemsWithSystem = 'PE5',
               Expertise_ExpertOnMusic = 'Ex1',
```

```
              Expertise_ListensALot = 'Ex2',
              Expertise_MusicLover = 'Ex3')

lavaanPlot(model = exp.sam, labels = labels, node_options = list(shape =
"box", fontname = "Times"),
          edge_options = list(color = "grey"), coefs = TRUE, stand = TRUE,
covs = TRUE, stars = c("regress","covs",'latent'), digits = 3,
          sig = .05)
```

List of Abbreviations

AFML	Adaptive Functions of Music Listening (Groarke & Hogan, 2018)
AHEC	Arts and Humanities Ethics Committee
AIC	Akaike Information Criterion
API	Application Programming Interface
AVD	Arousal, Valence, Depth
AVE	Average Variance Explained
BFI	Big-Five Inventory
BIC	Bayes Information Criterion
BPM	Beats Per Minute
CAMRS	Context-Aware Music Recommender System
CD	Compact Disc
CFA	Confirmatory Factor Analysis
CFF	Consensus Functions Framework
CFI	Comparative Fit Index
CI	Confidence Interval(s)
CSV	Comma-separated Values
DE	Direct Effect
DIAMONDS	Duty, Intellect, Adversity, Mating, pOsitivity, Negativity, Deception, Sociality (Rauthmann et al., 2014)
EFA	Exploratory Factor Analysis
ESM	Experience Sampling Method
ESR	Experience Sampling Report
FML	Function(s) of Music Listening

FSR	Factor Score Regression
GEMS	Geneva Emotions in Music Scale (Zentner et al., 2008)
HTTP	Hypertext Transfer Protocol
ID	Identifier
IE	Indirect Effect
IRT	Item Response Theory
ISE	Irrelevant Sound Effect
ItPF	Intention to Provide Feedback (Knijnenburg et al., 2012)
JSON	JavaScript Object Notation
KMO	Kaiser-Meyer-Olkin (measure of sampling adequacy)
M	Mean
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
ML	Maximum Likelihood
MLM	Robust Maximum Likelihood (Satorra-Bentler variant)
MLR	Robust Maximum Likelihood (Yuan and Bentler variant)
MMR	Music in Mood Regulation (Saarikallio, 2008)
MRS	Music Recommender System
MUSIC	Mellow, Unpretentious, Sophisticated, Intense, Contemporary (Rentfrow et al., 2012)
NA	Not Applicable
NLP	Natural Language Processing
OSA	Objective System Aspect (Knijnenburg et al., 2012)
PCA	Principal Component Analysis

PRQ	Perceived Recommendation Quality (Knijnenburg et al., 2012)
PSE	Perceived System Effectiveness (Knijnenburg et al., 2012)
reCAPTCHA	Completely Automated Public Turing Test to tell Computers and Humans Apart
RMSEA	Root Mean Square Error of Approximation
RQ	Research Question
SAM	Structural-After-Measurement
SD	Standard Deviation
SE	Standard Error
SEM	Structural Equation Model(-ing)
SEMA	Smartphone Ecological Momentary Assessment (Koval et al., 2019)
SMN	Semantic Multinomial
SSA	Subjective System Aspect (Knijnenburg et al., 2012)
SSPC	System-Specific Privacy Concern (Knijnenburg et al., 2012)
STOMP	Short Test of Music Preferences (Rentfrow et al., 2003)
TE	Total Effect
TIE	Total Indirect Effect
TLI	Tucker-Lewis Index
TrTe	Trust in Technology (Knijnenburg et al., 2012)
URL	Uniform Resource Locator
UX	User Experience

Reference List

Abdi, H., & Williams, L. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433-459. DOI: 10.1002/wics.101

Adomavicius, G., & Tuzhilin, A. (2015). Context-Aware Recommender Systems. In: F. Ricci, L. Rokach, and B. Shapira, ed., Recommender Systems Handbook, 2nd ed. *Springer*, 191-226.

Afchar, D., Melchiorre, A.B., Schedl, M., Hennequin, R., Epure, E.V., and Moussallam, M. (2022). Explainability in Music Recommender Systems. *AI Magazine*, 1(1), 190-208. DOI: 2201.10528

Agler, R., & De Boeck, P. (2017). On the Interpretation and Use of Mediation: Multiple Perspectives on Mediation Analysis. *Frontiers in Psychology*. 8(1984), DOI: 10.3389/fpsyg.2017.01984

Anderson, J.C., & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411–423. DOI: https://doi.org/10.1037/0033-2909.103.3.411

Andersson, G., & Yang-Wallentin, F. (2021). Generalized Linear Factor Score Regression: A Comparison of Four Methods. *Educational and Psychological Measurement*, 81(4), 617-643. DOI: 10.1177/0013164420975149

Arnett, J.J. (1995). Adolescents' uses of media for self-socialization. *Journal of youth and adolescence*, *24*(5), 519-533. DOI: https://doi.org/10.1007/BF01537054

Bailes, F. (2007). The prevalence and nature of imagined music in the everyday lives of music students. *Psychology of Music*, 35(4), 555-570. DOI: 10.1177/0305735607077834

Baltrunas, L., Kaminskas, M., Ludwig, B., Moling, O., Ricci, F., Aydin, A., Lüke, K., & Schwaider, R. (2011). InCarMusic: Context-Aware Music Recommendations in a Car. In Huemer C., Setzer T. (eds) E-Commerce and Web Technologies. EC-Web 2011. *Lecture Notes in Business Information Processing*, 85. *Springer*. DOI: https://doi.org/10.1007/978-3-642-23014-1_8

Balwant, P.T., Birdi, K., Stephan, U., & Topakas, A. (2019). Transformational instructor-leadership and academic performance: a moderated mediation model of student engagement and structural distance. *Journal of Further and Higher Education*, 43(7), 884-900. DOI: https://doi.org/10.1080/0309877X.2017.1420149

Bansal, J., Flannery, M.B., & Woolhouse, M.H. (2020). Influence of personality on music-genre exclusivity. *Psychology of Music*, 49(5), 1356-1371. DOI: 10.1177/0305735620953611

Barbour, R. (2014). Introducing Qualitative Research (2nd ed.). *SAGE*.

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal or Personality and Social Psychology*, 51(6), 1173-1182.

Barone, M.D., Bansal, J., & Woolhouse, M.H. (2017). Acoustic Features Influence Musical Choices Across Multiple Genres. *Frontiers in Psychology*, 8(931), 1-14. DOI: 10.3389/fpsyg.2017.00931

Bauer, D., McNeish, D., Baldwin, S., & Curran, P. (2020). Analyzing Nested Data: Multilevel Modeling and Alternative Approaches. In A. Wright & M. Hallquist (eds.), *The Cambridge Handbook of Research Methods in Clinical Psychology*. Cambridge University Press, 426-443. DOI: 10.1017/9781316995808.03

Beauducel, A., & Herzberg, Y.P. (2006). Structural Equation Modeling On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling*, 12(2), 186-203. DOI: 10.1207/s15328007sem1302_2

Beavers, A.S., Lounsbury, J.W., Richards, J.K., Huck, S.W., Skolits, G.J., and Esquivel, S.L. (2013). Practical Considerations for Using Exploratory Factor Analysis in Educational Research. *Practical Assessment, Research and Evaluation*, 18(6). DOI: https://doi.org/10.7275/qv2q-rk76

Behbehani, S., & Steffens, J. (2020). Musical DIAMONDS: The influence of situational classes and characteristic on music listening behavior. *Psychology of Music*, 49(6), 1532-1545. DOI: 10.1177/0305735620968910

Behne, K. (1997). The development of "Musikerleben" in adolescence. How and why young people listen to music. In I. Deliége & J. Sloboda, Perception and Cognition of Music (1st ed.). *Psychology Press*. DOI: https://doi.org/10.4324/9780203344262.

Benítez-Burraco, A., & Nikolsky, A. (2023). The (Co)Evolution of Language and Music Under Human Self-Domestication. *Human Nature*, (2):229-275. DOI: 10.1007/s12110-023-09447-1

Berlyne, D.E. (1971). *Aesthetics and psychobiology*. Appleton-Century-Crofts.

Blacking, J. (1973). How Musical is Man? (6th ed), 3-31. *University of Washington Press*.

Boateng, G.O., Neilands, T.B., Frongillo, E.A., Melgar-Quiñonez, H.R., & Young, S.L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health*, 6, 149. DOI: https://doi.org/10.3389/fpubh.2018.00149

Boer, D., Fischer, R., Strack, M., Bond, M.H., Lo, E., & Lam, J. (2011). How Shared Preferences in Music Create Bonds Between People: Values as the Missing Link. *Personality and Social Psychology Bulletin*, 37(9), 1159-1171. DOI: 10.1177/0146167211407521

Bogdanov, D., Haro, M., Fuhrmann, F., Xambó, A., Gómez, E., & Herrera, P. (2013). Semantic audio content-based music recommendation and visualisation. *Information Processing and Management*, 49, 13-33. DOI: https://dx.doi.org/10.1016/j.ipm.2012.06.004

Bollen, K. (1989). *Structural Equations with latent variables*. John Wiley & Sons.

Born, G., Morris, J., Diaz, F., & Anderson, A. (2021). Artificial Intelligence, Music Recommendation, and the Curation of Culture [White Paper]. *Canadian Institute for Advanced Research*.

Bourdieu, P. (1986). The Forms of Capital. In J. Richardson, Handbook of Theory and Research for the Sociology of Education, *Greenwood*, 241-258.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. DOI: 10.1191/1478088706qp063oa

Brown, S., & Krause, A.E. (2020). Freedom of Choice: Examining Music Listening as a Function of Favorite Music Format. *Psychomusicology: Music, Mind and Brain*, 30(2), 88-102. DOI: http://dx.doi.org/10.1037/pmu0000254

Brown, T.A. (2006). Confirmatory Factor Analysis for Applied Research. *The Guildford Press.*

Bull, M. (2006). Investigating the Culture of Mobile Listening: From Walkman to iPod. In K. O'Hara & B. Brown (eds.), *Consuming Music Together: Social and Collaborative*

Aspects of Music Consumption Technologies. Springer, 131-149. DOI: https://doi-org.libproxy.york.ac.uk/10.1007/1-4020-4097-0_7

Bullock, J.G., Green, D.P., & Ha, S.E. (2010). Yes, But What's the Mechanism? (Don't Expect an Easy Answer). *Journal of Personality and Social Psychology,* 98 (4), 550-558. DOI: 10.1037/a0018933

Caprara, G.V., Vecchione, M., Barbaranelli, C., & Fraley, R.C. (2007). When Likeness Goes with Liking: The Case of Political Preference. *Political Psychology*, 28(5), 609-632.

Celma, Ò., Cano, P., & Herrera, P. (2006). Search Sounds: An audio crawler focused on weblogs. In *Proceedings of the 7th International Conference on Music Information Retrieval*. Victoria, Canada. Retrieved 21 June 2021.

Chamorro-Premuzic, T., & Furnham, A. (2007). Personality and music: Can traits explain how people use music in everyday life? *The British Psychological Society*, 98, 175-185. DOI: 10.1348/000712606X111177

Cheng, J., & Chang, W. (2023). httpuv: HTTP and WebSocket Server Library. R package version 1.6.9. Available at: https://cran.r-project.org/web/packages/httpuv/index.html.

Cheong, J., & MacKinnon, D.P. (2012). Mediation/Indirect Effects in Structural Equation Modeling. In R.H. Hoyle, *Handbook of Structural Equation Modeling*, 417-435. *The Guildford Press*.

Chou, C-P., & Huh, J. (2012). Model Modification in Structural Equation Modeling. In R.H. Hoyle, *Handbook of Structural Equation Modeling*, 232-246. *The Guildford Press*.

Clark, I.N., Baker, F.A., & Taylor, N.F. (2016). The modulating effects of music listening on health-related exercise and physical activity in adults: a systematic review and narrative synthesis. *Nordic Journal of Music Therapy*, 25(1), 76-104. DOI: 10.1080/08098131.2015.1008558

Clayton, M. (2009). The social and personal functions of music in cross-cultural perspective. In S. Hallam, I. Cross & M. Thaut (Ed.), *The Oxford Handbook of Music Psychology*, 35-44. *Oxford University Press*.

Collins, T., Tillmann, B., Barret, F.S., & Delbé, C. (2014). A Combined Model of Sensory and Cognitive Representation Underlying Tonal Expectations in Music: From Audio Signals to Behavior. *Psychological Review*, 121(1), 33-65. DOI: 10.1037/a0034695

Comrey, A.L. (1973). A First Course in Factor Analysis. *Academic Press*.

Cooper, M. (2018). The psychology of goals: A practice-friendly review. In M. Cooper & D. Law (Eds.), *Working with goals in counselling and psychotherapy*. Oxford University Press.

Cross, I. (2014). Music and communication in music psychology. *Psychology of Music*, 42(6), 809-819. DOI: 10.1177/0305735614543968

Cross, I. (2016). The Nature of Music and its Evolution. In S. Hallam, I. Cross & M. Thaut, *The Oxford Handbook of Music Psychology* (2nd ed.). Oxford University Press.

Cunningham, S.J., & Nichols, D.M. (2009). Exploring Social Music Behavior: An Investigation of Music Selection at Parties. In *ISMIR 2009 - Proceedings of the 10th International Conference on Music Information Retrieval*, 747-752.

Curran, P. J., & Hussong, A.M. (2009). Integrative Data Analysis: The Simultaneous Analysis of Multiple Data Sets. *Psychological Methods*, 14(2), 81-100. DOI: 10.1037/a0015914

Dancey, C., & Reidy, J. (2017). Statistics without Maths for Psychology (7th ed.). *Pearson Education Limited*.

Danckwerts, S., & Kenning, P. (2019). “It's MY Service, it's MY Music”: The role of psychological ownership in music streaming consumption. *Psychology & Marketing*, 36(9), 803-816. DOI: 10.1002/mar.21213

Darmody, A., & Zwick, D. (2020). Manipulate to empower: Hyper-relevance and the contradictions of marketing in the age of surveillance capitalism. *Big Data and Society*, 7(1), 1-12. DOI: 10.1177/2053951720904112

Davies, S. (2013). Music-to-listener emotional contagion. In T. Cochrane, B. Fantini & K. Scherer, *The Emotional Power of Music,* 1st ed., 169-176. *Oxford University Press*.

De Gemmis, M., Lops, P., Musto, C., Narducci, F., & Semerarim G. (2015). Semantics-Aware Content-Based Recommender Systems. In: F. Ricci, L. Rokach, and B. Shapira, ed., Recommender Systems Handbook, 2nd ed. *Springer*, 119-159. DOI: 10.1007/978-1-4899-7637-6

DeVellis, R.F. (2017). *Scale Development: Theory and Applications* (4th ed.). SAGE.

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An Alternative for SEM? *Methodology*, 13, 31-38. DOI: 10.1027/1614-2241/a000130

de Rooij, M., & Weeda, W. (2020). Cross-validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, 3(2), 248-263. doi:10.1177/2515245919898466

de Rooij, M., Karch, J.D., Fokkema, M., Bakk, Z., Pratiwi, B.C., & Kelderman, H. (2022). SEM-based out-of-sample predictions. *Structural Equation Modeling: A Multidisciplinary Journal*, 30(1), 132-148, DOI: 10.1080/10705511.2022.2061494.

De Winter, J.C.F., Dodou, D., & Wieringa, P.A. (2009). Exploratory Factor Analysis With Small Sample Sizes. *Multivariate Behavioral Research*, 44, 147-181. DOI: 10.1080/00273170902794206

Dey, A.K. (2000). Providing Architectural Support for Building Context-Aware Applications (PhD). Georgia Institute of Technology.

Dibben, N., & Williamson, V. (2007). An exploratory survey of in-vehicle music listening. *Psychology of Music*, 35(4), 571–590. DOI: 10.1177⁄0305735607079725

Di Noia, T., Tintarev, N., Fatourou, P., & Schedl, M. (2022). Recommender Systems under European AI Regulations. *Communications of the ACM*, 65(4), 69-73. DOI: 10.1145/3512728

DiStefano, C., Zhu, M., Mîndrilã, D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research, and Evaluation*. 14(20). DOI: https://doi.org/10.7275/da8t-4g52

Dittmar, H. (2008). *To have is to be? Psychological functions of material possessions*. In H. Dittmar, Consumer Culture, Identity and Well-Being. *Psychology Press*.

Dolan, C.V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal Mathematical and Statistical Psychology*, 47(2), 309-326.

Downie, S. (2004). The Scientific Evaluation of Music Information Retrieval Systems: Foundations and Future. *Computer Music Journal*, 28(2). DOI: 10.1162/014892604323112211

Du Toit, S.H.C., & Du Toit, M. (2008). Multilevel Structural Equation Modeling. In J. de Leeuw & E. Meijer, *Hanbook of Multilevel Analysis* (1st eds.). *Springer*.

Eaton, P., Frank, B., Johnson, K., & Willoughby, S. (2019). Comparing exploratory factor models of the Brief Electricity and Magnetism Assessment and the Conceptual Survey of Electricity and Magnetism. *Physical Review Physics Education Research*, *15*(2), 020133. DOI: 10.1103/PhysRevPhysEducRes.15.020133

Edwards, M.C., Wirth, R.J., Houts, C.R., & Xi, N. (2012). Categorical Data in the Structural Equation Modeling Framework. In R.H. Hoyle (ed.), *Handbook of Structural Equation Modeling*, 195-208. *The Guildford Press*.

Egermann, H., Kopiez, R., & Altenmüller, E. (2013). The Influence of Social Normative and Informational Feedback on Musically Induced Emotions in an Online Music Listening Setting. *American Psychological Association*, 23(1), 21-32. DOI: 10.1037/a0032316

Eisele, G., Kasanova, Z., & Houben, M. (2022). Questionnaire design and evaluation. In I. Myin-Germeys & P. Kuppens, *The Open Handbook of Experience Sampling Methodology* (2nd ed.).

Evans, J.R., & Mathur, A. (2018). The value of online surveys: A look back and a look ahead. *Internet research*, 28(4), 854-887. DOI: 10.1108/IntR-03-2018-0089

Fairchild, A.J., & McDaniel, H.L. (2017). Best (but-oft-forgotten) practices: Mediation analysis. *American Journal of Clinical Nutrition*, 105(6), 1259-1271. DOI: 10.3945/ajcn.117.152546

Felstead, A., & Reuschke, D. (2023). A flash in the pan or a permanent change? The growth of homeworking during the pandemic and its effect on employee productivity in the UK. *Information Technology & People*, *36*(5), 1960-1981. DOI: 10.1108/ITP-11-2020-0758

Ferrer, R., & Eerola, T. (2011). Semantic structures of timbre emerging from social and acoustic descriptions of music. *EURASIP Journal on Audio, Speech, and Music Processing*, 11.

Ferrer, R., Eerola, T., & Vuoskoski, J.K. (2013). Enhancing genre-based measures of music preference by user-defined liking and social tags. *Psychology of Music*, 41(4), 499-518. DOI: 10.1177/0305735612440611

Ferwerda, B., Graus M.P., Vall, A., Tkalcic, M., & Schedl. M. (2017). How Item Discovery Enabled by Diversity Leads to Increased Recommendation List Attractiveness. In *Proceedings of the Symposium on Applied Computing*. 1693–1696. DOI: 10.1145/3019612.3019899.

Fetters, M.D., Curry, L.A., & Creswell, J.W. (2013). Achieving integration in mixed methods designs- principles and practices. *Health services research*, 48(2), 2134-2156. DOI: 10.1111/1475-6773.12117

Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics (5th ed.). *SAGE*.

Finnäs, L. (1989). A Comparison between Young People’s Privately and Publicly Expressed Musical Preferences. *Psychology of Music*, 17, 132-145.

Fokkema, M., Iliescu, D., Grieff, S., & Ziegler, M. (2022). Machin Learning and Prediction in Psychological Assessment. *European Journal of Psychological Assessment*, 38(3), 165-175. DOI: https://doi.org/10.1027/1015-5759/a000714

Fornell, C., & Larcker, D.F. (1981). Structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.

Fox, J. (2016). Applied Regression Analysis and Generalized Linear Models (3rd ed.). *SAGE*.

Franěk, M., van Noorden, L., & Režný, L. (2014). Tempo and walking speed with music in the urban context. Frontiers in psychology, 5, 1361. DOI: https://doi.org/10.3389/fpsyg.2014.01361

Fricke, K.R., Greenberg, D.M., Rentfrow, P.J., & Herzberg, P.Y. (2021). Measuring musical preferences from listening behavior: Data from one million people and 200,000 songs. *Psychology of Music*, 49(3), 371-381. DOI: 10.1177/0305735619868280

Friedlander, J.P. (2019). Mid-Year 2019 RIAA Music Revenues Report. *Recording Industry Association of America*.

Friedlander, J.P. (2020). Mid-Year 2020 RIAA Music Revenues Statistics. *Recording Industry Association of America*.

Friedlander, J.P. (2021). Mid-Year 2021 RIAA Music Revenues Statistics. *Recording Industry Association of America*.

Galbraith, S., Daniel, J.A., & Vissel, B. (2010). A study of clustered data and approaches to its analysis. *Journal of Neuroscience*, *30*(32), 10601-10608. DOI: 10.1523/JNEUROSCI.0362-10.2010

Garner, B.R. (2014). iPod use and the perception of social introversion. *Leisure Studies*, 33(1), 22–31. DOI: 10.1080/02614367.2012.699976

Garrido, S., & Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: a literature review and experiment. *Music Perception*, 28(3), 279–296. DOI: 10.1525/mp.2011.28.3.279

Geng, R., Sun, R., Li, J., Guo, F., Wang, W., & Sun, G. (2021). The impact of firm innovativeness on consumer trust in the sharing economy: a moderated mediation model. *Asia Pacific Journal of Marketing*, 34(5). DOI: 10.1108/APJML-10-2020-0748

Gerber, M., Brand, S., Elliot, C., Holsboer-Trachsler, E., & Pühse, U. (2014). Aerobic Exercise, Ball Sports, Dancing, and Weight Lifting as Moderators of the Relationship between Stress and Depressive Symptoms: An Exploratory Cross-Sectional Study with Swiss University Students. *Perceptual and Motor Skills*, *119*(3), 679–697. DOI: https://doi.org/10.2466/06.PMS.119c26z4

Gershman, A., Meisels, A., Lüke, K.H., Rokach, L., Schclar, A., & Sturm, A. (2010). A decision tree based recommender system. *10th International Conference on Innovative Internet Community Systems (I2CS)–Jubilee Edition 2010–*.

Gibbs, H., & Egermann, H. (2021). Music-Evoked Nostalgia and Wellbeing During the United Kingdom COVID-19 Pandemic: Content, Subjective Effects, and Function. *Frontiers in Psychology*. DOI: 10.3389/fpsyg.2021.647891

Giles, D., Pietrzykowski, S., & Clark, K. (2007). The psychological meaning of personal record collections and the impact of changing technological forms. *Journal Of Economic Psychology*, 28(4), 429-443. DOI: 10.1016/j.joep.2006.08.002

Gillhofer, M., & Schedl, M. (2015). Iron Maiden While Jogging, Debussy for Dinner? In X. He, S. Luo, D. Tao, C, Xu, J. Yang and M.A. Hasan, *Multimedia Modeling*. *Springer*.

Götting, M.C. (2022) *Share of music streaming subscribers worldwide in the 1st quarter of 2021, by company*, Statista. Available at: https://www.statista.com/statistics/653926/music-streaming-service-subscriber-share/ (Accessed: October 12, 2022).

Gove, W.R., & Geerken, M.R. (1977). Response Bias in Surveys of Mental Health: An Empirical Investigation. *American Journal of Sociology*, 82(6), 1289-1317. DOI: 144.32.240.50

Greasley, A. (2008). Engagement with music in everyday life: An in-depth study of adults' musical preferences and listening behaviour (PhD). University of Keele.

Greasley, A., & Lamont, A. (2011). Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae*, 15(1), 45–7. DOI: 10.1177/1029864910393417

Greb, F., Scholtz, W., & Steffens, J. (2018a). Personal and situational influences on the functions of music listening. *Psychology of Music*, 46(6), 763-794. DOI: https://doi.org/10.1177/0305735617724883

Greb, F., Steffens, J., & Scholtz, W. (2018b). Understanding music-selection behavior via statistical learning: Using the percentile-Lasso to identify the most important factors. *Music and Science*, 1, 1-17. DOI: 10.1177/2059204318755950

Greb, F., Steffens, J., & Scholtz, W. (2019). Modeling Music-Selection Behavior in Everyday Life: A Multilevel Statistical Learning Approach and Mediation Analysis of Experience Sampling Data. *Frontiers in Psychology*, 10, 1-20. DOI: 10.3389/fpsyg.2019.00390

Greenberg, D.M., Kosinski, M., Stillwell, D.J., Monteiro, B.L., Levitin, D.J., & Rentfrow, P.J. (2016). The Song Is You: Preferences for Musical Attribute Dimensions Reflect Personality. *Social and Psychological Personality Science*, 7(6), 567-605. DOI: 10.1177/1948550616641473

Griffin, M., Martino, R.J., LoSchiavo, C., Comer-Carruthers, C., Krause, K.D., Stults, C.B., & Halkitis, P.N. (2022). Ensuring survey research data integrity in the era of internet bots. *Quality & Quantity*, 56, 2841-2852. DOI: https://doi.org/10.1007/s11135-021-01252-1

Groarke, J.M., & Hogan, M.J. (2016). Enhancing wellbeing: An emerging model of the adaptive functions of music listening. *Psychology of Music*, 44(4), 769-791. DOI: 10.1177/0305735615591844

Groarke, J.M., & Hogan, M.J. (2018). Development and Psychometric Evaluation of the Adaptive Functions of Music Listening Scale. *Frontiers in Psychology*, 9(516). DOI: 10.3389/fpsyg.2018.00516

Grosche, P., Müller, M., & Serrà, J. (2012). Audio Content-Based Music Retrieval. In M. Müller, M. Goto & M. Schedl, *Multimodal Music Processing*, 157-174. *Schloss Dagstuhl*. Retrieved 21 June 2021. DOI: http://www.dagstuhl.de/dagpub/978-3-939897-37-8.

Gunzler, D., Chen, T., Wu, P., & Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai Archives of Psychiatry*, 25(6), 390-394. DOI: 10.3969/j.issn.1002-0829.2013.06.009

Güvendir, M.A., & Özkan, Y. Ö. (2022). Item Removal Strategies Conducted in Exploratory Factor Analysis: A Comparative Study. *International Journal of Assessment Tools in Education*, *9*(1), 165-180. DOI: https://doi.org/10.21449/ijate.827950

Haake, A.B. (2011). Individual music listening in workplace settings: an exploratory survey of offices in the UK. *Musicae Scientiae*, 15(1), 107–129. DOI: 10.1177/1029864911398065

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2014). *Multivariate Data Analysis* (7th ed.). Pearson Education.

Hallett, R., & Lamont, A. (2015). How do gym members engage with music during exercise? *Qualitative Research in Sport, Exercise and Health*, 7(3), 411–427. DOI: 10.1080/2159676X.2014.949835

Hallett, R., & Lamont, A. (2021). An interpretive phenomenological analysis of exercisers' use of music during workouts. *Psychology of Music*, 49(2). DOI: 10.1177/0305735619854527

Hansen, C., Hansen, C., Maystre, L., Mehrotra, R., Brost, B., Tomasi, F., & Lalmas, M. (2020). Contextual and Sequential User Embeddings for Large-Scale Music Recommendation. In *Fourteenth ACM Conference on Recommender Systems* (RecSys '20), September 22–26, 2020, 53-62. DOI: https://doi. org/10.1145/3383313.3412248

Hansen, N.C., & Keller, P.E. (2021). Oxytocin as an allostatic agent in the social bonding effects of music. *Behavioral and Brain Sciences*, 44(75). DOI: 10.1017/S0140525X200001235

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Elements of Statistical Learning* (2nd ed.) Springer. DOI: https://doi.org/10.1007/978-0-387-84858-7

Hayes, A.F., & Preacher, K.J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67(3), 451-470. DOI: 10.1111/bmsp.12028

Hektner, J., Schmidt, J., & Csikszentmihalyi, M. (2007). *Experience Sampling Method*. SAGE.

Henry, N., Kayser D., & Egermann, H. (2021). Music in Mood Regulation and Coping Orientations in Response to COVID-19 Lockdown Measures Within the United Kingdom. *Frontiers in Psychology*. DOI: 10.3389/fpsyg.2021.647879

Herzog, M., Lepa, S., Egermann, H., Schoenrock, A., & Steffens, J. (2020). Towards a common terminology for music branding campaigns, *Journal of Marketing Management*, 36(1-2), 176-209, DOI: 10.1080/0267257X.2020.1713856

Heye, A., & Lamont, A. (2010). Mobile listening situations in everyday life: The use of MP3 players while travelling. *Musicae Scientiae*, *14*(1), 95-120. DOI: https://doi.org/10.1177/102986491001400104

Hill, W.T., & Palmer, J.A. (2014). Affective responses to music without recognition: Beyond the cognitivist hypothesis. *The New School Psychology Bulletin*, *11*(1), 42-49.

Hinkin, T.R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods*, 1(1), 104-121.

Hodges, D.A. (2016). Bodily Responses to Music. In S. Hallam, I. Cross & M. Thaut, *The Oxford Handbook of Music Psychology* (2nd ed.). Oxford University Press.

Hofstede, G. (1980). Culture's consequences: International differences in work-related values. Beverly Hills, California: *Sage*.

Hu, R., & Pu, P. (2010). A study on user perception of personality-based recommender systems. In User Modeling, Adaptation, and Personalization: 18th International Conference, UMAP 2010, 291-302. *Springer*.

Huang, F.L. (2016). Alternatives to Multilevel Modeling for the Analysis of Clustered Data *The Journal of Experimental Education*, 84(1), 175-196. DOI: 10.1080/00220973.2014.952397

Huang, F.L. (2018). Multilevel Modeling Myths. *American Psychological Association*, 33(3), 492-499, DOI: http://dx.doi.org/10.1037/spq0000272492

Huron, D. (2011). Why is sad music pleasurable? A possible role for prolactin. *Musicae Scientiae*, 15(2), 146–158.

Hussong, A., Curran, P.J., & Bauer, D.J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology*, 9, 61-89. DOI: 10.1146/annurev-clinpsy-050212-185522

Hutchby, I. (2001). Technologies, Texts and Affordances. *Sociology*, 35(2), 441-456. DOI: 10.1017/S0038038501000219
Iacobucci, D. (2008). Mediation Analysis. *SAGE*.
Jackson, J.E. (1991). *A User's Guide to Principal Components*. Wiley-Interscience.
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning* (8th ed.). Springer.
Jelenchick, L.A., Eickhoff, J.C., & Moreno, M.A. (2013). "Facebook Depression?" Social Networking Site Use and Depression in Older Adolescents. Journal of Adolescent Health, 52(1), 128-130, DOI: https://doi.org/10.1016/j.jadohealth.2012.05.008
Jentschke, S. (2016). The Relationship Between Music and Language. In S. Hallam, I. Cross & M. Thaut, *The Oxford Handbook of Music Psychology* (2nd ed.). Oxford University Press.
Jin, Y., Tintarev, N., & Verbert, K. (2018). Effects of personal characteristics on music recommender systems with different levels of controllability. In *Proceedings of the* 12th ACM Conference on Recommender Systems, 13-21. DOI: https://doi.org/10.1145/3240323.3240358
Jolliffe, I.T. (2002). *Principal Component Analysis* (2nd ed.). Springer.
Jorgensen, T.D., Pornprasertmanit, S., Schoemann, A.M., & Rosseel, Y. (2022). semTools: Useful tools for structural equation modeling. Retrieved from https://CRAN.R-project.org/package=semTools
Juslin, P.N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. Juslin & J. A. Sloboda (eds), Music and emotion: Theory and Research, 309-337. *Oxford University Press*.
Juslin, P.N., & Laukka, P. (2004). Expression, perception and induction of musical emotions: a review and a questionnaire study of everyday listening. *Journal of New Music Research*, 33, 217–238.
Juslin, P.N., & Sloboda, J. (2011). *The Oxford Handbook of Music and Emotion*. Oxford University Press.
Juslin, P.N., & Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behavioural and Brain Sciences*, 31, 600-612.
Juslin, P.N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion*, 8 (5), 668-683, DOI: 10.1037/a0013505
Juslin, P.N., Liljeström, S., Västfjäll, D., & Lundqvist, L. (2010). How does music evoke emotions? Exploring the underlying mechanisms. *Handbook of Music and Emotion: Theory, Research, Applications*. DOI: 10.1093/acprof:oso/9780199230143.001.0001
Juslin, P.N. (2016). How does music induce emotions? In S. Hallam, I. Cross & M. Thaut, *The Oxford Handbook of Music Psychology* (2nd ed.). Oxford University Press.
Kaiser, H.F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, 34, 111-117. DOI: https://doi.org/10.1177/001316447403400115
Kamis, A., & Davern, M.J. (2004). Personalizing to product category knowledge: exploring the mediating effect of shopping tools on decision confidence. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.* IEEE.
Karageorghis, C.I., & Terry, P.C. (2009). The psychological, psychophysical and ergogenic effects of music in sport: a review and synthesis. In A. Bateman and J. Bale (Eds.), *Sporting sounds: Relationships between sport and music* (pp. 13-36). Routledge.

Karreman, A., Laceulle, O., Hanser, W., & Vingerhoets, A. (2017). Effects of emotion regulation strategies on music-elicited emotions: an experimental study explaining individual differences. *Personality and Individual Differences*, 114, 36–41. DOI: 10.1016/j.paid.2017.03.059

Kaminskas, M., & Ricci, F. (2012). Contextual music information retrieval and recommendation: State of the art challenges. *Computer Science Review*, 6(2-3), 89-119. DOI: 10.1016/j.cosrev.2012.04.002

Kay, M., Nelson, G.L., & Heckler, E.B. (2016). Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4521-4532. DOI: https://doi.org/10.1145/2858036.2858465

Kelloway, E.K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behaviour*, 16(3), 215-224.

Kline, P. (1994). An Easy Guide to Factor Analysis. *Routledge*.

Kline, R.B. (2016). Principles and Practice of Structural Equation modeling (4th ed.) *The Guildford Press*.

Knees, P., Schedl, M., & Goto, M. (2019). Intelligent user interfaces for music discovery: The past 20 years and what's to come. In 20th International Society for Music Information Retrieval Conference. Delft.

Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User modeling and user-adapted interaction*, 22(4), 441-504. DOI: 10.1007/s11257-011-9118-4

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A.D. (2004). Music, language and meaning: brain signature of semantic processing. *Nature Neuroscience*, 7(3), 302-307. DOI: 10.1038/nn1197

Koelsch, S., Fritz, T., von Cramon, D.Y., Müller, K., & Friederici, A.D. (2006). Investigating emotion with music: an fMRI study. *Human Brain Mapping*, 27, 239-250.

Konečni, V. (1982). Social Interaction and musical preference. In D. Deutsch (ed.). *The Psychology of Music*, 497-516.

Kostek, B. (2005). *Perception-Based Data Processing in Acoustics* (1st ed.). Springer.

Kotsopoulou, A., & Hallam, S. (2010). The perceived impact of playing music while studying: age and cultural differences. *Educational Studies*, 36(4), 431-440. DOI: 10.1080/03055690903424774

Koval, P., Hinton, J., Dozo, N., Gleeson, J., Alvarez, M., Harrison, A., Vu, D., Susanto, R., Jayaputera, G., & Sinnott, R. (2019). SEMA3: Smartphone Ecological Momentary Assessment, Version 3. Available at: http://www.sema3.com

Krause, A.E., North, A.C., & Hewitt, L.Y. (2014). Music Selection Behaviors in Everyday Listening. *Journal of Broadcasting & Electronic Media*, 58(2), 306-323. DOI: 10.1080/08838151.2014.906437

Krause, A.E., North, A.C., & Hewitt, L.Y. (2015). Music-listening in everyday life: Devices and choice. *Psychology of Music*, 43(2), 155-170. DOI: 10.1177/0305735613496860

Krause, A.E., North, A.C., & Hewitt, L.Y. (2016). The Role of Location in Everyday Experiences of Music. *Psychology of Popular Media Culture*, 5(2), 232-257. DOI: 10.1037/ppm0000059

Krause, A.E., & North, A.C. (2017). Pleasure, arousal, dominance, and judgements about music in everyday life. *Psychology of Music*, 45(3), 355-374. DOI: 10.1177/03057356166664214

Krause, A.E., Mackin, S., Mossman, A., Murray, T., Oliver, N., & Tee, V. (2020). Conceptualizing Control in Everyday Music Listening: Defining Dominance. *Music & Science*, 3, 1-13. DOI: 10.1177/2059204320931643

Krause, A.E., Dimmock, J., Rebar, A.L., & Jackson, B. (2021). Music Listening Predicted Improved Life Satisfaction in University Students During Early Stages of the COVID-19 Pandemic. *Frontiers in Psychology*. DOI: 0.3389/fpsyg.2020.631033

Krause, A.E., & Brown, S.C. (2021). A uses and gratifications approach to considering the music formats that people use most often. *Psychology of Music*, 49(3). DOI: 10.1177/0305735619880608

Kuch, M., & Wöllner, C. (2021). On the Move: Principal Components of the Functions and Experiences of Mobile Music Listening. *Music & Science*, 4, 1-14. DOI: 10.1177/20592043211032852

Lamere, P. (2008). Social Tagging and Music Information Retrieval. *Journal of New Music Research*, 37(2), 101-114. DOI: 10.1080/09298210802479284

Lamont, A., Greasley, A., & Sloboda, J. (2016). Choosing to Hear Music: Motivation, Process and Effect. In S. Hallam, I. Cross & M. Thaut, *The Oxford Handbook of Music Psychology* (2nd eds.). Oxford University Press.

Latham, G.P., & Locke, A. (1991) Self-regulation through Goal Setting. *The Academy of Management Review*, 50, 212-247. DOI: 10.2307/258875

Lathia, N., Rachuri, K.K., Mascolo, C., & Rentfrow, P.J. (2013). Contextual dissonance: Design bias in sensor-based experience sampling methods. *UbiComp 2013 – Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 183-192. DOI: 10.1145/2493432.2493452

Laukka, P., & Quick, L. (2013). Emotional and motivational uses of music in sports and exercise: a questionnaire study among athletes. *Psychology of Music*, 41(2), 198–215. DOI: https://doi.org/10.1177/0305735611422507

Law, S.F., & Le, A.T. (2023). A systematic review of empirical studies on trust between universities and society. *Journal of Higher Education Policy and Management*, 45(4) 393-408. DOI: https://doi.org/10.1080/1360080X.2023.2176598

Ledesma, R.D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical assessment, research, and evaluation*, 12(2), 1–11. DOI: https://doi.org/10.7275/wjnc-nm63

Lepa, S., Herzog, M., Steffens, J., Schoenrock, A., & Egermann, H. (2020). A computational model for predicting perceived musical expression in branding scenarios. *Journal of New Music Research*, 49(4), 387-402, DOI: 10.1080/09298215.2020.1778041

Levitin, D. (2006). This is Your Brain on Music. *Atlantic Books*.

Lew, M.S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-Based Multimedia Information Retrieval: State of the Art and Challenges. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2(1), 1-19. DOI: https://doi.org/10.1145/1126004.1126005

Lex, E., Kowald, D., Seitlinger, P., Tran, T.N.T., Felfering, A., & Schedl, M. (2021). Psychology-informed Recommender Systems. *Foundations and Trends in Information Retrieval*, 15(2), 134-242. DOI: http://dx.doi.org/10.1561/1500000090

Lex, E., & Schedl, M. (2022). Psychology-informed Recommender Systems: A Human-Centric Perspective on Recommender Systems. In *Proceedings of the 2022 ACM*

SIGIR Conference on Human Information Interaction and Retrieval (CHIR '22), 367-368. DOI: 10.1145/3498366.3505841
Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological methods*, 24(4), 452-467. DOI: 10.1037/met0000230
Lin, L-C., Huang, P-H., & Weng, L-J. (2017). Selecting Path Models in SEM: A Comparison of Model Selection Criteria, *Structural Equation Modeling: A Multidisciplinary Journal*, 24(6), 855-869. DOI: 10.1080/10705511.2017.1363652
Liu, Q., Xiang, B., Chen, E., Ge, Y., Xiong, H., Bao, T., & Zheng, Y. (2012). Influential seed items recommendation. In *Proceedings of the sixth ACM conference on Recommender systems*, 245-248. DOI: https://doi.org/10.1145/2365952.2366005
Lobera, J., Fernández Rodríguez, C.J., & Torres-Albero, C. (2020). Privacy, Values and Machines: Predicting Opposition to Artificial Intelligence. *Communication Studies*, 71(3), 448-465. DOI: 10.1080/10510974.2020.1736114
Loehlin, J.C., & Beaujean, A.A. (2017). Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis (5th eds.). *Routledge*.
Lonsdale, A.J., & North, A.C. (2009). Musical Taste and Ingroup Favouritism. *Group Processes & Intergroup Relations*, 12(3), 319-327. DOI: 10.1177/1368430209102842
Lonsdale, A. J., & North, A.C. (2011). 'Why do we listen to music? A uses and gratifications analysis.', *British Journal of Psychology*, 102(1), 108–134. DOI: 10.1348/000712610X506831
Lotto, A., & Holt., L. (2011). Psychology of auditory perception. *WIREs Cognitive Science*, 2 (5), 479-489. DOI: https://doi.org/10.1002/wcs.123
Loviglio, J. (2022). The Traffic In Feelings: The car-radio assemblage. In M. Indgren & J. Loviglio (eds.). *The Routledge Companion to Radio and Podcast Studies*. Routledge.
Ludewig, M., Kamehkhosh, I., Landia, N., & Jannach, D. (2018). Effective nearest-neighbor music recommendations. In *Proceedings of the ACM Recommender Systems Challenge 2018*, 1-6. DOI: https://doi.org/10.1145/3267471.3267474
Lundqvist, L., Carlsson, F., Hilmersson, P., & Juslin, P. (2009). Emotional responses to music: Experience, expression, and physiology. *Psychology of Music*, 37(1), 61-90. DOI: https://doi.org/10.1177/0305735607086048
Maloney, L. (2017). Music as Water: The Functions of Music from a Utilitarian Perspective. *AVANT*, 8 (Special Issue), 57-67. DOI: 10.26913/80s02017.0111.0006
Maloney, L. (2019). *Music Like Water: Exploring the Functions of Music Through Thematic Bibliometric Analysis and Comparative ESM Study* (PhD). University of York.
Maloney, L., O'Neill, K., & Gray, J. (2021). Alone Together: A Mixed-Methods Analysis of Spotify Listening During the COVID-19 Pandemic. *Journal of Music, Health, and Wellbeing*, 1-16.
Maskey, R., Fei, J., & Nguyen, H. (2018). Use of Exploratory Factor Analysis in Maritime Research. *The Asian Journal of Shipping and Logistics*, 34(2), 91-111. DOI: https://doi.org/10.1016/j.ajsl.2018.06.006
Matsunaga, M., (2010). How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International Journal of Psychological Research*, 3(1), 97-110. DOI: 10.21500/20112084.854
McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(5), 750-773. DOI: 10.1080/10705511.2016.1186549

McNeish, D., Stapleton, L.M., Silverman, R.D. (2017). On the Unnecessary Ubiquity of Hierarchical Linear Modeling. *Psychological Methods*, 22(1), 114-140, DOI: http://dx.doi.org/10.1037/met0000078

McNeish, D.M., & Harring, J.R. (2017). Clustered data with small sample sizes: Comparing the performance of model-based and design-based approaches. *Communications in Statistics – Simulation and Computation*, 46(2), 855-869. DOI: 10.1080/03610918.2014.983648

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. The MIT Press.

Meng, M., Steinhardt, S., Schubert, A. (2018). Application Programming Interface Documentation: What Do Software Developers Want? *Journal of Technical Writing and Communication*, 48(3), 295-330. DOI: 10.1177/0047281617721853

Merkle, E.C., You, D., & Preacher, K.J. (2016). Testing Nonnested Structural Equation Models. *Psychological Methods*, 21(2), 151-163. DOI: http://dx.doi.org/10.1037/met0000038

Merriam, A. (1964). The Anthropology of Music. *Northwestern University Press*.

Miles, J.N.V., Kulesza, M., Ewing, B., Shih, R.A., Tucker., J.S., & D'Amico, E.J. (2015). Moderated mediation analysis: An illustration using the association of gender with delinquency and mental health. *Journal of Criminal Psychology*, 5(2), 99-123. DOI: 10.1108/JCP-02-2015-0010

Millecamp, M., Htun, N.N., Jin. Y., & Verbert, K. (2018). Controlling Spotify Recommendations: Effects of Personal Characteristics on Music Recommender User Interfaces. In *Proceedings of the 26th Conference on user modeling, adaptation and personalization*, 101-109. DOI: https://doi.org/10.1145/3209219.3209223

Miotto, R., & Lanckriet, G. (2012). A Generative Context Model for Semantic Annotation and Retrieval. *IEEE Transaction on Audio, Speech, and Language Processing*, 20(4), 1096-1108. DOI: 10.1109/TASL.2011.2172423

Moen, E.L., Fricano-Kugler, C.J., Luikart, B.W., & O'Malley, A.J. (2016). Analyzing Clustered Data: Why and How to Account for Multiple Observations within a Study Participant? *PLoS ONE*, 11(1). DOI: https://doi.org/10.1371/journal.pone.0146721

Moreno, M.A., Jelenchick, L., Koff, R., & Eickhoff, J. (2012). Depression and Internet Use among Older Adolescents: An Experience Sampling Approach. *Psychology*, 3(9), 743-748. DOI: https://dx.doi.org/10.4236/psych.2012.329112

Morse, J.M. (2003). Principles of mixed methods and multimethod research design. *Handbook of mixed methods in social and behavioral research*, *1*, 189-208.

Mwakalonge, J., Siuhi, S., & White, J. (2015). Distracted walking: Examining the extent to pedestrian safety problems. *Journal of traffic and transportation engineering (English edition)*, 2(5), 327-337. DOI: https://doi.org/10.1016/j.jtte.2015.08.004

Nattiez, J. (1990). Music and Discourse: Toward a Semiology of Music, translated by Carolyn Abbate. Princeton: *Princeton University Press*.

Nevitt, J., & Hancock, G.R. (2004). Evaluating Small Sample Approaches for Model Test Statistics in Structural Equation Modeling. *Multivariate Behavioral Research*, 39(3), 439-478. DOI: 10.1207/S15327906MBR3903_3

Nguyen, T. T., Harper, F.M., Terveen, L., & Konstan. J.A. (2018). User personality and user satisfaction with recommender systems. *Information Systems Frontiers*. 20(6), 1173–1189. DOI: 10.1007/s10796-017-9782-y

Nitzl, C., Roldan, J.L., & Cepeda, G. (2016). Mediation analysis in partial least squares path modelling. *Industrial Management and Data Systems*, 116(9), 1849-1864. DOI: 10.1108/IMDS-07-2015-0302

North, A.C., & Hargreaves, D.J. (2000). Musical preferences during and after relaxation and exercise. *American Journal of Psychology*, 113, 43–67. DOI: https://doi-org.libproxy.york.ac.uk/10.2307/1423460

North, A.C., Hargreaves, D.J. & Hargreaves, J.J. (2004). Uses of music in everyday life. *Music Perception: An Interdisciplinary Journal*, 22, 41–77. DOI: https://doi.org/10.1525/mp.2004.22.1.41

North, A.C., & Hargreaves, D.J. (2008). *The Social and Applied Psychology of Music*. Oxford University Press.

Nye, C.D., & Drasgow, F. (2011). Assessing Goodness of Fit: Simple rules of thumb simply do not work. *Organizational Research Methods*, 14(3), 548-570. DOI: 10.1177/1094428110368562

Oberski, D. (2014). *lavaan.survey*: An R package for Complex Survey Analysis of Structural Equation Models. *Journal of Statistical Software*, 57(1), 1-27. DOI: https://doi.org/10.18637/jss.v057.i01

O'Neill, K., & Egermann, H. (2022). Development of the Social Experience of a Concert Scales (SECS): The Social Experience of a Live Western Art Music Concert Influences People's Overall Enjoyment of an Event but not Their Emotional Response to the Music. *Music & Science*, 5, 1-22. DOI: https://doi.org/10.1177/20592043221106478

Oxford University Press. goal, *n.d. Oxford English Dictionary*. Retrieved from https://www.oed.com/view/Entry/79557?rskey=OPzCBc&result=1#eid

Pearl, J. (2012). The Causal Foundations of Structural Equation Modeling. In R.H. Hoyle (ed.), *Handbook of Structural Equation Modeling*, 68-91. *The Guildford Press*.

Perham, N., & Vizard, J. (2010). Can Preference for Background Music Mediate the Irrelevant Sound Effect? *Applied Cognitive Psychology*, 25, 625-631. DOI: 10.1002/acp.1731

Perry, J.L., Nichols, A.R., Clough, P.J., & Crust, L. (2015). Assessing Model Fit: Caveats and Recommendations for Confirmatory Factor Analysis and Exploratory Structural Equation Modeling, *Measurement in Physical Education and Exercise Science*, 19(1), 12-21. DOI: 10.1080/1091367X.2014.952370

Pichl, M., Zangerle, E., & Specht, G. (2015). Towards a context-aware music recommendation approach: What is hidden in the playlist name? In *2015 IEEE international conference on data mining workshop (ICDMW)*, 1360-1365. IEEE. DOI: 10.1109/ICDMW.2015.145

Pierce, J.L., Kostova, T., & Dirks, K.T. (2003). The State of Psychological Ownership: Integrating and Extending a Century of Research. *Review of General Psychology*, 7(1), 84–107. DOI: https://doi.org/10.1037/1089-2680.7.1.84

Ployhart, R.E., & Oswald, F.L. (2004). Applications of Mean and Covariance Structure Analysis: Integrating Correlational and Experimental Approaches. *Organizational Research Methods*, 7(1), 27-65. DOI: 10.1177/1094428103259554

Pollet, T.V., Stulp, G., Henzi, S.P., & Barrett, L. (2015). Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data. *American Journal of Primatology*, 77(7), 727-749. DOI: 10.1002/ajp.22405

Porta, M. (2014). Recall Bias. A Dictionary of Epidemiology (5th ed.). *Oxford University Press*.

Preacher, K.J., Rucker, D.D., & Hayes, A.F. (2007). Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behavioral Research*, 41(1), 185-227. DOI: 10.1080/00273170701341316

Preacher, K.J., & Kelley, K. (2011). Effect Size Measures for Mediation Models: Quantitative Strategies for Communicating Indirect Effects. *Psychological Methods*,16(2), 93-115. DOI: 10.1037/a0022658

Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, 157-164. DOI: https://doi.org/10.1145/2043932.2043962

Punch, K.F. (2014). Introduction to Social Research: Quantitative & Qualitative Approaches (3rd ed.). *SAGE*.

R Core Team. (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. Available at: https://www.R-project.org/

Randall, W.M., & Rickard, N.S. (2013) Development and Trial of a Mobile Experience Sampling Method (m-ESM) for Personal Music Listening. *Music Perception*, 31(2), 157-170. DOI: https://www.jstor.org/stable/10.1525/mp.2013.31.2.157

Randall, W.M., Rickard, N.S. & Vella-Brodrick, D.A. (2014). Emotional outcomes of regulation strategies used during personal music listening: A mobile experience sampling study. *Music Scientiae*, 18(3), 275-291. DOI: 10.1177/1029864914536430

Randall, W.M., & Rickard, N.S. (2017). Personal Music Listening: A Model of Emotional Outcomes Developed Through Mobile Experience Sampling. *Music Perception*, 34(5), 501-514. DOI: https://doi.org/10.1525/MP.2017.34.5.501

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C., Sherman, R., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A Taxonomy of Major Dimensions of Situation Characteristics. *Journal of Personality and Social Psychology*, 107 (4), 677-718. DOI: https://doi.org/10.1037/a0037250

Raykov, T., Marcoulides, G.A., & Patelis, T. (2013a). Saturated Versus Just Identified Models: A Note on Their Distinction. *Educational and Psychological Measurement*, 73(1), 162-168. DOI: 10.1177/0013164412457367

Raykov, T., Lee, C-L., Marcoulides, G.A., & Chang, C. (2013b). A Commentary on the Relationship Between Model Fit and Saturated Path Models in Structural Equation Modeling Applications. *Educational and Psychological Measurement*, 73(6), 1054-1068. DOI: 10.1177/0013164413487905

Reddy, S., & Mascia, J. (2006). Lifetrak: Music In Tune With Your Life. In *HCM'06 - Proceedings of the 1st ACM International Workshop on Human-Centered Multimedia*, 25–34. DOI: https://doi.org/10.1145/1178745.1178754

Rendle, S. (2010). Factorization Machines. In *IEEE International Conference on Data Mining*, 995-1000. DOI: 10.1109/ICDM.2010.127

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*, 84 (6), 1236-1256. DOI: 10.1037/0022-3514.84.6.1236

Rentfrow, P.J., & Gosling, S.D. (2006). Message in a Ballad: The Role of Musical Preferences in Interpersonal Perception. *Psychological Science*, 17(3), 236-242. DOI: https://doi.org/10.1111/j.1467-9280.2006.01691.x

Rentfrow, P.J., Goldberg, L.R., & Levitin, D.J. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology*, 100(6), 1139-1157. DOI: 10.1037/a0022406

Rentfrow, P.J., Goldberg, L.R., Stillwell, D.J., Kosinski, M., Gosling, S.D., & Levitin, D.J. (2012). The song remains the same: A replication and extension of the MUSIC model. *Music Perception: An Interdisciplinary Journal*, 30(2), 161–185. DOI: 10.1525/mp.2012.30.2.161

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings ACM Conference on Computer-Supported Cooperative Work*, 175–186. DOI: https://doi.org/10.1145/192844.192905

Rhemtulla, M., Brosseau-Liard, P.E., & Savalei, V. (2012). When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categrorical SEM Estimation Methods Under Suboptimal Conditions. *Psychological Methods*, 17(3), 354-373. DOI: 10.1037/a0029315

Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender Systems: Introduction and Challenges. In: F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, (2nd eds.). *Springer*, 1-34. DOI: 10.1007/978-1-4899-7637-6_1

Robinson, M.A. (2018). Using multi-item psychometric scales for research and practice in human resource management. *Human Resource Management*, 57(3), 739-750. DOI: 10.1002/hrm.21852

Robitzsch, A. (2020). Why Ordinal Variables Can (Almost) Always be Treated as Continuous Variables: Clarifying Assumptions of Robust Continuous and Ordinal Factor Analysis Estimation Methods. *Frontiers in Education*, 5. DOI: 10.1037/a0029315

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. DOI: https://doi.org/10.18637/jss.v048.i02

Rosseel, Y. (2017). *Multilevel structural equation modeling with lavaan*. [Workshop], Zürich, Switzerland. Retrieved from: https://scholar.google.co.uk/scholar?q=Rosseel,+Y.+(2017).+Multilevel+structural+equation+modeling+with+lavaan.+Presentation,+Z%C3%BCrich.&hl=en&as_sdt=0&as_vis=1&oi=scholart

Rosseel, Y., & Loh, W.W. (2022). A Structural After Measurement Approach to Structural Equation Modeling. *Psychological Methods*. Advance online publication. DOI: https://doi.org/10.1037/met0000503

Rucker, D.D., Preacher, K.J., Tormola, Z.L., & Petty, R.E. (2011). Mediation Analysis in Social Psychology: Current Practices and New Recommendations. *Social and Personality Psychology Compass*, 5(6), 359-371. DOI: 10.1111/j.1751-9004.2011.00355.x

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. DOI: https://doi.org/10.1037/h0077714

Russell, P.A. (1997). Musical tastes and society. In D.J. Hargreaves and A.C. North (Eds.), *The social psychology of music* (pp. 161–187). Oxford: Oxford University Press.

Saarikallio, S., & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music*, 35(1), 88-109. DOI: https://doi.org/10.1177/0305735607068889

Saarikallio, S. (2008). Music in Mood Regulation: Initial Scale Development. *Musicae Scientiae*, 12(2), 291-309. DOI: https://doi.org/10.1177/102986490801200206

Saarikallio, S. (2011). Music as emotional self-regulation throughout adulthood. *Psychology of Music*, 39(3), 307-327. DOI: 10.1177/0305735610374894

Sanders, E.A. (2011). *Multilevel Analysis Methods for Partially Nested Cluster Randomized Trials* (PhD). University of Washington.

Saravanou, A., Tomasi, F., Mehrotra, R., & Lalmas, M. (2021). Multi-Task Learning of Graph-based Inductive Representations of Music Content. In *Proceedings of the 22nd International Society for Music Information Retrieval Conference*, 602-609.

Sardeshmukh, S.R., & Vandenberg, R.J. (2017). Integrating Moderation and Mediation: A Structural Equation Modeling Approach. *Organizational Research Methods*, 20(4), 721-745. DOI: 10.1177/1094428115621609

Savage, P., Louis, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, 44(59). DOI: 10.1017/S0140525X20000333

Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. *Frontiers in Psychology*, 4. DOI: 10.3389/fpsyg.2013.00511

Schäfer, T. (2016). The Goals and Effects of Music Listening and Their Relationship to the Strength of Music Preference. *PloS ONE*, 11(3). DOI: 10.1371/journal.pone.0151634

Schäfer, T., & Mehlhorn, C. (2017). Can personality traits predict musical style preferences? A meta-analysis. *Personality and Individual Differences*, 116, 265-273. DOI: 10.1016/j.paid.2017.04.061

Schäfer, K., Saarikallio, S., & Eerola, T. (2020). Music May Reduce Loneliness and Act as Social Surrogate for a Friend: Evidence from and Experimental Listening Study. *Music & Science*, 3, 1-16. DOI: 10.1177/2059204320935709

Schedl, M., Flexer, A., & Urbano, J. (2013). The neglected user in music information retrieval research. *Journal of Intelligent Information Systems*, 41(3), 523-539. DOI: 10.1007/s10844-013-0247-6

Schedl, M., Gómez, E., & Urbano, J. (2014). Music Information Retrieval: Recent Developments and Applications. *Foundations And Trends in Information Retrieval*, *8*(2-3), 127-261. DOI: https://doi.org/10.1561/1500000042

Schedl, M., Zamani, H., Chen, C-W., Deldjoo, Y., & Elahi, M. (2018). Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, 7, 95-116. DOI: 10.1007/s13735-018-0154-2

Schedl, M., Knees, P., McFee, B., & Bogdanov, D. (2022). Music Recommendation Systems: Techniques, Use Cases, and Challenges. In: Ricci, F., Rokach, L., Shapira, B. (eds) *Recommender Systems Handbook*. DOI: https://doi.org/10.1007/978-1-0716-2197-4_24

Scherer, K.R., & Zentner, M.R. (2001). Emotional Effects of Music: Production Rules. In P.N. Juslin & J.A. Sloboda (ed.), *Music and emotion: theory and research*. Oxford University Press.

Schmitt, T.A., Sass, D.A., Chappelle, W., & Thompson, W. (2018). Selecting the "best" factor structure and moving measurement validation forward: An illustration. *Journal of personality assessment*, 100(4), 345-362. DOI: 10.1080/00223891.2018.1449116

Schubert, E. (2007). The influence of emotion, locus of emotion and familiarty upon preference in music. *Psychology of Music*, 35(3), 499-515. DOI: 10.1177/0305735607072657

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289-310. DOI: 10.1214/10-STS330

Silverman, M. (2020). Music-based affect regulation and unhealthy music use explain coping strategies in adults with mental health conditions. *Community Mental Health Journal,* 56(5), 939–946. DOI: 10.1007/s10597-020-00560-4

Sinclair, G., & Tinson, J. (2017). Psychological ownership and music streaming consumption. *Journal Of Business Research*, 71, 1-9. DOI: https://doi.org/10.1016/j.jbusres.2016.10.002
Skånland, M.S. (2011). Use of MP3 players as a coping resource. *Music and Arts in Action*, 3(2), 15–33. DOI
Skinner, C., & de Toledo Vieira, M. (2007). Variance estimation in the analysis of clustered longitudinal survey data. *Survey Methodology*, 33(1), 3-12.
Sloboda, J.A., O'Neill, S.A., & Ivaldi, A. (2001). Functions of music in everyday life: An exploratory study using the experience sampling method. *Musicae Scientiae*, 5(1), 9–32. DOI:10.1177/ 102986490100500102
Small, C. (1998). Musicking: The meanings of Performing and Listening. *Wesleyan University Press*.
Smid, S.C., McNeish, D., Miočević, M., & van de Schoot, R. (2020). Bayesian Versus Frequentist Estimation for Structural Equation Models in Small Sample Contexts: A Systematic Review. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1),131-161, DOI: 10.1080/10705511.2019.1577140
Smid, S.C., & Rosseel, Y. (2020). SEM with small samples. In M. Miočević & R. van de Schoot (eds.), *Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners*, 241–254. Routledge.
Smid, S.C., & Winter, S.D. (2020). Dangers of the Defaults: A Tutorial on the Impact of Default Priors When Using Bayesian SEM With Small Samples. *Frontiers in Psychology*. 11:611963. DOI: 10.3389/fpsyg.2020.611963
Song, Y., Dixon, S., Pearce, M.P., & Halpern, A.R. (2016). Perceived and Induced Emotion Responses to Popular Music. *Music Perception: An Interdisciplinary Journal*, 33(4), 472-492. DOI: 10.1525/mp.2016.33.4.472
Spector, P. (1992). Conducting the Item Analysis. In P. Spector, Summated Rating Scale Construction, 30-46. *SAGE*. DOI: https://dx.doi.org/10.4135/9781412986038.
Stapleton, L.M., McNeish, D.M., & Yang, J.S. (2016). Multilevel and single-level models for measured and latent variables when data are clustered. *Educational Psychologist*, 51(3-4), 317-330. DOI: https://doi.org/10.1080/00461520.2016.1207178
Stylos, J., Faulring, A., Yang, Z., & Myers, B.A. (2009). Improving API Documentation Using API Usage Information. *In Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing*. IEEE. 119–126. DOI: 10.1109/VLHCC.2009.5295283
Suhr, D.D. (2005). Principal Component Analysis vs. Exploratory Factor Analysis. *SUGI 30 proceedings*, 203-230.
Swain, J.P. (1996). The Rang of Musical Semantics. *The Journal of Aesthetics and Art Criticism*, 54(2), 135-152. DOI: https://www.jstor.org/stable/431086
Takama, Y., Zhang, J., & Shibata, H. (2021). Context-aware Music Recommender System Based on Implicit Feedback. *Transactions of the Japanese Society for Artificial Intelligence*, 36(1), 1-10. DOI: 10.1527/tjsai.36-1_WI2-D
Tamir, M., Vishkin, A., & Gutentag, T. (2020). Emotion Regulation is Motivated. *Emotion*, 20(1), 115-119. DOI: http://dx.doi.org/10.1037/emo0000635
te Grotenhuis, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., & Konig, R. (2017). When size matters: advantages of weighted effect coding in observational studies. *International Journal of Public Health*, 62, 163-167. DOI: 10.1007/s00038-016-0901-1

Ter Bogt, T., Mulder, J., Raaijmakers, Q., & Gabhainn, S.N. (2010). Moved by music: A typology of music listeners. *Psychology of Music*, 39(2), 147-163. DOI: 10.1177/0305735610370223

Thompson, C., Antal, D., Parry, J., Phipps, D., Wolff, T., & Holsenbeck, S. (2022). spotifyr: R Wrapper for the 'Spotify' Web API. Available at: https://CRAN.R-project.org/package=spotifyr

Threadgold, E., Marsh, J.E., McLatchie, N., & Ball, L.J. (2019). Background music stints creativity: Evidence from compound remote associate tasks. *Applied Cognitive Psychology*, 33, 873-888. DOI: https://doi.org/10.1002/acp.3532

Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G.R.G. (2008a). Five approaches to collecting tags for music. In *ISMIR 2008 - Proceedings of the 9th International Conference on Music Information Retrieval*, 225-230.

Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. (2008b). Semantic Annotation and Retrieval of Music and Sound Effects. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2), 467-476.

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys*, 50(6), 1-40. DOI: 10.1145/3123988

Van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. *Advances in neural information processing systems*, 26(2).

Van den Tol, A., & Edwards, J. (2015). Listening to sad music in adverse situations: How music selection strategies relate to self-regulatory goals, listening effects and mood enhancement. *Psychology of Music*, 43(4), 473-494. DOI: 10.1177/0305735613517410

Van Goethem, A., & Sloboda, J. (2011). The functions of music for affect regulation. *Music Scientiae*, 15(2), 208-228. DOI: 10.1177/1029864911401174

Van Sonderen, E., Sanderman, R., & Coyne, J.C. (2013). Ineffectiveness of Reverse Wording of Questionnaire Items: Let's Learn from Cows in the Rain. *PLoS ONE*, 8(7). DOI: https://doi.org/10.1371/journal.pone.0068967

Vinay, A., Lerch, A., Leslie, G. (2021). Mind the Beat: Detecting Audio Onsets From EEG Recordings of Music Listening. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 231-235. DOI: 10.1109/ICASSP39728.2021.9414245

Wang, X., Rosenblum, D., & Wang, Y. (2012). Context-Aware Mobile Music Recommendation for Daily Activities. In *MM 2012 – Proceedings of the 20th ACM International Conference on Multimedia*, 99-108. DOI: https://doi.org/10.1145/2393347.2393368

Wang, Z., Qin, M., Chen, M., Jia, C. (2017). Hiding Fast Flux Botnet in Plain Email Sight. In: Lin, X., Ghorbani, A., Ren, K., Zhu, S., Zhang, A. (eds) Security and Privacy in Communication Networks. *SecureComm 2017*. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. *Springer*. DOI: https://doi.org/10.1007/978-3-319-78816-6_14

Wang, D., Deng, S., & Xu, G. (2018). Sequence-based context-aware music recommendation. *Information Retrieval Journal*, 21(2), 230-252. DOI: https://doi.org/10.1007/s10791-017-9317-7

Watkins, M. (2018). Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology*, 44(3), 219-246. DOI: 10.1177/0095798418771807

Weber, H. (2009). Taking your favorite sound along: Portable audio technologies for mobile music listening. In K. Bijsterveld & J. van Dijck (Eds.), *Sound Souvenirs: Audio Technologies, Memory and Cultural Practices*, 69–83. Amsterdam University Press.

Widaman, K. (2018). On Common Factor and Principal Component Representations of Data: Implications for Theory and for Confirmatory Replications. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(6), 829-847. DOI: 10.1080/10705511.2018.1478730

Williams, A. (2007). Portable Music & Its Functions. New York: *Peter Lang*.

Worthington, R., & Whittaker, T. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist*, 34(6), 806-838. DOI: 10.1177/0011000006288127

Wu, J-Y., & Kwok, O-M. (2012). Using SEM to Analyze Complex Survey Data: A Comparison between Design-Based Single-Level and Model-Based Multilevel Approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(1), 16-35. DOI: 10.1080/10705511.2012.634703

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons from Machine Learning. *Perspectives on Psychological Science*, 12(6), 1100-1122. DOI: 10.1177/1745691617693393

Zentner, M., Grandjean, D., & Scherer, K.R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8(4), 494–521. DOI: https://doi.org/10.1037/1528-3542.8.4.494

Zhao, X., Lynch. J.G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, 37(2), 197-206. DOI: 10.1086/651257

Zheng, Y., Mobasher, B., & Burke, B.D. (2013). "The Role of Emotions in Context-aware Recommendation". In *Decisions@RecSys*. 21–28.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75-89. DOI: 10.1057/jit.2015.5