Sense of Belonging and Language Acquisition: a Pilot RCT Examining Effects of Group Musicking in Adult Migrants Learning Norwegian

Johanna Magdalena Engh

Master's thesis in music therapy

Faculty of Fine Arts, Music, and Design

University of Bergen

Spring semester 2023

Abstract

This master's thesis uses a parallel Pilot-RCT design to explore a possible connection between group musicking, sense of belonging and second language acquisition in adults learning Norwegian in a Norwegian University. Based on a rationale of earlier research and several theories from developmental psychology, music therapy, pedagogy, linguistics and more, the concepts of musicking, sense of belonging and second language acquisition are directly and indirectly linked. Participants (n=24), who had to be older than 18 and in the Norwegian course of the University of Bergen, were randomly assigned by stratified randomisation to either a group (n=13) receiving group musicking in addition to standard language classes, or the control group (n=11) which received nothing outside of their standard language classes. No blinding was done. The primary outcome of the study, Sense of Belonging, was measured at baseline and after 9 weeks of weekly group musicking sessions. An ANCOVA showed a mean difference between groups of -2.84 (95% Confidence Interval: -14.40 — 8.71), with a p-value of 0.604, favouring the control group. The results are not significant, but further research needs to be conducted. There is a high risk of a type two error, considering positive narrative feedback. A large proportion of participants dropped out. This pilot study provides a good basis for a full-size RCT.

Acknowledgments

I wish to express my gratitude to:

My thesis supervisor **Christian Gold** for your constant support and guidance and taking the time to teach me details where my knowledge was lacking (such as statistics and R). You are an inspiration to music therapists and music therapy students.

My thesis outline supervisor **Simon Gilbertson** for helping me come up with the general idea of this study.

All the **participants of the study**. Without your participation none of this would have been possible.

The **Norwegian course teachers** at the university. Your input and help to recruit were a privilege, and I am very thankful for that.

Marie for helping me with methodology, revision, and inspiration.

Nikolaus for your help with revision and programming issues. Thank you for your patience when helping me understand statistical concepts.

Richard for your help with revision.

Erlend for your amazing support throughout the year.

My colleague **Haakon** for listening to countless rants on theories and statistics during writing hours, and in turn supplying entertainment with your own rants.

And lastly the **air purifier** for keeping me mostly sane and the air mostly clean in the basement room where the writing spots for the master's students were located.

Table of Content

List of Tables and Figures.

Appendix Overview

See attached separate documents:

Appendix 1: Consent forms

Appendix 2: Scripts from RStudio

Appendix 3: Full output of ANCOVAs and t-tests

Appendix 4: Datasets excluding background data

Appendix 5: Application to NSD

Appendix 6: NSD approval

Appendix 7: SOBI questionnaire

Appendix 8: CEFR questionnaire

Appendix 9: Background questionnaire

1 Introduction

Musicking, sense of belonging and language acquisition are concepts that heavily intertwine through various theories and previous research. These are themes important in a migrant context, yet little research has been done linking musicking, sense of belonging and language acquisition to a migrant context. This thesis aims to address that link, by showing the pre-existing links between the different contexts and why it makes sense to link the context with practical examples. By linking these concepts, describing the importance of them, and giving examples, the introduction chapter lays a theoretical groundwork for a Pilot-RCT, conducted with 24 participants who were recruited from the Norwegian course at the University of Bergen.

1.1 Target Group

Migration, as a global challenge, has become more prevalent throughout the last century. The international organization for migration of the United Nations Systems defines "migrants", describing of course the individuals that make up migrations, as "An umbrella term, not defined under international law, reflecting the common lay understanding of a person who moves away from his or her place of usual residence, whether within a country or across an international border, temporarily or permanently, and for a variety of reasons" (International Organization for Migration (IOM), 2019, p. 132). On a world-wide basis, around 3.6% of the total population were migrants in 2020 (Migration, 2021). Migration to Norway has increased a considerable amount since the 1950's (Laue et al., 2021). In 2023, around 877 000 people, which amounts to 16% of the population, were migrants (Statistisk Sentralbyrå, 2023b). Around 60 000 people moved to Norway within the last year (Statistisk Sentralbyrå, 2023a).

Migrant health as a field of study is a complex field which needs more research and mapping (Laue et al., 2021). The needs of this population vary, depending on economic background, reason for migration, geographical background. As the general population group "migrants in Norway" is so diverse, migrants as a group face many challenges and have many different

needs. In any case, risks for mental distress and disorders have been reported to be increasing (Kale & Hjelde, 2017). Having a migrant status can affect both physical and mental health in a negative way (Chou, 2007; Henderson et al., 2017). The sense of belonging of migrant students can affect mental health, and is often lower than in students without a migration background (Marksteiner et al., 2019). Meeting these needs and supporting migrants during these challenges is important, especially considering the large percentage of the population that are migrants. One of largest issues when trying to achieve health equity in a diverse context is the language barrier (Kale & Kumar, 2012). To be able to meet the health needs of this population group, one often needs to face challenges of communication, so supporting functional language programs also supports population health.

Language and communication are major steppingstones towards functional integration into the local culture. To this date the government of Norway has no program that is for all immigrants, but has adopted a legal requirement for all municipalities to have an introduction program for refugees between the ages of 18 and 55 (Regjeringen, 2021). This is based on the integration law ("Integreringsloven," 2020), which states that the introduction program needs to be adapted to the individual with individual goals, and needs to include at least teaching Norwegian, social studies, a course in life management and an element that is directed towards work or education. Participants of the course that have children under 18 need to participate in parent guidance. After the introduction program, statistics show that around 60% of the participants in the programs get jobs or education (Statistisk Sentralbyrå, 2023b). A larger part of the population has rights to education, both in Norwegian and primary education, and have a right to get this education without being proficient in Norwegian and for the program to be adapted to their needs. This is according to the education law ("Opplæringslova," 1998). In addition to the legally required programs other resources exist for migrants. For example, several ideological and private companies organize their own Norwegian classes and migrant help programs, and companies hiring work from outside of Norway often offer Norwegian classes.

Music therapy with migrants is in some respects well researched. A literature review by Henderson et al. (2017) showed that music participation promotes social wellbeing, stress reduction, enhances mental health and supports mental health, while arguing that not enough

research has been done on the topic. Music participation by the target population can promote social inclusion, help develop a sense of identity and increase self-esteem and confidence (Belfiore & Bennett, 2007; Chew, 2009; Ruud, 2010; Sæther, 2008). Music participation can bridge cultural gaps, reducing prejudice, discrimination, and aggression between groups (Enge & Stige, 2022; Krüger & Diaz, 2023). The research shows that the potential of music therapy and musicking in a migrant contest is promising, but there is a clear need for more research.

1.2 Sense of Belonging and Musicking

Hagerty et al. (1992, p. 173) define *sense of belonging* as "the experience of personal involvement in a system or environment so that persons feel themselves to be an integral part of that system or environment". The definition builds upon an earlier similar sounding definition, where sense of belonging is defined as a "sense of personal involvement in a social system so that persons feel themselves to be an indispensable and integral part of the system" (Anant, 1966). In educational settings, sense of belonging has its own separate definition: "[sense of belonging in educational environments is a] students' sense of being accepted, valued, included and encouraged by others (teacher and peers) in the academic classroom setting and of feeling oneself to be an important part of the life and activity of the class. More than simple perceived liking or warmth, it also involves support and respect for personal autonomy and for the student as an individual" (Goodenow, 1993, p. 25). Belonging has been found to be a fundamental human need which is necessary for motivation and well-being (Barringer et al., 2022; Baumeister & Leary, 1995). As such, the importance of sense of belonging is quite pronounced. For example in the context of language learning, several studies show that sense of belonging can have an effect on learning abilities both with children and students (Allen & Bowles, 2012; Allen et al., 2018; Caligiuri et al., 2020; Glass & Westmont, 2014; Henderson et al., 2017; Pittman & Richmond, 2007; Sari, 2012). Some research within music therapy shows positive results of music therapy on sense of belonging in adult migrants (Ebberts, 1994; Hunt, 2005; Ledger & Baker, 2007; Pollack & Namazi, 1992; Rio, 2002).

Clarification of several related terms could provide an even better understanding of the term "sense of belonging". For example, McMillan and Chavis created a theory of sense of community, which they define as "a feeling that members have of belonging, a feeling that members matter to one another and to the group, and a shared faith that members' needs will be met through their commitment to be together" (McMillan, 1976; McMillan & Chavis, 1986, p. 9). The concept of community is larger than sense of belonging, including membership and the ability to affect a society, integration and a shared emotional connection (Stige & Aarø, 2011). As part of McMillan's theory and definition, is the emphasis on the importance of belonging and membership and define it as "the sense of belonging and identification involves the feeling, belief, and expectation that one fits in the group and has a place there, a feeling of acceptance by the group, and a willingness to sacrifice for the group" (McMillan & Chavis, 1986, p. 10). Their connection between belonging and identity is a theme sometimes found in several disciplines in social sciences, and is sometimes used interchangeably (Amit & Bar-Lev, 2015).

1.3 Second Language Acquisition and Musicking

Every language we learn beyond our first language is a second language (Ellis, 1989; Krashen, 1981). Tomasello defines language acquisition as the process of developing the ability "to perceive, produce and use words to communicate" (Tomasello, 2003). The process of acquiring the language "second language acquisition", a term that is broader than learning (Hoque, 2017). Second language learning is defined as referring "to the formal learning of a language in the classroom" (Hoque, 2017), and is the pedagogical field on how language can be taught and learned. Second language acquisition happens everywhere and every time a person acquires parts of the language, whether it be vocabulary, pronunciation, subtle social cues within the language or prosody, the melody and rhythm of language (Boutsen, 2003). Second language learning facilitates for second language acquisition, but focusing on second language acquisition emphasizes the importance of learning "by doing", through everyday interactions and communication in addition to language learning.

1.3.1 Central Theories to Language Acquisition and Musicking

Language and its acquisition are themes that are central in many theories of development and motivation. To use musicking in a context of learning or acquisition it is beneficial to have an overview over central theories, as it can give insight as to how music can be a facilitator of motivation, learning and thereby language acquisition. Though many of the theories are mainly about the first language we learn, an understanding of how we learn our first language can help us learn a second language. All of the theories can be applied to a teaching style that includes music, which examples in this subchapter describe. Of course, musicking can be connected to a myriad more specialized theories (such as Deci and Ryan's theory on belonging and autonomy which leads to stronger intrinsic motivation (2017)), but in this chapter, the aim is to connect some of the most general cognitive development theories to musicking and language learning.

Lev Vygotsky's theory on cognitive development is sociocultural and includes language as an essential tool to cognitive development (Vygotsky, 1962). In Vygotsky's theory, language (as a root of culture) is essential to learning. We learn through culture and community, and language is key to communicate said culture. Vygotsky's theory also includes the theory about the "zone of proximal development". What lies within the zone of proximal development is just beyond what we are able to learn by ourselves and what we have learned, but we can learn it with help from others. For example, we might not be able to learn a new song by ourselves even though we know other songs and know how to sing, but with a teacher (or other help) who guides us through the song phrase by phrase, we can learn any piece of music. Pronunciation is also difficult to learn but can be learned with the help from a teacher.

Skinner's theory has a social operant behaviour perspective on human learning and cognitive development (Skinner, 1963). For example, through imitation of imitation of language, the sounds which are more similar to language get positive reactions and are thereby rewarded and reinforced. In this way learning of language, according to Skinner, happens simultaneously socially and by operant conditioning. This process of operant learning is similar to instrumental learning (Staddon & Cerutti, 2003). This theory is in practice directly transferrable from learning language to learning music and singing. Learning a song often

happens through imitation and when that is coupled with positive feedback from the teacher of the song, that becomes an example of this theory in practice.

Within social learning, Bandura's theory on social learning builds upon the operant learning theory of Skinner (Bandura, 1977; Mills, 2014). Instead of focusing mainly on operant learning, Bandura's theory has a larger focus on social learning through observation and imitation. Human behaviour, such as language, is learned through observation, modelling, and imitation of other behaviour in one's environment, with motivation which stems partially from self-efficacy. Self-efficacy is a very important motivational factor, which is informed through self-reflection and mastery experiences (Bandura, 1997; Mills, 2014). In a musicking context, songs are often taught through imitation and repetition of musical phrases. Singing in a new language can provide a new arena of achieving a feeling of mastery while exploring the new language (Valvatne, 2007).

Piaget's theory on cognitive development focuses more on stages and schemas of learning (Piaget, 1971). Language in development is a reflector of the logical stage of learning, and as such language has different purposes through the different stages in children's development. The learning process itself, according to Piaget's theory', happens through assimilating and fitting new data into the already learned things. For example, when singing a song which is familiar in one language in a new language, that would assimilate the new language learned through the song into the previously learned melody and meaning of the song.

Noam Chomsky's theory on linguistics describes language and language learning as being an innate and biological part of the human mind (Chomsky, 2006). Chomsky's theory focuses on brain structures and biological inheritance of language. Over time and aging, brain structures such as synapses strengthen through cognitive development, growth and practice and enable language learning. Musicking enables strengthening of synapses and brain plasticity (Dalla Bella, 2016; Levitin, 2006; Olszewska et al., 2021). Since music and musicking uses a wide area of the brain, music and language share several brain structures (Brown et al., 2006; Janata & Grafton, 2003; Peretz & Zatorre, 2005; Schön et al., 2010).

Tomasello's usage-based theory on learning and language acquisition describes a social interaction leading to learning, and that learning happens through usage of what we are trying to learn, as in "learning by doing" (Tomasello, 1999, 2003, 2010). Through learning and using the different parts of speech, especially in a social context, we acquire the language that is being learned. In Tomasello's theory, language is essential to be able to learn from others. The theory builds on earlier theories such as the theories by Vygotsky (1962) and Piaget (1971). In a musicking context this means that through a musicking session, it would create an arena where participants can learn and use language in a safe social context.

In many of these theories, motivation and especially intrinsic motivation is a common topic. Motivation to learn is very important to keep focus and a steady learning curve. Brown (1973), who talks about affective variables in learning a language says: "The self-knowledge, self-esteem, and self- confidence of the language learner could have everything to do with success in learning a language". As music can be engaging, affectionate and can be adapted to many people, it can support intrinsic motivation for learning over a long period of time (Good et al., 2015). Additionally, motivation can be facilitated through social activities in learning "by doing". Through musicking and singing in the learned language there is a level of language immersion in a social context which promotes usage of the language and social language, which is mentioned in many of these theories.

1.3.2 Music as a Facilitator for Learning a Language

Musicking is a common tool for language learning, both in pedagogical and therapeutic fields. A considerable amount of research and multiple theories advocate for the use of music in foreign language learning (for example Engh, 2013; Good et al., 2015; Medina, 1993; Piri, 2019; Schoepp, 2001). Though there seems to be a substantial amount of research supporting its efficacy in a learning environment, Engh (2013) found that it is used quite little in practice in learning environments. Engh suggests that not only is more research in the field needed, but practical materials for teaching with the use of music are needed for a successful common implementation of music in language pedagogy. Music as a tool for language learning and rehabilitation is also common in other settings. Melodic intonation therapy (MIT) is a methodical intervention designed to rehabilitate language after a stroke, by using several activities to activate several brain regions and build new neural pathways (Albert et al.,

1973). Developmental Speech and Language Training through Music (DSLM) is a music therapy group intervention designed to enhance communicative skills in children with autism (Blythe laGasse, 2014).

Neurological music therapy research theorizes musical brain structures and how music affects the brain. Some overlap in brain structures for language and music such as neural networks and shared cortical activations could suggest a practical connection between music and language (Blythe laGasse, 2014; Brean & Skeie, 2019; Brown et al., 2006; Good et al., 2015; Koelsch et al., 2002; Maess et al., 2001; Schön et al., 2010). Musical training has a positive effect on verbal intelligence and attention to language melody (Lathroum, 2011; Moreno et al., 2011). Musical activity also adds plasticity of the brain structures and can improve memory and attention span (Brean & Skeie, 2019; Koelsch et al., 2002).

Intrinsic motivation is an important factor when learning anything according to several central theories on motivation (Bandura, 1997; Ryan & Deci, 2017). By introducing learning material musically, which might be more pleasurable, it might be more engaging and lead to more motivation (Good et al., 2015; Paquette & Rieg, 2008; Sandberg, 2009). For example, it might be easier for many people to learn long lists (such as countries or chemical elements) by learning a song that lists all items on that list in the lyrics, than by just listing the items. This method of memorizing through song could be compared to mnemonics.

1.4 Group Musicking

Christopher Small defined the verb Musicking as: "To music is to take part, in any capacity, in a musical performance, whether by performing, by listening, by rehearsing or practicing, by providing material for performance (what is called composing), or by dancing" (Small, 1998, p. 9). This concept or definition is important in this thesis, because we want to explore the effect of the entire process around creating music.

Though music therapy is not used in this paper, the paper is heavily connected to music therapy and uses rationale from music therapy as a theoretical framework. A definition of music therapy from 2014 by Kenneth Bruscia is: "Music therapy is a reflexive process wherein the therapist helps the client to optimize the client's health, using various facets of music experience and the relationships formed through them as the impetus for change. As defined here, music therapy is the professional practice component of the discipline, which informs and is informed by theory and research" (Bruscia, 2014, p. 36). This emphasizes the reflexive process between the therapist and client, using various aspects of music experience and relationships formed through them as the driving force for change towards optimizing the client's health. In the context of this thesis, language development can be seen as the sought after "change," and various music experiences and relationships built through them are utilized to facilitate language learning and a sense of belonging. Although the focus of the sessions is not on a therapeutic goal in this educational setting, the connection between musicking, sense of belonging, and language learning is highly relevant to the field of music therapy.

Community music therapy (CoMT) connects music therapy and musicking to a larger community perspective. Community music therapy has a broader definition of what music therapy can be, and gives music therapy a larger context to perform in. The approach emerged in the early 2000s, from a need to shift the focus within music therapy from an individual to an individual within a community (Ansdell, 2002; Steele, 2016). While community music therapy as an approach enables music therapists to shift therapy to a community perspective, it has been criticized to be too vague and overlapping with other approaches within music therapy (Steele, 2016; Wood, 2016). As described in more detail above, sense of community and sense of belonging overlap, showing that a sense of belonging is necessary to building community. Therefore, building a sense of belonging for large parts of the population which need this focus is essential to community music therapy.

To facilitate social belonging and language acquisition through music, an adequate way is to use group musicking. By musicking in a group, we use both the benefits of musicking and group interaction, a combination that might be beneficial to language acquisition and sense of

belonging in the target group, which can benefit from both facilitating of language acquisition and building sense of belonging.

1.5 The Aims and Hypothesis of This Study

Based on the previous reasoning, a question clearly formulating the aims of this study is:

"In adult migrants, what effect does group musicking have on sense of belonging and second language acquisition, when compared to no specific activity?"

As written here the question is a PICO-question, which reflects some of the basic components of an RCT; the participants or population of the study, the intervention that is tested, what the control group gets for comparison, and what the measured outcomes are. All of these parts are key elements of a transparent and clear reporting of a trial. For this research question this means:

P - Population of interest/Participants – Adult migrants learning Norwegian

I - Intervention - Group musicking

C - Comparison – The control group gets their usual Norwegian classes, and no specific other intervention.

O - Outcomes measured: Sense of belonging as a primary outcome and self-perceived knowledge of language as a secondary outcome

The primary outcome of this study will be the data about sense of belonging, whereas second language acquisition will be the secondary outcome. A primary outcome is essential for a trial and should be selected before the start of the study (Bradt, 2012; Stanley, 2007), which I have done for this study. Sense of belonging is the primary outcome for analysis because group musicking might be able to supplement normal language learning with an additional and beneficial direction of fostering social belonging in the target group, while the normal language learning is already a great and effective language method.

The question states the purpose of this thesis: to explore a possible effect group musicking might have on sense of belonging and language learning in adult migrants. With the reasoning from the introduction, musicking with a focus on language learning and social connectedness might have a positive effect on sense of belonging and language learning. However, to state a hypothesis on the result of the thesis, I do not expect a significant or recognizable difference in either outcome due to the very small sample size and the time constraints given.

If musicking can aid in second language learning, it can also be implemented in societal structures for language learning, community building, and supporting well-being by fostering a sense of belonging. In other words, group musicking in a migrant context can be a cost-effective and sustainable way to address several pertinent issues. While the intervention in this study may not be labelled as music therapy, it still highlights the potential impact that music can have on language learning and social connectedness.

2 Methods

Choosing a suitable method to address a research question is a major part of a master's thesis, but presenting it clearly is at least as important as the method itself. This section presents the method of this study in detail, following the CONSORT 2010 statement to ensure clarity and transparency (Juszczak et al., 2019; Moher et al., 2010; Schulz et al., 2010). First, the trial design will be presented with reasoning to justify why it is a suitable design. The next part, moving on to the practical parts of the trial design, is about recruitment & randomization of participants, followed by a brief summary of ethical considerations and anonymity for the participants. The group musicking sessions will be described with the intervention protocol that is planned. Finally, the questionnaires that are the method to collect the data are described, followed by the methods of statistical analysis.

2.1 Trial Design: a Pilot RCT

To investigate the impact of group musicking on adult learners of Norwegian, various trial designs could be considered. One common and straightforward design is a randomized controlled trial (RCT), which involves recruiting participants and randomly assigning them to either an experimental group (receiving the intervention, in this case, group musicking sessions) or a control group (receiving standard language teaching). The outcome measures are then compared between the two groups to determine if there is a significant difference. This type of design is referred to as a parallel trial since both groups run in parallel.

For a high level of certainty, reliability, and level of evidence quantitative research is recommended, ultimately to ensure that as many as possible can benefit from what is being researched. Among quantitative research methods, randomized controlled trials (RCTs) are generally recognized as being "the most reliable method determining effectiveness" and is the only way of establishing causality out of all research methods (Campbell et al., 2000). In music therapy there are many examples of RCTs, among them by Baker et al. (2019); Chen et al. (2016); Ghetti (2013); Ridder et al. (2013); van Bruggen-Rufi et al. (2017). However there are not enough RCTs conducted within music therapy, at least in some fields within music

therapy such as in the psychiatric music therapy literature base (Silverman, 2010). To promote more RCTs in the field of music therapy, Bradt (2012) has published detailed guidelines for conducting RCTs in the field of music therapy. Only very few music therapy master's students have conducted RCTs or similar designs previously. Among these there are even fewer that have been published; among them are Ulrich et al. (2007) about group music therapy for schizophrenic patients with 37 participants, and Werner et al. (2017), comparing the effects of group music therapy to the effects of recreational group singing in a 117 elderly nursing home residents.

To explore the potential of a full-size RCT, a smaller design can suffice. It needs the key elements of a full-size RCT, but uses a smaller sample size and could be over a shorter timeframe, requiring fewer resources, but still exploring feasibility in a meaningful way and informing for optimized use of resources in a full-size RCT. Feeley et al. (2009) argue for the importance of a Pilot-RCT, outlining several purposes that differ from full-size RCTs. Pilot RCTs are used to assess practical questions concerning methods and formats of a future full-size RCT. Campbell et al. (2000) shows through their figures how exploratory trials are necessary to complete definitive randomized controlled trials for the most effective way of obtaining reliable evidence. Other terms for "pilot-RCT" and "full-size RCT" are also "exploratory" RCT and "confirmatory" RCT, emphasizing the purpose of exploring methodology and other facets for a full-size RCT. This also means that the outcome measured resembles that of a full-size RCT, but a Pilot-RCT is rather about finding out if a full-size RCT is doable, similar to a feasibility study. The difference between a Pilot RCT and a feasibility study is that a Pilot-RCT mimics a full-size RCT with an outcome, while a feasibility study explores it statistically and theoretically (Eldridge et al., 2016).

Figure 1 provides a simple illustration of a parallel trial design. While there are many ways to carry out each step, the core components of the study must be present to generate reliable statistical results. Although the current study is a Pilot-RCT, the design is mostly the same as an RCT, with the main difference being a smaller group of participants.

Figure 1. Graphical abstract of trial design
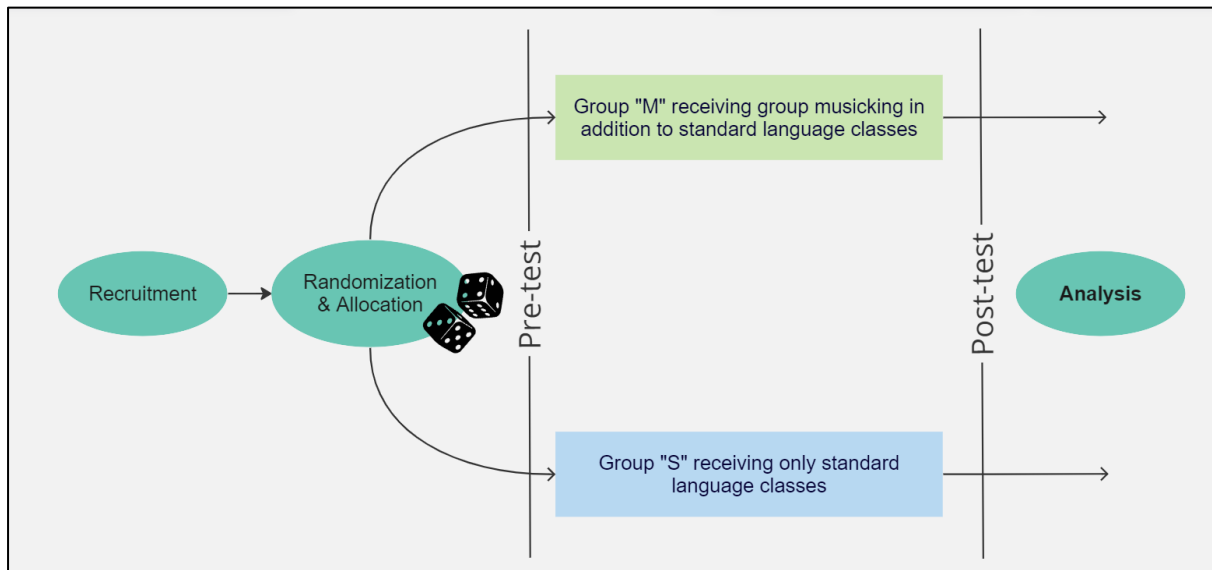
Note. The general schematic design of this trial.

For this trial, a parallel design was chosen as the simplest and most straightforward approach. Participants were recruited from an eligible source and randomly assigned to either the experimental group, which received group musicking sessions, or the control group, which received standard care. An initial outcome measurement was conducted at the beginning of the trial to establish baseline values, and then group musicking sessions were conducted with the experimental group. At the end of the study, a final outcome measurement was conducted to evaluate any differences between the two groups.

One variation of this model which was considered and ruled out for this trial was a crossover trial. A crossover trial has the same design in the beginning of the trial, but after a while of treatment/non-treatment the two groups switch, and the control group is now the test group while the test group is the control group. Before switching the outcomes are measured, and then the outcomes are measured once again at the end of the trial. An advantage of this design would have been the opportunity for all participants to get group musicking, which might have led to greater retention of participants, as many who initially signed up for the study were excited about using music for their learning process. However, it is not only very time consuming, but it also carries a risk of "carryover effects". It is difficult to separate the effects of the first part of the trial without a "wash-out" period between the two (Johnson, 2010; Lim

& In, 2021), which was also not possible due to time constraints. Another issue is that the effects on sense of belonging might be time-dependent and may carry a period-effect, which could make a crossover design suboptimal (Lim & In, 2021). Due to these disadvantages a crossover design was ruled out and deemed as impractical for this study. As a lot of participants were excited to learn by musicking and to attempt better retainment of participants in the control group I chose to have open and free group musicking sessions following the project. Anyone could participate in these open sessions, given that they were after the second round of questionnaires.

2.2 Recruitment and Randomization

To recruit participants an information letter was sent out and/or handed out to students of Norwegian courses, in cooperation with the department responsible for the courses and also with the Norwegian course teachers. Around 600 students, professors and refugees attend the courses each year. Most attendees are connected to the university, but people from outside the university can sign up, and in recent years the courses have been offered for free to refugees. Many people showed an interest to participate, but either discontinued contact after the initial contact, were ineligible to participate due to age (being under 18) or decided they did not have time to participate. Contact was either established by e-mail or in person. In a way, the exclusion process was part of the recruitment process, as participants who were not eligible ceased contact. There were two eligibility criteria. Participants had to be enrolled in the Norwegian courses of the university and had to be older than 18 years old. As these are very few eligibility criteria and the recruitment process was very targeted this was an intuitive way to recruit participants. Since the recruitment happened through the language program at the university, the participants all happened to be affiliated with the university and were highly educated people. The individuals who decided they wished to participate signed a consent form, received an anonymous participant number, and answered the first round of questionnaires. The recruitment process was continuous for 2 weeks.

Small sample sizes are a frequent criticism of literature reviews in music therapy (Bradt, 2012; Silverman, 2010). Effect size is an important factor in determining sample sizes for

trials, to achieve significant results (Gold, 2004). When trying to measure an effect that has a small effect size, a large sample size is needed to be a reliable and significant result. Reducing sample size, reduces the chance of finding a significant result. Therefore a sample size calculation is important, to have a large enough sample size to achieve trustworthy results within reasonable limits without studying more participants than required (Kadam & Bhalerao, 2010). There are a few online tools to conduct a sample size calculation. As part of a sample size calculation an anticipated mean, using the primary outcome and examining earlier studies within the same population. I did not find an earlier study done on adult migrants using SOBI-P and was therefore unable to perform this sample size calculation confidently. Though a sample size calculation is recommended for pilot studies (Bradt, 2012), a pilot trial can also help guide future sample size calculations (Cocks & Torgerson, 2013).

The numbers, together with their gender, were sent to my supervisor, who performed a stratified randomization to separate participants into a test-group and a control group. A stratified randomization means that the numbers have one extra factor (in this case the gender of the participant) and based on this factor smaller batches of the participants get randomly assigned to the test and control group, to divide the factor evenly between the groups. In this case the smaller batches were a size of four participants per group. This process of randomisation was performed by my supervisor Christian Gold, who had no contact with participants, to ensure allocation concealment. Allocation concealment is a way of ensuring an unbiased and equivalent way of allocation and randomization, avoiding systematic biases such as selection and confounding biases (Bradt, 2012; Schulz, 2001). This requires certainty that the recruiters have no prior knowledge of group assignment (Suresh, 2011), and is necessary because clinicians and researchers might try to influence the allocation to receive better results (Gold, 2015). Properly executed randomization will help ensure evenly distributed demographics, and subjects in various groups should not differ in any systematic way (Suresh, 2011), though random error can still occur. This is especially true for smaller sample sizes, as just a few data points can change the descriptive statistics of the sample. However, because these random errors occur by chance and not choice, the randomization process remains valid and supports scientific accuracy (Schulz & Grimes, 2002). As recruitment was continuous for two weeks, some more participants were included after the first randomization and a second randomization with only the second group was performed.

Blinding or masking is a technique used to reduce bias after allocation. This can involve blinding any contributing part of the study, such as participants, intervention providers, data analysts, or statisticians (Bradt, 2012; Gøtzsche, 1996). In research on music therapy or musicking, it can be challenging to blind several groups. Giving participants a placebo that emulates the researched intervention would quickly turn it into group musicking again. Another way to blind participants is to make them unaware of the intervention or the hypothesis of the study. Participants could not be unaware of the hypothesis of the study in this particular case due to Norwegian ethical guidelines, which require informing participants about the process of the researched intervention and the study. Additionally, it was impossible to blind the intervention provider, the statistician, and the data analysts since they were all the same person.

Participants were contacted through their given contact information to give information about their given group and session time for group musicking. After allocation and contacting all participants were part of analysis according to intention-to-treat (ITT) principles.

2.3 Anonymity and Safety - Ethical Procedures

All participants signed consent sheets including details of the purpose of the study, details on who is responsible for the study, what participation involved for participants, how participants could withdraw, and how data is anonymized. The consent form, which can be viewed in Appendix 1, was based on a template by the Norwegian centre for research data ("Norsk senter for forskningsdata" NSD). The project, including the proposed plan and consent form was approved by NSD in the beginning of September 2022. The application and plan proposal is viewable in Appendix 5, and the approval from NSD is viewable in Appendix 6.

To ensure the safe handling and a respectful way of treating all participants, I am following a few sets of ethical guidelines. The main ethical set of guidelines I am following is the set of

general national research guidelines by the National Research Ethics Committees of Norway (De nasjonale forskningsetiske komiteene, 2019). The main categories of principals these guidelines are divided into are "respect", "good consequences", "fairness" and "integrity", and further divided into 14 subcategories that are described in detail. The titles of the 14 subcategories are as follows: Quest for truth, academic freedom, quality, voluntary informed consent, confidentiality, impartiality, integrity, good reference practice, collegiality, institutional responsibility, availability of results, social responsibility, global responsibility, and laws and regulations. The national research ethics committees also have more specified guidelines for Social Sciences and Humanities, and for Medical and health research that are also taken into consideration for this study. These guidelines also take into account the European Code of Conduct for Research Integrity, the Vancouver protocol and the declaration of Helsinki.

The University of Bergen (UiB) also have their own research ethics guidelines (Universitetet i Bergen, 2019). UiBs ethical guidelines are based on the ethical guidelines by the national research ethics committee and include 10 rules or guidelines: Academic freedom and responsibility, expertise, integrity and reliability, responsibility for academic supervision, transparency, projects that are relevant to many people, multiple authors, conflicts of interest and impartiality, and conflict resolution.

For a responsible and clear conducting and reporting of the trial I am following the CONSORT-statement (Juszczak et al., 2019; Moher et al., 2010; Schulz et al., 2010) to ensure that all important parts are reported.

In practice for this study, ethical considerations include especially the target group, that are potentially in a vulnerable position. As such it is extra important to treat the participants with respect, integrity, clarity, and openness. Due to potential language difficulties, making sure everything is understood is of essence. Therefore, I had the consent form available in several languages, and kept in touch with interested people, asking if they need clarifications of terms or translations. Since many participants in the control group wanted to participate in the group musicking sessions, I also found it was unfair to make group musicking unavailable to

them. In line with fairness, I organized a few group musicking sessions after the trial was finished, to enable the participants and more people to receive the group musicking sessions.

For safety and anonymity, the data was collected on a secure online program called SurveyXact by Rambøll, which is affiliated with the University. Already at this stage the data was anonymized with anonymous participant numbers. From the online questionnaire program, it was loaded into an external desktop through the SAFE program of the university, and the online data was deleted as soon as possible. The data was kept completely safe and not identifiable.

2.4 Group Musicking Sessions

In total, nine sessions lasting one hour each were scheduled, starting at the end of September and continuing on a weekly basis until the end of November. The sessions were held in a group room at the humanistic faculty of the university, near the location of the standard language classes to enable easy navigation. I held all the sessions as part of an internship within my master's degree study program. Participants generally attended sessions as often as possible, with absences only occurring due to illness or other uncontrollable circumstances. Occasionally, participants faced time conflicts with their schedules, but were able to attend other sessions on the same day since the sessions had similar structures and plans. The participants were not divided by language level, making it possible for them to attend any session that was convenient for them.

In music therapy the topic of treatment manuals is a controversial and discussed topic. As music therapy is a very reflexive and flexible method as described in Bruscia's definition of music therapy (Bruscia, 2014), standardizing and defining exact ways of intervention is considered limiting of the therapeutic process. However, replicability and standardization is important for quantitative research because we need to have an understanding of the exact intervention we are measuring the effect of. Rolvsjord et al. (2005) suggest writing protocols using approaches based on therapeutic principles rather than set activities or "a fixed set of techniques". Since this intervention is not based on music therapy approaches, the need to

have a therapeutic alliance and process is not relevant and the protocol can be more specific. The guidelines for reporting of music interventions by Robb et al. (2011) are more fitting for this study. Together with the Template for Intervention Description and Replication (TIDieR) checklist (Hoffmann et al., 2014) these set the groundwork for how the intervention is reported as planned.

Key principles and theoretical groundworks I use in the interventions come from my education as a music therapist. Additionally, to having a focus on language learning and group participation as the study rationale suggests, interventions were focused on being resource-oriented, reflexive, and participatory. These are a part of the larger acronym PREPARE (which stands for: participatory, resource-oriented, ecological, performative, activist, reflective, and ethics-driven) from the theory of community music therapy (Stige & Aarø, 2011).

Each session followed a two-part structure, with a warm-up phase and a main activity. The activities in the sessions were designed to be both pre-planned and feedback-oriented, providing the participants with several options for activities during each session. The pre-planning process was informed by previous feedback received from the participants, ensuring that the activities were tailored to their preferences and needs. This approach allowed for flexibility in the sessions, with the group choosing the activity that they wanted to focus on, while also incorporating elements of structured planning.

The warm-up activities were based on choir warm-ups. The sessions started out with stretches and bodily warm-up activities for about 20 minutes, mostly in Norwegian or a mix of Norwegian and English, depending on the group's understanding of the language. These lasted a few minutes only, before going over into a vocal warm-up. The vocal warm-ups were focused on either vowels or consonants that are uncommon outside the Norwegian language, such as "æ", "ø", "å" or the Norwegian "y" and rolling "r". To clarify, during these activities the vowels were sung, and shifted notes either staying on one vowel for several notes or switching between vowels for each note that was sung. These activities made it possible to focus and linger on the sounds of the language, making it possible to practice different sounds

together and in repetition. These warm-ups enabled not only physical and vocal warming up, but also might be enabling warming up to the group socially, as some sources suggest choir singing facilitates social bonding and peer support (Kreutz, 2014; Tamplin et al., 2013).

During the second half of the sessions, the focus was on a vocabulary basis, with translating songs or writing songs in Norwegian. An activity example is singing songs with texts, which were chosen based on the participants' feedback. These included well-known Norwegian songs such as "Ja, vi elsker" (the Norwegian national anthem), "Hurra for deg" (a Norwegian birthday song), and a few lullabies. Usually the songs were requested by participants and can be categorized as participant-selected music. Additionally, participants had the opportunity to write new lyrics to familiar melodies or to create their own original songs. As the year progressed and the holiday season approached, the group also practiced Norwegian yuletide songs.

An overview of the session can help for clarity and can make it easier to replicate the sessions, and the activities in the sessions. All activities were led by me, with an example for the participants first and then repeating the activity together.

1 session (60 minutes):

1st part (about 20-25 minutes):

- Starting with physical activity and warm-up – stretching and then shaking body parts.
- Breathing exercises.
- Voice warm-up with focus on support of the voice and diaphragm.
- Voice warm-up with focus on vowels and consonants. For example: singing a scale on a "difficult" vowel (such as y) or switching between vowels that sound similar, with a focus on the "difficult" vowel (For example: one long note on A, singing one note lower on Æ, and one note lower on E). It is important to take the time on each vowel to achieve a sense of mastery.

2nd part possible activities (about 35-40 minutes, depending on the time left):

- Learning and singing a song and translating and understanding the text and adding percussion when the song is learned enough.

- Song writing: Either by lyric replacement or free composition (as in Silverman, 2010, p. 176). The participants choose a theme, then create a "word bank" together. To write verses and chorus the words from the word bank get arranged into phrases and then a verse/chorus.

All materials used during sessions are a guitar for accompanying the songs, a speaker for listening to songs when analysing the lyrics, a laptop for screening the lyrics and playing the music, printed lyrics for when songs were planned, a blackboard for writing lyrics when song writing, and small percussion instruments (shakers and a guiro) for accompaniment. All songs written during the sessions were printed out and handed out at the next session.

The sessions also encouraged participants to participate actively by playing rhythmical instruments and/or guitar, following inspiration from the methodology of melodic intonation therapy (MIT) (Albert et al., 1973). The activities in the second half of the sessions were designed to be interactive and engaging, allowing participants to connect with each other and the language through music.

2.5 Self-report Questionnaires

In this thesis, two self-report measures were used to gather data: the Sense of Belonging Instrument (SOBI) and a variation of the Common European Framework of Reference for Languages (CEFR). To administer the questionnaires, participants were sent an email with a link to an online questionnaire tool called Survey-Xact by Rambøll. The university had an agreement with Survey-Xact, so this was the best option to ensure data privacy and security. The first round of questionnaires was completed by participants within a nine-day period at the end of September, before the first session. The second round of questionnaires was completed within a two-week period at the end of November to the beginning of December, after the last session. The use of an online questionnaire tool allowed for efficient data collection and reduced the risk of data loss. By administering the questionnaires before and after the intervention, the data collected was able to capture any changes in participants' sense of belonging and language proficiency.

2.5.1 The Sense of Belonging Instrument

SOBI is a scale developed by Hagerty and Patusky in 1995, consisting of 2 parts, SOBI-P (psychological state) and SOBI-A (antecedents), together 27 items. All items in SOBI-P are scored on a 4-point Likert scale ranging from 1: Strongly disagree to 4: strongly agree. SOBI-P has 18 items, whereof 17 are reversed items. Item number 4 is *positive,* while all other items are *negative*. The negative items are reversed in the scoring, which makes the final total score positive. In SOBI-A, all items are positive. The total scores can range from 18 to 72 (SOBI-P) and from 9 to 36 (SOBI-A).

SOBI-P addresses the psychological state of being, categorized by "fit" (example: "I often wonder if there is any place on earth where I really fit in") and "valued involvement" (example: "I would like to make a difference to people or things around me, but I don't feel that what I have to offer is valued"). Part 2, SOBI-A addresses antecedents or precursors to sense of belonging. Reliability and validity have been demonstrated (Hagerty & Patusky, 1995). The SOBI questionnaire is viewable in appendix 7.

2.5.2 The Common European Framework of Reference for Languages

I did not find an established/validated instrument to measure an individual's language level. I did find a study using non-standardized methods of measuring correct pronunciation among children repeating singing phrases (Good et al., 2015), but this was not quite fitting for this study. As this was in an educational setting, availability of test results might have been a more reliant way of measuring "successful" language development. Test results were not available to this thesis, due to the very sensitive personal information in addition to these, and the nature of this master's thesis. Test results may also not be as broad of a way of measurement as one's self-perceived language level. Instead, this study relied on self-perceived language level as a means of measurement. While self-reporting is subjective and may be biased, individuals have a good understanding of their day-to-day language usage. Self-confidence, which is correlated with language learning, can also affect daily language usage (Gardner et al., 1997). Therefore, reporting a positive result in self-perceived language level can be a useful indicator of successful language development in this context. A possible

bias in this is not knowing enough of something to be aware of how little you know, as in the Dunning-Kruger effect (Dunning, 2011).

To measure the language levels of participants in this study, a suitable method needed to be found or developed. After thorough research, it was determined that repurposing the Common European Framework of Reference for languages (CEFR) as a self-report questionnaire was the most effective approach. CEFR was initially established to meet the growing need for language learning in the European Union, as member states were obligated to promote language learning and encourage communication between populations. Through article 2 of the European Cultural Convention from 1954 member states of the Council of Europe have to "encourage the study by its own nationals of the languages, history and civilisation of the other Contracting Parties" and "endeavour to promote the study of its language or languages, history and civilisation in the territory of the other Contracting Parties" (Council of Europe, 1954, p. 2). This led to common language learning objectives, which eventually led to the development and publishing of CEFR as a common ground for how we learn, teach and assess languages (Council of Europe, 2001). In 2020 this was updated with the CEFR – Companion Volume (Council of Europe, 2020). CEFR has developed to be a widely recognized tool used by language learners, educators, and employers alike, to place language levels in a common European framework, thereby promoting effective communication across borders.

CEFR is usually presented as a grid, and is a tool to help teachers, students, employers and more to place their language levels in a common European framework. In the grid, the persons filling out the grid may place themselves on a scale between A1 (lowest language proficiency) to C2 (fluent language proficiency), or 1-6, within 3 main categories (understanding, speaking, writing) divided into 5 subcategories (listening, reading; spoken interaction, spoken production; writing) for a way to measure language proficiency. The Main Categories are understanding, speaking, and writing, further divided into the subsections of listening, reading, spoken interaction, spoken production and writing. The description of the fluency levels are detailing what one can do, as for example "I can use simple phrases and sentences to describe where I live and people I know" for the lowest level of spoken production and "I can present a clear, smoothly-flowing description or argument in

a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points" for the highest level.

To transform the CEFR into a questionnaire with scoring, I turned the grid of the original 5 CEFR subcategories into 5 multiple-choice items. Each categorical line of the grid was one item, and the individual boxes in the grids with the descriptions of language proficiency were used as the response categories, without altering the text. For scoring I turned these into a Likert scale ranging from 1-6. An additional option of "none of the above" was also added for cases where language proficiency fell below A1, and this option was scored as 0. The total scores of the CEFR questionnaire can range from 0 to 30. Table 1 shows how the CEFR levels coincide with the scoring. The CEFR questionnaire is viewable in appendix 8.

Table 1. Interpretation of CEFR and scoring in this study

Score	Corresponding CEFR level
0 (min.)	none
5	A1 (Breakthrough)
10	A2 (Waystage)
15	B1 (Threshold)
20	B2 (Vantage)
25	C1 (Advanced)
30 (max.)	C2 (Mastery)

Note: A table showing how scoring of the CEFR results corresponds to the language level of CEFR.

2.6 Statistical Analysis

The statistical analysis in this thesis consists of descriptive statistics to describe properties of the sample and statistical inference to draw conclusions from the sample about the larger population (Gauthier & Hawley, 2015). In this thesis, the goal is to draw a tentative conclusion about people learning a second language, using the data from the 24 participants, and to inform a future larger trial. It's advised to have a statistical plan and finished protocol

ahead of commencing the study, but due to time limitations of the study program this was not possible.

2.6.1 Data Preparation and Software

I exported the data from the online questionnaire program onto a safe remote desktop on Microsoft Excel and via that onto RStudio.

Excel was used for data storage and sorting the data so that every questionnaire had the same sequence of participants, sorted from the lowest ID number to the highest. (Here, I discovered an error that had occurred when handing out the participant numbers, where two participants were assigned the same participant number. Only one of these participants answered the follow-up and was not possible to identify unambiguously. Therefore, the follow-up data of both had to be deleted.) For every questionnaire there was one Excel file, and each file had columns for each question. The columns were re-named, and a variable was added for the assigned group. Another variable was added for whether participants had completed the follow-up or not.

The prepared Excel files were read into RStudio. On RStudio I used three scripts (see appendix 2 for the full scripts): One for reading and preparing the data, one for the analysis of baseline variables, and one for the analysis of follow-up data. When preparing the data, I reversed the items in SOBI-P that needed reversing, gave the values in CEFR the values required for analysis (recoded value 7 to 0), and then created a combined dataset holding all the scores from each pre- and post-test. I used the second script to analyse the baseline data. The third script contains the descriptive and inferential analyses for pre and post values.

2.6.2 Main Principles

Once participants were randomized, they were all analysed, ignoring withdrawal, noncompliance, protocol deviations or other factors happening after randomization, on a Intention-to-treat (ITT) basis (Gupta, 2011). In short terms ITT is described as "once randomized, always analysed" (Kruse et al., 2002). In this project this means that all

participants who were assigned to the music intervention were analysed as such, regardless of whether they received the intervention or not. The participants who were in the control group were also analysed as such, and factors such as dropping out of the language program had no impact on their being included in the analysis. Intention-to-treat is recommended by guidelines for clinical trials, such as the CONSORT-statement (Moher et al., 2010; Schulz et al., 2010). All participants who had follow-up data were analysed, however participants who were lost to follow-up and had missing data points could not be included and were excluded in the analysis of outcomes.

In principle it is recommended to mask the statistician as to which group is the treatment group vs the control group (Gold, 2015). Unfortunately, this was not possible, due to the nature of the study having only one investigator. I did the statistical analysis along with almost everything else (except for randomization, which my supervisor did for allocation concealment), and was not blinded or masked.

I used a two-sided 5% significance level and reported estimates of effects with 95% confidence intervals (CIs). Confidence Intervals provide a margin of error, providing a frame which might include the population mean (Stinerock, 2018). Larger sample sizes have a higher chance of including the population mean. A 95% confidence interval has a good chance to include the population mean based on our data. Higher confidence intervals have a higher chance of including the means of the population in general and would rise in accuracy, but would be less precise (Crash Course, 2018). The significance levels are two-sided because that could show whether treatment is harmful or beneficial. If the results fall within the two sides, that will indicate evidence against the null-hypothesis.

2.6.3 Descriptive Analysis

The study employed a range of methods to compare groups for primary and secondary outcomes. Initially, a descriptive analysis of Mean (M) and standard deviation (SD) and percentages (n(%)) was conducted and compared to the whole subset of participants, which were then presented side by side in a table. Boxplots were also created to visualize the

numbers for each test and group, either connected to the assigned groups or to whether they completed the follow-up.

However, dropouts from the study may introduce attrition bias and potentially skew the data. Attrition bias is the selection bias that occurs through the differences of the participants who drop out of the study (Nunan et al., 2018).To address this issue, the researchers performed a Fisher's exact test for each variable in the background table, comparing not the groups but whether participants were lost to follow-up or not. This test helped to determine whether completers differed systematically from those who dropped out and to check for potential attrition bias. The Fisher's exact test calculates the possibility that the observed differences between the two groups are due to random chance, providing a p-value that can inform further analysis and interpretation of the results.

2.6.4 Inferential Analysis

To determine the impact of both the assigned group and baseline scores on post-test results, an analysis of covariance (ANCOVA) was conducted at a two-sided 5% significance level to test the null hypothesis of no difference between groups. ANCOVA is a statistical method that considers several factors that could influence the results when measuring effects. ANCOVA utilizes a continuous dependent variable, which in this case is the outcome at post-test, a covariate, which is the baseline score of the outcome, and a categorical independent variable, which is the assigned group. The covariate helps to account for any potential differences in the baseline scores between the groups, which could affect the outcome. The measurement must be between groups and not within, to see a real effect comparison (Gold, 2015). ANCOVA differs from a t-test, which compares means of two groups, or an analysis of variance (ANOVA), which compares means across multiple groups (Mishra et al., 2019). This approach is more powerful and provides a better estimate of the effects because it compares the mean after considering several factors that might influence the result.

While the study also conducted an independent sample t-test for comparison and better understanding of the numbers, the ANCOVA was interpreted as the primary result.

Additionally, to facilitate comparison and better understanding of the findings and the process of analysing, an independent sample t-test was also conducted and reported, although the ANCOVA was interpreted as the primary result.

3 Results

This section includes the results of the study. The results are stated clearly and show what the results were, without discussing meaning, as that is the purpose of the section containing the discussion.

3.1 Flow of Participants

Figure 2 illustrates the participant flow throughout the 9-week experimental period. Of 24 individuals initially assessed for eligibility, all were randomized and assigned to either the group receiving group musicking or to the group only receiving standard language classes. The participants, once allocated, were placed into groups based on the time that fit their schedule. The musicking group included 13 participants, while the group getting the standard language classes had 11, a slightly uneven allocation resulting from the randomization. During the follow-up period, 6 of the 13 participants (46%) in the musicking group were lost to follow-up due to not completing the questionnaires, while only 2 of the 11 (18%) in the standard group were lost to follow-up. The lost participants in the musicking group may have changed their contact information, as they could not be reached. There was a pronounced difference between the number of participants in each group who completed the post-test. In the musicking group, only 54% (7 out of 13) responded to the post-test, whereas in the control group 82% (9 out of 11) provided post-test data. This discrepancy is explored further later.

Figure 2. Flow chart

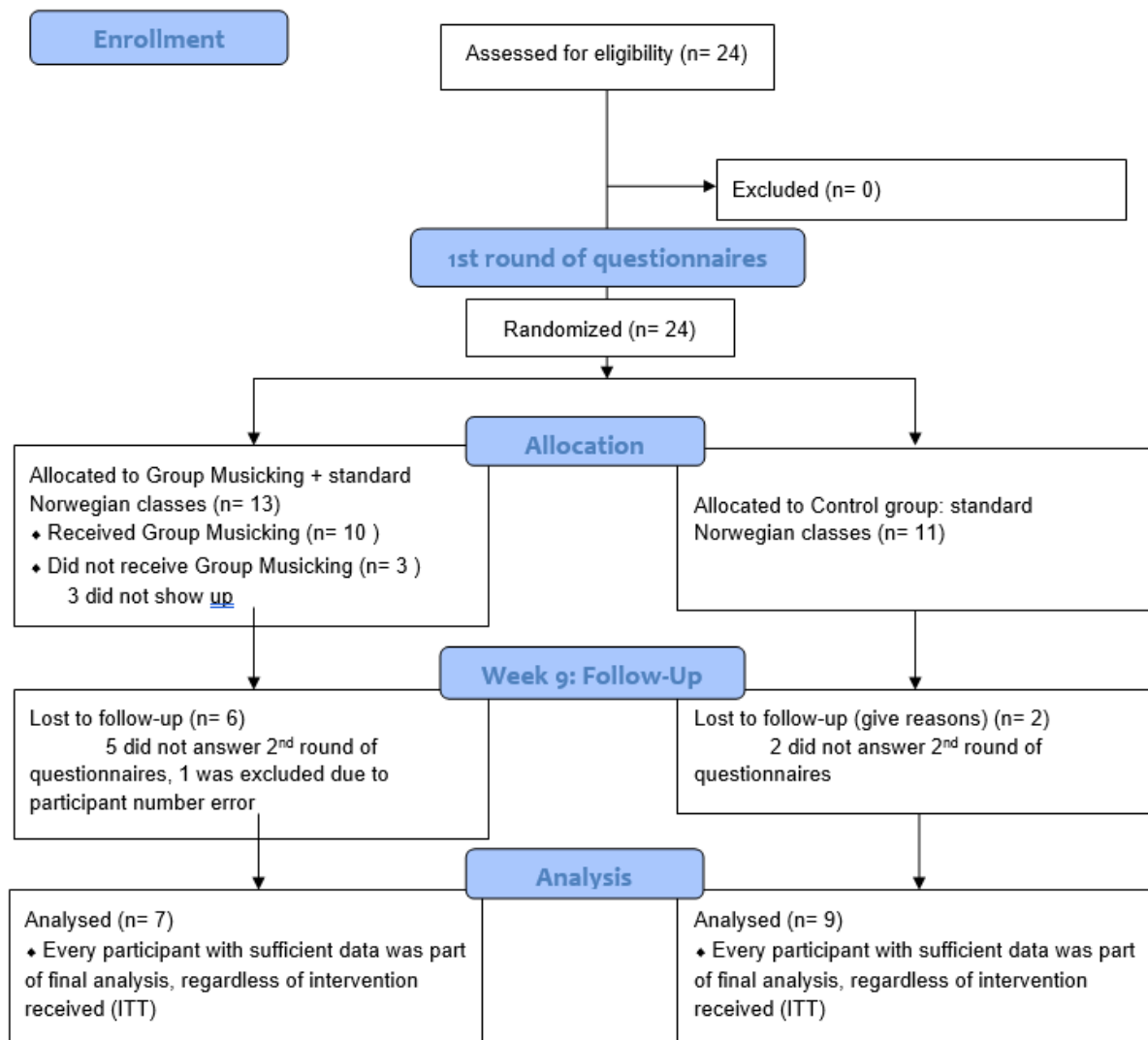*Figure 1: Flow chart reporting enrolment, allocation, follow-up, and analysis of participants. ITT=intention-to-treat. CONSORT 2010*

As all participants with available data were analyzed based on the intention-to-treat principle, the final sample analyzed consisted of seven in the musicking group and nine in the standard group. Overall, 16 out of 24 participants (67%) who were initially randomized were included in the analysis.

Table 2 presents the baseline characteristics of all study participants. Most participants, around 60%, were female and all were aged between 18 and 55, with the largest age group being between 35 and 45 years (38%). Most participants, also around 60%, are from Europe outside Scandinavia. The values of demographic characteristics were similar between the two groups. None of the participants reported being purely monolingual, but most reported being able to speak one language fluently and several non-fluently. There were slightly more participants in the category of one fluent and several non-fluent languages in the group doing group musicking sessions.

There are six variables of reasons for learning. Participants could report several reasons as their motivation and report a potential reason that falls outside the given categories in "other". The "other" reported reasons include two potential long stays in Norway, one who wished to be able to talk with a family member, and one who learned out of courtesy and interest. In the category of "long intended stay" with 9 participants in the group doing musicking sessions and 5 in the group getting only standard language classes a small random error or uneven number can be seen.

The last category about the length of stay in Norway showed that most participants had either lived in Norway for less than 3 months or more than one year. In this category, the group doing group musicking had slightly more participants who had lived in Norway for more than a year than the group getting only standard language classes. Variations that differ in the participant groups, such as those in the last three categories, which happen after randomization, are not unlikely to occur with numerous variables and small participant groups, as just a few participants can significantly affect the demographic characteristics. This is one reason why a larger sample size might be beneficial, as a larger population size would reduce the probability of random errors (Gold, 2015). Overall, and considering the small sample size, the values of demographic characteristics are similar between the two groups.

Table 2. Baseline characteristics

Variable	All (N=24)	Music (N=13)	Standard (N=11)
	n (%)	n (%)	n (%)
Gender			
female	15 (63)	8 (62)	7 (64)
male	9 (38)	5 (38)	4 (36)
other	0 (0)	0 (0)	0 (0)
Age group			
18-25	7 (29)	4 (31)	3 (27)
25-35	5 (21)	3 (23)	2 (18)
35-45	9 (38)	4 (31)	5 (45)
45-55	3 (13)	2 (15)	1 (9)
55-65	0 (0)	0 (0)	0 (0)
Older	0 (0)	0 (0)	0 (0)
NA	0 (0)	0 (0)	0 (0)
Area of origin			
Scandinavia	0 (0)	0 (0)	0 (0)
Europe outside Scandinavia	15 (63)	8 (62)	7 (64)
Asia	3 (13)	2 (15)	1 (9)
N. America	3 (13)	1 (8)	2 (18)
S. America	1 (4)	0 (0)	1 (9)
Africa	2 (8)	2 (15)	0 (0)
Oceania	0 (0)	0 (0)	0 (0)
NA	0 (0)	0 (0)	0 (0)
Languages spoken			
One	0 (0)	0 (0)	0 (0)
One fluent + several non-fluent	13 (54)	8 (62)	5 (45)
Several fluent	3 (13)	2 (15)	1 (9)
Several fluent + not fluent	8 (33)	3 (23)	5 (45)
NA	0 (0)	0 (0)	0 (0)

Reasons for learning[1]			
Just for fun	5 (21)	2 (15)	3 (27)
Long intended stay	14 (58)	9 (69)	5 (45)
Communicate at work/study	5 (21)	2 (15)	3 (27)
Communicate socially	14 (58)	7 (54)	7 (63)
Compulsory	1 (4)	0 (0)	1 (9)
Other	4 (17)	2 (15)	2 (18)
Length in Norway			
<1 month	0 (0)	0 (0)	0 (0)
<3 months	10 (42)	6 (46)	4 (36)
<6 months	1 (4)	0 (0)	1 (9)
<12 months	4 (17)	1 (8)	3 (27)
>one year	9 (38)	6 (46)	3 (27)
NA	0 (0)	0 (0)	0 (0)

[1]Participants could select several options.

Although table 3 may appear similar to the previous table, it presents different information. Instead of comparing the demographic characteristics of the group musicking participants with those who solely participated in standard language classes, it compares the demographics of the participants who completed the second questionnaire with those who did not. Since a substantial number of participants withdrew from the study, I sought to identify any possible common factors or trends among them. We can observe that the numbers are not consistently uniform, but the discrepancies which are there can be attributed to random error. We do see that for example participants giving "I intend to stay in Norway for a long time" were more likely to drop out of the study than others. The participants having stayed in Norway for more than a year before the study show the same tendency. To determine whether the differences are statistically significant, I performed a Fisher's exact test to calculate a p-value. The p-value indicates that none of the differences are statistically significant, with only a few instances remotely approaching significance. For example, all participants who had stated that they were learning the language "just for fun" completed both rounds of questionnaires. Additionally, nearly all participants who had lived in Norway for less than three months before the first session also completed both rounds of questionnaires.

Table 3. Baseline characteristics of those who completed versus those who dropped out

Variable	Completed (N=16)	Dropped out (N=8)	p-value
	n (%)[1]	n (%)[1]	
Gender			0.657
female	9 (60)	6 (40)	
male	7(78)	2 (22)	
other	0	0	
Age group			0.868
18-25	5 (71)	2 (29)	
25-35	4 (80)	1 (20)	
35-45	5 (67)	4 (44)	
45-55	2 (67)	1 (33)	
55-65	0	0	
Older	0	0	
NA	0	0	
Area of origin			0.324
Scandinavia	0	0	
Europe outside Scandinavia	11 (73)	4 (27)	
Asia	2 (66)	1 (33)	
N. America	2 (66)	1 (33)	
S. America	1 (100)	0 (0)	
Africa	0 (0)	2 (100)	
Oceania	0	0	
NA	0	0	
Languages spoken			0.360
One	0	0	
One fluent, several not fluent	10 (77)	3 (23)	
Several fluent	1 (33)	2 (67)	
Several fluent + not fluent	5 (63)	3 (38)	
NA5	0	0	

Reasons for learning[1]			
Just for fun	5 (100)	0 (0)	0.130
Long intended stay	9 (64)	5 (36)	1
Communicate at work/study	4 (80)	1 (20)	0.631
Communicate socially	11 (79)	3 (21)	0.204
Compulsory	1 (100)	0 (0)	1
Other	3 (75)	1 (25)	1
Length in Norway			0.126
<1 month	0	0	
<3 months	9 (90)	1 (10)	
<6 months	0 (0)	1 (100)	
<12 months	2 (50)	2 (50)	
>one year	5 (56)	4 (44)	
NA	0	0	
Group			0.211
M	7 (54)	6 (46)	
S	9 (82)	2 (18)	

Note. P-values are based on Fisher's exact test. [1]Participants could choose multiple items, so a p-value is calculated for each item. [1]Percentages are of participants within group.

Another big difference to see in the comparison of the participants who dropped out vs the participants who did not, is how long participants had lived in Norway before the baseline was set. Participants who had lived in Norway for longer than three months, are relatively evenly divided between the set who dropped out vs. the set who completed the study and the sessions. However, almost all participants that had lived in Norway less than 3 months completed the study.

3.2 Effects of Interventions

To analyse the effects of the interventions we will look at the observed values or descriptive statistics, including a calculation of means, standard deviation and changes and inferential statistics including results from an ANCOVA calculation and a t-test.

3.2.1 Observed Values

Table 4 shows the descriptive statistics – observed mean and standard deviation of the SOBI-P, SOBI-A and CEFR at the pre-test and the post-test.

The initial mean values of SOBI-P are comparable between the group doing music sessions and the group receiving standard language classes only during the pre-test stage. The post-test results reveal that the mean SOBI-P score in the music-group is lower than that in the group receiving standard language classes, indicating that the sense of belonging was worse among those who participated in musicking. Despite this, the difference in mean scores between the two groups is not substantial, with the standard deviation being around 10-11 points, while the mean hovers around 53. It's worth noting that a large standard deviation indicates a significant degree of variation between individual scores, which could potentially impact the interpretation of the results. Given the relatively small differences in mean scores and the high degree of variability observed, it is essential to exercise caution when interpreting the findings.

Alike the values from SOBI-P, the mean scores for SOBI-A in both groups were similar during the pre-test phase. In the post-test phase, the mean score for the group doing group musicking sessions was lower than the mean score of the group receiving standard language classes. The differences in the mean scores between the two groups were relatively small, with changes of less than 2 points up or down from the pre-test phase. Again, the standard deviation was relatively large, which is notable given the range of scores between 25 and 30. In particular, the post-test scores for the group receiving standard language classes had a large standard deviation, indicating that the individual scores were quite spread out around the mean.

Table 4. Observed values of scores before and after the interventions

Variable	All		Music		Standard	
	n	M (SD)	n	M (SD)	n	M (SD)
SOBI-P						
pre	23	52.74 (9.66)	12[1]	52.92 (9.72)	11	52.55 (10.06)
post	16	53.31 (10.36)	7	51.71 (10.61)	9	54.56 (10.63)
SOBI-A						
pre	23	28.48 (2.84)	12[1]	28.33 (3.31)	11	28.64 (2.38)
post	16	28.38 (3.9)	7	26.71 (2.56)	9	29.67 (4.39)
CEFR						
pre	24	10.96 (4.35)	13	10.85 (4.3)	11	11.09 (4.61)
post	16	13.44 (4.23)	7	11.14 (4.78)	9	15.22 (2.86)

Note. [1] One participant only responded to CEFR in round one and not SOBI. n = number of participants in this group, M = mean, SD = standard deviation.

The results of the CEFR self-reported language levels indicate that at the pre-test, the group doing music sessions reported slightly lower language levels compared to the group receiving standard language classes. However, both groups show an increase in language levels from pre to post-test, with the group doing group musicking sessions showing a smaller increase of only a few decimals, while the group only receiving standard language classes shows a larger increase of 4 points. Despite these mean differences, it is important to note that the standard deviation is quite large, indicating a wide variation in individual responses. This means that there is a considerable amount of variability within each group, and the mean differences between groups may not be statistically significant. It is also worth noting that self-reported language levels may not necessarily reflect actual language proficiency, as they are subjective and can be influenced by factors such as self-perception and confidence.

To examine the data set in more detail, we can view the pre- and post-values and the change between them individually, as shown in Table 5. I am doing this in addition to analyse the means of the whole group above, to better understand how data sets are presented and what

Table 5. Table showing pre- and post-values and the change between them.

		SOBI-P			SOBI-A			CEFR		
		Pre	Post	Change	Pre	Post	Change	Pre	Post	Change
1	m	59	58	-1	29	29	0	6	4	-2
2	m	43	46	3	26	26	0	4	12	8
4	m	37	38	1	28	23	-5	5	5	0
5	m	46	54	8	25	26	1	10	12	2
7	m	52	NA	NA	25	NA	NA	16	NA	NA
10	m	NA	NA	NA	NA	NA	NA	7	NA	NA
12	m	61	NA	NA	36	NA	NA	11	NA	NA
13	m	64	NA	NA	30	NA	NA	15	NA	NA
14	m	58	NA	NA	27	NA	NA	17	NA	NA
17	m	41	40	-1	25	26	1	14	15	1
19	m	68	66	-2	32	31	-1	14	14	0
21	m	56	60	4	27	26	-1	11	16	5
24	m	50	NA	NA	30	NA	NA	11	NA	NA
3	s	70	69	-1	32	36	4	13	13	0
6	s	50	52	2	29	30	1	9	13	4
8	s	66	71	5	32	36	4	17	21	4
9	s	53	53	0	28	26	-2	14	13	-1
11	s	53	57	4	28	30	2	6	13	7
15	s	56	NA	NA	27	NA	NA	5	NA	NA
16	s	37	41	4	25	23	-2	19	17	-2
18	s	53	NA	NA	27	NA	NA	7	NA	NA
20	s	38	43	5	30	31	1	10	15	5
22	s	46	46	0	31	29	-2	8	18	10
23	s	56	59	3	26	26	0	14	14	0

Note. Table showing individual scores and the changes between them as part of the larger data set. m – musicking group, s – standard language classes only group, NA – no data available.

[... continued from previous page]

the observed values mean in practice. Table 5 first shows the scores of participants who received group musicking, and then the scores of the participants who only received standard language classes. The table highlights the large proportion of data missing, especially from the participants who only received group musicking. This very large proportion of missing data points is an issue which leads to high uncertainty of all conclusions that can be drawn from the collected data.

The table shows us that most changes in each category are very little, except for a few outliers. In SOBI-P for example we see that most changes are at about 4 or 5 data points or below, except for a few outliers. These are small changes considering the mean scores at around 50. Most changes are positive, and only a data points have no change or are slightly negative. In table 4 it was possible to see these as part of the mean, as the changes in SOBI-P are mostly positive, and there are a few more negative changes in the group receiving the group musicking sessions. In SOBI-A the changes are almost all at 2 data points or less, remaining mostly consistent except for a few outliers. The outliers in the group only receiving standard language classes are positive, and the outlier in the group receiving group musicking sessions is negative. In CEFR, the language measure, the numbers vary a bit more. A few participants report that they improved their language skills quite substantially, while others report that their language levels even changed negatively.

Figure 3 presents boxplots as a graphical representation of the observed values of the study. The boxplots provide a visual representation of the data distribution and allow for the detection of outliers and skewness, which may be obscured in other types of graphs. These boxplots depict the distribution of the data using the median as a central tendency measure, the box as the middle 50%, the lines as the quartiles furthest from the median, and the dots as outliers. The pre-test numbers exhibit a relatively even distribution, which is typical of a normal numerical vector. However, the post-test boxplots display significant changes and deviate from the usual boxplot appearance. For instance, the SOBI-A post-test boxplot for group m demonstrates a median that is at the same position as the lower 50%, with the exception of an outlier. This deviation can be attributed to the low number of participants in the group receiving group musicking sessions, where four participants account for more than 50% of the sample. A similar pattern is observed in the group that received standard language classes in the CEFR post-test, which has a slightly higher number of participants.

Figure 3. Boxplots showing a visual representation of observed values.
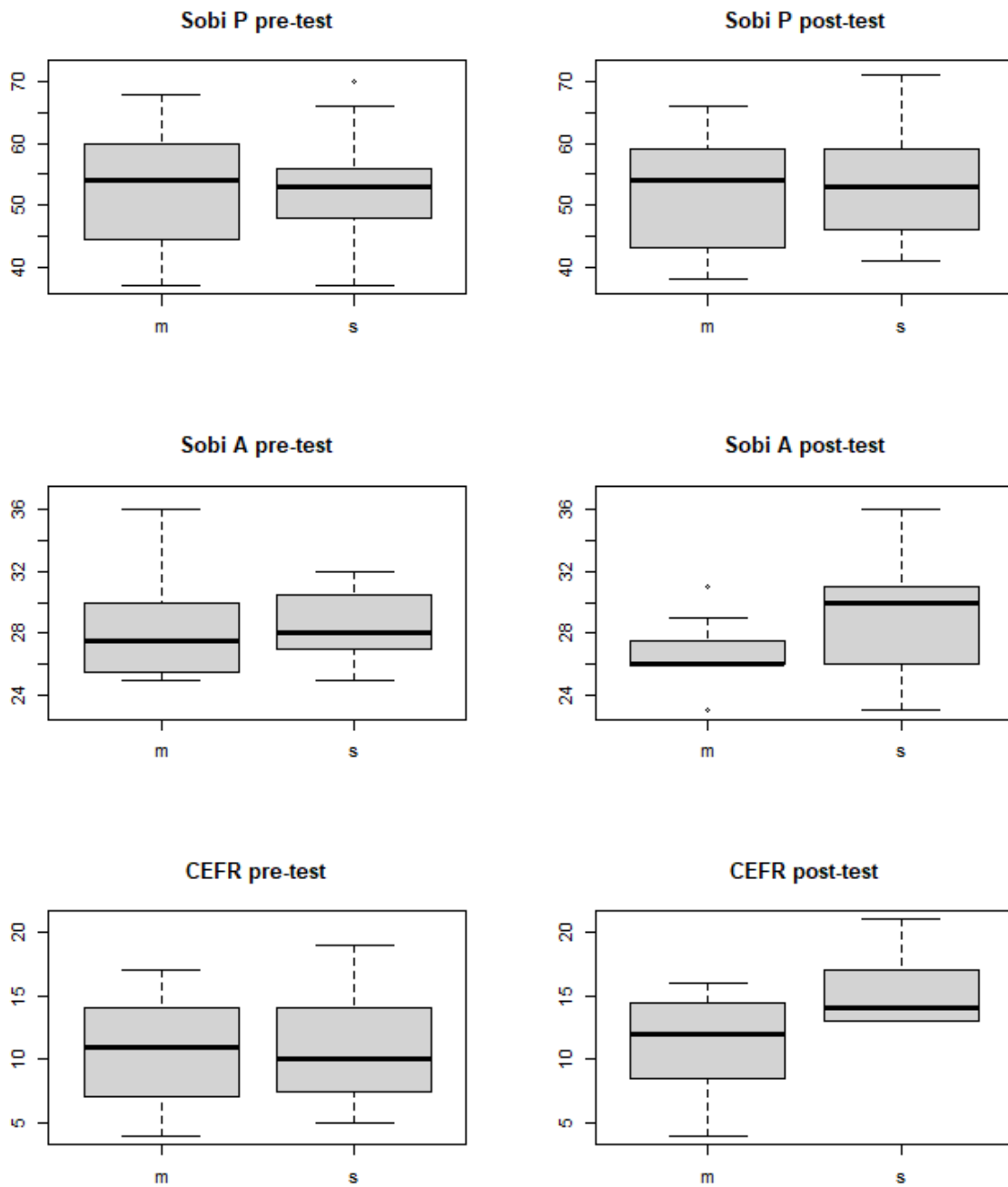
Note. Boxplots showing the median, quartiles and outliers of the scores of participants. m = group receiving group musicking sessions, s = group receiving only standard language classes.

3.2.2 Statistical Inference: ANCOVAs

The effect estimates from the ANCOVAs with 95% confidence intervals and p-values are presented in Table 6 (the full ANCOVA tables are in Appendix 3). The Table shows the mean difference of the scores of SOBI-P, SOBI-A and CEFR, with 95% confidence intervals. The mean differences are all relatively small and in the negatives. with a Confidence interval ranging from negatives to positive. The numbers being in the negatives means that the group receiving group musicking had a lower outcome.

There are no significant p-values among the results, with the least significant p-value being in SOBI-P at 0.541. The difference in SOBI-A has a p-value of 0.393, which is not significant. Although the p-value in CEFR approaches a significant value at 0.175, it is still far away from a significance. Significant values range from 0.1 to 0.001, but a common p-value determined as significant is 0.05 (Andrade, 2019; Cesana, 2018; Lakens, 2015). However in more recent times, the P-value has become more controversial, with some papers disallowing the publishing of a p-value, as it is "too easy to pass" as a test (Woolston, 2015). The P-value shows the probability value of the results and their distribution within the group, giving us a value to judge the results from. In ANCOVA, the value is calculated with respect to a factor that might affect the results, in this case the data from the pre-test.

Table 6. ANCOVA estimates

Variable	Mean difference (95% CI)	p-value
SOBI-P	-0.90 (-3.73, 1.92)	0.541
SOBI-A	-1.07 (-3.45, 1.31)	0.393
CEFR	-2.53 (-5.97, 0.92)	0.175

Note. The effect estimates from ANCOVA.

For comparison and learning, the effect estimates from a t-test calculation is presented in Table 7. The mean differences presented are larger than the mean differences from the ANCOVAs, and the confidence intervals are further apart. The p-value is less significant in the SOBI-P results, but slightly more significant in the SOBI-A and CEFR results. In CEFR the p-value approaches significance, which would be evidence against the null-hypothesis (group musicking has no effect on second language acquisition) and could in that hypothetical case (if significant and if a t-test was the main outcome) have suggested that group musicking has a negative effect on second language acquisition as measured.

Table 7. T-test estimates

Variable	Mean difference (95% CI)	p-value
SOBI-P	-2.84 (-14.40, 8.71)	0.604
SOBI-A	-2.95 (-6.74, 0.83)	0.116
CEFR	-4.08 (-8.68, 0.52)	0.076

3.3 Experiences with Conducting the Interventions

Not every participant assigned to group musicking was able to participate in all sessions. One session (the last session) was cancelled due to illness. A few participants did not show at all, and most times participants did not show it was due to illness or other uncontrollable circumstances. Some of the participants also started a week later due to the continuous participant recruitment and randomization process. Participation in sessions ranged from 0 sessions to 8, with a mean of 4.08 (SD = 3.12, Mdn = 5). Three participants did not show up, and most of the ones that did show up showed incentive to participate in all sessions, which can be seen in the numbers.

Participants had showed eagerness to participate, but some participants got ill a few times, and a few reported they had to look after their children during an ongoing kindergarten-strike. This matches the age group with the most dropouts (highest drop-out rate [50%] among those aged 35-45; [Table 3] who may be most likely to have small children) and is possibly a contributing factor in the attrition of the study. Some participants reported that they had dropped out of the language classes. No data is available as to exactly how many dropped out of the language course entirely. Very little contact was kept with participants not coming to the music sessions, except for follow-up of follow-up e-mails to the participants assigned to music sessions.

Since some participants did not show up, the session group size of 3-4 participants was not a good group size. When participants were ill, this was very noticeable and made the groups feel "unsteady".

As the intervention was new and created for this study alone, feedback on the intervention is important for potentially conducting a future full-sized RCT. Narrative feedback from the participants in the group musicking include:

- positive feedback on the warm-ups; participants reported that it was nice to practice pronunciation of difficult or new sounds regularly.
- It appeared easier to remember words for longer when they are a part of a learned song.
- Using rhythmic instruments while singing in the language being learned was useful, as "there was no time to think about the words" and it was surprising when the words came automatically in these instances.
- During the experimental time I had hurt both of my hands and could not play instruments as both my hands were bandaged. In these sessions participants had the main instruments. Narrative feedback on this was that it was "useful to not have time to think about the words", "it was challenging" and "it was fun". This was a paradoxical positive result of my injury.

Otherwise, the sessions were conducted as planned.

3.4 Adverse Events

One of the participants in the control group had a non-related adverse event (a sports accident).

None of the participants reported any study-related adverse events. It is still important to mention potential adverse results and a need to report possible adverse is identified (Geretsegger et al., 2017; Silverman et al., 2020). Possible harms with musicking include music induced harms such as listed in Silverman et al. (2020). This includes harms like hearing damage, negatively impacted cognitive tasks, re-experienced trauma, music-induced cravings in people with substance use disorders, mood worsening and even social isolation and group rumination.

4 Discussion

Though none of the effect results are significant, the results leave a lot of room for discussion. How do the results fit together with previous research? What conclusions *can* be drawn from a "non-result"? What do these results and the study mean for the future? How could the study have been conducted better? How could the results have been better, or what would have been necessary for the results to be significant? Questions like these arise from the results of the study, and this discussion section will try to address and answer these questions.

4.1 Findings

None of the findings are mathematically significant. Though the general means of the group receiving standard language classes are more positive than the general means of the group receiving sessions of group musicking in addition to the language classes, the individual scores are mostly similar apart from a few outliers. It is difficult to know if this is random, or if these numbers are direct results of the intervention, especially with a small sample size as in this study. The sessions only lasted 9 weeks, which may not be enough to see a big change in language acquisition, considering how long it takes to learn a language. Since there is a lot of previous research showing positive outcomes (such as Good et al., 2015; Piri, 2019) it is likely that it is quite random, but more research needs to be conducted. The effect results show this as part of the p-value, which is not significant in any of the calculations.

Likely the main factor affecting the outcome is the attrition. Since almost half of the group music session participant had no follow-up data, the attrition most likely affects the outcome numbers and outcome means. Comparing the full set of baseline data to the full set of outcomes is not the same data set, and it is difficult to draw any conclusions at all based on this comparison. Based on that the participants who had lived in Norway longer who were more likely to have a social network and language knowledge, had a bigger percentage who dropped out, it is highly likely that there is an attrition bias. The dropouts were also more in the group who got group musicking, which makes it more likely that the attrition bias affects

the group musicking participants. It also adds another data point to discuss, which is that the sessions appear to be best for people who have moved more recently. There could be several reasons for the high attrition. One could be that the musicking sessions simply were not appealing, may it be due to the personal style of the person holding the intervention, or that it was not motivating as it might not have been challenging enough to such highly educated people. The attrition could also have been of time limitations of the participants, since both groups went to the regular Norwegian Course.

This session design and the measured items show little outcome. However, the focus of the sessions and the items measured don't really coincide. This means that this exact design of sessions might not be affecting sense of belonging and size of vocabulary in second language acquisition. Narrative feedback from the participants however showed an appreciation of practicing the sounds of the language, implicating a high risk of a type two error. A type two error in an RCT is the assumption that there is no effect from an intervention when there is no effect measured. Narrative feedback showed that participants appreciated the pronunciation warm-ups, which is consistent with previous research by Good et al. (2015). The measures used in this study did not measure pronunciation. To find out if there is an effect from the intervention, and what effects there possibly are, a lot more research needs to be conducted.

Since the target group of "adult migrants" has such a large variation, having a very small sample size from one educational institution reflects only a very small portion of the target group. This group of participants was recruited from a university language course, and as such has a higher degree of education than average. This again creates some limitation, but adds to already previous research done on migrants, that mostly has included migrants without higher education (Chou, 2007; Grünigen et al., 2010; Henderson et al., 2017; Sæther, 2008).

4.2 Strengths and Weaknesses

The present study has a small sample size, which may limit the generalizability of the findings. While some researchers have suggested that a sample size of 12 participants per

group is sufficient for a pilot study (Cocks & Torgerson, 2013), the high attrition rate in the present study raises concerns about the validity of any conclusions drawn. The purpose of a pilot study is to estimate effect sizes and calculate sample sizes for future full-scale randomized controlled trials (RCTs); however, with such a high attrition rate, this is not possible. Additionally, the uneven distribution of attrition among the groups introduces potential bias in the outcomes. The study faced challenges in recruitment due to limited availability and a busy target group, which resulted in a lower-than-anticipated number of participants. Despite these limitations, the study provides important insights into the feasibility and potential benefits of group musicking as a language learning intervention. Further research with larger and more diverse samples is needed to confirm these findings and establish the effectiveness of group musicking as a language learning intervention.

The attrition rate observed in the study was high, and although some of the reasons for this were unavoidable, there were also some factors that could have been addressed. For example, it was challenging to reach some of the participants for the second round of questionnaires, as they may not have checked their university email if they were no longer affiliated with the institution. Additionally, some participants were only present for one semester, and they may have already completed their courses or moved away, making it difficult to follow up with them. However, there are some potential solutions that could have been implemented to reduce attrition rates. Approaches could have been to arrange a meet-up for participants, which may have incentivized them to fill out the second round of questionnaires, or to have several ways to contact the participants. This could have been particularly effective for participants who were only present for a short period. Additionally, providing participants with gentle reminders or follow-up emails may have helped to keep them engaged and motivated to complete the study. The participants were also very busy, maybe an expected feature of the recruited group of professors and university students. Overall, while some attrition is expected in any study, efforts should be made to minimize it, particularly in cases where it may impact the ability to draw valid conclusions or estimate sample sizes for future research.

The study was conducted as a part of a master's thesis, which means that only one person was responsible for conducting the entire study, except for the language classes that were held

independently. Due to the limited resources, it was not possible to have several conductors of the intervention, recruiters, statisticians, or analysts, which created challenges in minimizing bias. As a result, blinding was not possible, and there were only a few practical measures that could be taken to reduce potential sources of bias. This also made the study more vulnerable to external factors, which is evidenced by one session being cancelled due to illness. Having multiple conductors of interventions and analysts in a study can increase the study's overall robustness and decrease the risk of bias. With more people involved in the study, there are more checks and balances in place, which can help to ensure that the results are reliable and accurate. However, it is also important to balance this with the need for consistency in the study design and implementation.

The study is not registered prospectively in a trial registry, such as Clinicaltrials.gov, which is used most often (Zarin et al., 2017). It will be published in BORA UiB, the university-wide publishing tool, but it does not have a registration number. The protocol was also not fully developed, there was no full statistical analysis plan before the study commenced, and none of these are published or viewable. Trial registration and published protocols ahead of commencing the trial is recommended and essential for researchers to verify the conducted study against its original plan (Gold, 2019).

The internal reliability of the measures Is also a limitation of the study. Though SOBI-P, which is the main result of this study, the SOBI-A validity and consistency is not as supported as SOBI-P and Hagerty and Patusky had initially suggested to develop SOBI-A further to enhance it (Hagerty & Patusky, 1995). The third measure, CEFR, was transformed from a categorizing grid (Council of Europe, 2001) to a questionnaire by me for the purpose of this study. It had no usual way of scoring it for a scientific study before. Therefore, no validity or consistency was tested for it. Though it is used internationally as a way of categorizing language proficiency, it might not pick up on other parts of the language that might have been affected by the intervention. Should this be developed further as a measure or questionnaire to be used in further research, I recommend adding parts or questions for pronunciation, the actual amount of usage of the language, and a part for self-efficacy for language-usage.

Due to a limited amount of statistical knowledge and time restrictions only few statistical calculations were done. I did not conduct any sensitivity analyses for missing data, such as last observation carried forward (LOCF) (Streiner, 2014) or multiple imputation (MI) (Lydersen, 2022). With a large attrition this could have added important data points to discuss.

A strength of the study is that the results are all available and reproducible. As such, the limited statistical knowledge does not have to remain a limitation, as all data is available for further analyzation, and to reproduce the results from this thesis. The individual results are shared completely anonymously as part of a larger data set in Results (Table 5), and the scripts to reproduce the results are shared in Appendix 2.

Further small details that could be done better are about the stratifying factor and wording in the questionnaires. I decided to use gender as a stratifying factor when randomizing in this case because it seemed intuitive and because I had a very short time frame to decide what should be picked (less than 5 minutes), however in hindsight I would suggest using a factor of what motivation lies behind learning the language as a future stratifying factor. This is based on the result of participants with different motivational factors to learn the language dropping out of the study. Other possible future stratifying factors that could be better than gender is either language level, education level or area of origin.

The questionnaires were created within a very short timeframe due to strict time limitations and had a few unclear wordings. With more time the questionnaires could have been worded more clearly and given more precise data. This is especially the case with the background data questionnaire (as seen in Appendix 9), which includes the overlapping age groups (18-25, 25-35 etc instead of <18-25 or 18-<25). Given enough time to review the questionnaires, this could have been easily fixed.

In all, the study has very many limitations due to it being a master's thesis conducted over a very short time frame, by only one person in a new setting. This leads to a very limited

amount of available data and a high risk of bias. As this study was a pilot-RCT however, the goal is mainly not to achieve reliable effect results, but also to assess the feasibility of the design, intervention, and measures.

4.3 Implications for Practice

Considering that migration is a global challenge, finding a cost efficient, sustainable, user friendly and motivating way of social integration and supporting this large part of the population is a solution currently needed in large parts of the world. Music therapy, when implemented right, could be part of this solution (Krüger & Diaz, 2023). With a view on social integration and community like in CoMT, while potentially providing a practical arena for learning a language group musicking could be a perfect addition to the existing system that exists today. Though the results in this study are all statistically insignificant, the narrative feedback shows that the intervention is not completely out of place. Interventions should be improved, and having more of a therapeutic focus with a community perspective as a CoMT background would support would be beneficial. A successful implementation of musicking with the target group with a focus on group building, social belonging and language learning could be a cost-effective way to promote effective inclusion of the population group.

Considering that most participants who had lived three months or less in Norway before the beginning of the study completed the study, perhaps the design of the sessions is most suitable for this group. In contrast, around half of the participants that had stayed in Norway longer than three months before the study started dropped out of the study, indicating less interest in the sessions or the research on them.

4.4 Implications for Future Research

As this is a data set analysed within a lot of time limitations, a lot of further possible analyses can be done with the existing dataset. For this dataset a few further possible analyses can be done. A reliability analysis of the CEFR scale could provide insight if the scores across the

CEFR scale are even. An analysis of individual CEFR items to see if effect scores are different for different categories within the CEFR scale, as CEFR levels are normally defined within each category (Council of Europe, 2001). Sensitivity analysis could also provide important insight to the robustness of the results (Thabane et al., 2013). Though some data is available, the individual background data is not available for anonymity reason.

This study lays a groundwork for future full-size RCTs. For this, this exploratory study can provide an insight to the feasibility of the study design, the measures used and the interventions of the study. For a full-size study a few changes can recommended, along with the obvious requirements of having a bigger sample size and more researchers or contributors to the study. Several persons performing interventions are recommended, to make sure that personal style is not an influencing factor. The study design could benefit from having a second comparison group which does for example group reading such as used in the MIDDEL project (Rasing et al., 2022), as a comparator of the musicking factor in the study. Based on the background info on the participants who dropped out (in Table 3), a stratifying factor of motivation for learning the language or of age group might be a better factor than gender, as dropouts appeared to correlate with different motivational factors. Other future research could also provide further insight, such as a mixed-methods study combining interviews of participants in depth with numerical values to optimize the intervention design, as another pre-cursor to a full-size RCT. Qualitative data could ensure that further quantitative research is accurate and measuring precise data. Interviewing participants could identify further outcomes that could be measured, to avoid Type 2 errors. Thematic analysis could further support intervention development.

The measures used in this study could undergo more reliability testing, and a few outcomes could be measured in addition to sense of belonging and language level. Potential other outcomes that could be measured include self-confidence or self-efficacy in using the language, as that could affect both language usage and motivation for learning (Bandura, 1977). Other social factors than sense of belonging could also be good measures to include.

The feasibility of intervention is important to mention since this is the first time this intervention was used. I recommend using larger groups than 3-4 participants per group, as it was very noticeable when only a few participants were ill, which might have given an impression of group disruption to the group and might have affected the outcome of sense of belonging. Depending on the target group, timing in the day is a very important factor to think about. What time of the day does the group have time most often? For parents with young children this might be early in the day, unless they have work at that time. Timing during the day and week have to be considered for the group musicking sessions. The length of the sessions of around one hour was a good length, since it was easy to fit into a schedule, and to divide into the two parts. The activities within the sessions were engaging, and received positive feedback, especially the choir warm-ups and the participatory accompaniment. A focus on speaking the language being learned with a lot of gestures for clarification of meaning is beneficial to learning the language. Depending on where future interventions are held, they might be able to have more of a music therapeutic focus.

5 Conclusions

This study was a Pilot RCT exploring both the effect of group musicking on sense of belonging and language learning in adult migrants learning Norwegian and exploring the feasibility of the study design to explore the effect. Using a rationale of a lot of theories and previous research connecting musicking, sense of belonging and second language acquisition it is reasonable to expect a positive result of group musicking on sense of belonging and language acquisition in the target group.

There were no significant findings of effect results of the effect of group musicking on sense of belonging and language learning. The effect result of the primary outcome of SOBI-P was a mean difference of -0.90 (-3.73, 1.92), reported with a 95% Confidence interval, and a p-value of 0.604. SOBI-A had a mean difference of -1.07 (-3.45, 1.31) and a p-value of 0.393, and CEFR had a mean difference of -2.53 (-5.97, 0.92), with a p-value of 0.175. The chance of a type 2 error however is high, especially considering positive narrative feedback results. There needs to be further research on additional outcomes with the intervention used with the target group.

For a future full-size RCT on the subject, I recommend a similar study design but with a few changes. More contributing researchers and people conducting the intervention are needed to ensure less possible bias. Other practical changes are also recommended. The measures used to measure the effect should be validated further. Further analysis of the existing data can be done.

6 Presented in Contexts Outside of the Normal Curriculum

January 11th, 2023: Seminar at the Norwegian Courses at the University of Bergen with the Norwegian teachers - progress report.

January 3rd, 2023: Grieg Academy Sharp (GA#): - Presenting the thesis' theoretical background and methods.

References

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of neurology, 29*(2), 130-131. https://doi.org/10.1001/archneur.1973.00490260074018

Allen, K.-A., & Bowles, T. (2012, 01/01). Belonging as a Guiding Principle in the Education of Adolescents. *Australian Journal of Educational and Developmental Psychology, 12*, 108-119.

Allen, K., Kern, M. L., Vella-Brodrick, D., Hattie, J., & Waters, L. (2018, 2018/03/01). What Schools Need to Know About Fostering School Belonging: a Meta-analysis. *Educational Psychology Review, 30*(1), 1-34. https://doi.org/10.1007/s10648-016-9389-8

Amit, K., & Bar-Lev, S. (2015). Immigrants' Sense of Belonging to the Host Country: The Role of Life Satisfaction, Language Proficiency, and Religious Motives. *Social Indicators Research, 124*(3), 947-961. https://doi.org/10.1007/s11205-014-0823-3

Anant, S. S. (1966). The need to belong. *Canada's Mental Health, 14*, 21-27.

Andrade, C. (2019, May-Jun). The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian J Psychol Med, 41*(3), 210-215. https://doi.org/10.4103/ijpsym.Ijpsym_193_19

Ansdell, G. (2002). Community Music Therapy & The Winds of Change. *Voices: A World Forum for Music Therapy, 2*. https://doi.org/10.15845/voices.v2i2.83

Baker, F. A., Tamplin, J., Clark, I. N., Lee, Y.-E. C., Geretsegger, M., & Gold, C. (2019). Treatment fidelity in a music therapy multi-site cluster randomized controlled trial for people living with dementia: The MIDDEL project intervention fidelity protocol. *Journal of Music Therapy, 56*(2), 125-148. https://doi.org/10.1093/jmt/thy023

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychol Rev, 84*(2), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1997). *Self-efficacy : the exercise of control*. Freeman.

Barringer, A., Papp, L. M., & Gu, P. (2022). College students' sense of belonging in times of disruption: prospective changes from before to during the COVID-19 pandemic. *Higher Education Research & Development*, 1-14. https://doi.org/10.1080/07294360.2022.2138275

Baumeister, R. F., & Leary, M. R. (1995, May). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol Bull, 117*(3), 497-529.

Belfiore, E., & Bennett, O. (2007, 2007/05/01). Rethinking the social impacts of the arts. *International Journal of Cultural Policy, 13*(2), 135-151. https://doi.org/10.1080/10286630701342741

Blythe laGasse, A. (2014). Developmental speech and language training through music (DSLM). *Handbook of neurologic music therapy*, 196.

Boutsen, F. (2003, 03/01). Prosody: The Music of Language and Speech. *ASHA Leader, 5*, 7-9. https://doi.org/10.1044/leader.FTR1.08042003.6

Bradt, J. (2012, Summer). Randomized Controlled Trials in Music Therapy: Guidelines for Design and Implementation. *J Music Ther, 49*(2), 120-149. https://doi.org/10.1093/jmt/49.2.120

Brean, A., & Skeie, G. O. (2019). *Musikk og hjernen : om musikkens magiske kraft og fantastiske virkning på hjernen*. Cappelen Damm.

Brown, H. D. (1973). Affective variables in second language acquisition. *Language Learning, 23*(2), 231-244. https://doi.org/10.1111/j.1467-1770.1973.tb00658.x

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *Eur J Neurosci, 23*(10), 2791-2803. https://doi.org/10.1111/j.1460-9568.2006.04785.x

Bruscia, K. E. (2014). *Defining Music Therapy*. University Park: Barcelona.

Caligiuri, P., DuBois, C. L. Z., Lundby, K., & Sinclair, E. A. (2020, 2020/12/01). Fostering international students' sense of belonging and perceived social support through a semester-long experiential activity. *Research in Comparative and International Education, 15*(4), 357-370. https://doi.org/10.1177/1745499920954311

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000, Sep 16). Framework for design and evaluation of complex interventions to improve health. *BMJ, 321*(7262), 694-696. https://doi.org/10.1136/bmj.321.7262.694

Cesana, B. M. (2018, 07/06). What p-value must be used as the Statistical Significance Threshold? P. *Biomedical Journal of Scientific & Technical Research, 6*. https://doi.org/10.26717/BJSTR.2018.06.001359

Chen, X.-J., Hannibal, N., & Gold, C. (2016). Randomized trial of group music therapy with Chinese prisoners: impact on anxiety, depression, and self-esteem. *International journal of offender therapy and comparative criminology, 60*(9), 1064-1081. https://doi.org/10.1177/0306624X15572795

Chew, R. (2009). Community-based arts organizations: A new center of gravity. *Washington, DC: Americans for the Arts*.

Chomsky, N. (2006). *Language and mind* (Third edition. ed.). Cambridge University Press.

Chou, K.-L. (2007, 2007/02/01/). Psychological distress in migrants in Australia over 50 years old: A longitudinal investigation. *Journal of Affective Disorders, 98*(1), 99-108. https://doi.org/10.1016/j.jad.2006.07.002

Cocks, K., & Torgerson, D. J. (2013). Sample size calculations for pilot randomized trials: a confidence interval approach. *Journal of Clinical Epidemiology, 66*(2), 197-201. https://doi.org/10.1016/j.jclinepi.2012.09.002

Council of Europe. (1954). *European Cultural Convention*.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching and assessment* (Modern Languages Division, Ed.). Cambridge university press. https://rm.coe.int/1680459f97

Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching and assessment - Companion Volume* (Language Policy Programme, Ed.). Council of Europe Publishing. http://www.coe.int/lang-cefr

Crash Course, S. (2018). *Crash course statistics. Confidence intervals* Place of publication not identified, Crash Course Statistics.

Dalla Bella, S. (2016). Music and brain plasticity. *The Oxford handbook of music psychology*, 325-342.

De nasjonale forskningsetiske komiteene. (2019). *Generelle forskningsetiske retningslinjer*. Retrieved 6. Juni 2022 from https://www.forskningsetikk.no/retningslinjer/generelle/

Dunning, D. (2011). Chapter five - The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 44, pp. 247-296). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-385522-0.00005-6

Ebberts, A. G. (1994). *The effectiveness of three types of music therapy interventions with persons diagnosed with probable dementia of the Alzheimer's type who display agitated behaviors* [Master's thesis, University of Kansas, Music and Dance].

Eldridge, S. M., Lancaster, G. A., Campbell, M. J., Thabane, L., Hopewell, S., Coleman, C. L., & Bond, C. M. (2016). Defining Feasibility and Pilot Studies in Preparation for Randomised Controlled Trials: Development of a Conceptual Framework. *PLoS One, 11*(3), e0150205. https://doi.org/10.1371/journal.pone.0150205

Ellis, R. (1989). *Understanding second language acquisition* (Vol. 31). Oxford university press Oxford.

Enge, K. E. A., & Stige, B. (2022, 2022/01/01). Musical pathways to the peer community: A collective case study of refugee children's use of music therapy. *Nordic Journal of Music Therapy, 31*(1), 7-24. https://doi.org/10.1080/08098131.2021.1891130

Engh, D. (2013, 01/09). Why Use Music in English Language Learning? A Survey of the Literature. *English Language Teaching, 6*, 113-127. https://doi.org/10.5539/elt.v6n2p113

Feeley, N., Cossette, S., Côté, J., Héon, M., Stremler, R., Martorella, G., & Purden, M. (2009). The importance of piloting an RCT intervention. *Canadian Journal of Nursing Research Archive*, 84-99.

Gardner, R. C., Tremblay, P. F., & Masgoret, A.-M. (1997). Towards a Full Model of Second Language Learning: An Empirical Investigation. *The Modern Language Journal, 81*(3), 344-362. https://doi.org/10.1111/j.1540-4781.1997.tb05495.x

Gauthier, T. D., & Hawley, M. E. (2015). Chapter 5 - Statistical Methods. In B. L. Murphy & R. D. Morrison (Eds.), *Introduction to Environmental Forensics (Third Edition)* (pp. 99-148). Academic Press. https://doi.org/10.1016/B978-0-12-404696-2.00005-9

Geretsegger, M., Mössler, K. A., Bieleninik, Ł., Chen, X.-J., Heldal, T. O., & Gold, C. (2017). Music therapy for people with schizophrenia and schizophrenia-like disorders. *Cochrane Database of Systematic Reviews, 2017*(5). https://doi.org/10.1002/14651858.cd004025.pub4

Ghetti, C. M. (2013). Effect of music therapy with emotional-approach coping on preprocedural anxiety in cardiac catheterization: A randomized controlled trial. *Journal of Music Therapy, 50*(2), 93-122. https://doi.org/10.1093/jmt/50.2.93

Glass, C., & Westmont, C. (2014, 01/31). Comparative effects of belongingness on the academic success and cross-cultural interactions of domestic and international

students. *International Journal of Intercultural Relations, 38*, 106–119. https://doi.org/10.1016/j.ijintrel.2013.04.004

Gold, C. (2004). The Use of Effect Sizes in Music Therapy Research. *Music therapy perspectives, 22*(2), 91-95. https://doi.org/10.1093/mtp/22.2.91

Gold, C. (2015). Quantitative Psychotherapy Outcome Research: Methodological Issues. In (pp. 537-558). Vienna: Springer Vienna. https://doi.org/10.1007/978-3-7091-1382-0_26

Gold, C. (2019, 2019/01/01). Randomised controlled trials: Why the small print matters. *Nordic Journal of Music Therapy, 28*(1), 1-3. https://doi.org/10.1080/08098131.2019.1548796

Good, A., Russo, F., & Sullivan, J. (2015, 04/01). The efficacy of singing in foreign-language learning. *Psychology of Music, 43*, 627-640. https://doi.org/10.1177/0305735614528833

Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools, 30*(1), 79-90. https://doi.org/10.1002/1520-6807(199301)30:1<79::AID-PITS2310300113>3.0.CO;2-X

Grünigen, R., Perren, S., Nägele, C., & Alsaker, F. D. (2010). Immigrant children's peer acceptance and victimization in kindergarten: The role of local language competence. *British Journal of Developmental Psychology, 28*(3), 679-697. https://doi.org/10.1348/026151009x470582

Gupta, S. K. (2011, Jul). Intention-to-treat concept: A review. *Perspect Clin Res, 2*(3), 109-112. https://doi.org/10.4103/2229-3485.83221

Gøtzsche, P. C. (1996, 1996/08/01/). Blinding during data analysis and writing of manuscripts. *Controlled Clinical Trials, 17*(4), 285-290. https://doi.org/10.1016/0197-2456(95)00263-4

Hagerty, B., & Patusky, K. (1995). Developing a Measure Of Sense of Belonging. *Nursing research, 44*, 9-13. https://doi.org/10.1097/00006199-199501000-00003

Hagerty, B. M., Lynch-Sauer, J., Patusky, K. L., Bouwsema, M., & Collier, P. (1992). Sense of belonging: A vital mental health concept. *Archives of psychiatric nursing, 6*(3), 172-177. https://doi.org/10.1016/0883-9417(92)90028-h

Henderson, S., Cain, M., Istvandity, L., & Lakhani, A. (2017). The role of music participation in positive health and wellbeing outcomes for migrant populations: A systematic

review. *Psychology of Music, 45*(4), 459-478.
https://doi.org/10.1177/0305735616665910

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D.
    G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M.,
    McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of
    interventions: template for intervention description and replication (TIDieR) checklist
    and guide. *BMJ : British Medical Journal, 348*, g1687.
    https://doi.org/10.1136/bmj.g1687

Hoque, M. (2017). An Introduction to the Second Language Acquisition. In *Language
    Acquisition* (pp. 1-23). EDRC.

Hunt, M. (2005). Action Research and Music Therapy: Group Music Therapy with Young
    Refugees in a School Community. *Voices : a world forum for music therapy, 5*(2).
    https://doi.org/10.15845/voices.v5i2.223

International Organization for Migration (IOM). (2019). *Glossary on Migration* (ISSN
    1813-2278). https://publications.iom.int/system/files/pdf/iml_34_glossary.pdf

Janata, P., & Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for
    behaviors related to sequencing and music. *Nature neuroscience, 6*(7), 682-687.
    https://doi.org/10.1038/nn1081

Johnson, D. E. (2010). Crossover experiments. *Wiley Interdisciplinary Reviews:
    Computational Statistics, 2*(5), 620-625. https://doi.org/10.1002/wics.109

Juszczak, E., Altman, D. G., Hopewell, S., & Schulz, K. (2019). Reporting of Multi-Arm
    Parallel-Group Randomized Trials: Extension of the CONSORT 2010 Statement.
    JAMA, 321(16), 1610-1620. https://doi.org/10.1001/jama.2019.3087

Kadam, P., & Bhalerao, S. (2010, Jan). Sample size calculation. *Int J Ayurveda Res, 1*(1), 55-
    57. https://doi.org/10.4103/0974-7788.59946

Kale, E., & Hjelde, K. H. (2017). Mental health challenges of immigrants in Norway, a
    literature review 2009–17. *health, 370*(9590), 859-877.

Kale, E., & Kumar, B. N. (2012). Challenges in healthcare in multi-ethnic societies:
    communication as a barrier to achieving health equity. *Children, 3*, 295-308.
    https://doi.org/10.5772/38259

Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D.
    (2002). Bach speaks: A cortical" language-network" serves the processing of music.
    Neuroimage, 17(2), 956-966.

Krashen, S. (1981). Second language acquisition. *Second Language Learning, 3*(7), 19-39.

Kreutz, G. (2014). Does singing facilitate social bonding? *Music and Medicine, 6*(2), 51-60. https://doi.org/https://doi.org/10.47513/mmd.v6i2.180

Kruse, R. L., Alper, B. S., Reust, C., Stevermer, J. J., Shannon, S., & Williams, R. H. (2002). Intention-to-treat analysis: who is in? Who is out? *Journal of Family Practice, 51*(11), 969-971.

Krüger, V., & Diaz, E. (2023, 2023/06/01/). The potential to meet the needs of refugees and other migrants through music therapy. *The Lancet Regional Health - Europe, 29*, 100637. https://doi.org/https://doi.org/10.1016/j.lanepe.2023.100637

Lakens, D. (2015). On the challenges of drawing conclusions from p-values just below 0.05. *PeerJ, 3*, e1142. https://doi.org/10.7717/peerj.1142

Lathroum, L. M. (2011). *The role of music perception in predicting phonological awareness in five- and six-year-old children* [Doctoral dissertation,

Laue, J., Diaz, E., Eriksen, L., & Risør, T. (2021). Migration health research in Norway: a scoping review. *Scandinavian Journal of Public Health*, 14034948211032494. https://doi.org/10.1177/14034948211032494

Ledger, A. J., & Baker, F. A. (2007, 2007/05/01). An investigation of long-term effects of group music therapy on agitation levels of people with Alzheimer's Disease. *Aging & Mental Health, 11*(3), 330-338. https://doi.org/10.1080/13607860600963406

Levitin, D. J. (2006). *This is Your Brain on Music: The Science of a Human Obsession*. Dutton. https://books.google.no/books?id=prV4UrZ2df0C

Lim, C. Y., & In, J. (2021, Aug). Considerations for crossover design in clinical study. *Korean J Anesthesiol, 74*(4), 293-299. https://doi.org/10.4097/kja.21165

Lov om grunnskolen og den vidaregåande opplæringa (opplæringslova), Kunnskapsdepartementet (1998). https://lovdata.no/pro/NL/lov/1998-07-17-61

Lov om integrering gjennom opplæring, utdanning og arbeid (integreringsloven), Arbeids- og inkluderingsdepartementet (2020). https://lovdata.no/pro/NL/lov/2020-11-06-127

Lydersen, S. (2022). Multippel imputering av manglende data. *Tidsskrift for Den norske legeforening*. https://doi.org/10.4045/tidsskr.21.0772

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature neuroscience, 4*(5), 540-545. https://doi.org/DOI: 10.1038/87502

Marksteiner, T., Janke, S., & Dickhäuser, O. (2019). Effects of a brief psychological intervention on students' sense of belonging and educational outcomes: The role of students' migration and educational background. *Journal of School Psychology, 75*, 41-57. https://doi.org/https://doi.org/10.1016/j.jsp.2019.06.002

McMillan, D. (1976). *Sense of community: An attempt at definition*. George Peabody College for Teachers.

McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of community psychology, 14*(1), 6-23.

Medina, S. (1993). The Effect of Music on Second Language Vocabulary Acquisition Graduate Education Department California State University from; http://www. oocities. org. *ESLmusic/articles/article01. htm l*.

Migration, I. O. f. (Ed.). (2021). *Global Migration Indicators 2021*. International Organization for Migration. https://publications.iom.int/books/global-migration-indicators-2021.

Mills, N. (2014). Self-Efficacy in Second Language Acquisition. In S. Mercer & M. Williams (Eds.), *Multiple perspectives on the self in SLA* (Vol. 73, pp. 6-22). United Kingdom: Multilingual Matters.

Mishra, P., Singh, U., Pandey, C. M., Mishra, P., & Pandey, G. (2019). Application of student's t-test, analysis of variance, and covariance. *Ann Card Anaesth, 22*(4), 407-411. https://doi.org/10.4103/aca.ACA_94_19

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ, 340*, c869. https://doi.org/10.1136/bmj.c869

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological science, 22*(11), 1425-1433.

Nunan, D., Aronson, J., & Bankhead, C. (2018, Feb). Catalogue of bias: attrition bias. *BMJ Evid Based Med, 23*(1), 21-22. https://doi.org/10.1136/ebmed-2017-110883

Olszewska, A. M., Gaca, M., Herman, A. M., Jednoróg, K., & Marchewka, A. (2021). How Musical Training Shapes the Adult Brain: Predispositions and Neuroplasticity [Review]. *Frontiers in Neuroscience, 15*. https://doi.org/10.3389/fnins.2021.630829

Paquette, K., & Rieg, S. (2008). Using Music to Support the Literacy Development of Young English Language Learners. *Early Childhood Education Journal, 36*, 227-232. https://doi.org/10.1007/s10643-008-0277-9

Peretz, I., & Zatorre, R. J. (2005). Brain Organization for Music Processing. *Annual Review of Psychology, 56*(1), 89-114. https://doi.org/10.1146/annurev.psych.56.091103.070225

Piaget, J. (1971). The theory of stages in cognitive development. In *Measurement and Piaget.* (pp. ix, 283-ix, 283). McGraw-Hill.

Piri, S. (2019). The Role of Music in Second Language Learning. *Studies in Literature and Language, 17*, 75-78.

Pittman, L. D., & Richmond, A. (2007). Academic and Psychological Functioning in Late Adolescence: The Importance of School Belonging. *The Journal of Experimental Education, 75*(4), 270-290. https://doi.org/10.3200/JEXE.75.4.270-292

Pollack, N. J., & Namazi, K. H. (1992). The Effect of Music Participation on the Social Behavior of Alzheimer's Disease Patients1. *Journal of Music Therapy, 29*(1), 54-67. https://doi.org/10.1093/jmt/29.1.54

Rasing, N. L., Janus, S. I. M., Kreutz, G., Sveinsdottir, V., Gold, C., Nater, U. M., & Zuidema, S. U. (2022). The Impact of Music on Stress Biomarkers: Protocol of a Substudy of the Cluster-Randomized Controlled Trial Music Interventions for Dementia and Depression in ELderly Care (MIDDEL). *Brain Sciences, 12*(4), 485. https://doi.org/https://doi.org/10.3390/brainsci12040485

Regjeringen, D. n. (2021). *Introduksjonsprogram*. Retrieved 4. juni 2022 from https://www.regjeringen.no/no/tema/innvandring-og-integrering/asd/Verkemiddel-i-integreringsarbeidet/introduksjonsprogram/id2343472/

Ridder, H. M., Stige, B., Qvale, L. G., & Gold, C. (2013). Individual music therapy for agitation in dementia: an exploratory randomized controlled trial. *Aging Ment Health, 17*(6), 667-678. https://doi.org/10.1080/13607863.2013.790926

Rio, R. (2002). Improvization with the elderly: moving from creative activities to process-oriented therapy. *The Arts in Psychotherapy, 4*(29), 191-201. https://doi.org/10.1016/S0197-4556(02)00156-9

Robb, S. L., Carpenter, J. S., & Burns, D. S. (2011). Reporting Guidelines for Music-based Interventions. *Journal of Health Psychology, 16*(2), 342-352. https://doi.org/10.1177/1359105310374781

Rolvsjord, R., Gold, C., & Stige, B. (2005, 01/01). Research Rigour and Therapeutic Flexibility: Rationale for a Therapy Manual Developed for a Randomised Controlled Trial. *Nordic Journal of Music Therapy Nordic Journal of Music Therapy, 14*, 15-32. https://doi.org/10.1080/08098130509478122

Ruud, E. (2010). *Music therapy : a perspective from the humanities*. Barcelona Publishers.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory : basic psychological needs in motivation, development, and wellness*. Guilford Press.

Sandberg, H. F. (2009). *Get moving! The effects of music and movement on student attention and engagement*. Arizona State University.

Sari, M. (2012). Sense of school belonging among elementary school students. *Çukurova University Faculty of Education Journal, 41*(1).

Schoepp, K. (2001). Reasons for using songs in the ESL/EFL classroom. *The internet TESL journal, 7*(2), 1-4. http://iteslj.org/Articles/Schoepp-Songs.html

Schulz, K. F. (2001). Assessing allocation concealment and blinding in randomised controlled trials: why bother? *Evidence Based Nursing, 4*(1), 4-6. https://doi.org/10.1136/ebn.4.1.4

Schulz, K. F., Altman, D. G., Moher, D., & Group*, C. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of internal medicine, 152*(11), 726-732.

Schulz, K. F., & Grimes, D. A. (2002). Generation of allocation sequences in randomised trials: chance, not choice. *The Lancet, 359*(9305), 515-519. https://doi.org/https://doi.org/10.1016/S0140-6736(02)07683-3

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J. L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *Neuroimage, 51*(1), 450-461. https://doi.org/10.1016/j.neuroimage.2010.02.023

Silverman, M. J. (2010). Applying levels of evidence to the psychiatric music therapy literature base. *The Arts in Psychotherapy, 37*(1), 1-7. https://doi.org/10.1016/j.aip.2009.11.005

Silverman, M. J., Gooding, L. F., & Yinger, O. (2020). It's...Complicated: A Theoretical Model of Music-Induced Harm. *Journal of Music Therapy, 57*(3), 251-281. https://doi.org/10.1093/jmt/thaa008

Skinner, B. F. (1963). Operant behavior. *The American psychologist, 18*(8), 503-515. https://doi.org/10.1037/h0045185

Small, C. (1998). *Musicking : The Meanings of Performing and Listening*. University Press of New England.

Staddon, J. E. R., & Cerutti, D. T. (2003). Operant Conditioning. *Annual Review of Psychology, 54*(1), 115-144. https://doi.org/10.1146/annurev.psych.54.101601.145124

Stanley, K. (2007). Design of Randomized Controlled Trials. *Circulation, 115*(9), 1164-1169. https://doi.org/10.1161/circulationaha.105.594945

Statistisk Sentralbyrå. (2023a). *09817: Innvandrere og norskfødte med innvandrerforeldre, etter innvandringskategori, landbakgrunn, statistikkvariabel og år*. Retrieved March 6th, 2023 from https://www.ssb.no/statbank/sq/10079569

Statistisk Sentralbyrå. (2023b). *Innvandrere og norskfødte med innvandrerforeldre*. Retrieved March 6th, 2023 from https://www.ssb.no/befolkning/innvandrere/statistikk/innvandrere-og-norskfodte-med-innvandrerforeldre

Steele, M. E. (2016). How Can Music Build Community? Insight from Theories and Practice of Community Music Therapy. *Voices : a world forum for music therapy, 16*(2). https://doi.org/10.15845/voices.v16i2.876

Stige, B., & Aarø, L. E. (2011). *Invitation to community music therapy*. Routledge.

Stinerock, R. (2018). *Statistics with R: A Beginner's Guide*. SAGE Publications. https://books.google.no/books?id=T7pDDwAAQBAJ

Streiner, D. L. (2014). Statistics Commentary Series: Commentary #3—Last Observation Carried Forward. *Journal of Clinical Psychopharmacology, 34*(4). https://journals.lww.com/psychopharmacology/Fulltext/2014/08000/Statistics_Commentary_Series__Commentary__3_Last.2.aspx

Suresh, K. (2011, Jan). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci, 4*(1), 8-11. https://doi.org/10.4103/0974-1208.82352

Sæther, E. (2008). When minorities are the majority: voices from a teacher/researcher project in a multicultural school in Sweden. *Research Studies in Music Education, 30*(1), 25-42. https://doi.org/10.1177/1321103x08089888

Tamplin, J., Baker, F. A., Jones, B., Way, A., & Lee, S. (2013). 'Stroke a Chord': The effect of singing in a community choir on mood and social engagement for people living with aphasia following a stroke. *NeuroRehabilitation, 32*, 929-941. https://doi.org/10.3233/NRE-130916

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., Thabane, M., Giangregorio, L., Dennis, B., Kosa, D., Debono, V. B., Dillenburg, R., Fruci, V., Bawor, M., Lee, J., Wells, G., & Goldsmith, C. H. (2013). A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Medical Research Methodology, 13*(1), 92. https://doi.org/10.1186/1471-2288-13-92

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

Tomasello, M. (2003). *Constructing a language : a usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M. (2010). *Origins of human communication* (Vol. 2008). MIT Press.

Ulrich, G., Houtmans, T., & Gold, C. (2007). The additional therapeutic effect of group music therapy for schizophrenic patients: a randomized study. *Acta Psychiatrica Scandinavica, 116*(5), 362-370. https://doi.org/https://doi.org/10.1111/j.1600-0447.2007.01073.x

Universitetet i Bergen. (2019). *Etikk og Regelverk*. Retrieved 6. Juni 2022 from https://www.uib.no/foransatte/17298/etikk-og-regelverk

Valvatne, H. (2007). Å vokse opp med mer enn ett språk. I H. Valvatne, & M. Sandvik (Red.). *Barn, språk og kultur, 2*, 287-328.

van Bruggen-Rufi, M. C., Vink, A. C., Wolterbeek, R., Achterberg, W. P., & Roos, R. A. (2017). The effect of music therapy in patients with Huntington's disease: A randomized controlled trial. *Journal of Huntington's Disease, 6*(1), 63-72.

Vygotsky, L. S. (1962). *Thought and language*. MIT press.

Werner, J., Wosch, T., & Gold, C. (2017, 2017/02/01). Effectiveness of group music therapy versus recreational group singing for depressive symptoms of elderly nursing home residents: pragmatic trial. *Aging & Mental Health, 21*(2), 147-155. https://doi.org/10.1080/13607863.2015.1093599

Wood, S. (2016). *A matrix for community music therapy practice*. Barcelona Publishers.

Woolston, C. (2015, 2015/03/01). Psychology journal bans P values. *Nature, 519*(7541), 9-9. https://doi.org/10.1038/519009f

Zarin, D. A., Tse, T., Williams, R. J., & Rajakannan, T. (2017, Jan 26). Update on Trial Registration 11 Years after the ICMJE Policy Was Established. *N Engl J Med, 376*(4), 383-391. https://doi.org/10.1056/NEJMsr1601330