Manuscript for

Learning and connecting through songs: a proof-of-concept study with newcomers

Fidaa Akrout[1,2], Dawn Merrett[1,2], Hsun-Yi Liao[1,2], Isabelle Héroux[3] and Isabelle Peretz[1,2]

[1] Department of psychology, University of Montreal, Quebec, Canada

[2] International Laboratory for Brain, Music and Sound Research (BRAMS), University of Montreal, Quebec, Canada

[3] Department of music, UQAM

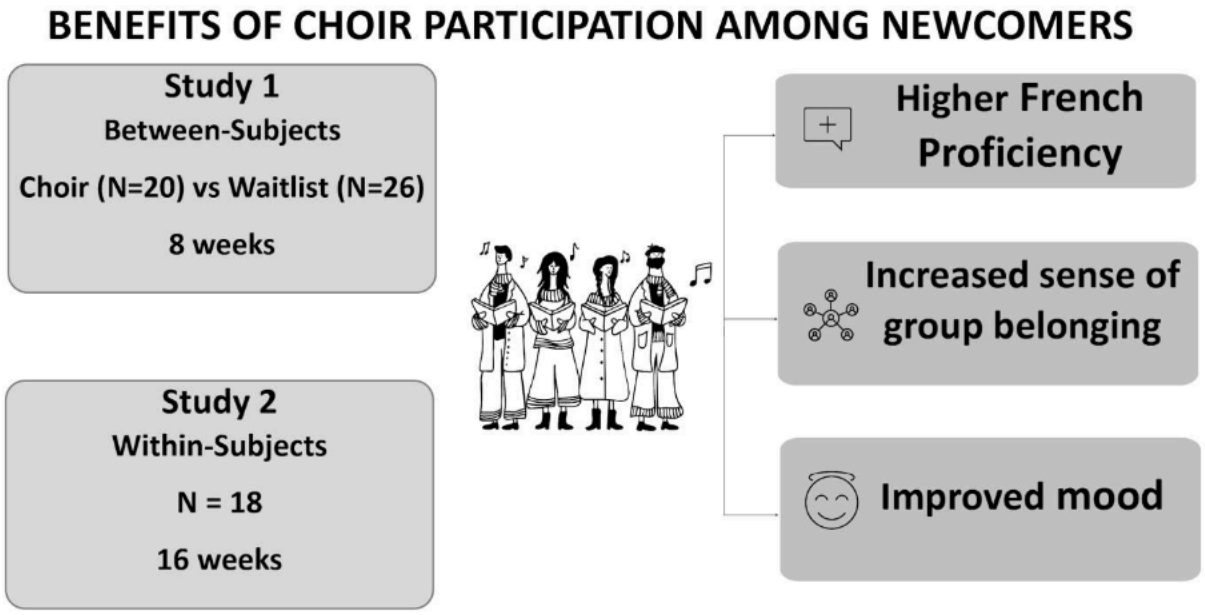
***Corresponding author*:** Isabelle.Peretz@umontreal.ca

Abstract

Group singing may offer an effective means of addressing challenges faced by newcomers in learning another language and integrating into an unfamiliar society. The effect of choir participation was compared here to a wait-list control group, on both qualitative (Study 1) and quantitative (Studies 1 and 2) measures of French proficiency, social connectedness, and well-being. The assessments took place before and after an 8-week period. In Study 1, 20 newcomers participated in the choir, and 26 were assigned to a wait-list control group. The latter group later joined the choir and was included in Study 2. As expected, choir participation led to improvements in French proficiency, social connection, and mood, relative to the control group. Moreover, human ratings of French production strongly correlated with machine learning–based assessments, highlighting the sensitivity of a simple AI tool for evaluating second-language proficiency despite the diversity of accents. These encouraging results lay the groundwork for a randomized controlled trial.

Keywords: Group singing, second language acquisition, mood, social integration, immigration

Graphical Abstract



Introduction

Immigration has long been a driving force of Canada's population growth, with newcomers now representing nearly one in four residents[1,2]. As a nation built on diversity, Canada takes pride in welcoming people from around the world and offering them opportunities to build new lives and contribute to society. However, the successful integration of newcomers remains a challenge—particularly when they lack proficiency in one of the country's official languages[3,4]. This is especially true in Québec, Canada's Francophone province, which welcomes over 30,000 immigrants annually[2], 40% of whom arrive with little or no knowledge of French[5].

A basic command of French is essential for accessing healthcare and other social services in Québec. In response, the provincial government has made French lessons widely available and, in some cases, mandatory[5]. Yet, fewer than half of all newcomers enroll in these programs[6]. Clearly, additional or alternative strategies are needed. We propose group singing as an engaging approach that not only addresses communication barriers but also enhances social integration and psychological well-being.

Learning a language through song is a common and intuitive practice. Many educators integrate familiar songs—like *Twinkle, Twinkle Little Star*[7]—to support second language (L2) acquisition[8]. Indeed, singing slows down speech and enhances pronunciation[9–13]. In addition, group singing enhances vocabulary acquisition[10,11,14–16], word recall, grammar[17] and translation skills[9,11,12] when compared to other active control conditions such as poetry recitation[9], visual arts[12], or music listening[18]. While one study found no advantage of singing over speaking lyrics when assessing grammar skills[16], the majority support the utility of singing for L2 learning.

Nonetheless, existing studies present several limitations. First, language assessments often focus narrowly on the lyrics learned, limiting generalizability[9–13,15–17]. Second, few studies control for test-retest effects through the use of parallel forms or comparable assessment tools. Third, most studies target homogeneous populations, leaving it unclear whether group singing benefits individuals with diverse linguistic backgrounds and varying levels of L2 proficiency.

Beyond language learning, group singing is associated with increased well-being[19–27] and stronger social bonds[19,22–24,28–31]. Multiple studies have reported immediate mood benefits after singing sessions lasting from 20 to 60 minutes[19,21,27,31,32], with effects surpassing those of solo singing[25,27], conversation[31], chatting[21], music listening[26], creative writing, crafting[23] or silence[19]. Although many participants were experienced choir members—potentially biasing results—some studies mitigated this by forming new choirs for research purposes and tracking them over longer periods (6–7 months)[20,22,23]. These studies found comparable benefits relative to other group activities such as crafting, creative writing[22,23], or routine daily tasks[20].

In the context of integration, singing together can help forge social bonds: participants often report feeling more connected to others after choir participation than after other group activities[19,22–24,28]. In a pioneering study, Anshel & Kipper[30] found that choir members displayed more trust and cooperation in games like the Prisoner's Dilemma than those engaged in group music listening or film viewing. At the very least, group singing appears to act as a social "ice-breaker", helping to reduce initial unfamiliarity and increase comfort among strangers[19,22,23,28,29]. Qualitative research further suggests that singing together fosters a sense of personal and collective achievement[33], group solidarity[34], and the formation of new social networks[35,36].

Surprisingly, few studies have investigated choir singing among immigrants or refugees. The only quantitative study we found involved a choir designed to teach Finnish through song[13]. Participants showed improved pronunciation, and qualitative data revealed enhanced comprehension and greater confidence and motivation. However, this study lacked a comparison group, making it difficult to isolate the causes of these benefits.

In sum, although prior findings are promising, no controlled study has yet evaluated group singing as a structured intervention addressing both linguistic and social integration. Assessing its effects on language skills, social connection, and well-being in a single study offers a unique opportunity to evaluate its multifunctional potential for newcomer integration—and to understand how these outcomes may be interrelated.

To address this gap, we conducted two related studies evaluating French proficiency, social connectedness, and mood before and after an 8-week choir intervention compared to a waitlist control condition.

STUDY 1: COMPARISON BETWEEN A CHOIR AND WAIT GROUP OF NEWCOMERS

Methods & Material

Participants: Participants were eligible if they: (1) were adult newcomers (immigrants or refugees) who had lived in Québec for less than three years, (2) intended to reside permanently in Canada, and (3) had basic written and spoken English to understand the consent form and study instructions. Recruitment was conducted primarily through social media, French language classes, and word-of-mouth.

After attrition, 46 eligible participants formed two groups: 20 in the choir group and 26 in the control group. Due to recruitment challenges, planned randomization was not possible. The choir group began singing in October 2023, while recruitment for the control group continued until March 2024. Control participants were informed they would be invited to join the choir after the 8-week wait period (see Study 2).

Of the 23 participants who attended the choir's first session, 20 were present at the final session, with an average attendance of 18 across the eight weeks. In the control group, 41 participants attended the first session, and 26 returned for the second.

Most participants had university-level education in their countries of origin and had lived in Québec for less than one year. They came from 14 countries—Ukraine, Taiwan, Iran, China, Colombia, Cuba, Mexico, Egypt, Germany, Israel, Brazil, Spain, India, and the Philippines—with Ukraine being the most represented. Over half were enrolled in French language classes, consistent with Québec's immigration requirements. The choir group had more formal musical training and choir experience than the control group, whereas the control group had a higher average level of French, as measured by their French class level (see **Table 1**).

Table 1. Demographic Characteristics of Participants

	Choir Group n (%)	Control Group n (%)
n	20	26
Average age (Range)	37.5 (21-60)	39.4 (23-61)
Gender		
Females	16 (75)	20 (76.92)
Males	4 (25)	6 (23.08)
Educational level		
Secondary	3 (5)	2 (7.69)
University	17 (15)	24 (92.31)
Country		
Ukraine	8 (40)	7 (26.92)
Taiwan	5 (25)	0
Iran	1 (5)	3 (11.54)
China	2 (10)	4 (15.83)
Colombia	1 (5)	2 (7.69)
Others	4 (40)	10 (38.46)
Living in Quebec (years)		
< 1	13 (65)	19 (73.08)
1 - 2	7 (35)	6 (23.08)

2 - 3	0	1 (3.85)
French courses (level*)	11 (60)	20 (76.92)
Beginner	9 (81.81)	5 (25)
Intermediate	2 (18.18)	10 (50)
Advanced	0	5 (25)
Musical experience		
Number of participants (%)	14 (70)	12 (46.15)
Mean years	11.5	2.6

* as defined by language classes[37]

This research was approved by the Educational and Psychological Research Ethics Committee of the University of Montréal, and all participants gave written informed consent.

Procedure

During the first session (week 1), participants came as a group but were tested individually in separate rooms. They completed questionnaires assessing their expectations as well as self-reported French proficiency and language exposure. In addition, participants completed sentence repetition and image description tasks on a computer, guided by an examiner.

Following these assessments, participants from both the choir and wait group engaged in a group singing session led by the choir director. They were then offered refreshments during an informal "snack break," during which their interactions were video-recorded (these recordings are not analyzed in the present study). After the break, participants completed two additional measures: a mood questionnaire using visual analogue mood scales[38] and a group belonging questionnaire (the Inclusion of Other in Self Scale; IOS[39]).

This procedure was repeated for both groups eight weeks later, with the exception of the expectations questionnaire, which was administered only during the first session. Semi-structured interviews were conducted with choir participants at the end of the study (see **Figure 1** for a schema of the experimental procedure).

Choir sessions were held at the Music Department of the Université du Québec à Montréal (UQAM) every Saturday for two hours over an eight-week period. Each session began with vocal and physical warm-up exercises, followed by the learning of French songs, such as Gilles Vigneault's *Gens du pays* (see Table S1). Led by a professional French-speaking choir director, each session typically included approximately 40 minutes of actual singing. The remaining time was devoted to explaining song lyrics, reviewing correct pronunciation, and discussing the

cultural background of each piece. Each week also included a snack break, offering participants an opportunity to socialize informally.

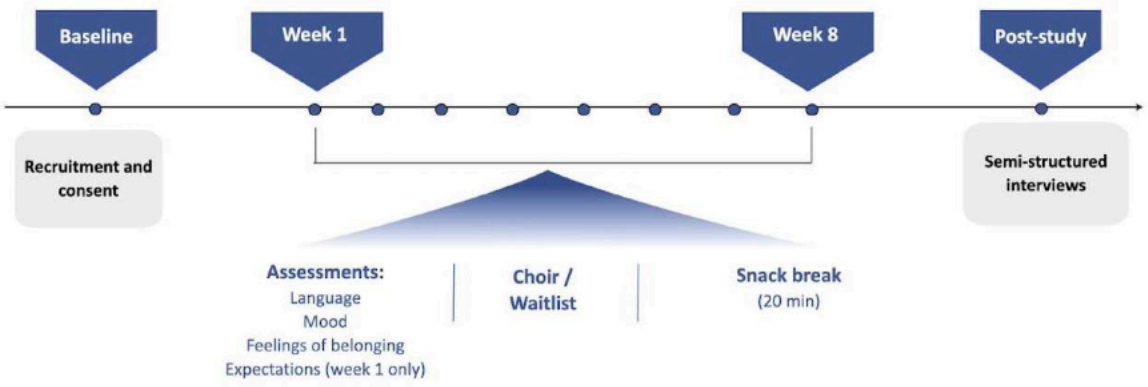


Figure 1. Experimental design of Study 1

Material

Pronunciation/Intelligibility. Pronunciation was assessed using a sentence repetition task[40] consisting of 16 pre-recorded sentences, each containing 8 to 15 words and syllables. The sentences were recorded by a native Quebec French speaker, and the instructions were delivered by a native English speaker. Participants were instructed to repeat each sentence after a beep. The task took approximately 10 minutes to complete.

Participant responses were evaluated using two methods: (1) Two native Quebec French speakers, blind to group and session, rated the number of correctly pronounced syllables in each sentence. A syllable was marked as incorrect if it was omitted, reversed, invented, or mispronounced (phonetically or lexically). If a participant began speaking before the beep, that response received a score of zero. (2) The audio recordings were also processed using AI-based speech recognition software (https://www.veed.io/). Words that were correctly transcribed by the algorithm were considered intelligible and given a score of 1. The total number of correctly detected words was divided by the total number of target words in order to produce a percentage score.

Self-Reported French Proficiency. Participants rated their French-speaking, understanding, and reading abilities on an 8-point Likert scale (from "none" to "excellent") at both Week 1 and Week 8.

Feelings of Group Belonging. The *Inclusion of Other in Self* (IOS) scale[39] was used to measure perceived group belonging. This visual tool presents a series of Venn-like diagrams with

progressively overlapping circles representing the self and a group. Participants selected the diagram that best reflected their sense of closeness to the group. Scores range from 1 (no overlap) to 7 (almost complete overlap).

Mood. Visual analogue mood scales[38] were used to assess participants' emotional states. Each subscale presented a 10 cm line anchored by two cartoon faces: one neutral and the other emotional (happy, sad, confused, energetic, afraid, tired, or angry). Participants marked a point on the line to indicate how strongly they felt each mood. Scores ranged from 0 (neutral) to 10 (maximum intensity), enabling detection of subtle mood changes.

Expectations. Three 11-point Likert scales were used at baseline to assess participants' expectations regarding choir participation. Instructions read: "On the following scale, indicate how much you agree or disagree with the statement." The statements were:
(1) "I think that singing together would help me to connect with others."
(2) "I think that singing together would make me feel good."
(3) "I think that singing together would help me learn French."
Scores ranged from 0 (strongly disagree) to 10 (strongly agree).

Semi-Structured Interviews. At the end of the 8-week period, online semi-structured interviews were conducted individually with 13 choir participants who were available. Interviews lasted approximately 15 minutes and followed a predefined interview grid covering topics such as recruitment, language learning, social integration, obstacles, and facilitators. Given participants' diverse language backgrounds and the use of English during interviews, questions were emailed in advance to aid preparation. Six participants submitted written responses beforehand, facilitating a smoother interview process. Prior to each session, consent to record was confirmed, and participants were briefed on the main themes. The interviewer adapted the question order based on participants' engagement and used clarifications or follow-ups as needed (e.g., when responses were limited or unclear).

Data analysis

All quantitative analyses were conducted using R software (version 4.3.3)[41]. For each variable, we assessed the assumptions of normality, homoscedasticity, and linearity. Missing data were minimal, with no variable exceeding 5% missingness. Given the repeated-measures design, linear mixed-effects models (LMMs) were used, which are well-suited for handling correlated observations and are robust to violations of normality[42]. Models were fitted using the lmerTest package with restricted maximum likelihood (REML) estimation.

Each model included Time (Week 1, Week 8), Group (Choir, Control), and their interaction (Time × Group) as fixed effects, and a random intercept for each participant to account for inter-individual variability. P-values for fixed effects were computed using Satterthwaite's approximation for degrees of freedom, as implemented in the lmerTest package. Post hoc

pairwise comparisons were performed using the emmeans package, with custom contrasts comparing pre- and post-test scores within each group. Holm's correction was applied to adjust for multiple comparisons. All reported means are estimated marginal means (EMMs).

We also conducted a covariate analysis to assess whether choir participation influenced improvements in French proficiency, while controlling for participants' baseline French level. A separate LMM, also estimated using REML via lmerTest, included Time**,** Group**,** French level (coded on a continuous scale from 0 to 8, where 0 indicated no French classes and 1–8 reflected increasing levels of formal instruction), and the Time × Group interaction as fixed effects, with a random intercept for participants. P-values were again calculated using Satterthwaite's approximation.

To explore the relationship between self-assessed proficiency and objective performance on the language task, we pooled pre- and post-test data and conducted a Spearman rank-order correlation**,** due to non-normal distribution of the data.

Qualitative Data

Qualitative data were analyzed using Interpretative Phenomenological Analysis (IPA)[43], which focuses on participants' lived experiences and the meaning they attribute to them. Verbatim interview transcripts were lightly corrected for language errors while preserving participants' original expressions. An initial reading of each transcript was conducted to identify overarching themes. Given the semi-structured nature of the interviews, coding was organized by topic axis.

For each axis, relevant content was extracted, and recurring themes were identified. Statements were then grouped by theme, and categories and subcategories were created as necessary. Finally, we calculated the number of participants who mentioned each theme or subtheme to gauge prevalence and thematic saturation.

Results

Self-Assessment of French Proficiency

Linear mixed-effects models revealed significant Group × Time interactions for self-rated speaking**,** understanding**,** and reading proficiency in French (all $p < .05$). Post hoc comparisons of estimated marginal means (EMMs) indicated that choir participants showed significant improvements in all three domains between Week 1 and Week 8 (all $p < .05$). In contrast, control participants did not exhibit significant changes in self-assessed proficiency over the same period (all $p > .05$).

Importantly, at baseline (Week 1), choir participants reported significantly lower proficiency scores than control participants across all measures (all $p < .05$), highlighting a steeper trajectory of perceived improvement among choir members (see **Figure 2**).

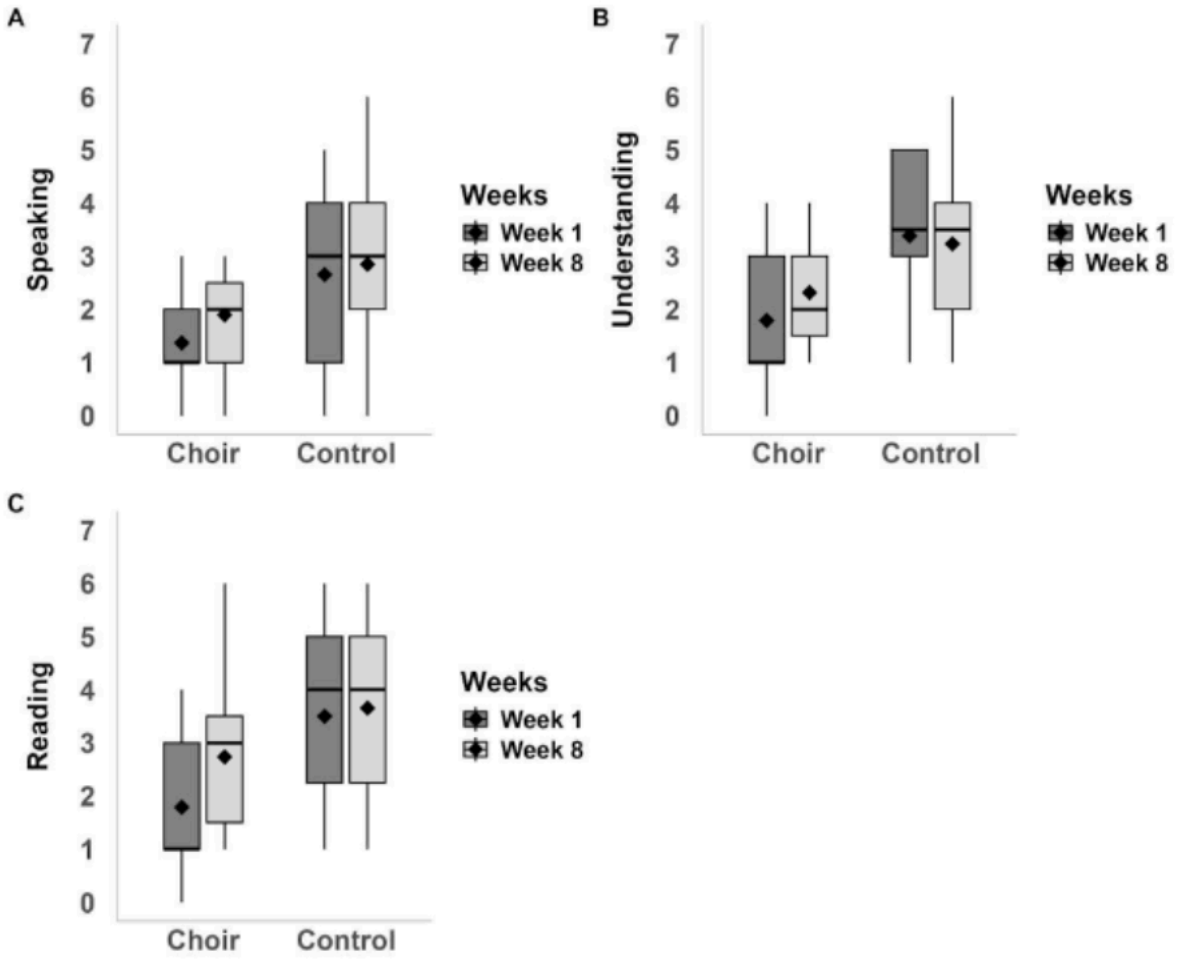


Figure 2. Boxplots of self-reported French proficiency scores. *Self-ratings of participants' speaking (A), understanding (B), and reading skills (C) in French, measured on an 8-point Likert scale (0 = "none", 7 = "excellent proficiency"). The horizontal line within each box represents the median; boxes indicate the interquartile range (IQR), and whiskers extend to the most extreme values within 1.5 times the IQR. Diamonds represent the means.*

Pronunciation/Intelligibility

To assess reliability, Intraclass Correlation Coefficients (ICCs) were calculated to evaluate agreement between the two human raters and between their average scores and the automated transcription. Results indicated excellent inter-rater reliability for both the pre-test (ICC(A,1) = .96, $F(18,13) = 65$, $p < .001$) and post-test (ICC(A,1) = .97, $F(18,18.9) = 88.1$,

$p < .001$). Agreement between the automated transcription and the mean human ratings was also high at both time points—pre-test (ICC(A,1) = .79, $F(28,20.4) = 10.2$, $p < .001$) and post-test (ICC(A,1) = .80, $F(28,28.3) = 8.92$, $p < .001$)—suggesting that the automated method is consistent with human evaluation. Given this consistency and its objectivity, all subsequent analyses were based on the automated transcription scores.

Average intelligibility scores in the choir group remained relatively stable from Week 1 (46%) to Week 8 (43%). Similarly, in the control group, scores showed minimal change, from 54% to 48% across the same period. A linear mixed-effects model revealed no significant main effects of Time ($\beta = -3.16$, $p = .319$), Group ($\beta = 8.37$, $p = .289$), or a Time × Group interaction ($\beta = -2.92$, $p = .482$).

It is worth noting that intelligibility scores exhibited substantial variability across participants (see **Figure 3**).

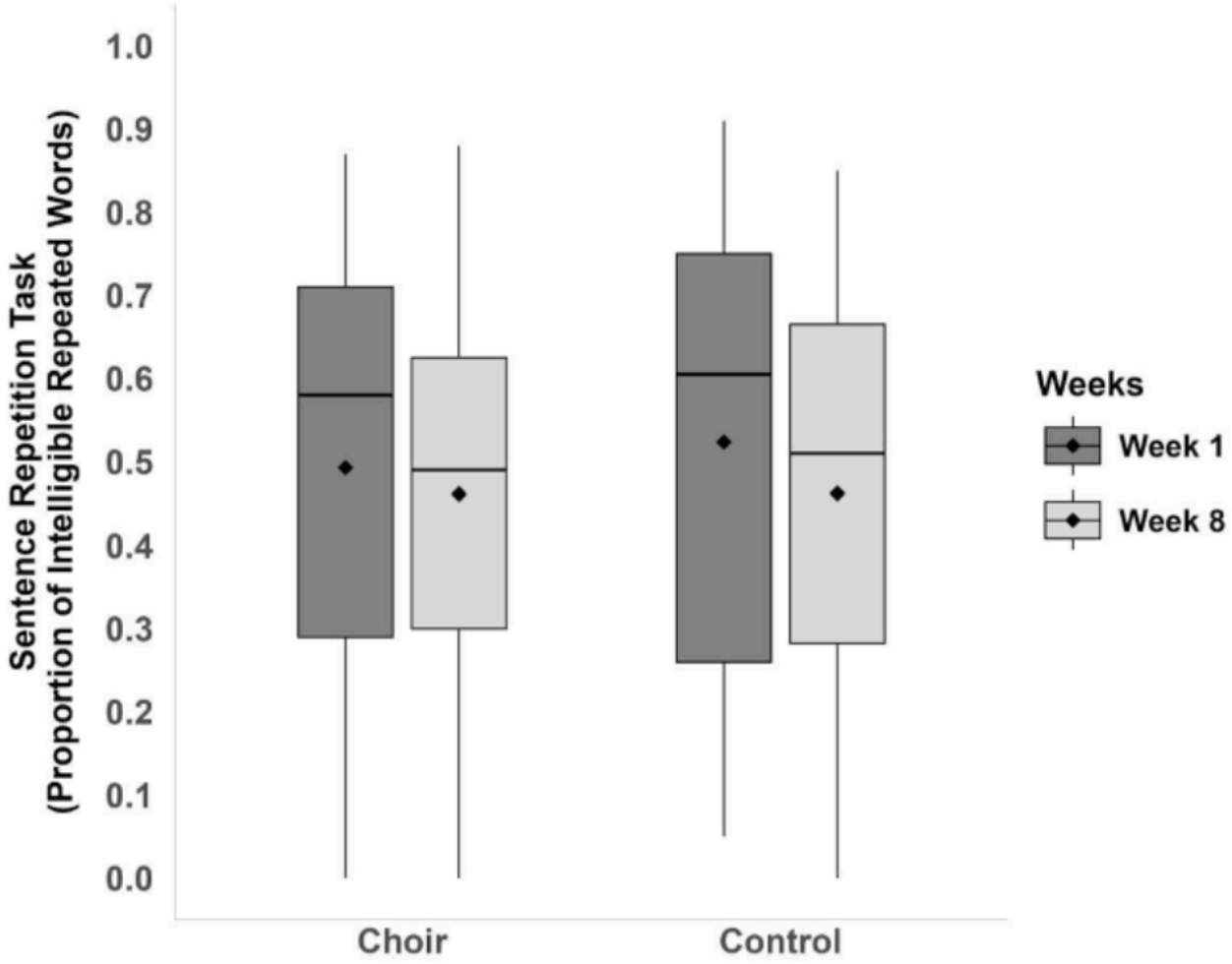


Figure 3. Boxplot of the proportion of intelligible words in the Sentence Repetition Task (SRT). *Proportion of correctly repeated (intelligible) words. The horizontal line within each box represents the median; boxes indicate the interquartile range (IQR), and whiskers extend to the most extreme values within 1.5 times the IQR. Diamonds represent the means.*

Impact of Initial French Proficiency Level

A linear mixed-effects model revealed that initial French proficiency level was not a statistically significant predictor of participants' performance on the sentence repetition task ($\beta = 3.11$, $p = .079$). However, there was a general main effect of Time ($F(1,43) = 5.01$, $p = .03$), indicating that all participants' pronunciation performance changed over the 8 week-period, regardless of their starting level.

Correlation Between Self-Reported Speaking Proficiency and Sentence Repetition Task Performance

A positive and significant correlation was observed between participants' self-assessed French speaking proficiency and their performance on the sentence repetition task (SRT) when data from both groups and time points were pooled. To test the robustness of this relationship, separate correlation analyses were conducted for pre- and post-assessment scores. In both cases, significant positive correlations were found across groups, supporting the link between subjective and objective measures of spoken language proficiency (see **Figure 4**).

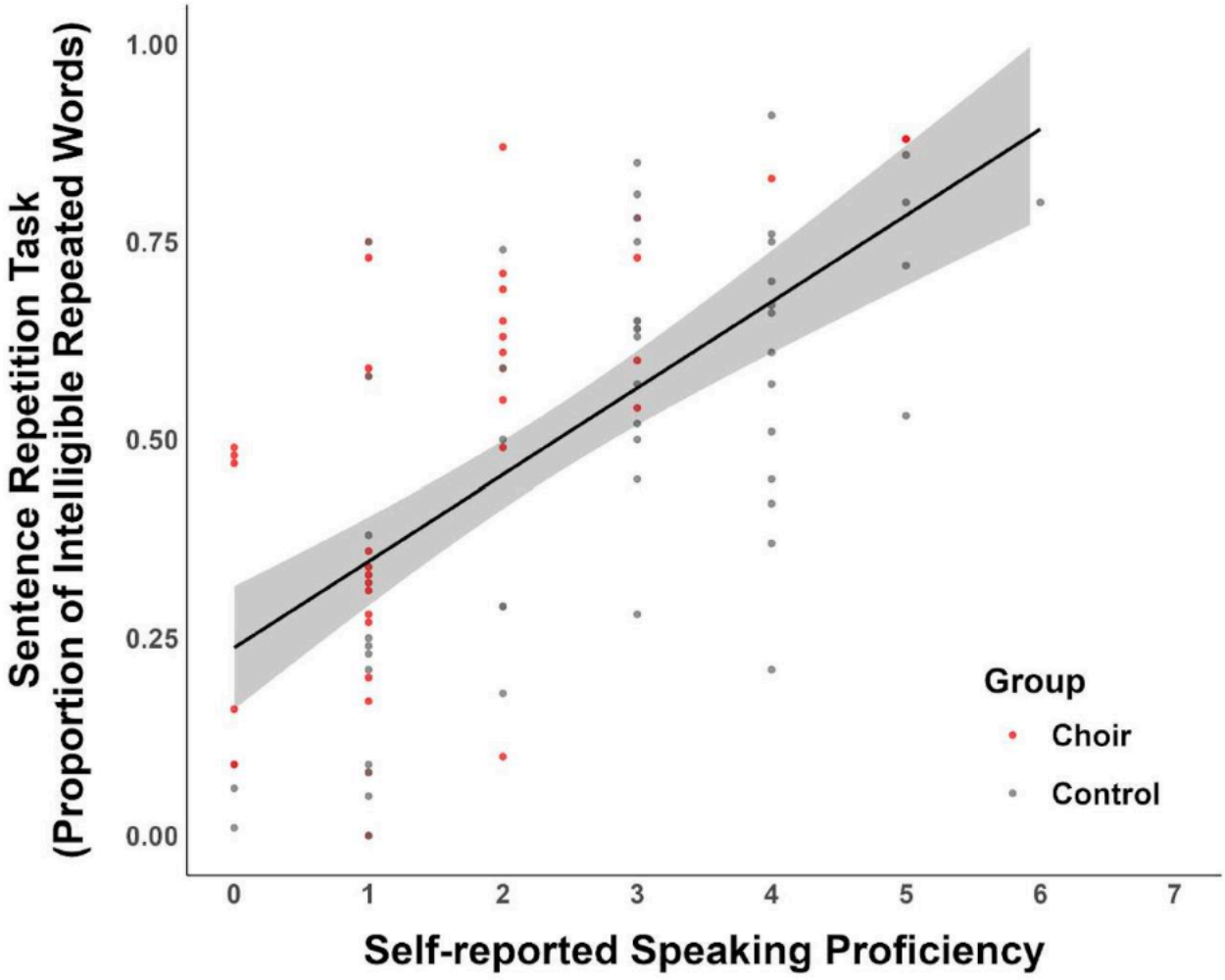


Figure 4. Correlation between participants' self-assessed French speaking proficiency and their performance in the Sentence Repetition Task (SRT). *The X-axis shows self-reported French speaking proficiency, rated on a scale from 0 (none) to 7 (excellent). The Y-axis represents the percentage of intelligible words repeated in the SRT.*

Sense of group belonging

A significant Time × Group interaction was found for participants' sense of group belonging (β = –1.450, SE = 0.446, $t(44) = –3.251$, $p = .002$). Analysis of estimated marginal means (EMMs) indicated a significant increase in the choir group's sense of group belonging, rising from moderate perceived overlap between "self" and "others" at Week 1 ($M = 3.35$, SE = 0.38) to relatively strong overlap at Week 8 ($M = 4.80$, SE = 0.38; β = –1.450, SE = 0.335, $t(44) = –4.324$, $p < .001$). In contrast, the control group showed no change in their sense of group belonging between sessions (Week 1: $M = 4.08$, SE = 0.33; Week 8: $M = 4.08$, SE = 0.33; β = 0, SE = 0.29, $t(44) = 0$, $p = 1$; see **Figure 5**)

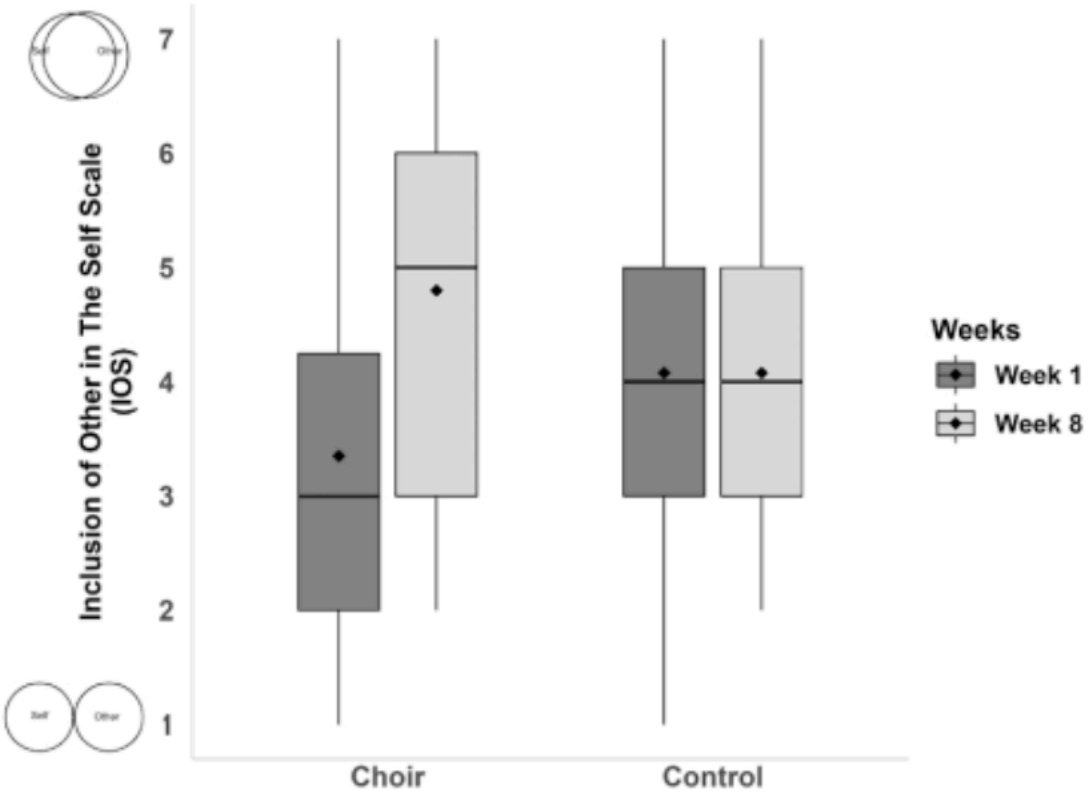


Figure 5. Boxplot of group belonging scores before and after the choir and waiting period. *Scores range from 1 (complete separation between "self" and "others") to 10 (complete overlap). The horizontal line within each box represents the median; boxes indicate the interquartile range (IQR), and whiskers extend to the most extreme values within 1.5 times the IQR. Diamonds represent the means.*

Mood

A significant Time × Group interaction was observed for happiness (β = –1.553, SE = 0.730, $t(44) = –2.128$, $p = .039$), indicating that changes in happiness scores differed between groups over time. Follow-up analyses using estimated marginal means (EMMs) revealed a significant

increase in happiness among choir participants, from Week 1 (M = 7.10, SE = 0.53) to Week 8 (M = 8.38, SE = 0.53; β = –1.280, SE = 0.548, $t(44) = -2.333$, $p = .02$). In contrast, control participants showed no significant change in happiness (Week 1: M = 8.13, SE = 0.47; Week 8: M = 7.85, SE = 0.57; see **Figure 6**).

No significant changes were observed in any of the other emotional states measured—energetic, angry, sad, confused, afraid, and tired—in either group (all $p > .05$).

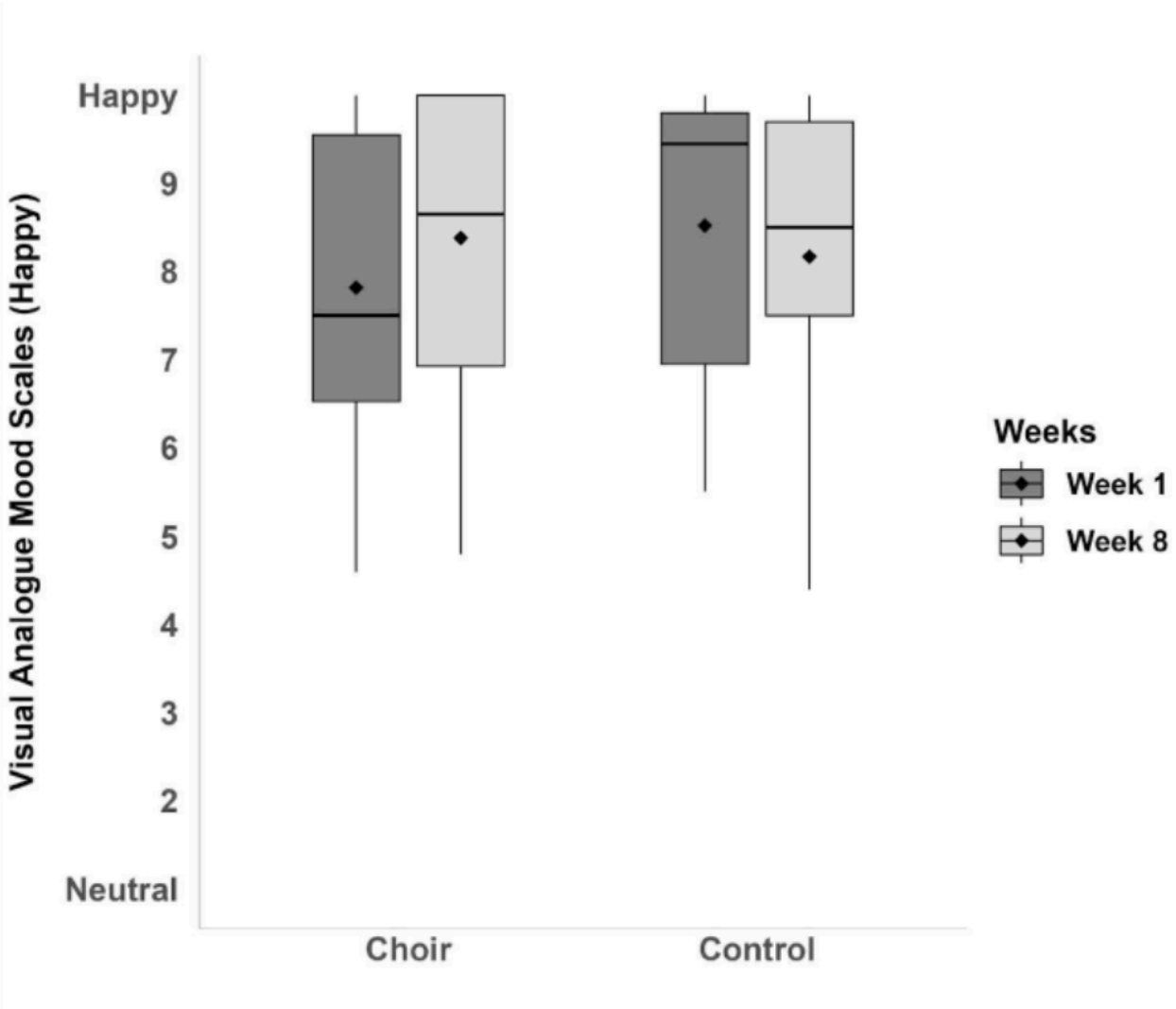


Figure 6. Boxplot of self-reported happiness scores before and after the choir and waiting period. *Scores range from 1 (absence of emotion) to 10 (intense happiness). The horizontal line within each box represents the median; boxes indicate the interquartile range (IQR), and whiskers extend to the most extreme values within 1.5 times the IQR. Diamonds represent the means.*

Expectations

All participants reported high expectations that choir participation would positively impact their sense of belonging, mood, and French proficiency, with median ratings of 10 out of 10 for all three variables. No significant differences were observed between the choir and control groups.

Semi-structured interviews

French Acquisition and Proficiency

A majority of participants (8 out of 13) reported having recommended the choir to other newcomers who were not fluent in French. More than half (7 participants) noted improvements in their French vocabulary and pronunciation. As one participant explained:

"Now I understand how I should pronounce and when I should use and how I should build a sentence." (Interviewee 10)

Two participants specifically highlighted vocabulary gains, while two others emphasized improved pronunciation. Several participants also reported enhanced speaking and comprehension skills:

"When I'm surrounded by people or when I'm in the metro, I started to understand words—like what they're talking about in the metro, or some posters I can read; and I can understand what they mean. I can catch what they mean by one word, and when I ask someone if the meaning I guessed is correct, they say yes!" (Interviewee 10)
"Sometimes I read a French text, and I find a word that I've seen in the choir." (Interviewee 12)

Participants described a shift in their relationship to the French language—from feelings of *fear*, *complexity*, and *avoidance* to *confidence* and *courage*. For example:

"I've started to understand that French is not something I should be afraid of. French should be afraid of me because I just need to focus on forgetting my bad experience with it and start to speak no matter what I need to do with it." (Interviewee 10)

The choir environment was described as "more relaxed," "more joyful," and "recharging for learners" compared to traditional French classes, which were sometimes seen as stressful or boring.

"When you sing, you don't have to think a lot." (Interviewee 5)

Social Integration and Well-Being

Almost all participants (12 out of 13) said they would recommend choir participation to newcomers as a means of integrating into society. Twelve reported that singing in a group helped them connect with others. The choir was often described as a "society" where participants could

interact with people from different cultures. Three exchanged contact information; four made new friends, and six formed "singing colleagues."

Several participants emphasized the value of having Francophone members to encourage the use of French:

"I think some of the native speakers did encourage us to speak French. Once I tried to speak in French, I got really positive feedback from them. So I feel really good during the break time." (Interviewee 3)

Many initially gravitated toward others from their own cultural background, but gradually opened up to the larger group as the atmosphere became more welcoming:

"The atmosphere makes you feel open… and it's really good for communication." (Interviewee 2)

Feasibility

Six participants felt that singing and language learning were well balanced in the choir sessions:

"It's a good balance, because even when we are singing, we are learning the pronunciation." (Interviewee 11)

Several noted that the balance depended on the complexity of the song:

"I think that it depends on the song; if the song is simple, we can sing more than focus on understanding the words. And if it's complex, we should spend more time on understanding and learning the new words." (Interviewee 9)

Participants identified several facilitators that supported their continued participation. The most frequently cited was the convenient location near a metro station. Other positive aspects included: the professionalism of the choir director, the two-hour session length, the coffee break, the opportunity to learn about the historical background of songs, and the focus on vocabulary, pronunciation, and the diversity and emotional positivity of the song selection. Vocal techniques and warm-up exercises were also appreciated.

The primary obstacle reported was the short eight-week duration of the program, which was viewed as insufficient for experiencing significant gains in French learning. Some participants also expressed a desire for a shared goal, such as a mini-concert, to enhance motivation and group cohesion:

"Having a mini concert can be a great thing to think about. It will be easier for us to bond if we have a common goal." (Interviewee 2)

Other challenges included the complexity of certain melodies—particularly for those with no prior musical background:

"Probably you should think about the songs and melodies. Because when the melody is too complicated, I need to work more on the melody than on my pronunciation or the song itself. To talk to each other, we need something in common. The participation in choir is not enough to find something in common to speak about." (Interviewee 14)

Additional concerns included the short duration of breaks (20–25 minutes), the lack of English translations, and scheduling conflicts for a few participants.

Discussion

The results indicate that participating in a choir for eight weeks significantly enhanced positive affect and feelings of connectedness compared to a wait-list group. The findings align with previous research demonstrating the psychosocial benefits of choir participation[19,20,23,23–25,29]. The quantitative findings are further supported by qualitative data, in which participants described the choir as a small "society" that fostered social interaction, new friendships, and a sense of belonging among "singing colleagues". These reports are consistent with earlier studies on the social impact of group singing[33–36].

In addition, and in line with earlier work[13], qualitative responses highlighted perceived improvements in French acquisition—particularly in vocabulary, pronunciation, speaking, comprehension, and reading skills. Participants also reported a shift in their relationship with the language, from anxiety and avoidance to motivation and confidence. These experiences were corroborated by self-assessment data, which showed significant gains in self-rated French proficiency over time. Notably, participants' self-assessed proficiency was positively correlated with their objective performance on the sentence repetition task, suggesting that participants had a realistic understanding of their language abilities.

However, unlike prior studies[9,11,12], we did not observe a significant effect of choir participation on pronunciation as measured by the sentence repetition task. Both groups exhibited slightly negative trends. To better understand this finding, we conducted an item difficulty analysis, which revealed that the test items in Version B were more challenging than those in Version A. This discrepancy may explain the decline in scores and suggests the need for more balanced test versions in future research.

Covariate analyses further showed that initial French proficiency did not significantly predict performance on the sentence repetition task. This suggests that the choir intervention was beneficial regardless of participants' starting proficiency level and supports its relevance for learners with varying linguistic backgrounds and skill levels. While task difficulty may have

introduced some variability in outcomes, the overall pattern underscores the potential of choir participation to support L2 development across a broad range of abilities.

Taken together, these findings reinforce the psychosocial and language learning benefits of group singing. Choir participation emerges as an inclusive, engaging, and low-barrier approach to second language acquisition—one that simultaneously supports emotional well-being and social integration.

STUDY 2: WITHIN-SUBJECTS DESIGN: FROM WAITLIST TO CHOIR

Methods

Participants

Eighteen of the 26 control participants of Study 1 joined the choir after the 8 weeks of the waiting period. Here, we compared their data in the waiting period to their data after 8 weeks of choir participation. Participant demographics are presented in **Table 2**.

Table 2. Demographic characteristics of Participants

	n (%)
n	18
Average age (Range)	38.5 (23-54)
Gender	
Females	13 (81.2)
Males	3 (18.7)
Educational level	
Secondary	1 (6.2)
University	15 (93.7)
Country	
Ukraine	5 (31.25)
China	3 (18.75)
Iran	2 (12.50)
Brazil	2 (12.50)
Egypt	1 (6.25)
Mexico	1 (6.25)
Colombia	1 (6.25)

Living in Quebec (years)	
< 1	11 (68.7)
1 - 2	4 (25)
2 - 3	1 (6.2)
French courses (level)	14 (77.7)
Beginner	3 (21.4)
Intermediate	8 (57.1)
Advanced	3 (21.4)
Musical experience	
Number of participants (%)	9 (43.7)
Mean years	3.54

Procedure

Participants were tested three times: before (week 1) and after (week 8) the waiting period, and after eight weeks of choir (week 16). The procedure was the same as in Study 1 (see **Figure 7** for study's experimental design).

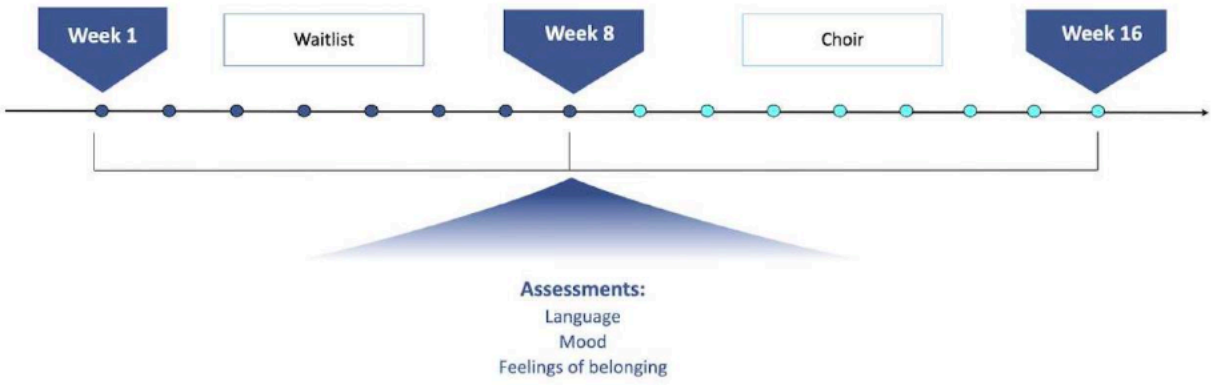


Figure 7. Experimental design of Study 2

Material

The same measures used in Study 1 were employed to assess participants' mood (Visual Analogue Mood Scales), sense of group belonging (Inclusion of Other in Self scale, IOS), and self-assessed French proficiency (questionnaire). In addition, a picture description task was included to obtain a connected speech sample for assessing oral production and fluency. This

task was adapted from the *Picnic Scene Test* in the Western Aphasia Battery (WAB)[44]. Participants were asked to describe the scene using their own words and phrases. The task was untimed and conducted on a computer; speech productions were recorded following the same procedure as in Study 1.

Data analysis

Data were analyzed using R software (version 4.3.3)[41]. Linear mixed-effects models (LMMs) were used, with Time (Week 1, Week 8, and Week 16) as a fixed effect and a random intercept for participants to account for repeated measures. Models were fitted using restricted maximum likelihood (REML), and Satterthwaite's method was used to approximate degrees of freedom for the significance testing of fixed effects. When applicable, post hoc comparisons were adjusted using the Holm correction.

To examine the effect of choir participation on language improvement while controlling for baseline French proficiency, an LMM was fitted with fixed effects for Time (Week 1, Week 8, Week 16), French level (coded from 0 to 8), and their interaction, along with a random intercept for each participant.

To assess the relationship between self-assessed French proficiency and actual performance on the picture description task, Pearson correlation coefficients were calculated separately at each time point.

Results

Self-reported French proficiency

Participants reported substantial improvements in their French speaking proficiency, increasing from Week 1 ($M = 2.59$, $SD = 1.28$) to Week 16 ($M = 4.00$, $SD = 1.46$; $\beta = 1.25$, $p < .001$, Holm-adjusted $p = .0001$). This gain was also significant between Week 8 ($M = 2.94$, $SD = 1.43$) and Week 16 ($\beta = 0.90$, $p = .003$, Holm-adjusted). They also reported significant gains in their French listening comprehension, with scores rising from Week 1 ($M = 3.32$, $SE = 0.38$) to Week 16 ($M = 4.07$, $SE = 0.38$; $p = .009$, Holm-adjusted $p = .02$), and from Week 8 ($M = 3.38$, $SE = 0.38$) to Week 16 ($p = .03$, Holm-adjusted). However, no significant changes were observed in their self-assessed French reading proficiency across the three time points ($p > .05$; see **Figure 8**).

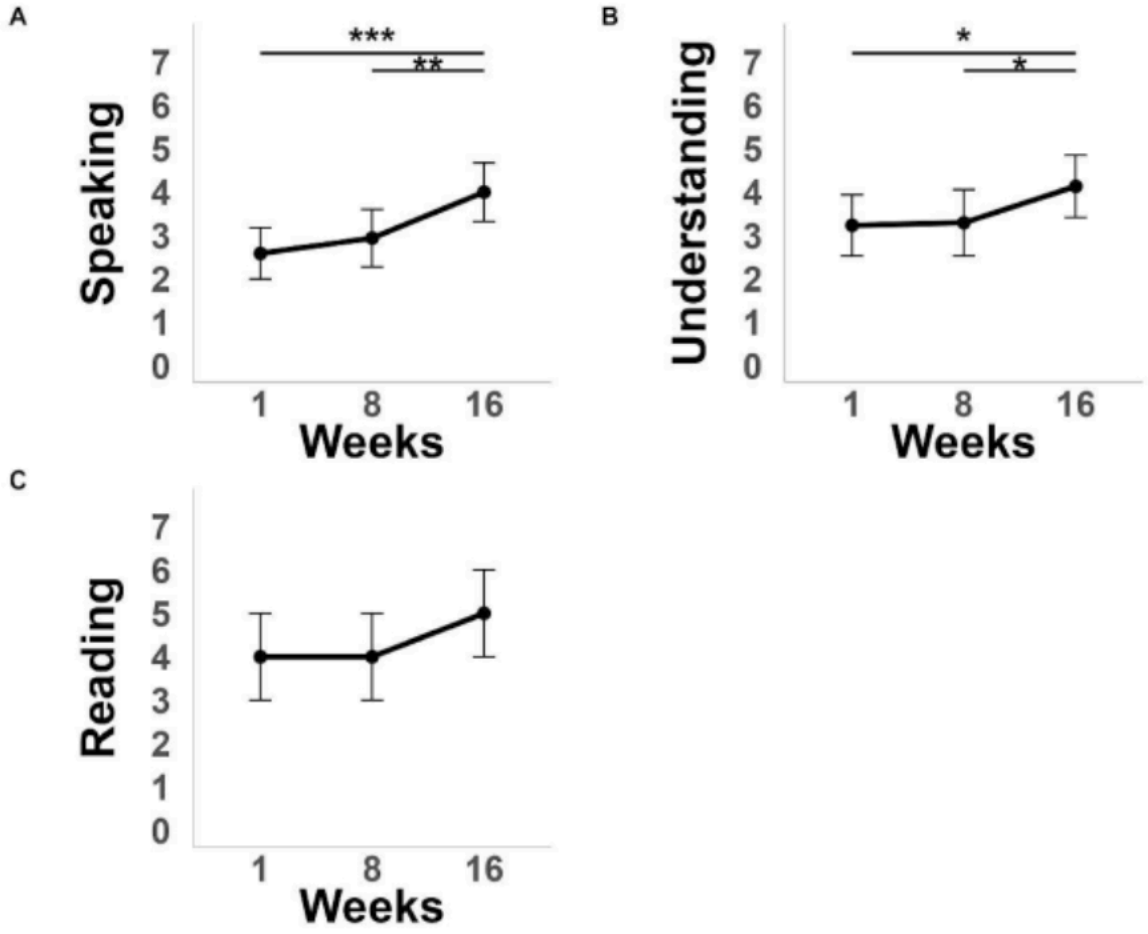


Figure 8. Self-assessments of French proficiency across sessions. *Mean scores for French speaking (A), understanding (B), and reading (C) (solid lines), along with standard error of the mean (SEM), at Week 1 (baseline), Week 8 (post-waitlist), and Week 16 (post-choir). Ratings were made on an 8-point Likert scale, where 0 indicates "none" and 7 indicates "excellent" French proficiency.*

Picture description

Participants' responses in the picture description task were transcribed using an AI-based speech recognition tool (https://www.veed.io/). The transcriptions were reviewed and corrected by the first author. Overall, 72% of the responses were accurately detected by the automated algorithm. The number of intelligible words produced by each participant is presented in **Figure 9**.

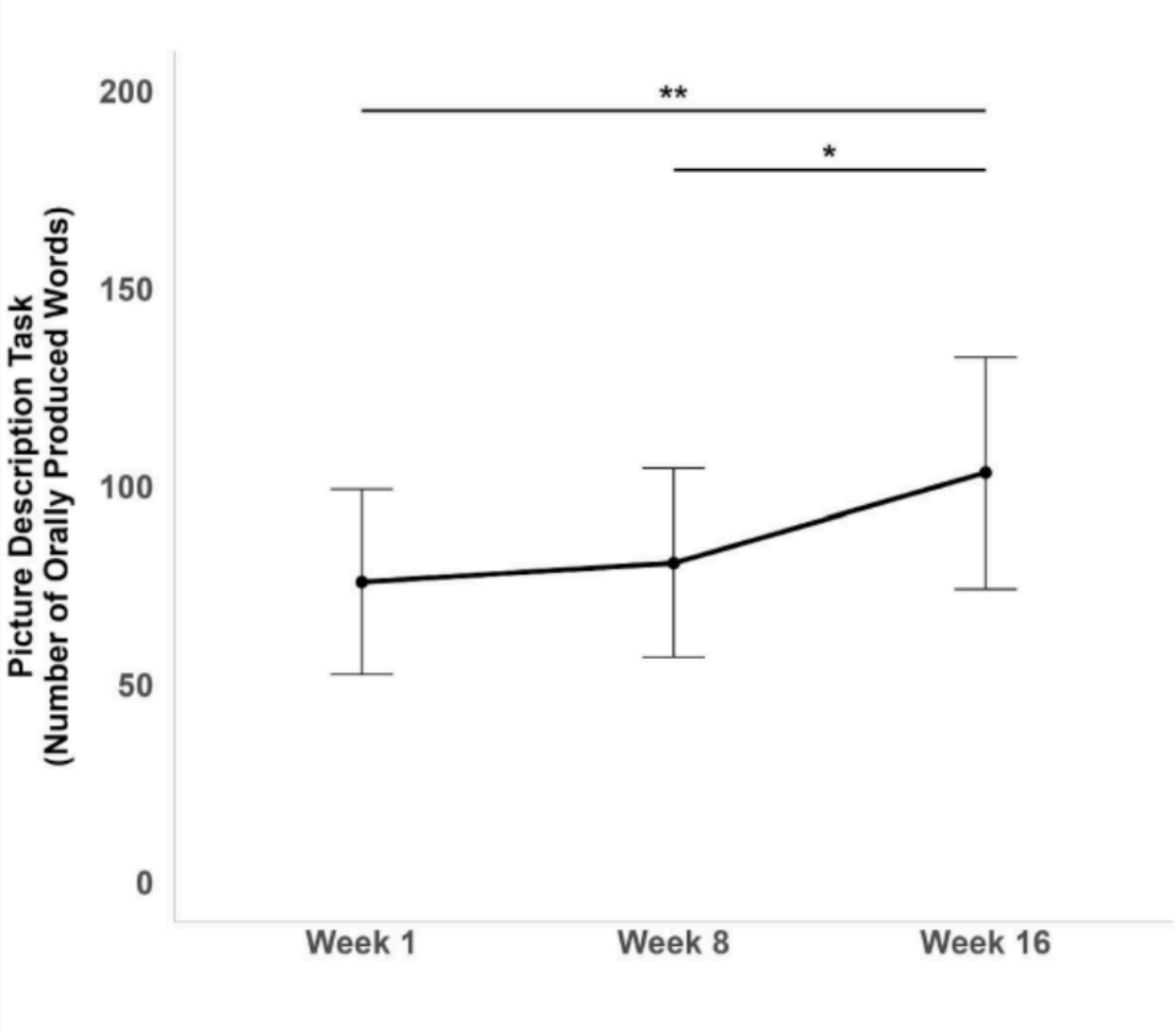


Figure 9. Picture description test. *Mean number of orally intelligible words produced in the Picnic Scene, along with the corresponding standard errors (SEM) at weeks 1 (baseline), 8 (post waitlist) and 16 (post choir) of study 2.*

As can be seen, participants produced significantly more words after the choir intervention at Week 16 ($M = 106$, $SD = 58.2$) compared to Week 1 ($M = 75.7$, $SD = 47$; $\beta = 30.22$, $p < .001$, Holm-adjusted $p = .0005$) and Week 8 of the waiting period ($M = 79.8$, $SD = 48.3$; $\beta = –26.06$, $p = .002$, Holm-adjusted). No significant difference was found between Week 1 and Week 8 ($\beta = 4.17$, $p = .56$, Holm-adjusted $p = 1.00$), suggesting that the increase in word production occurred after choir participation, rather than during the waiting period.

Results from the covariate analysis revealed that participants' initial French proficiency was a significant predictor of their performance on the picture description task ($p = .05$), indicating that individuals with higher baseline proficiency tended to produce more words overall. However, the interaction between Time and French level was not statistically significant at either Week 8 ($p = .33$) or Week 16 ($p = .93$). This suggests a negligible moderating effect of initial proficiency

on performance gains over time, implying that the benefits of choir participation were relatively consistent across varying proficiency levels.

Correlation analyses further supported this finding. A moderate and significant correlation was observed between participants' self-reported speaking proficiency and their actual performance on the picture description task ($R = .52$, $p = .0001$), indicating good alignment between subjective and objective measures.

Finally, when pooling data from Week 1 and Week 8 (i.e., during the waiting period), a moderate and significant positive correlation was found between scores on the sentence repetition task and the picture description task ($R = .56$, $p < .0001$; see **Figure 10**), further validating the consistency of language performance measures across different tasks.

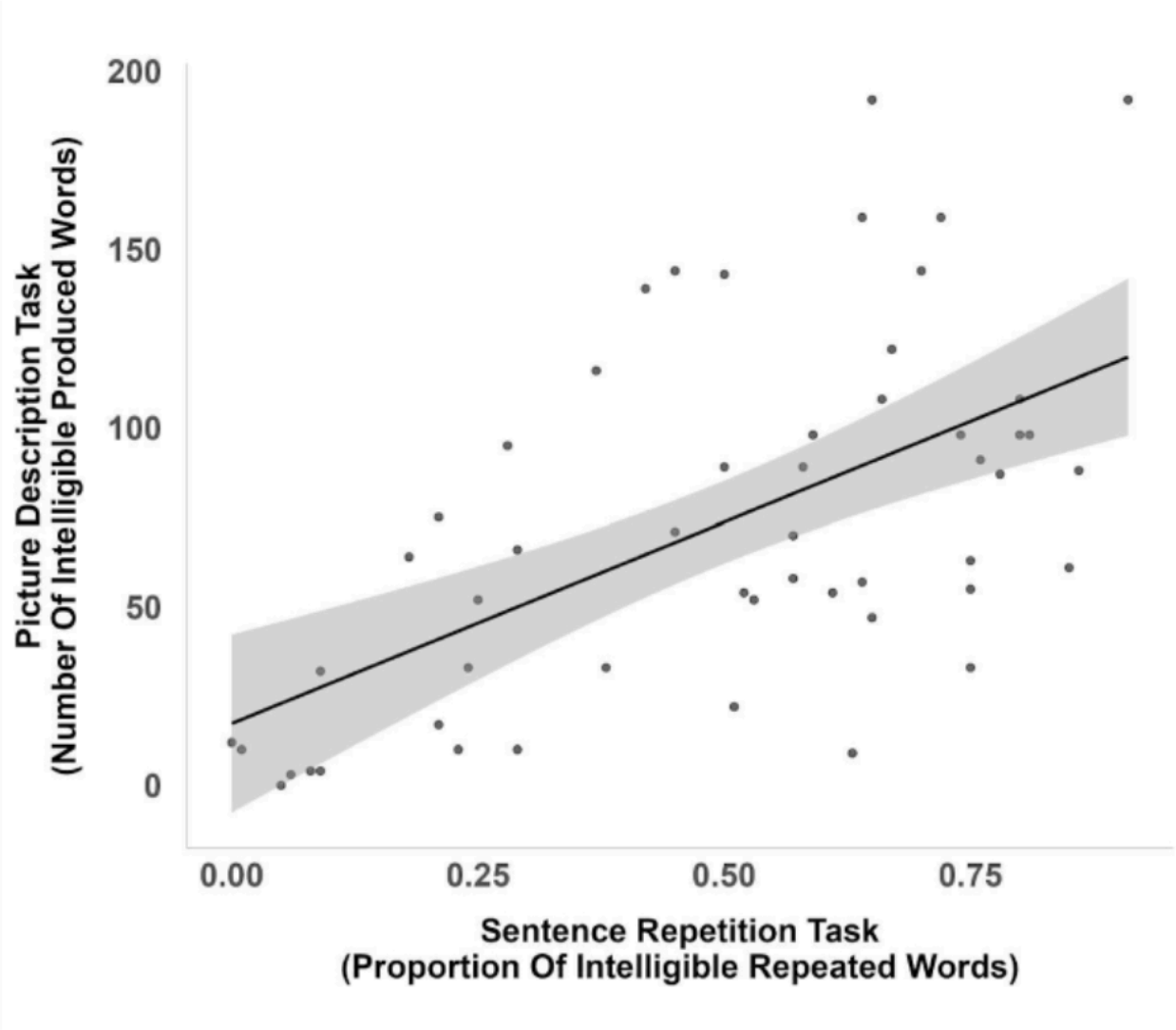


Figure 10. Correlation between the Sentence Repetition Task (SRT) and the Picture Description Task. *Each dot represents the scores of one participant.*

Sense of belonging

No significant change in IOS scores was observed over the waiting period (β = –0.18, p = .63). However, participants' sense of group belonging improved following choir participation (β = 0.79, p = .03). Pairwise comparisons revealed a significant increase in IOS scores from Week 1 (M = 4.23, SD = 1.71) to Week 16 (M = 5.02, SD = 1.39; β = 0.79, p = .03, Holm-adjusted). A significant improvement was also found between Week 8 of the waiting period (M = 4.06, SD = 1.52) and Week 16 of choir participation (β = 0.96, p = .02, Holm-adjusted; see **Figure 10**).

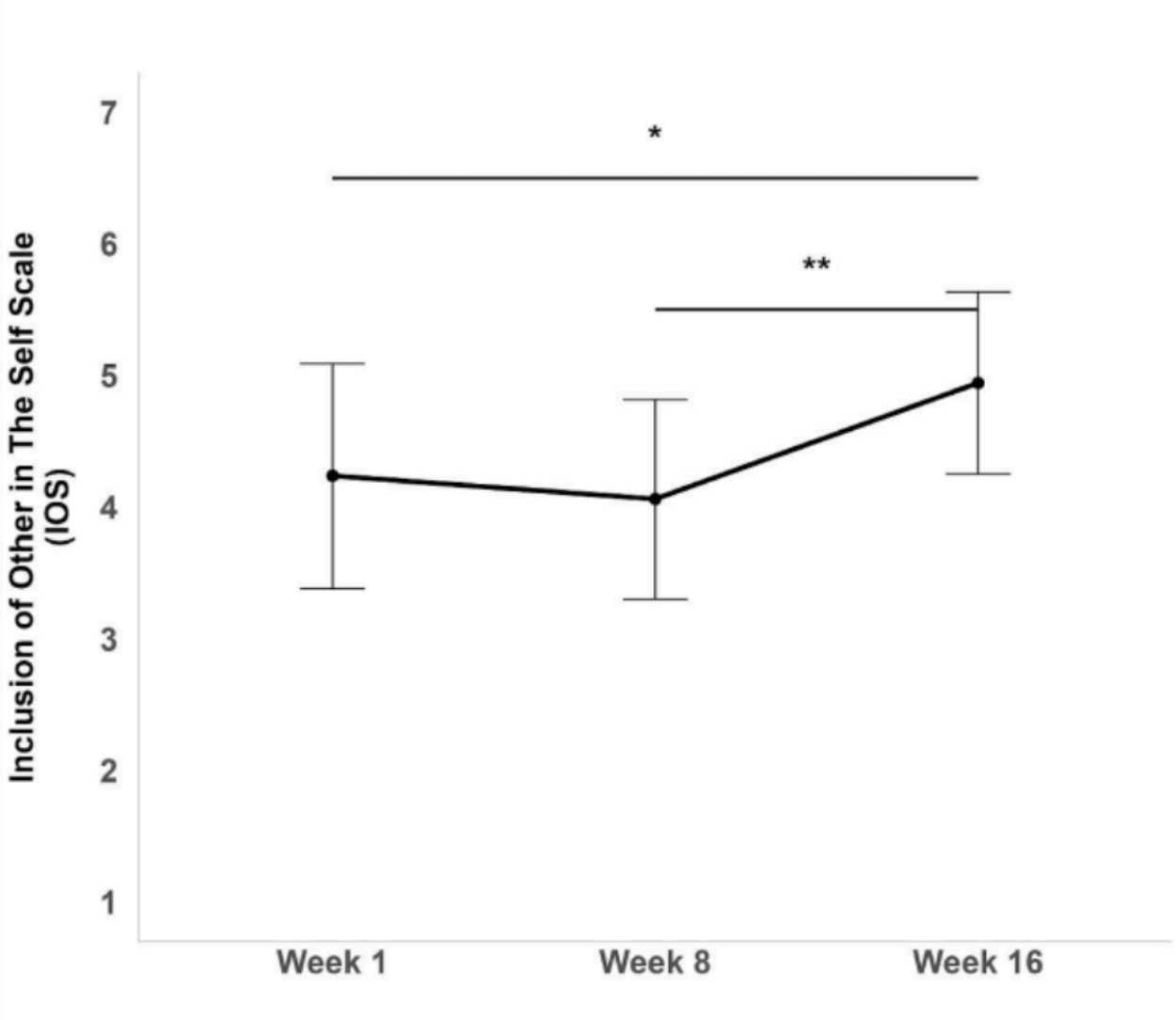


Figure 11. Sense of group belonging. *Mean scores and standard errors of the mean (SEM) of sense of group belonging scores at weeks 1 (baseline), 8 (post wait) and 16 (post choir) of the study*

Mood

Although participants reported feeling slightly happier at Week 1 (M = 8.37, SD = 2.62) than at Week 8 (M = 7.58, SD = 2.88) and Week 16 (M = 7.38, SD = 2.63), this difference was not

statistically significant (β = –0.39, p = .60). No significant changes were observed in other mood dimensions, including energetic, sad, confused, tired, and angry (all p > .05).

Discussion

In line with Study 1, results from the second study revealed a positive impact of choir participation on sense of group belonging[22–24,28–30] and self-reported French proficiency. The important finding is that the picture description task used here in Study 2 revealed significant improvements in participants' oral production, consistent with prior research[9,11,13,45]. No significant improvements in mood were observed though.

GENERAL DISCUSSION

Across two studies, we examined the impact of choir participation on newcomers' linguistic, emotional, and social integration. We hypothesized that singing in a group for eight weeks would enhance French proficiency, improve mood, and strengthen the sense of group belonging compared to a non-singing control period. The results largely supported these predictions.

In both studies, participants reported significant improvements in self-assessed French proficiency, particularly in speaking and understanding. These subjective gains were supported by qualitative accounts of increased confidence, vocabulary growth, and more positive attitudes toward the French language. To assess language skills objectively, we used a sentence repetition task in Study 1 and a picture description task in Study 2. While both tasks tapped into similar skills such as pronunciation, only the measure of spontaneous, connected speech proved sensitive to the effects of choir participation. Participants in Study 2 produced significantly more words after the group singing activity, suggesting enhanced verbal fluency. This aligns with previous research highlighting the language-learning benefits of singing[9–15,17] and extends these findings beyond song-specific vocabulary to broader linguistic competence.

Importantly, these language proficiency gains were captured using a semi-automated, AI-based speech transcription tool. The high agreement between automated and human ratings in our studies underscores the potential of machine learning tools as valid, efficient, and scalable alternatives to time-consuming human evaluation. Automated transcription not only reduces costs and effort but also minimizes risks of human bias and fatigue[46].

Both studies also revealed a consistent and significant improvement in participants' sense of group belonging following choir participation. Participants described the choir as a welcoming space that promoted connection, cultural exchange, and mutual support—what some called a "small society." These findings build on earlier work demonstrating the social bonding effects of group singing[19,22–24,28,29,33–36], and extend it to a new population: newcomers with diverse cultural

backgrounds and urgent integration needs. This adds to the literature, which has primarily focused on university students or amateur choristers[19,22–24,28].

Notably, the benefits of choir participation were observed across varying levels of French proficiency. Covariate analyses showed that baseline language skills did not significantly moderate the observed improvements, suggesting the accessibility of choir-based activities for individuals with a wide range of linguistic backgrounds. Moreover, significant correlations between self-reported and objective measures of language proficiency suggest that participants had a realistic understanding of their progress.

Contrary to expectations, however, no significant improvements in mood were observed in Study 2, despite the positive emotional effects seen in Study 1. Several factors may account for this discrepancy, including elevated baseline happiness scores in Study 2 and the possibility that direct self-evaluation of mood may be perceived as intrusive or culturally inappropriate by individuals in the midst of adjusting to a new country. Social desirability bias, known to be elevated among newcomers[47,48], along with potential questionnaire fatigue due to repeated testing, may have influenced responses. While self-reports are valuable and widely used in the literature, incorporating objective indicators of well-being in future research would be beneficial. Biomarkers such as cortisol and oxytocin may provide deeper insights into physiological responses to social bonding and well-being[25,27,29,49].

Future studies would benefit from incorporating more objective measures of both well-being and social communication. Including a non-musical active control group (e.g., a drama group, a conversation class) and implementing randomized group assignments would also strengthen causal conclusions. The present findings offer a strong rationale for pursuing these directions.

Taken together, these studies reinforce the potential of choir singing as a low-barrier, engaging, and multifaceted tool for newcomer integration. By simultaneously supporting language development, social connection, and emotional adjustment, choir participation emerges as a promising complement to more traditional integration strategies.

REFERENCES

1. Statistique Canada. (2023). Projections démographiques pour le Canada et ses régions sur les thèmes de l'immigration et de la diversité, 2016 à 2041: Aperçu des hypothèses et des scénarios. 17. https://www150.statcan.gc.ca/n1/pub/17-20-0001/172000012022001-fra.htm
2. Gouvernement du Canada, S. C. (2022, October 26). Près de 1 personne sur 4 au Canada est un immigrant, ce qui représente la plus forte proportion de la population en plus de 150 ans. https://www150.statcan.gc.ca/n1/daily-quotidien/221026/g-a001-fra.htm
3. Chiswick, B. R., & Miller, P. W. (2004). Language Skills and Immigrant Adjustment: What Immigration Policy Can Do! Available at SSRN 632341. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=632341
4. Pottie, K. (2007). Misinterpretation: Language proficiency, recent immigrants, and global health disparities. Canadian Family Physician, 53(11), 1899–1901.
5. Gouvernement du Québec. (2022). Projet de loi no 96: Loi sur la langue officielle et commune du Québec, le français. Éditeur Officiel Du Québec. https://www.publicationsduquebec.gouv.qc.ca/fileadmin/Fichiers_client/lois_et_reglements/LoisAnnuelles/fr/2022/2022C14F.PDF
6. Gouvernement de Québec. (2024). Services d'apprentissage du français. Gouvernement du Québec. https://www.quebec.ca/immigration/statistiques-recherches-immigration/francisation
7. Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. Psychology of Music, 43(5), 627–640. https://doi.org/10.1177/0305735614528833
8. Richards, R. G. (1975). Singing: A fun route to a second language. The Reading Teacher, 29(3), 283–285.
9. Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. Psychology of Music, 43(5), 627–640.
10. Baills, F., Zhang, Y., Cheng, Y., Bu, Y., & Prieto, P. (2021). Listening to songs and singing benefitted initial stages of second language pronunciation but not recall of word meaning. Language Learning, 71(2), 369–413.
11. Zhang, Y., Baills, F., & Prieto, P. (2023). Singing Songs Facilitates L2 Pronunciation and Vocabulary Learning: A Study with Chinese Adolescent ESL Learners. Languages, 8(3), 219.
12. Ludke, K. M. (2018). Singing and arts activities in support of foreign language learning: An exploratory study. Innovation in Language Learning and Teaching, 12(4), 371–386.
13. Lehtinen-Schnabel, J., & Levänen, S. (2024). Singing Language – Integrating Second Language Learning into Choir Practice. Music & Science, 7, 20592043241244542. https://doi.org/10.1177/20592043241244542

14. Davis, G. M., & Fan, W. (2016). English Vocabulary Acquisition Through Songs in Chinese Kindergarten Students. Chinese Journal of Applied Linguistics, 39(1). https://doi.org/10.1515/cjal-2016-0004
15. Chen, I.-S. J. (2020). Music as a mnemonic device for foreign vocabulary learning. English Teaching & Learning, 44(4), 377–395.
16. Busse, V., Jungclaus, J., Roden, I., Russo, F. A., & Kreutz, G. (2018). Combining Song—And Speech-Based Language Teaching: An Intervention With Recently Migrated Children. Frontiers in Psychology, 9, 2386. https://doi.org/10.3389/fpsyg.2018.02386
17. Busse, V., Hennies, C., Kreutz, G., & Roden, I. (2021). Learning grammar through singing? An intervention with EFL primary school learners. Learning and Instruction, 71, 101372. https://doi.org/10.1016/j.learninstruc.2020.101372
18. Alisaari, J., & Heikkola, L. M. (2016). Increasing fluency in L2 writing with singing. Studies in Second Language Learning and Teaching, 6(2), 271–292. https://doi.org/doi: 10.14746/ssllt.2016.6.2.5
19. Bullack, A., Gass, C., Nater, U. M., & Kreutz, G. (2018). Psychobiological effects of choral singing on affective state, social connectedness, and stress: Influences of singing activity and time course. Frontiers in Behavioral Neuroscience, 12, 223.
20. Coulton, S., Clift, S., Skingley, A., & Rodriguez, J. (2015). Effectiveness and cost-effectiveness of community singing on mental health-related quality of life of older people: Randomised controlled trial. The British Journal of Psychiatry, 207(3), 250–255. https://doi.org/10.1192/bjp.bp.113.129908
21. Kreutz, G. (2014). Does singing facilitate social bonding? Music and Medicine, 6(2), 51–60.
22. Pearce, E., Launay, J., Machin, A., & Dunbar, R. I. (2016). Is group singing special? Health, well-being and social bonds in community-based adult education classes. Journal of Community & Applied Social Psychology, 26(6), 518–533.
23. Pearce, E., Launay, J., & Dunbar, R. I. (2015). The ice-breaker effect: Singing mediates fast social bonding. Royal Society Open Science, 2(10), 150221.
24. Pearce, E., Launay, J., van Duijn, M., Rotkirch, A., David-Barrett, T., & Dunbar, R. I. (2016). Singing together or apart: The effect of competitive and cooperative singing on social bonding within and between sub-groups of a university Fraternity. Psychology of Music, 44(6), 1255–1273.
25. Schladt, T. M., Nordmann, G. C., Emilius, R., Kudielka, B. M., de Jong, T. R., & Neumann, I. D. (2017). Choir versus solo singing: Effects on mood, and salivary oxytocin and cortisol concentrations. Frontiers in Human Neuroscience, 11, 430.
26. Boyd, M., von Ranson, K. M., Whidden, C., & Frampton, N. (2020). Short-term effects of group singing versus listening on mood and state self-esteem. Psychomusicology: Music, Mind, and Brain, 30(4), 178.
27. Good, A., & Russo, F. A. (2022). Changes in mood, oxytocin, and cortisol following group and individual singing: A pilot study. Psychology of Music, 50(4), 1340–1347.

28. Pearce, E., Launay, J., MacCarron, P., & Dunbar, R. I. (2017). Tuning in to others: Exploring relational and collective bonding in singing and non-singing groups over time. Psychology of Music, 45(4), 496–512.
29. Weinstein, D., Launay, J., Pearce, E., Dunbar, R. I., & Stewart, L. (2016). Group music performance causes elevated pain thresholds and social bonding in small and large groups of singers. Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society, 37(2), 152.
30. Anshel, A., & Kipper, D. A. (1988). The influence of group singing on trust and cooperation. Journal of Music Therapy, 25(3), 145–155.
31. Bowling, D. L., Gahr, J., Ancochea, P. G., Hoeschele, M., Canoine, V., Fusani, L., & Fitch, W. T. (2022). Endogenous oxytocin, cortisol, and testosterone in response to group singing. Hormones and Behavior, 139, 105105.
32. Schafer, R., Bonakdarpour, B., Miller, S. S., & Miller, J. (2023). Teleconference group singing programs for well-being of healthy older adults and persons with neurocognitive disorders during the COVID-19 pandemic. Alzheimer's & Dementia, 19(S19), e077378. https://doi.org/10.1002/alz.077378
33. Jacob, C., Guptill, C., & Sumsion, T. (2009). Motivation for continuing involvement in a leisure-based choir: The lived experiences of university choir members. Journal of Occupational Science, 16(3), 187–193. https://doi.org/10.1080/14427591.2009.9686661
34. Lally, E. (2009). The power to heal us with a smile and a song: Senior Well-being, Music-based Participatory Arts and the Value of Qualitative Evidence. Journal of Arts & Communities, 1(1), 25–44. https://doi.org/10.1386/jaac.1.1.25_1
35. Tonneijck, H. I. M., Kinébanian, A., & Josephsson, S. (2008). An exploration of choir singing: Achieving wholeness through challenge. Journal of Occupational Science, 15(3), 173–180. https://doi.org/10.1080/14427591.2008.9686627
36. Louhivuori, J., Salminen, V.-M., & Lebaka, E. (2005). "Singing Together": A Crosscultural Approach to the Meaning of Choirs As a Community. In Cultural diversity in music education: Directions and challenges for the 21st century (pp. 81–92). Australian Academic Press Bowen Hills, Qld. https://search.informit.org/doi/abs/10.3316/INFORMIT.386981901084587
37. Desbiens, D., Laurier, M. D., & Leroux, J. (2011). Échelle québécoise des niveaux de compétence en français des personnes immigrantes adultes. Québec: Gouvernement Du Québec.
38. Stern, R. A., Arruda, J. E., Hooper, C. R., Wolfner, G. D., & Morey, C. E. (1997). Visual analogue mood scales to measure internal mood state in neurologically impaired patients: Description and initial validity evidence. Aphasiology, 11(1), 59–71.
39. Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. Journal of Personality and Social Psychology, 63(4), 596.

40. Gaillard, S., & Tremblay, A. (2016). Linguistic Proficiency Assessment in Second Language Acquisition Research: The Elicited Imitation Task. Language Learning, 66(2), 419–447. https://doi.org/10.1111/lang.12157
41. R Core Team. (2024). R: A language and environment for statistical computing (Version 4.3.3) [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/
42. Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. Methods in Ecology and Evolution, 11(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434
43. Antoine, P., & Smith, J. A. (2017). Saisir l'expérience: Présentation de l'analyse phénoménologique interprétative comme méthodologie qualitative en psychologie. Psychologie Française, 62(4), 373–385.
44. Boucher, J., Brisebois, A., Slegers, A., Courson, M., Désilets-Barnabé, M., Chouinard, A.-M., Gbeglo, V., Marcotte, K., & Brambati, S. M. (2022). Picture Description of the Western Aphasia Battery Picnic Scene: Reference Data for the French Canadian Population. American Journal of Speech-Language Pathology, 31(1), 257–270. https://doi.org/10.1044/2021_AJSLP-20-00388
45. Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. Memory & Cognition, 42(1), 41–52.
46. Liu, Y., Binti Ab Rahman, F., & Binti Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. Cogent Education, 12(1), 2466288. https://doi.org/10.1080/2331186X.2025.2466288
47. An, B. P. (2015). The role of social desirability bias and racial/ethnic composition on the relation between education and attitude toward immigration restrictionism. The Social Science Journal, 52(4), 459–467. https://doi.org/10.1016/j.soscij.2014.09.005
48. Janus, A. L. (2010). The Influence of Social Desirability Pressures on Expressed Immigration Attitudes*: Social Desirability Pressures and Immigration Attitudes. Social Science Quarterly, 91(4), 928–946. https://doi.org/10.1111/j.1540-6237.2010.00742.x
49. Beck, R. J., Cesario, T. C., Yousefi, A., & Enamoto, H. (2000). Choral singing, performance perception, and immune system changes in salivary immunoglobulin A and cortisol. Music Perception, 18(1), 87–106.

Acknowledgements

We thank Olga Kudriakova and Tiphaine Legrand for their assistance with recruiting participants. We also thank Mihaela Felezeu and research assistants from BRAMS and UQAM for their assistance with logistics and data collection. We are grateful to each of the newcomer participants and Francophone choristers for their enthusiastic participation.
Address correspondence to Isabelle Peretz, Department of Psychology, University of Montreal, QC, Canada. E-mail: Isabelle.Peretz@umontreal.ca

Funding

This research was funded by the Casavant Chair in Neurocognition of Music held by Isabelle Peretz, and by a subgrant from SingWell (www.singwell.ca) through their Partnership Grant "Group singing to support social wellbeing and communication in adults with communication disorders: A knowledge generation and mobilization partnership" (895-2021-1018) from the Social Sciences and Humanities Research Council (SSHRC) of Canada.

Author Contributions

IP, DM and IH designed the study. Data collection was performed by FA, H-YL, DM and research assistants. FA, DM and IP analyzed the data. FA, DM and IP wrote the manuscript.

Competing Interests

The authors declare no competing interests.

Data Availability

The datasets created during and/or analyzed during the current study are available from the corresponding author on reasonable request.