Scientific Data

https://doi.org/10.1038/s41597-026-06654-0

Article in Press

Music Ensemble: a large dataset on musicianship, cognition, and personality in musicians and nonmusicians

Received: 29 July 2025

Accepted: 19 January 2026

Cite this article as: Talamini, F., Grassi, M., Altoè, G. *et al.* Music Ensemble: a large dataset on musicianship, cognition, and personality in musicians and nonmusicians. *Sci Data* (2026). https://doi.org/10.1038/s41597-026-06654-0

Francesca Talamini, Massimo Grassi, Gianmarco Altoè, Elvira Brattico, Anne Caclin, Barbara Carretti, Véronique Drai-Zerbib, Laura Ferreri, Filippo Gambarota, Jessica Grahn, Lucrezia Guiotto Nai Fovino, Marco Roccato, Antoni Rodriguez-Fornells, Swathi Swaminathan, Barbara Tillmann, Peter Vuust, Jonathan Wilbiks, Marcel Zentner, Karla Aguilar, Christ B. Aryanto, Frederico C. Assis Leite, Aíssa M. Baldé, Deniz Başkent, Laura Bishop, Graziela Bortz, Fleur L. Bouwer, Axelle Calcus, Giulio Carraturo, Antonia Čerič, Antonio Criscuolo, Léo Dairain, Simone Dalla Bella, Oscar Daniel, Anne Danielsen, Anne-Isabelle de Parcevaux, Delphine Dellacherie, Verónica Detlefsen, Tor Endestad, Victor Cepero-Escribano, Juliana L. d. B. Fialho, Caitlin Fitzpatrick, Anna Fiveash, Juliette Fortier, Noah R. Fram, Eleonora Fullone, Stefanie Gloggengießer, Lucia Gonzalez Sanchez, Reyna L. Gordon, Mathilde Groussard, Assal Habibi, Heidi M. U. Hansen, Eleanor E. Harding, Kirsty Hawkins, Steffen A. Herff, Veikka P. Holma, Kelly Jakubowski, Maria G. Jol, Aarushi Kalsi, Veronica Kandro, Rosaliina Kelo, Sonja A. Kotz, Gangothri S. Ladegam, Bruno Laeng, André Lee, Miriam Lense, César F. Lima, Simon P. Limmer, Chengran K. Liu, Paulina d. C. Martín Sánchez, Langley McEntyre, Jessica P. Michael, Daniel Mirman, Julieta Moltrasio, Daniel Müllensiefen, Niloufar Najafi, Jaakko Nokkala, Ndassi Nzonlang, Maria Gabriela M. Oliveira, Katie Overy, Andrew J. Oxenham, Edoardo Passarotto, Marie-Elisabeth Plasse, Herve Platel, Alice Poissonnier, Vasiliki Provias, Neha Rajappa, Pablo Ripolles, Michaela Ritchie, Italo R. Rodrigues Menezes, Rafael Román-Caballero, Paula Roncaglia, Wanda Rubinstein, Farrah Y.-A. Sa'adullah, Suvi Saarikallio, Daniela Sammler, Séverine Samson, E. Glenn Schellenberg, Nora R. Serres, L. Robert Slevc, Ragnya-Norasoa Souffiane, Florian J. Strauch, Hannah Strauss, Nicholas Tantengco, Mari Tervaniemi, Rachel Thompson, Renee Timmers, Petri Toiviainen, Laurel J. Trainor, Clara Tuske, Jed Villanueva, Claudia C. von Bastian, Kelly L. Whiteford, Emily A. Wood, Florian Worschech & Ana Zappa

Music Ensemble: a large dataset on musicianship, cognition, and personality in musicians and nonmusicians

Francesca Talamini[1†], Massimo Grassi[2†], Gianmarco Altoè[3], Elvira Brattico[4,5], Anne Caclin[6], Barbara Carretti[2], Véronique Drai-Zerbib[7], Laura Ferreri[8], Filippo Gambarota[3], Jessica Grahn[9,10], Lucrezia Guiotto Nai Fovino[2], Marco Roccato[2], Antoni Rodriguez-Fornells[11,12], Swathi Swaminathan[9,10], Barbara Tillmann[7], Peter Vuust[4,13], Jonathan Wilbiks[14], Marcel Zentner[1], Karla Aguilar[14], Christ B. Aryanto[15,16], Frederico C. Assis Leite[17], Aíssa M. Baldé[18], Deniz Başkent[19], Laura Bishop[20,21], Graziela Bortz[22], Fleur L. Bouwer[23], Axelle Calcus[24], Giulio Carraturo[5], Antonia Čerič[25], Antonio Criscuolo[26], Léo Dairain[24], Simone Dalla Bella[27,28,29,30], Oscar Daniel[31], Anne Danielsen[20,21], Anne-Isabelle de Parcevaux[7], Delphine Dellacherie[32], Verónica Detlefsen[33,34], Tor Endestad[35,20,36], Victor Cepero-Escribano[37,12], Juliana L. d. B. Fialho[22,38], Caitlin Fitzpatrick[9], Anna Fiveash[39,7], Juliette Fortier[27], Noah R. Fram[40], Eleonora Fullone[8], Stefanie Gloggengießer[25], Lucia Gonzalez Sanchez[12], Reyna L. Gordon[40,41], Mathilde Groussard[42], Assal Habibi[43], Heidi M. U. Hansen[35], Eleanor E. Harding[19], Kirsty Hawkins[44], Steffen A. Herff[45,46], Veikka P. Holma[47], Kelly Jakubowski[31], Maria G. Jol[23], Aarushi Kalsi[31], Veronica Kandro[48], Rosaliina Kelo[49], Sonja A. Kotz[26,50], Gangothri S. Ladegam[51], Bruno Laeng[35,20], André Lee[52,53], Miriam Lense[40], César F. Lima[18], Simon P. Limmer[1], Chengran K. Liu[51], Paulina d. C. Martín Sánchez[54], Langley McEntyre[40], Jessica P. Michael[51], Daniel Mirman[51], Julieta Moltrasio[55,33], Daniel Müllensiefen[44], Niloufar Najafi[23], Jaakko Nokkala[49], Ndassi Nzonlang[32], Maria Gabriela M. Oliveira[17], Katie Overy[56,57], Andrew J. Oxenham[58], Edoardo Passarotto[52], Marie-Elisabeth Plasse[6], Herve Platel[42], Alice Poissonnier[42], Vasiliki Provias[59,60], Neha Rajappa[58], Pablo Ripolles[59,60], Michaela Ritchie[14], Italo R. Rodrigues Menezes[17], Rafael Román-Caballero[54,61], Paula Roncaglia[62], Wanda Rubinstein[55,33], Farrah Y.-A. Sa'adullah[39], Suvi Saarikallio[47], Daniela Sammler[25,50], Séverine Samson[32,63], E. Glenn Schellenberg[18,64], Nora R. Serres[35,20], L. R. Slevc[48], Ragnya-Norasoa Souffiane[6], Florian J. Strauch[1], Hannah Strauss[1], Nicholas Tantengco[48], Mari Tervaniemi[49,65], Rachel Thompson[48], Renee Timmers[66], Petri Toiviainen[47,67], Laurel J. Trainor[68,69], Clara Tuske[48], Jed Villanueva[43], Claudia C. von Bastian[15], Kelly L. Whiteford[58], Emily A. Wood[68], Florian Worschech[52], Ana Zappa[12]

[1]Department of Psychology, University of Innsbruck, Innsbruck, Austria,[2]Department of General Psychology, University of Padua, Padua, Italy, [3]Department of Developmental Psychology and Socialization, University of Padua, Padua, Italy, [4]Center for Music in the Brain (MIB), Department of Clinical Medicine, Aarhus University, Denmark, [5]Department of Education, Psychology,

Communication, University of Bari Aldo Moro, Bari, Italy, [6]Université Claude Bernard Lyon 1, CNRS, INSERM, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, F-69500, Bron, France, [7]Université Bourgogne Europe, CNRS, LEAD UMR5022, 21000, Dijon, France, [8]Department of Brain & Behavioural Science, Università degli Studi di Pavia, Pavia, Italy, [9]Department of Psychology, Western University, London, Ontario, Canada, [10]Brain and Mind Institute, Western University, London, Ontario, Canada, [11]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain, [12]Department of Cognition, Development and Education Psychology, University of Barcelona / Institut d'Investigació Biomèdica de Bellvitge, Barcelona, Spain, [13]Royal Academy of Music Aarhus/Aalborg, Denmark, [14]Department of Psychology, University of New Brunswick Saint John, 100 Tucker Park Road, Saint John, NB, Canada, [15]Department of Psychology, University of Sheffield, Sheffield, United Kingdom, [16]Faculty of Psychology, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia, [17]Department of Psychobiology, Universidade Federal de São Paulo, São Paulo, Brazil, [18]Centre for Psychological Research and Social Intervention (CIS-Iscte), Iscte - Instituto Universitário de Lisboa, Lisboa, Portugal, [19]Department of Otorhinolaryngology / Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, [20]RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway, [21]Department of Musicology, University of Oslo, Oslo, Norway, [22]Department of Music, Universidade Estadual Paulista "Júlio de Mesquita Filho", São Paulo, Brazil, [23]Cognitive Psychology Unit, Institute of Psychology & Leiden Institute for Brain and Cognition, Leiden University, Leiden, Netherlands, [24]Center for Research in Cognitive Neuroscience (CRCN), Université libre de Bruxelles (ULB), Brussels, Belgium, [25]Research Group Neurocognition of Music and Language, Max Planck Institute for Empirical Aesthetics, Frankfurt/M., Germany, [26]Department of Neuropsychology & Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands, [27]International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, Canada, [28]Department of Psychology, University of Montreal, Montreal, Canada, [29]Centre for Research on Brain, Language and Music (CRBLM), Montreal, Canada, [30]Department of Cognitive Psychology, Vizja University, Warsaw, Poland, [31]Department of Music, Durham University, Durham, United Kingdom, [32]Département de Psychology, Laboratoire PSITEC ULR 4072, Université de Lille, Lille, France, [33]Centro de Investigaciones en Neurociencias y Neuropsicología, Universidad de Palermo, Buenos Aires, Argentina, [34]Facultad de Psicología, Universidad de Buenos Aires. Buenos Aires, Argentina, [35]Department of Psychology, Faculty of Social Sciences, University of Oslo, Oslo, Norway, [36]Department of Neuropsychology, Helgeland Hospital, Mosjøen, Norway, [37]Bridging Research in AI and Neuroscience (BRAIN), Computer Vision Center (CVC), Bellaterra, Spain, [38]Department of Psychology, Universidade Cruzeiro do Sul, São Paulo, Brazil, [39]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney,

Australia, [40]Department of Otolaryngology - Head & Neck Surgery, Vanderbilt University Medical Center, Nashville, United States, [41]Vanderbilt Genetics Institute, Vanderbilt University, Nashville, United States, [42]Department of Psychology, Université de Caen Normandie, Inserm, EPHE-PSL, PSL Université Paris, CHU de Caen, GIP Cyceron, U1077, NIMH, 14000 Caen, France, [43]Brain and Creativity Institute, Dornsife College of Letters Arts and Sciences, University of Southern California, United States, [44]Department of Psychology, Goldsmiths, University of London, London, United Kingdom, [45]Sydney Conservatorium of Music, The University of Sydney, Sydney, Australia, [46]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia, [47]Centre of Excellence in Music, Mind, Body and Brain, University of Jyväskylä, Jyväskylä, Finland, [48]Department of Psychology, University of Maryland, College Park, United States, [49]Centre of Excellence in Music, Mind, Body and Brain; Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland, [50]Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [51]University of Edinburgh, Edinburgh, United Kingdom, [52]Institute of Music Physiology and Musicians' Medicine, Hanover University of Music, Drama, and Media, Hanover, Germany, [53]Department of Neurology, TUM University Clinic Rechts der Isar, Technical University Munich, Germany, [54]Mind, Brain and Behavior Research Center (CIMCYC), University of Granada, Granada, Spain, [55]Unidad de Medicina traslacional, Laboratorio de deterioro cognitivo (UMTLABDCOG), San Martín, Buenos Aires, Argentina, [56]Reid School of Music, ECA, University of Edinburgh, Edinburgh, United Kingdom, [57]Edinburgh Neuroscience, University of Edinburgh, Edinburgh, United Kingdom, [58]Department of Psychology, University of Minnesota, Minneapolis, United States, [59]Music and Audio Research Laboratory (MARL), New York University, New York, United States, [60]Department of Psychology, New York University, New York, United States, [61]Department of Psychology, Neuroscience & Behaviour, McMaster University, [62]Cognitive Science and Artificial Intelligence Department, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands, [63]Institut Pasteur, Université Paris Cité, AP-HP, INSERM, CNRS, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, F-75012 Paris, France, [64]Department of Psychology, University of Toronto Mississauga, Mississauga, Canada, [65]Cognitive Brain Research Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland, [66]Department of Music, University of Sheffield, Sheffield, United Kingdom, [67]Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland, [68]Department of Psychology, Neuroscience & Behaviour, McMaster University, Canada, [69]McMaster Institute for Music and the Mind, McMaster University

†Correspondence should be addressed to Francesca Talamini (Francesca.Talamini@uibk.ac.at) or Massimo Grassi (massimo.grassi@unipd.it).

Abstract

The Music Ensemble dataset is a large-scale, cross-national database that provides detailed information about the musical, cognitive, personality, and demographic profiles of young adult musicians and nonmusicians. Data were collected from 1438 participants (aged 18–30) across thirty-five research sites in Europe, North America, South America, and Australia. Participants completed an in-person, in-lab battery of objective tests, including measures of verbal, visuospatial and musical short-term memory, executive functions (updating component), nonverbal reasoning, verbal comprehension, and music perception skills. The battery also included standardized and custom self-report questionnaires assessing music sophistication, music reward, personality traits, socioeconomic status, and demographic characteristics. Music Ensemble was preregistered, and the research protocol followed a standardized procedure across sites. The dataset also includes a large subsample of musicians and nonmusicians that are pair-matched for age, gender, and education (678 pairs). It enables well-powered investigations into the relationship between musical expertise and individual differences in cognition, personality, and demographic variables. It is also suitable for training in statistical and psychometric methods.

Background and Summary

In neuroscience and behavioral sciences, scientists are often interested in investigating experts. Experts are individuals who engage in extensive, long-term practice within a specific domain; classic examples, frequently examined in neuroscience and cognitive science, include London taxi drivers[1], Braille readers[2], chess players[3], and athletes[4,5]. A major research goal is to understand whether, and how, such prolonged and intensive experience leads to changes in brain structure and function, as well as in behavior and cognition. These effects are examined both in tasks directly related to the domain of expertise and in tasks that are far from it, in order to test the specificity or generality of changes arising from intensive practice. Of the numerous expert groups considered in the literature, musicians are commonly investigated. Musical expertise involves years of structured training that is cognitively, perceptually, and motorically demanding. Becoming a musician requires the development of fine motor control, advanced auditory discrimination, and mastery of the symbolic system of music notation, along with the principles of melody, harmony, and temporal organization, which can vary across musical cultures. Importantly, musicians must integrate these skills in real time during performance. Musical training is also formalized within dedicated educational institutions, and can culminate in advanced degrees such as bachelor's and master's diplomas. These features make musicians a particularly informative group for examining experience-dependent plasticity.

Understanding the characteristics of expert musicians as well as the relationship between this training and cognitive functioning remains a central question in cognitive psychology and neuroscience[6]. Musical expertise has been found to be related to a wide range of abilities and individual differences, including auditory perception, short-term memory, executive functions, and to personality traits[7–10]. These differences are often interpreted as training-related effects arising from the intensive and prolonged demands of musical practice[11,12], or as selection effects, in which pre-existing characteristics (e.g., cognitive abilities, socio-economic status, personality) influence who chooses to engage in musical training and ultimately becomes a musician[13,14]. Despite this ongoing debate about the origins of differences between musicians and nonmusicians, largely based on correlational evidence, the empirical literature on cognitive and personality correlates of musicianship remains fragmented. Studies frequently rely on small sample sizes, partly due to the limited accessibility of specialized populations such as expert musicians, and show substantial heterogeneity in both methodological approaches and measurement tools. Moreover, across studies musicians and nonmusicians can differ widely in their definitions and inclusion criteria. For example, a musician could be defined as having at least 6, 8 or 10 years of music training, with some studies looking for active playing musicians and some others only using past formal training as

inclusion criteria. The same problem, but reversed, occurs with the definition of nonmusician, which is often neglected. Nonmusicians are often "participants that are not musicians" regardless that many individuals receive some music training during life and/or play occasionally a music instrument. This lack of shared definitions and protocols limits the comparability of different study's results. To address these challenges, we present the Music Ensemble dataset, a large-scale, cross-sectional dataset collected from 1438 young adult participants across thirty-five research units of sixteen different countries. Participants were categorized as musicians (i.e., ≥ 10 years of music training and musically active) and nonmusicians (i.e., ≤ 2 years of music training and musically inactive) and completed a standardized battery of cognitive tasks and self-report questionnaires. The dataset includes measures of verbal, visuospatial, and musical short-term memory, executive functions, nonverbal reasoning, vocabulary knowledge, and music perception skills. Participants also completed self-report questionnaires assessing personality traits, socioeconomic background, music-related abilities, experiences and behaviors, and demographic characteristics. The dataset also includes a subsample of gender, age and education-matched pairs of musicians and nonmusicians. All data were collected through in-person, in-lab testing sessions using standardized protocols across sites. Testing materials and procedures were openly shared and coordinated via two central platforms, located in Padua (Italy) and Innsbruck (Austria). The dataset presented here enables the systematic comparisons between musicians and nonmusicians, facilitating broader research into the relationship between musical expertise and individual differences in cognition, musicality, personality and demographics. The present data descriptor provides a comprehensive overview of the full dataset, including its structure, variables, and how to use it. A subset of this data was analysed in a separate study investigating short-term memory differences between musicians and nonmusicians[15].

Methods

Ethics

The research complies with all relevant ethical regulations and was conducted in accordance with the Declaration of Helsinki. Overarching ethical approval (protocol n. 5305) was granted by the Ethics Committee (Comitato Etico della Ricerca Psicologica, Area 17) of the University of Padova, and it explicitly included permission to share the data, provided that they were properly anonymized. In cases where this central approval could not be applied, individual research units obtained local ethical clearance from their respective institutions. Specifically, the study protocol was approved by the Durham University Music Department Ethics Committee; the Research Ethics Board at the University of New Brunswick

(File #2023-016); the local Ethics Committee of the Université libre de Bruxelles (1668/2024); the Philosophy, Psychology and Language Sciences (PPLS) Research Ethics Committee of the University of Edinburgh (203-2324/2); the Brazilian National Ethical Committee via Plataforma Brasil (n. 70181123.8.0000.5411); the Ethics Review Committee Psychology and Neuroscience, Maastricht University (External approval 76B8BA7B9078DF27CF46CDC26FA9E5); the Ethics Council of the Max Planck Society (no. 2023_18); the Non-Medical Research Ethics Board of the University of Western Ontario; the Institutional Review Board of the University of Minnesota (0605S85872); the Institutional Review Board of Vanderbilt University (#162002); the Leiden University Psychology Research Ethics Committee (2023-11-24-F.L. Bouwer-V1-5105); the Ethics Committee of Research in Education and Psychology (CEREP) of the University of Montreal (study protocol n. 2023-4765); the Research Ethics Committee of the University of Sheffield (Alternative Ethics Approval no. 860); the Norwegian Agency for Shared Services in Education and Research SIKT (ref: 805630); the Central Ethics Review Board for non-WMO studies at the University Medical Center Groningen (Research Register number: 18645); the Board for Ethical Questions in Science of the University of Innsbruck (28/2023); the Institutional Review Board of New York University (#4157). Consent for participation and data sharing was obtained for all participants.

Privacy and De-identification

Data were anonymized prior to deposition. No direct identifiers such as names or contact information were collected or stored. Demographic variables were limited to non-identifying categories including gender, age, years of education, and country. The anonymized dataset was reviewed by the coordinating sites in Padua and Innsbruck to ensure compliance with GDPR.

Study design

The study design was developed by eighteen authors and conducted as a Registered Report[16]. The preregistration included hypotheses, sample size targets, primary analyses, and a research protocol[15]. Once the protocol was accepted as a Stage-1 Registered Report, it was then shared with the later-joining collection units. An overview of the process timeline is displayed in Figure 1. The data collection took part in 35 independent research units across 16 countries in Europe, North America, South America, and Australia (see Figure 2) and took place between March 2023 and July 2025. The data collection is now frozen. Each site followed a common, shared protocol which was designed to ensure methodological consistency and replicability across locations.

Figure 1

Overview Of The Study Process

[Figure 1 about here]

Figure 2

Heat Map Showing The Number Of Participants Collected In Each Participating Country

[Figure 2 about here]

Participants

Participants were young adults between the ages of 18 and 30. Recruitment was carried out locally by each site using university subject pools, social media, and/or public advertising. Participants were native speakers of the testing language and provided informed consent prior to participation. Participants received compensation for their participation according to each unit's standards (e.g., course credits, monetary compensation). Inclusion criteria were assessed by the experimenters at the recruiting stage. To be included in the musician group, participants were required to have at least ten years of formal music training (instrument or voice) in an official music school, conservatory or private lessons. Musicians had also to be actively engaged in musical practice at the time of testing, without any periods of interruption in the last year. Musicians could not be self-taught. Nonmusicians were defined as individuals with no more than two years of formal music training across their lifetime (excluding mandatory music classes at school), not actively practicing any instrument or singing at the time of testing or in the last five years prior to participation. Each collection unit was instructed to approximately match each musician to a nonmusician based on age (±1 year), gender, and years of education (±1 year). The dataset includes 1438 participants in total. Of these, 1274 are matched musicians and nonmusicians who completed all tasks. The remaining 164 participants (80 musicians and 84 nonmusicians) are unmatched; among them, 27 have one or two missing tasks. Demographic characteristics (age, gender, education) are reported separately for musicians and nonmusicians and for matched and unmatched participants in Table 1 and Table 2. Educational history included both general and musical schooling. In cases where music training was pursued in parallel with another form of education (e.g., conservatory and high school), overlapping years were counted as a single year of education.

Table 1

Demographic Details Of Matched Musicians And Nonmusicians

	Musicians		*Nonmusicians*	
Gender	*Age*	*Education*	*Age*	*Education*
Female (*N* = 371 pairs)	21.8 (3.0)	15.7 (2.1)	21.8 (2.9)	15.7 (2.0)
Male (*N* =264 pairs)	22.9 (3.5)	16.2 (2.3)	23.0 (3.4)	16.1 (2.3)
Non-binary (*N* = 2 pairs)	19.5 (2.1)	14.5 (2.1)	20.5 (2.1)	15.5 (2.1)

Note. Mean (SD). All matched participants completed all tasks.

Table 2

Demographic Details Of Unmatched Musicians And Nonmusicians

		Musicians			*Nonmusicians*	
Gender	*N*	*Age*	*Education*	*N*	*Age*	*Education*
Female	43	22.4 (3.7)	15.9 (3.0)	52	20.8 (2.7)	14.9 (1.5)
Male	37	23.8 (3.4)	16.9 (2.5)	32	21.7 (3.1)	15.5 (2.5)

Note. Mean (SD). Unmatched participants can include missing data.

Materials

Performance tasks

Digit span (verbal short-term memory). The digit span task is a common task to assess the span of verbal short-term memory[17]. It is a recall task in which the participant has to reproduce the to-be-remembered stimuli. In the current implementation, participants completed a forward digit span task using visually presented sequences of digits (1–9), rather than using auditory presentation, in order to avoid a potential auditory advantage for musicians. Digits appeared centrally on screen at a rate of one digit per second (500 ms on, 500 ms off). Participants recalled and typed the digits in the same order after each sequence. The task began with two-digit sequences, with two trials per length. If at least one response was correct, the sequence length increased by one. Individual digits were not repeated in sequences of up to nine digits; repetitions (non-adjacent) were allowed in longer sequences. Salient, easily chunkable patterns were excluded. The task stopped after two incorrect responses at the same sequence length. Sequences were fixed across participants and capped at 18 digits. An auditory beep signalled the start and end of each trial. The task was repeated in two consecutive blocks of trials with a second, unique set of sequences. The task was preceded by a familiarization test, consisting of two trials with three-digit sequences. The total score was the mean of the scores of the two blocks of trials. The block score was computed as the sum of correctly recalled sequences. The task was implemented in jsPsych[18].

Spatial span (visuospatial short-term memory). The spatial span task is an adaptation of the Corsi block task, a common measure of visuospatial short-term memory[17,19]. Like the digit span, it is a recall task in which the participant has to reproduce the to-be-remembered stimuli. In the current implementation, participants completed a forward matrix span task using a 4×4 grid. On each trial, a dot appeared sequentially in different cells of the grid at a rate of one dot per second (500 ms on, 500 ms off). After the final dot disappeared, participants reproduced the sequence by clicking the corresponding grid positions in the same order. The task began with two-position sequences, with two trials per sequence length. If at least one trial was correct, the sequence length increased by one. The task ended after two incorrect trials at the same length. Sequences were fixed across participants and designed to avoid easily chunked spatial patterns. Up to 16 positions, no spatial location was repeated; for longer sequences, non-consecutive repetitions were allowed. The task had a maximum sequence length of 32 to prevent ceiling effects and was administered twice in two consecutive blocks of trials with different sets of sequences. Auditory cues signalled the beginning and end of each trial. The task was preceded by a familiarization test, consisting of two trials with three-position sequences. The total score was the mean of the scores of

the two blocks of trials. The block score was computed as the sum of correctly recalled sequences. The task was implemented in jsPsych.

Melody span (musical short-term memory). The Melody span task was designed ad-hoc to tap into the capacity of musical short-term memory, similarly to span tasks[17] but using a recognition paradigm. Participants completed a same/different discrimination task involving melodies composed of piano tones from the C major diatonic scale (C4–C5). On each trial, a standard melody was presented, followed after a 2-second delay by a comparison melody of the same length. Participants judged whether the two melodies were identical or different. Melodies started at two notes in length and increased incrementally, with four trials per length. To discourage response strategies, each set of trials included a varying number of "different" responses (1–3 out of 4). The task ended after the third incorrect response. Melodies were fixed across participants and could be up to 40 notes long to avoid ceiling effects. Each melody was signalled by a visual cue and labelled on screen. The task was administered twice in two consecutive blocks of trials using distinct sets of pseudo-random, non-tonal melodies that included ascending and descending intervals. In comparison trials, differences were introduced by reversing two non-adjacent notes, ensuring that identical notes were never adjacent. The task was preceded by a familiarization test consisting of two trials of three-note melodies and one trial of four-note melodies. The total score was the mean of the scores of the two blocks of trials. The block score was computed as the sum of correctly recognized sequences (both same and different trials). The task was implemented in jsPsych.

2-back task (executive functions). To assess the updating component of executive functions, participants performed a visual 2-back task. Participants viewed a continuous stream of 22 capital letters drawn from a fixed set (C, D, G, K, P, Q, T, V). Each letter was displayed for 500 ms at the center of the screen, followed by a 1500 ms blank screen, resulting in a 2000 ms onset-to-onset interval. The task required participants to press a response key whenever the current letter matched the one presented two positions earlier. The task was preceded by a familiarization block and was repeated five times in successive blocks of trials, with each block containing six target trials. The total score was computed as the sum of the correct responses (max 30) minus the false alarms. The task was implemented in jsPsych.

All stimuli and the softwares' scripts created with jsPsych are available and shared in the OSF repository of the project[20]. Users can find the "jsPsych scripts" folder included in the "digital materials" folder of the "Files" section of the OSF project (see Figure 3, left).

Figure 3

Folder tree of the "Files" section of the OSF project (left) and of the file "analysis.zip" (right).

[Figure 3 about here]

Note. The file "analysis.zip" can be found in the OSF folder "Data and analyses scripts"[20].

Raven matrices (nonverbal reasoning). To assess nonverbal reasoning, we administered a digitalized version of the Raven Advanced Progressive Matrices[21], a widely used, language-independent test of abstract, nonverbal reasoning. Participants completed a series of matrices by choosing an option among 6 different possibilities. They had 10 minutes to complete as many matrices as possible. The task consisted of a single block of 46 items, created by concatenating the final 10 items of Set I with all 36 items of Set II. The first two items from Set I were presented as practice trials. The total score was computed as the sum of correct answers. The task was implemented in LimeSurvey.

WAIS vocabulary (verbal comprehension). The Vocabulary subtest of the Wechsler Adult Intelligence Scale (WAIS), was used to assess participants' vocabulary knowledge, which is part of the verbal comprehension index of the WAIS. In most countries, the WAIS-IV version[22] (maximum 30 items) was used. In Portuguese-speaking countries, where the WAIS-IV was unavailable, the most recently validated version, WAIS-III[23] (maximum 33 items), was administered instead. The task requires participants to give a definition for words of increasing difficulty. Scores were computed following the WAIS official manual: precise, correct definitions scored 2 points, less precise, correct definitions scored 1 point, and wrong definitions scored 0. The total score corresponds to the sum of these points. The task was discontinued after 3 wrong answers. The scores obtained by the two different WAIS versions were then transformed into the proportion of correct responses out of the total possible, to allow comparison across all participants. The WAIS vocabulary was administered orally in the native language of the experimenter and the participant.

Mini-PROMS (music perception skills). Music perception skills were assessed with a short version of the Profile of Music Perception Skills (Mini-PROMS[24]). The PROMS was originally designed to objectively assess music perception skills in both musically-trained and untrained individuals across a wide range of musical dimensions[25]. With Melody, Tuning, Accent, and Tempo, the Mini-PROMS comprises four of the original nine PROMS dimensions. In each trial, participants listen twice to a reference stimulus followed by a comparison stimulus, and rate whether they are identical or different. To indicate the confidence of their answer, participants can choose between "definitely same" "probably same", "I don't know", "probably different" and "definitely different". Subtest and total scores are derived as the sum of correct answers, by scoring 2 points for correct answers with high confidence, 1

point for correct answers with lower confidence, and 0 points for incorrect or “I don’t know” answers. The total score corresponds to the sum of the scores divided by 2. The Melody subtest assesses the ability to detect differences between monophonic melodic sequences of varying complexity. The Tuning subtest involves comparing diatonic piano chords, with different trials featuring a frequency shift in one of the middle notes. The Accent subtest requires evaluating rhythmic patterns in which beat accents are manipulated through intensity changes. In the Tempo subtest, participants compare the tempo (beats per minute) of pairs of stimuli, including both single and multilayer structures with different instruments and rhythms. Each subtest is preceded by a practice trial to familiarize participants with the task. The task was implemented in LimeSurvey.

Self-report measures

Custom questions. All participants were asked to provide demographic information, including gender, age, and the type of non-musical education they were pursuing or had completed (e.g., type of bachelor's degree). Musicians were asked to report the number of years of formal music training received in an official music school, conservatory, or through private lessons; the number of hours they practiced per day at the time of testing; the instrument(s) they played; whether they sang; and whether their musical activity over the past 12 months had been continuous without interruptions. If a musician played multiple instruments, these questions were answered focussing on their primary instrument. All participants, regardless of group, were also asked whether they had relative pitch. They additionally responded to a musical identity question taken from the Ollen Musical Sophistication Index[26]: “Which title best describes you?”, with six possible responses ranging from “Nonmusician” to “Professional musician.” The custom questionnaire also included a question about self-perceived expertise, asking participants to rate their expertise (on a 1–6 scale) in domains such as art (excluding musical performance), sports, games, and other areas they could input. The full list of the custom questions is reported as supplementary information.

Gold-MSI (musical sophistication). The Goldsmiths Musical Sophistication Index (Gold-MSI) is a 38-item self-report questionnaire designed to measure individual differences in musical sophistication within the general Western population[27]. It captures both common and more specialized musical behaviors, including interest in and engagement with music. The questionnaire is divided into five subscales: active engagement (e.g., time and resources devoted to music), perceptual abilities (self-rated music listening skills), self-rated singing abilities, musical training (extent of formal instruction), and emotional engagement (e.g., the ability to describe emotional experiences related to music). The response to each item is provided with a 7-point Likert scale. The task was implemented in LimeSurvey.

eBMRQ (reward from music). The extended Barcelona Music Reward Questionnaire (eBMRQ[28]) is a 24-item self-report instrument developed to assess individual differences in the sensitivity to musical reward. It expands on the original BMRQ[29] by adding musical absorption (tendency to become deeply immersed in musical experiences) to the five previously validated subscales: music seeking (propensity to pursue music-related experiences), emotion evocation (capacity of music to elicit strong emotions), mood regulation (use of music to influence mood), sensory-motor (automatic bodily responses to music), and social reward (music's role in enhancing social interaction). The response to each item is provided with a 5-point Likert scale. The task was implemented in LimeSurvey.

Big-Five Inventory-2 (personality). Personality traits were assessed using the Big Five Inventory–2 (BFI-2[30]), a 60-item self-report questionnaire that measures five broad personality dimensions: Open-mindedness, Conscientiousness, Extraversion, Agreeableness, and Negative Emotionality. The response to each item is provided with a 5-point Likert scale. Responses to the items can also be used to calculate the following Facet Scales: Sociability; Assertiveness; Energy Level; Compassion; Respectfulness; Trust; Organization; Productiveness; Responsibility; Anxiety; Depression; Emotional Volatility; Intellectual Curiosity; Aesthetic Sensitivity; Creative Imagination. The task was implemented in LimeSurvey.

Hollingshead four factor index (socioeconomic status). Socioeconomic status was assessed using the Hollingshead Four Factor Index[31], which gathers information about participants' family background, including parental education, occupational status, and marital status (e.g., married, separated). In addition to these items, participants were also asked to report their own educational level and occupational status. The questions (including possible conditionals) were implemented in LimeSurvey and the participant responded autonomously to the questions. The experimenter could help the participant in the categorization of their occupation with a grid of options.

The following google page provides a working-demo version of the full experiment, with all tasks and questionnaires (identical to those described here): https://sites.google.com/view/the-memory-experiment-clone/home. We recommend google chrome and a stable (e.g., wired) internet connection to run the experiment.

Translations

For the Melody span, Digit Span, Matrix span, Raven matrices, and 2-back task, instructions were shown in English for all participants, and verbal instructions were presented before each task by experimenters in the appropriate local language. For the remaining measures, a multi-step translation process was

implemented. First, the availability of existing translations was checked for each measure. Translations from both formal publications and informal sources, such as preprints or resources publicly available on OSF, were considered. After compiling all available materials, research units were asked to check the accuracy of existing item translations and to provide translations for any items that had not yet been translated. For the latter, research units were instructed to obtain independent translations from two bilingual speakers and to consolidate these into a single common version, which was then to be subjected to back-translation, final review, and piloting. To facilitate this process, a shared Google spreadsheet was set up for each language version, containing the English original alongside any existing translations for each measure in separate tabs. Detailed translation instructions as well as the link to the respective spreadsheet were sent to corresponding lab members for each language. Detailed information on the translation sources for the individual measures is presented below.

Mini-PROMS. A formal translation of the Mini-PROMS was available for German from the original publication[24]. Additionally, French, Italian, Portuguese, and Spanish versions have been generated by various collaborators for individual research projects. For the present project, existing translations were reviewed by Music Ensemble units, and any missing information was supplemented through the standardized translation process. Finnish, Dutch, and Portuguese (Brazil) translations were fully provided by the respective research units.

Gold-MSI. Formal translations were available for German[32] and Portuguese[33]. For the latter, minimal adaptations in phrasing were made for the instrument-played and the absolute pitch items that are not part of the subscale scoring. Dutch items were retrieved from the official Gold-MSI Scoring App[27]. For the French language version, the translation of Degrave and Deonder[34] was used as a basis. French collaborators added translations for three items not covered by the initial translation and modified the phrasing for 17 questions to make the items clearer. Finnish, Italian, and Spanish language versions were independently translated by the respective research units. Because the Spanish used in Argentina differs slightly from that used in Spain, the Spanish version developed there was modified with a few small wording changes for the Argentinian site.

eBMRQ. A formal translation for the eBMRQ was available for Spanish from the original publication[28]. The Spanish version was slightly adapted for the Argentinian site to account for regional differences in Spanish between the two countries. Additionally, BMRQ items of a French validation[35] were extended by adding independent translations of the remaining four AIMS items by the research units to obtain the full eBMRQ. Language versions for Dutch, Finnish, German, Italian, and Portuguese were independently translated by the respective Music Ensemble units.

BFI-2. Translations of the BFI-2 were drawn from the OSF storage[20] of the International Situations Project (ISP[36]) for all language versions. The German translation of the ISP was identical to the one used in an independent evaluation of the BFI-2 based on a representative sample of 1,224 German subjects[37]. In contrast to the original English version and all other translations used in the present project, the German items were phrased as individual statements (e.g., "I am someone who is outgoing, social") rather than being an appendix to the instruction question (e.g., "is outgoing, social" as a response to the instruction "I am someone who…"). Items of the ISP Dutch version were almost identical to the Dutch adaption[38] by Denissen et al. based on more than 800 participants, with only 4 items showing minimal differences in word order across both versions. The French ISP version, however, significantly differed from the version used in the independent evaluation by Lignier et al.[39].

Hollingshead. For the Hollingshead Four-Factor Index[31], education levels and occupational titles were independently translated from English by the respective research units for all languages.

WAIS. The WAIS vocabulary subtest was organized by each research unit independently, drawing on the respective official WAIS language version.

The translations of all measures used to collect the present data are available via OSF[20], in the "LimeSurvey questionnaires" folder contained in the "Digital Materials" folder of the "Files" section of the OSF project (see Figure 3, left).

Procedure

Each participant completed the testing session individually in a quiet laboratory of the research unit. Before the testing, participants were asked questions about their psychophysiological state (e.g., drug/alcohol intake, amount of sleep in the past 24 hours, and current sickness). If the participant reported to have used any drug and/or taken an excessive amount of alcohol in the 24 hours preceding the experiment, or reported insufficient sleep (i.e., < 6 hours), or sickness, they were not allowed to continue with the experiment. Participants first completed the three short-term memory tasks: melody span, digit span and spatial span. The order of these three tasks was counterbalanced within each unit, and within each group (i.e., musicians and nonmusicians) following a latin square. Subsequently, the remaining tasks were taken in a fixed order: the Raven test, the WAIS vocabulary, the 2-back task, the Mini-PROMS test, and then the self-report questionnaires. The questionnaires were administered in the following order: custom experimental questions, Gold-MSI, eBMRQ, BFI-2, Hollingshead four factor index.

Testing was carried out using secure platforms such as LimeSurvey or jsPsych/JATOS and data storage was centralized in Innsbruck (for the data gathered with LimeSurvey), and in Padua (for the data gathered with jsPsych/JATOS). The data gathered with the WAIS vocabulary were saved locally by the individual units. Task instructions were presented in the local language of the testing site except of the text elements in the jsPsych tasks which appear only in English. For these latter tasks, the experimenter manual included detailed instructions in the local language, which were read aloud to participants in their mother tongue prior to each task. Piloting confirmed that all participants understood the tasks correctly. Experimenters and participants had to be native speakers of the local language of the testing site. Tasks implemented auditorily (melody span and PROMS, see below) were delivered with headphones whereas the remaining tasks were delivered via computer screen, or orally by the experimenter (WAIS vocabulary). The experiment session lasted approximately 2 hours.

Data Records

All materials are publicly hosted on the Open Science Framework repository[20]. The dataset includes task-based performance data, self-report questionnaire data and demographic data, organized across multiple files and formats. The finalized datasets are available in CSV, Excel, and R formats (.rds extension) and are accompanied by a comprehensive data dictionary and data processing scripts. Figure 4 displays a visual summary of the dataset's content.

Figure 4

Visual representation of the dataset's content

[Figure 4 about here]

Within the OSF repository, the folder titled "Data and analyses scripts" contains a ZIP archive called "analysis.zip". Readers interested in the full data should download this ZIP archive. Upon extraction, we recommend opening the README file, which gives an overview of the various materials and where they are located. The folder tree of the content of the zip archive is reported in Figure 3 (left).

Data files can be found in the "data" folder (see Figure 3, right). The "data" folder contains four directories in which raw and clean data can be found: "clean", "raw-after-selection", "raw-before-selection", and "raw-before-fixing". The "clean" folder contains the full dataset (with both matched and

unmatched participants, and also the participants with incomplete data) named “dat-clean-full”. This dataset includes all demographic variables (i.e., age, gender, education, SES) and processed variables (i.e., span scores, 2-back task scores, mini-PROMS scores, questionnaire total scores and questionnaire subscales scores), see Methods section for details about how the various scores were computed and Supplementary Table 1 for the range of these scores (in Supplementary Information). The file “dat_clean” contains a subset of the data analyzed in the original Music Ensemble publication[15], with 600 matched pairs of musicians and nonmusicians. The data dictionary of these two files is reported as Supplementary Information and available in the Excel file “dictionary.xlsx” that is available in the folder “files” (see Figure 3, right).

The “raw-after-selection” folder includes separate files corresponding to the outputs of each measure (i.e., spans, n-back, mini-PROMS, raven matrices, and questionnaires). Each file contains single items and single trial responses of all participants. Readers interested in the single item-level or single trial-level responses are recommended to use these files.

The “raw-before-selection” contains the same files of the “raw-after-selection” folder, but with more columns corresponding to the technical output columns from LimeSurvey (e.g., timing metadata that tracks time spent on individual questions or sections). Finally, the “raw-before-fixing” contains the raw data with errors that need to be fixed (e.g., ID misspells), thus we recommend the reader not to use the files in this folder. The scripts to pre-process, clean, and merge the data are provided in the folder “scripts”.

Data from the melody span, digit span, spatial span, and 2-back tasks (those generated using the jsPsych platform) are stored in four separate files, each in long format. In these files, each row represents a single trial (i.e., one observation for one participant), and each column corresponds to a variable. All four datasets include a common set of metadata fields: date and time of data collection, country, research unit, ID of the experimenter, participant group (musician or nonmusician), subject ID, age, gender, and browser used. Block and trial numbers provide trial-level positioning within the task sequence. Additional task-specific columns indicate the stimulus presented, the participant’s response, the correct response, and an accuracy score (coded as “1” for correct, “0” for incorrect). In the 2-back dataset, an additional column flags whether the trial was a target that required a response.

Data collected via LimeSurvey are distributed across three files: one for the mini-PROMS, one for the Raven matrices, and one comprehensive file for all self-report questionnaires. These files are structured in a wide format, with each row representing one participant, and the various columns representing individual items or summary scores. All LimeSurvey files contain initial columns about the country, the

unit, the experimenter's ID, the participant's ID, gender, age, and group (musician or nonmusician). The files contain both automatically generated metadata and study-relevant variables; only the latter are described here.

The mini-PROMS dataset includes columns about item-level responses in text format for each of the four subtests (in text format, e.g., "Definitely the same") as well as subtests' total score (e.g., melody subtest score), and the general total score (sum of subtests' scores).

The Raven dataset contains item-level correctness data for both Series I and Series II. Correct answers are coded as 1; incorrect answers as 0.

The questionnaire dataset compiles responses from all self-report measures, namely the custom questions, the Gold-MSI, the eBMRQ, the BFI-2 and the Hollingshead Four Factor Index. Columns include only item-level responses (in text format).

The data dictionary of the melody span, digit span, spatial span, and 2-back tasks, mini-PROMS, Raven, and questionnaires is available in the file "dictionary.xlsx" included in the "files" folder (see Figure 3, right). Table 3 gives an overview of the main data files included on OSF, with details about file size, number of rows and columns, and type of content.

Table 3

Overview Of The File Main Data Files.

File name	*File size (MB)*	*Rows*	*Columns*	*Primary key*	*Description*
dat-clean-full.csv	0.35	1438	49	subject_ID	Clean dataset with all demographic and task summary variables (matched and unmatched participants).
melody_span.csv	25.3	85635	16	subject_ID + block_number + trial_number	Trial-level data for the melody span task.
digit_span.csv	11.1	40970	16	subject_ID + block + trial	Trial-level data for the digit span task.
spatial_span.c	10.1	37412	15	NBlock + NTrial	Trial-level data

sv					for the spatial span task.
back.csv	49.6	189288	16	subject_ID + block_number + trial_number	Trial-level data for the 2-back task.
proms.csv	1.3	1438	58	subject_ID	PROMS item-level responses.
raven.csv	0.35	1438	59	subject_ID	Raven's Progressive Matrices item-level responses.
questionnaires.csv	3.8	1438	203	subject_ID	All questionnaire items (Gold-MSI, eBMRQ, BFI-2, Hollingshead, and custom questions).

Notes. Data files are provided also in xlsx and rds format. The characteristics of the xlsx and rds files are identical except for the file size.

Technical Validation

Data cleaning

Following data collection, we conducted an initial screening of the dataset using predefined quality control procedures. To ensure consistency, we first verified that demographic variables (group, gender, and age) and alphanumeric participant IDs were identical across all task and questionnaire files.

We then proceed to identify incomplete sessions. All tasks, with the exception of the WAIS Vocabulary subtest, were administered using web-based platforms. Although the dataset was designed to include only complete records, a small number of participants have missing data from one or more tasks. In total, twenty-seven participants recruited were identified as incomplete for at least one task. Two task-specific issues were identified and addressed during the quality control process. In the melody span task, an implementation error was discovered after data collection had begun: in Block 2, comparison melodies differed from the standard in multiple notes rather than in two consecutive notes, as originally intended. This issue was corrected in mid-September 2023. Sixty-seven participants completed the task using the erroneous version of Block 2. These participants can be easily identified in the dataset (see Usage Notes below). A similar issue occurred with the BFI-2 questionnaire, in which two items were missing from the

initial implementation: item 1 (Extraversion: "I am someone who is outgoing, sociable") and item 33 (Conscientiousness: "I am someone who keeps things neat and tidy"). These items were added in mid-September 2023. Participants who completed the earlier version were recontacted to provide the missing responses; all but twelve participants replied. For the remaining twelve, the missing values were imputed using the neutral response, corresponding to the midpoint of the BFI-2 Likert scale. To identify these participants who had these missing items, readers have to open the "missing_item_BFI2" file contained in the "fixing" folder (see Figure 3, right). The Excel file shows the IDs of all participants with missing items, indicating which items could be collected subsequently and which had to be imputed instead.

For the Hollingshead Four Factor Index, the socioeconomic status score could not be computed when both parents were reported as having no income. This applied to 13 musicians and 10 nonmusicians. In line with the preregistered plan, a score of 8 (i.e., the lowest possible value on the Hollingshead scale) was assigned to these cases. In addition, many participants were university students without personal income, which made it impossible to compute their individual socioeconomic status. These fields were left blank in the dataset. Table 4 reports an overview of the missing data (or data with possible issues).

Table 4

Overview Of The Missing Data / Errors

Type of datum	*N*	*Issue/reason*	*How to identify the datum in the dataset*
Unmatched nonmusicians	84	Unit did not find the match.	Missing/blank value in column "subject_ID of the matched participant"
Unmatched musicians	80	Unit did not find the match.	Missing/blank value in column "subject_ID of the matched participant"
Melody span	6	Experimenter lapse and/or network issues.	Missing/blank value in column "melody_span"
Melody span (participants with block-1 only)	67	Error in the stimuli.	"Block1" in the column "melody_span_score"
Digit span	5	Experimenter lapse and/or network issues.	Missing/blank value in column "digit_span"
Spatial span	5	Experimenter lapse and/or network issues.	Missing/blank value in column "spatial_span"
2-Back	4	Experimenter lapse	Missing/blank value in column "nback"

		and/or network issues.	
WAIS vocabulary	12	Experimenter lapse.	Missing/blank value in column “waisvocabulary”
BFI-2	14	Response to item 1 & item 33 missing.	See file “missing_items_BFI.xls” in “fixing” folder

Measurements validation

To assess the internal consistency of the three self-report standardized questionnaires (BFI-2, eBMRQ and Gold-MSI) Cronbach’s alpha was computed. Additionally, to assess site-level heterogeneity, we estimated intraclass correlations (ICCs) using mixed-effects models with a random intercept for the testing unit.

Concerning the BFI-2, the internal consistency was assessed for the five core personality traits. All subscales showed acceptable-to-good reliability. Extraversion: 0.82; Agreeableness: 0.77; Conscientiousness: 0.86; Negative emotionality: 0.89; Open-mindedness: 0.84. ICCs were very small for all BFI subscales (Extraversion = .02; Agreeableness = .04; Conscientiousness = .02; Negative Emotionality = .04; Open-mindedness = .02), indicating that less than 5% of score variance was attributable to site differences. The Gold-MSI was also assessed separately for the five subfacets as well for the general sophistication index. All subscales showed acceptable to excellent reliability. Musical Training: 0.95; Active Engagement: 0.8; Perceptual Abilities: 0.89; Singing Abilities: 0.82; Emotions: 0.74; General Sophistication Index: 0.94. ICCs were also very small for all eBMRQ subscales (Active Engagement = .02; Perceptual Abilities = .00; Emotions = .02; Musical Training = .00; General Sophistication Index = .00;), indicating minimal site-level variance. The eBMRQ reliability was computed separately for the six subscales and for the total reward score. Subscales’ reliability ranged from questionable to good. Music Seeking: 0.67; Emotion Evocation: 0.79; Mood Regulation: 0.81; Sensory-Motor: 0.75; Social Reward: 0.68; Musical Absorption: 0.83. The total score showed an excellent reliability: 0.92. ICCs were very small for all eBMRQ subscales (Music Seeking = .00; Emotion Evocation = .03; Mood Regulation = .01; Sensory-Motor = .03; Social Reward = .01; Musical Absorption = .01; Total eBMRQ Score = .01). This indicates that less than 4% of score variance was attributable to site differences.

To evaluate the robustness of reliability estimates to changes in test length, we applied the Spearman–Brown prophecy formula. For each scale, we estimated the expected reliability if the scale were hypothetically shortened by 20% or lengthened by 20%. The resulting test-length–corrected reliability

values are reported in three Supplementary Tables (Supplementary Tables 2, 3, 4). Across instruments, corrected reliability coefficients showed similar patterns to the observed alphas, indicating that the internal consistency of the subscales would remain stable under reasonable variations in test length.

Because the questionnaires were administered in multiple languages and some translated versions did not have prior official validation (see Translations section), we report reliability (Cronbach's alpha) separately by language (Tables 5, 6, and 7).

Table 5

Bfi-2 Reliability (Cronbach's Alpha) By Language

Language	*Extraversion*	*Agreeableness*	*Conscientiousness*	*Negative emotionality*	*Open-mindedness*
Dutch (N = 104)	0.82	0.75	0.82	0.89	0.85
English (N = 531)	0.85	0.77	0.87	0.90	0.84
Finnish (N = 45)	0.84	0.85	0.89	0.84	0.85
French (N = 237)	0.80	0.74	0.86	0.88	0.82
German (N = 127)	0.85	0.76	0.85	0.89	0.83
Italian (N = 126)	0.78	0.78	0.85	0.91	0.83
Norwegian (N = 42)	0.83	0.73	0.86	0.88	0.86
Portuguese (N = 91)	0.79	0.78	0.85	0.88	0.85
Spanish (N = 135)	0.74	0.83	0.87	0.88	0.84

Table 6

Gold-MSI Reliability (Cronbach's Alpha) By Language

Language	*Active engagement*	*Perceptual abilities*	*Musical training*	*Emotions*	*Singing abilities*	*General sophisticati on index*
Dutch (N = 104)	0.87	0.91	0.94	0.85	0.88	0.94
English (N = 531)	0.85	0.87	0.95	0.76	0.84	0.94
Finnish (N = 45)	0.92	0.94	0.98	0.84	0.93	0.97
French (N = 237)	0.85	0.90	0.97	0.68	0.76	0.93
German (N = 127)	0.86	0.90	0.96	0.75	0.90	0.95
Italian (N = 126)	0.86	0.87	0.97	0.73	0.82	0.94
Norwegian (N = 42)	0.83	0.91	0.95	0.72	0.87	0.96
Portuguese (N = 91)	0.84	0.86	0.95	0.72	0.84	0.93
Spanish - Argentina (N = 42)	0.88	0.88	0.97	0.59	0.80	0.95
Spanish - Spain (N = 93)	0.81	0.91	0.95	0.55	0.88	0.95

Notes. Spanish is reported separately for Spain and Argentina due to minor wording modifications that account for regional differences between the two countries.

Table 7

eBMRQ Reliability (Cronbach's Alpha) By Language

Language	*Music seeking*	*Emotion evocation*	*Mood regulation*	*Sensory-motor*	*Social reward*	*Musical absorption*	*Total score*
Dutch (N = 104)	0.70	0.80	0.83	0.76	0.73	0.86	0.93
English (N = 531)	0.65	0.78	0.81	0.71	0.67	0.81	0.91

Finnish (N = 45)	0.73	0.85	0.82	0.79	0.85	0.90	0.95
French (N = 237)	0.70	0.81	0.79	0.78	0.68	0.83	0.92
German (N = 127)	0.66	0.83	0.81	0.76	0.68	0.89	0.93
Italian (N = 126)	0.77	0.71	0.81	0.75	0.62	0.84	0.92
Norwegian (N = 42)	0.54	0.73	0.85	0.82	0.65	0.80	0.92
Portuguese (N = 91)	0.62	0.80	0.87	0.72	0.64	0.84	0.91
Spanish - Argentina (N = 42)	0.72	0.59	0.63	0.69	0.66	0.82	0.88
Spanish - Spain (N = 93)	0.57	0.79	0.68	0.80	0.71	0.84	0.90

Notes. Spanish is reported separately for Spain and Argentina due to minor wording modifications that account for regional differences between the two countries.

To provide validity of the cognitive tasks, we will report correlations and descriptive statistics. According to models of intelligence, cognitive abilities should be positively correlated with each other[40,41]. In our sample, small positive correlations emerged among the various cognitive measures, with the largest correlations emerging between the scores of the span tasks, and between Raven matrices and the 2-back task. Table 8 provides a full correlation matrix between the various cognitive tasks. The distribution of scores of the cognitive tasks that were purpose-built for the present work are shown in Table 9.

Table 8

Correlations Among The Cognitive Measures

	1.	**2.**	**3.**	**4.**	**5.**	**6.**
1. Digit span	—					
2. Spatial span	0.303	—				

3. Melody span	0.337	0.204	—		
4. 2-back	0.212	0.214	0.172	—	
5. Raven Matrices	0.217	0.297	0.197	0.309	—
6. WAIS Vocabulary	0.144	0.047	0.169	0.234	0.273

Note. All participants (Musicians and Nonmusicians, $N = 1438$). All correlations are significant at $p < .001$.

Table 9

Distribution Of Scores Across The Purpose-built Tasks

Task	Min	Q1	Median	Q3	Max
Digit span	4	8.5	10	11.5	19.5
Spatial span	1.5	7.5	8.5	10	27.5
Melody span	4.5	20	24	29	47
2-back task	-40	14	20	24	30

Finally, we also expected a correlation between age and years of education, which emerged: $r = 0.74$, $p < .001$.

Group validation

To assess the validity of the selection of musicians and nonmusicians, we have compared the two groups on variables that assess objective musical skills and subjective (self-report) musical experience. Specifically, it is expected that musicians perform better than nonmusicians in the mini-PROMS test (subscales and total score) and report higher values in the Gold-MSI. Table 10 shows means, standard deviations, and group comparisons (via t-tests) of the musicians and nonmusicians in the mini-PROMS and Gold-MSI. As shown in the table, the two groups differed significantly on all musical variables.

Table 10

Group Differences On Music Perception Skills And Music Sophistication

	Nonmusicians		Musicians		
	Mean	SD	Mean	SD	Comparison
mini-PROMS Melody	4.98	1.65	7.43	1.35	p <.001, d = 1.62
mini-PROMS Tuning	4.26	1.38	6.19	1.31	p <.001, d =1.43
mini-PROMS Beat	5.09	1.59	6.47	1.49	p <.001, d = 0.90
mini-PROMS Speed	5.33	1.41	6.24	1.20	p <.001, d = 0.69
mini-PROMS Total	19.66	4.36	26.32	3.65	p <.001, d = 1.65
GMSI active engagement	32.81	9.68	46.38	7.33	p <.001, d = 1.58
GMSI musical training	12.94	5.93	40.88	4.37	p <.001, d = 5.36
GMSI perceptual abilities	41.02	8.27	54.66	5.95	p <.001, d = 1.89
GMSI singing abilities	24.28	7.45	36.29	6.77	p <.001, d = 1.69
GMSI emotions	31.66	5.56	36.79	4.04	p <.001, d = 1.05
GMSI general sophistication	56.55	14.56	99.26	11.74	p <.001, d = 3.23

Note. GMSI = Gold-MSI. Group comparisons were performed via a series of independent samples t-tests. Only p-values and Cohen's d resulting from this comparison are reported.

Usage Notes

To facilitate the navigation on OSF and the identification of the desired data, Figure 3 displays a tree diagram of the various folders and files. Once the "analysis" zip folder is downloaded, users can find a data dictionary file (named "dictionary.xlsx" and contained in the "files" folder), which reports in detail the name and description of all variables of each data file present in the "data" folder. The dictionary of the main data files (i.e., "dat-clean-full" and "dat_clean") is also reported as Supplementary Information (Supplementary Table 5). The data dictionary includes multiple sheets, each corresponding to each data

file. Within each sheet, all variables contained in that file are listed, along with a column named “keep”, which indicates whether each variable was retained after processing the raw data or after merging files (in the case of the clean dataset). Values corresponding to 1 identify variables that were kept, and variables corresponding to 0 identify variables that were removed. Variables removed in this process include, for example, technical output columns generated by LimeSurvey, or - in the clean dataset - variables that were not part of the preregistered analyses[15]. Users who wish to generate an aggregated dataset that includes additional variables can still do so by starting from the raw data files and modifying the provided processing scripts to avoid these exclusions.

In the full aggregated dataset (“dat-clean-full”), it is possible for users to filter the data and exclude particular cases. Specifically, for users interested only in the matched participants, the column “subject_ID of the matched participant” allows for filtering the data and excluding empty cells, which correspond to unmatched participants. This can be done by opening the data file on software such as Excel or similar, applying a filter to that variable, and unselecting empty cells. This will produce a subset of the data containing only matched participants. Alternatively, users can employ any other computation software such as R and simply remove all rows containing NAs in the “subject_ID of the matched participant” variable.

If users want to exclude the participants that had the wrong block 2 of the music short-term memory task (N = 67), they may do so by filtering the variable called “melody_span_score” which specifies if the melody score was calculated on the two correct blocks or only on the first block. Specifically, values corresponding to “block12” indicate that the score was computed on the two blocks, whereas values corresponding to “block1” indicate that the score was computed only on the first block. As for the previous case for matched and unmatched participants, users can apply a filter by selecting only the cases of interest via programs such as excel or similar.

Recommended analysis and common pitfalls.

The dataset has an intrinsic multilevel structure. Participants were recruited and tested across multiple countries and research units, and several cognitive tasks (e.g., short-term memory and executive function measures) include data at the trial level. Analyses should therefore account for this nested structure, for example by modelling “research unit” as a random effect or by applying hierarchical or multilevel approaches. Users should also note that some measures, particularly those assessing musical abilities and/or musical identity (e.g., Gold-MSI subscales and Mini-PROMS scores), are moderately to highly correlated. These inter-correlations may cause redundancy or multicollinearity in statistical analyses and

should be considered when selecting variables or building composite scores. Addressing these aspects appropriately will enhance the robustness and interpretability of secondary analyses.

Reuse potential.

The present dataset can be used to compare groups of expert musicians and nonmusicians across a wide range of variables, including personality traits and cognitive abilities. Beyond simple group comparisons, the dataset makes it possible to investigate complex relationships among musical, cognitive, demographic, and personality variables. Questionnaire data can be analyzed not only at the scale level but also at the item level, allowing for more fine-grained examination of specific constructs. Cognitive measures can similarly be re-scored in multiple ways - for example, by computing span length instead of total accuracy - so that alternative indices of cognitive functioning can be compared between groups or correlated with musical and nonmusical characteristics. Several questionnaires also allow for the extraction of distinct subscales or facets (e.g., different personality traits within the BFI), offering opportunities to address more nuanced research questions about the correlates of musicianship and musical abilities.

The dataset also offers strong potential for methodological and statistical research. Researchers could systematically compare how different analytic strategies influence the conclusions drawn from the same data, reflecting contemporary concerns about analytic flexibility and researcher degrees of freedom[42] . The dataset is particularly well suited for examining the robustness of findings across alternative statistical models; this could be done, for example, by performing a multiverse analysis, which evaluates how different defensible analytic choices lead to convergent or divergent outcomes[37].

Limitations.

The dataset contains some inconsistencies. Some participants were categorized via recruitment as either musicians and nonmusicians but their answers in the self-report questionnaires did not reflect the original intent of the recruitment (e.g., musicians with at least 10 years of formal music training). Because participant recruitment was handled locally by each unit, it is not possible to determine whether these differences arose from mistakes in the recruitment process, mistakes by the participants (including misinterpretation of questions), or both. To support different analytical approaches, users are free to rely either on the original classification provided in the variable “group” or on participants’ self-reported musical background (e.g., Gold-MSI scores and custom music-training items). Users may also compare both sources and, if desired, apply their own inclusion criteria or redefine groups based on their research needs.

Another limitation of the dataset is its relatively narrow age range and international focus, with participants primarily between 18 and 30 years old and largely drawn from highly educated, industrialised, and relatively high-income societies. This restricts the generalizability of findings to other age groups and populations.

Although the dataset includes multiple variables describing musicians' activities, such as instrument type, hours of practice, and age of onset, some potentially relevant information is missing. Notably, details about the genre of music played (e.g., classical, jazz), the context of performance (solo versus ensemble), and whether participants were actively performing are not available. Similarly, potential expertise of nonmusicians in domains outside music was assessed using custom questions, which may indicate whether individuals consider themselves experts in certain fields, but do not provide comprehensive information such as years of experience or hours dedicated to practice. Moreover, this dataset includes only formally trained musicians, so any findings related to musicianship may not generalize to self-taught musicians.

As noted in the Translation section, some questionnaire versions were not validated at the time of data collection in certain languages. For these languages, localized versions were created by individual units specifically for this project. Although reliability indices suggest acceptable to excellent internal consistency for total scores and most subscales, sample sizes for some language-specific analyses may be limited, and therefore further validation is recommended before drawing firm conclusions.

Data Availability

All data files are publicly hosted on the Open Science Framework repository[20]. As reported above, in the OSF page of the project, the folder "Digital materials" includes the translations of the questionnaires (folder "LimeSurvey"). The folder "Data and analysis scripts" includes one zip file and one "readme" file (see Figure 3, left). The zip file includes all data (both raw and elaborated).

Code Availability

All codes to process the data, and experiment scripts are publicly hosted on the Open Science Framework repository[20]. Data processing codes can be found in the folder "Data and analysis scripts", and are contained in the zip file, the same mentioned in the Data Availability section. The experiment codes (i.e., jsPsych scripts) can be found in the folder "Digital materials" and then in the folder "jsPsych".

Acknowledgements

Elvira Brattico and Peter Vuust were supported by Danish National Research Foundation, project number 117; Fleur L. Bouwer was supported by Veni grant (VI.Veni.201G.066) awarded by the Dutch Research Council (NWO). Anna Fiveash was supported by the Australian Research Council (DE220100783). Author list and contributions were compiled with Tenzing[43].

Author Contributions

Conceptualization: Francesca Talamini, Massimo Grassi, Anne Caclin, and Barbara Carretti.
Data curation: Francesca Talamini, Massimo Grassi, Gianmarco Altoè, and Filippo Gambarota.
Formal analysis: Francesca Talamini and Filippo Gambarota.
Investigation: Anne Caclin, Lucrezia Guiotto Nai Fovino, Karla Aguilar, Christ B. Aryanto, Frederico C. Assis Leite, Aíssa M. Baldé, Laura Bishop, Graziela Bortz, Fleur L. Bouwer, Axelle Calcus, Giulio Carraturo, Antonia Čerič, Antonio Criscuolo, Léo Dairain, Simone Dalla Bella, Oscar Daniel, Anne-Isabelle de Parcevaux, Verónica Detlefsen, Victor Cepero-Escribano, Juliana L. d. B. Fialho, Caitlin Fitzpatrick, Anna Fiveash, Juliette Fortier, Noah R. Fram, Eleonora Fullone, Stefanie Gloggengießer, Lucia Gonzalez Sanchez, Reyna L. Gordon, Mathilde Groussard, Eleanor E. Harding, Kirsty Hawkins, Steffen A. Herff, Kelly Jakubowski, Maria G. Jol, Aarushi Kalsi, Veronica Kandro, Rosaliina Kelo, Sonja A. Kotz, Gangothri S. Ladegam, Bruno Laeng, André Lee, Miriam Lense, César F. Lima, Simon P. Limmer, Chengran K. Liu, Paulina d. C. Martín Sánchez, Langley McEntyre, Jessica P. Michael, Julieta Moltrasio, Niloufar Najafi, Jaakko Nokkala, Ndassi Nzonlang, Katie Overy, Andrew J. Oxenham, Edoardo Passarotto, Marie-Elisabeth Plasse, Herve Platel, Alice Poissonnier, Vasiliki Provias, Neha Rajappa, Pablo Ripolles, Michaela Ritchie, Italo R. Rodrigues Menezes, Rafael Román-Caballero, Paula Roncaglia, Farrah Y.-A. Sa'adullah, Suvi Saarikallio, Nora R. Serres, L. R. Slevc, Ragnya-Norasoa Souffiane, Florian J. Strauch, Nicholas Tantengco, Mari Tervaniemi, Rachel Thompson, Renee Timmers, Laurel J. Trainor, Clara Tuske, Jed Villanueva, Claudia C. von Bastian, Kelly L. Whiteford, and Emily A. Wood.
Methodology: Francesca Talamini, Massimo Grassi, Gianmarco Altoè, Elvira Brattico, Anne Caclin, Barbara Carretti, Véronique Drai-Zerbib, Laura Ferreri, Filippo Gambarota, Jessica Grahn, Antoni Rodriguez-Fornells, Swathi Swaminathan, Barbara Tillmann, Peter Vuust, Jonathan Wilbiks, and Marcel Zentner.
Project administration: Francesca Talamini and Massimo Grassi.
Resources: Francesca Talamini, Massimo Grassi, Elvira Brattico, Anne Caclin, Véronique Drai-Zerbib, Laura Ferreri, Jessica Grahn, Antoni Rodriguez-Fornells, Swathi Swaminathan, Barbara Tillmann, Peter

Vuust, Jonathan Wilbiks, Marcel Zentner, Christ B. Aryanto, Frederico C. Assis Leite, Aíssa M. Baldé, Graziela Bortz, Fleur L. Bouwer, Axelle Calcus, Antonio Criscuolo, Simone Dalla Bella, Anne Danielsen, Delphine Dellacherie, Tor Endestad, Anna Fiveash, Reyna L. Gordon, Mathilde Groussard, Assal Habibi, Eleanor E. Harding, Kirsty Hawkins, Steffen A. Herff, Kelly Jakubowski, Sonja A. Kotz, Bruno Laeng, André Lee, Miriam Lense, César F. Lima, Daniel Mirman, Daniel Müllensiefen, Katie Overy, Andrew J. Oxenham, Edoardo Passarotto, Alice Poissonnier, Pablo Ripolles, Rafael Román-Caballero, Paula Roncaglia, Wanda Rubinstein, Suvi Saarikallio, Daniela Sammler, Séverine Samson, L. R. Slevc, Mari Tervaniemi, Renee Timmers, Petri Toiviainen, Laurel J. Trainor, Claudia C. von Bastian, Florian Worschech, and Ana Zappa.

Software: Massimo Grassi, Anne Caclin, Lucrezia Guiotto Nai Fovino, Marco Roccato, Swathi Swaminathan, and Hannah Strauss.

Supervision: Francesca Talamini, Massimo Grassi, Elvira Brattico, Anne Caclin, Véronique Drai-Zerbib, Jessica Grahn, Antoni Rodriguez-Fornells, Barbara Tillmann, Jonathan Wilbiks, Deniz Başkent, Graziela Bortz, Fleur L. Bouwer, Axelle Calcus, Simone Dalla Bella, Delphine Dellacherie, Verónica Detlefsen, Tor Endestad, Anna Fiveash, Reyna L. Gordon, Mathilde Groussard, Eleanor E. Harding, Steffen A. Herff, Kelly Jakubowski, Sonja A. Kotz, André Lee, Miriam Lense, César F. Lima, Daniel Mirman, Julieta Moltrasio, Daniel Müllensiefen, Maria Gabriela M. Oliveira, Katie Overy, Andrew J. Oxenham, Pablo Ripolles, Rafael Román-Caballero, Wanda Rubinstein, Suvi Saarikallio, Daniela Sammler, Séverine Samson, E. G. Schellenberg, L. R. Slevc, Mari Tervaniemi, Renee Timmers, Petri Toiviainen, Laurel J. Trainor, and Florian Worschech.

Validation: Francesca Talamini and Massimo Grassi.

Visualization: Francesca Talamini, Massimo Grassi, Filippo Gambarota, Lucrezia Guiotto Nai Fovino, and Marco Roccato.

Writing - original draft: Francesca Talamini, Massimo Grassi, Véronique Drai-Zerbib, Laura Ferreri, César F. Lima, Julieta Moltrasio, Pablo Ripolles, and Hannah Strauss.

Writing - review & editing: Francesca Talamini, Massimo Grassi, Véronique Drai-Zerbib, Laura Ferreri, Anna Fiveash, Kelly Jakubowski, Sonja A. Kotz, André Lee, César F. Lima, Julieta Moltrasio, Katie Overy, Pablo Ripolles, Hannah Strauss, and Kelly L. Whiteford.

Competing Interests

The authors declare no competing interests.

References

1. Maguire, E. A. *et al.* Navigation-related structural change in the hippocampi of taxi drivers. *Proc. Natl. Acad. Sci.* **97**, 4398–4403, https://doi.org/10.1073/pnas.070039597 (2000).

2. Pascual-Leone, A. & Torres, F. Plasticity of the sensorimotor cortex representation of the reading finger in Braille readers. *Brain* **116**, 39–52, https://doi.org/10.1093/brain/116.1.39 (1993).

3. De Groot, A. D. *Thought and Choice in Chess*. (De Gruyter Mouton, 2014).

4. Chang, Y. Reorganization and plastic changes of the human brain associated with skill learning and expertise. *Front. Hum. Neurosci.* **8**, https://doi.org/10.3389/fnhum.2014.00035 (2014).

5. Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S. & Roberts, B. Are expert athletes 'expert' in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Appl. Cogn. Psychol.* **24**, 812–826, https://doi.org/10.1002/ACP.1588 (2010).

6. Hannon, E. E. & Trainor, L. J. Music acquisition: effects of enculturation and formal training on development. *Trends Cogn. Sci.* **11**, 466–472, https://doi.org/10.1016/j.tics.2007.08.008 (2007).

7. Micheyl, C., Delhommeau, K., Perrot, X. & Oxenham, A. J. Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* **219**, 36–47, https://doi.org/10.1016/j.heares.2006.05.004 (2006).

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

9. Degé, F., Kubicek, C. & Schwarzer, G. Music Lessons and Intelligence: A Relation Mediated by Executive Functions. *Music Percept.* **29**, 195–201, https://doi.org/10.1525/mp.2011.29.2.195 (2011).

10. Gjermunds, N., Brechan, I., Johnsen, S. Å. K. & Watten, R. G. Personality traits in musicians. *Curr. Issues Personal. Psychol.* **8**, 100–107, https://doi.org/10.5114/cipp.2020.97314 (2020).

11. Herholz, S. C. & Zatorre, R. J. Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron* **76**, 486–502, https://doi.org/10.1016/j.neuron.2012.10.011 (2012).

12. Moreno, S. & Bidelman, G. M. Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear. Res.* **308**, 84–97, https://doi.org/10.1016/j.heares.2013.09.012 (2014).

13. Schellenberg, E. G. & Lima, C. F. Music Training and Nonmusical Abilities. *Annu. Rev. Psychol.* **75**, 87–128, https://doi.org/10.1146/annurev-psych-032323-051354 (2024).

14. Corrigall, K., Schellenberg, E. G. & Misura, N. Music Training, Cognition, and Personality. *Front. Psychol.* **4**, https://doi.org/10.3389/fpsyg.2013.00222 (2013).

15. Grassi, M. *et al.* Do Musicians Have Better Short-Term Memory Than Nonmusicians? A Multilab Study. *Adv. Methods Pract. Psychol. Sci.* **8**, 25152459251379432, https://doi.org/10.1177/25152459251379432 (2025).

16. Chambers, C. D. Registered reports: a new publishing initiative at Cortex. *Cortex J. Devoted Study Nerv. Syst. Behav.* **49**, 609–610, https://doi.org/10.1016/j.cortex.2012.12.016 (2013).

17. Richardson, J. T. E. Measures of Short-Term Memory: A Historical Review. *Cortex* **43**, 635–650, https://doi.org/10.1016/s0010-9452(08)70493-3 (2007).

18. de Leeuw, J. R. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* **47**, 1–12, https://doi.org/10.3758/s13428-014-0458-y (2015).

19. Corsi, P. M. Human memory and the medial temporal region of the brain. vol. 34 891 (ProQuest Information & Learning, US, 1973).

20. Grassi, M. *et al.* Music Ensemble: comparing musical skills, cognition and personality of musicians and nonmusicians. *OSF* https://doi.org/10.17605/OSF.IO/Y97T3 (2025).

21. Raven, J. C., Raven, J. C., & John Hugh Court. *Advanced Progressive Matrices*. (HK Lewis London, 1962).

22. Wechsler, D. Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio TX NCS Pearson* **22**, 816–827 (2008).

23. Wechsler, D. *WAiS-Iii*. (Psychological Corporation San Antonio, TX, 1997).

24. Zentner, M. & Strauss, H. Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS. *Ann. N. Y. Acad. Sci.* **1400**, 33–45, https://doi.org/10.1111/nyas.13410 (2017).

25. Law, L. N. C. & Zentner, M. Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills. *PLoS ONE* **7**, e52508, https://doi.org/10.1371/journal.pone.0052508 (2012).

26. Ollen, J. E. A criterion-related validity test of selected indicators of musical sophistication using expert ratings. (The Ohio State University, 2006).

27. Müllensiefen, D., Gingras, B., Musil, J. & Stewart, L. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE* **9**, e89642, https://doi.org/10.1371/journal.pone.0089642 (2014).

28. Cardona, G., Ferreri, L., Lorenzo-Seva, U., Russo, F. A. & Rodriguez-Fornells, A. The forgotten role of absorption in music reward. *Ann. N. Y. Acad. Sci.* **1514**, 142–154, https://doi.org/10.1111/nyas.14790 (2022).

29. Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J. & Rodriguez-Fornells, A. Individual Differences in Music Reward Experiences. *Music Percept.* **31**, 118–138, https://doi.org/10.1525/mp.2013.31.2.11 (2013).

30. Soto, C. J. & John, O. P. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *J. Pers. Soc. Psychol.* **113**, 117–143, https://doi.org/10.1037/pspp0000096 (2017).

31. Hollingshead, A. A. *Four-Factor Index of Social Status.* (Yale University, New Haven, CT, 1975).

32. Schaal, N. K., Bauer, A.-K. R. & Müllensiefen, D. Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musikalischer Erfahrenheit anhand einer deutschen Stichprobe. *Music. Sci.* **18**, 423–447, https://doi.org/10.25656/01:12622 (2014).

33. Lima, C. F., Correia, A. I., Müllensiefen, D. & Castro, S. L. Goldsmiths Musical Sophistication Index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality and music preferences. *Psychol. Music* **48**, 376–388, https://doi.org/10.1177/0305735618801997 (2020).

34. Degrave, P. & Dedonder, J. A French translation of the Goldsmiths Musical Sophistication Index, an instrument to assess self-reported musical skills, abilities and behaviours. *J. New Music Res.* **48**, 138–144, https://doi.org/10.1080/09298215.2018.1499779 (2019).

35. Saliba, J. *et al.* French validation of the Barcelona Music Reward Questionnaire. *PeerJ* **4**, e1760, https://doi.org/10.7717/peerj.1760 (2016).

36. Baranski, E. *et al.* Translated Measures. https://osf.io/9zc7n/ (2019).

37. Danner, D. *et al.* Das Big Five Inventar 2. *Diagnostica* https://econtent.hogrefe.com/doi/10.1026/0012-1924/a000218 (2019).

38. Denissen, J. J. A., Geenen, R., Soto, C. J., John, O. P. & van Aken, M. A. G. The Big Five Inventory–2: Replication of Psychometric Properties in a Dutch Adaptation and First Evidence for the Discriminant Predictive Validity of the Facet Scales. *J. Pers. Assess.* **102**, 309–324, https://doi.org/10.1080/00223891.2018.1539004 (2020).

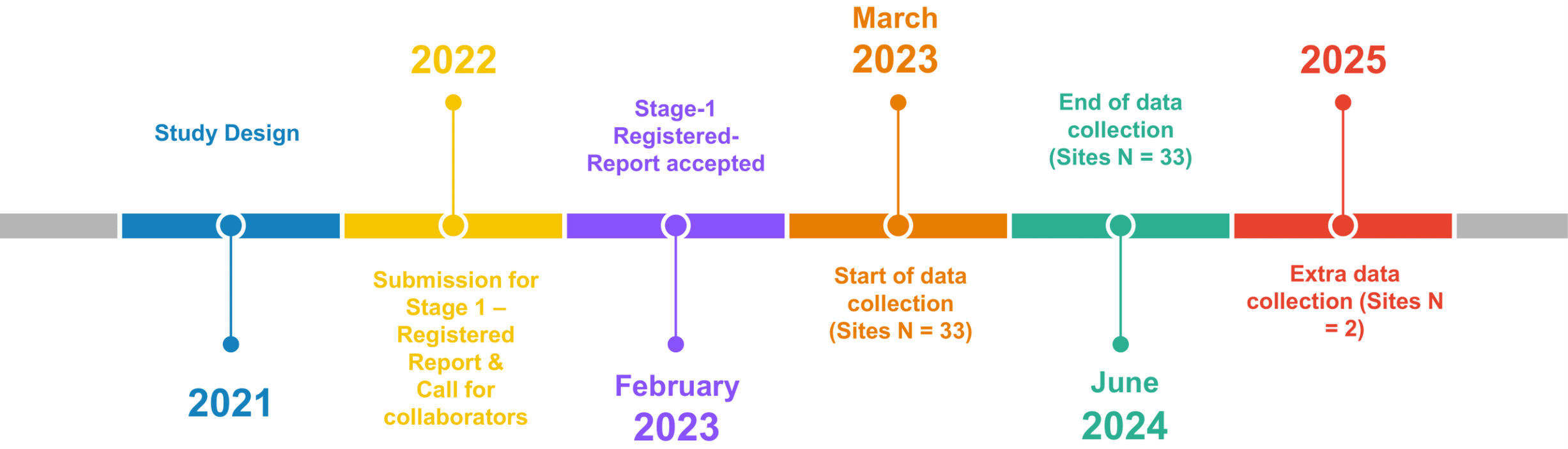
39. Lignier, B. *et al.* Factor structure, psychometric properties, and validity of the Big Five Inventory-2 facets: evidence from the French adaptation (BFI-2-Fr). *Curr. Psychol.* **42**, 26099–26114, https://doi.org/10.1007/s12144-022-03648-0 (2023).

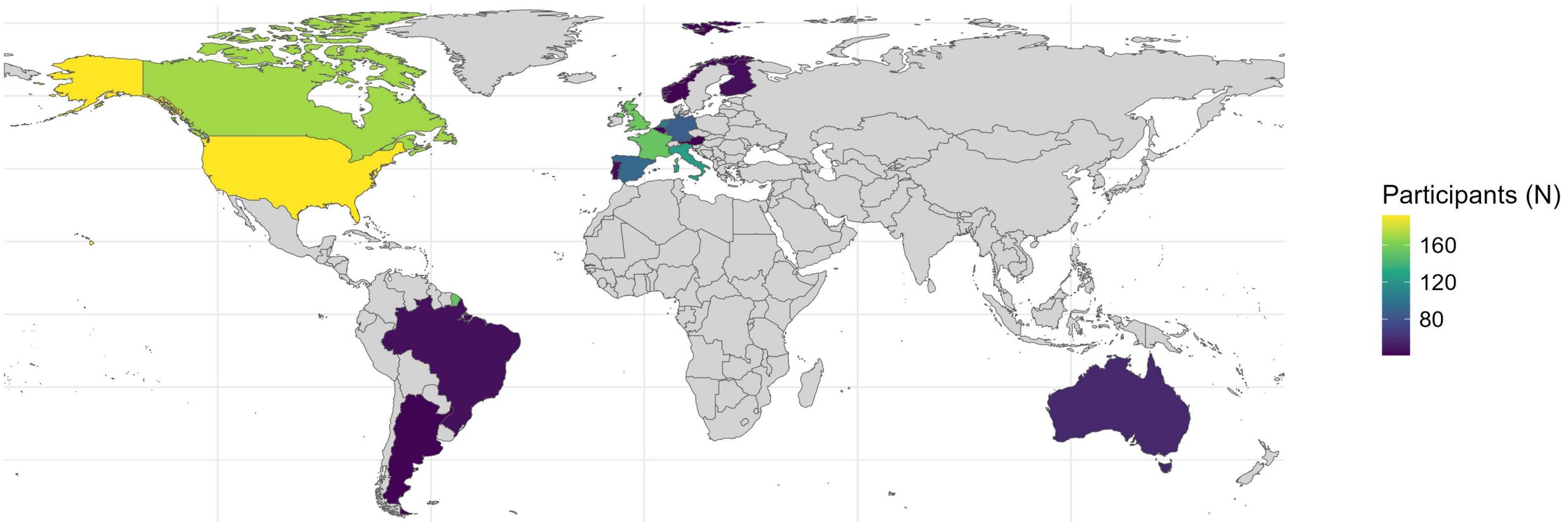
40. Van Der Maas, H. L. J. *et al.* A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychol. Rev.* **113**, 842–861, https://doi.org/10.1037/0033-295X.113.4.842 (2006).

41. Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S. & Engle, R. W. Attention control and process overlap theory: Searching for cognitive processes underpinning the positive manifold. *Intelligence* **91**, 101629, https://doi.org/10.1016/j.intell.2022.101629 (2022).

42. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366, https://doi.org/10.1177/0956797611417632 (2011).

43. Holcombe, A. O., Kovacs, M., Aust, F. & Aczel, B. Documenting contributions to scholarly articles using CRediT and tenzing. *PLOS ONE* **15**, e0244611, https://doi.org/10.1371/journal.pone.0244611 (2020).


Study Design
2021
2022
Submission for Stage 1 – Registered Report & Call for collaborators
Stage-1 Registered-Report accepted
February 2023
March 2023
Start of data collection (Sites N = 33)
End of data collection (Sites N = 33)
June 2024
2025
Extra data collection (Sites N = 2)

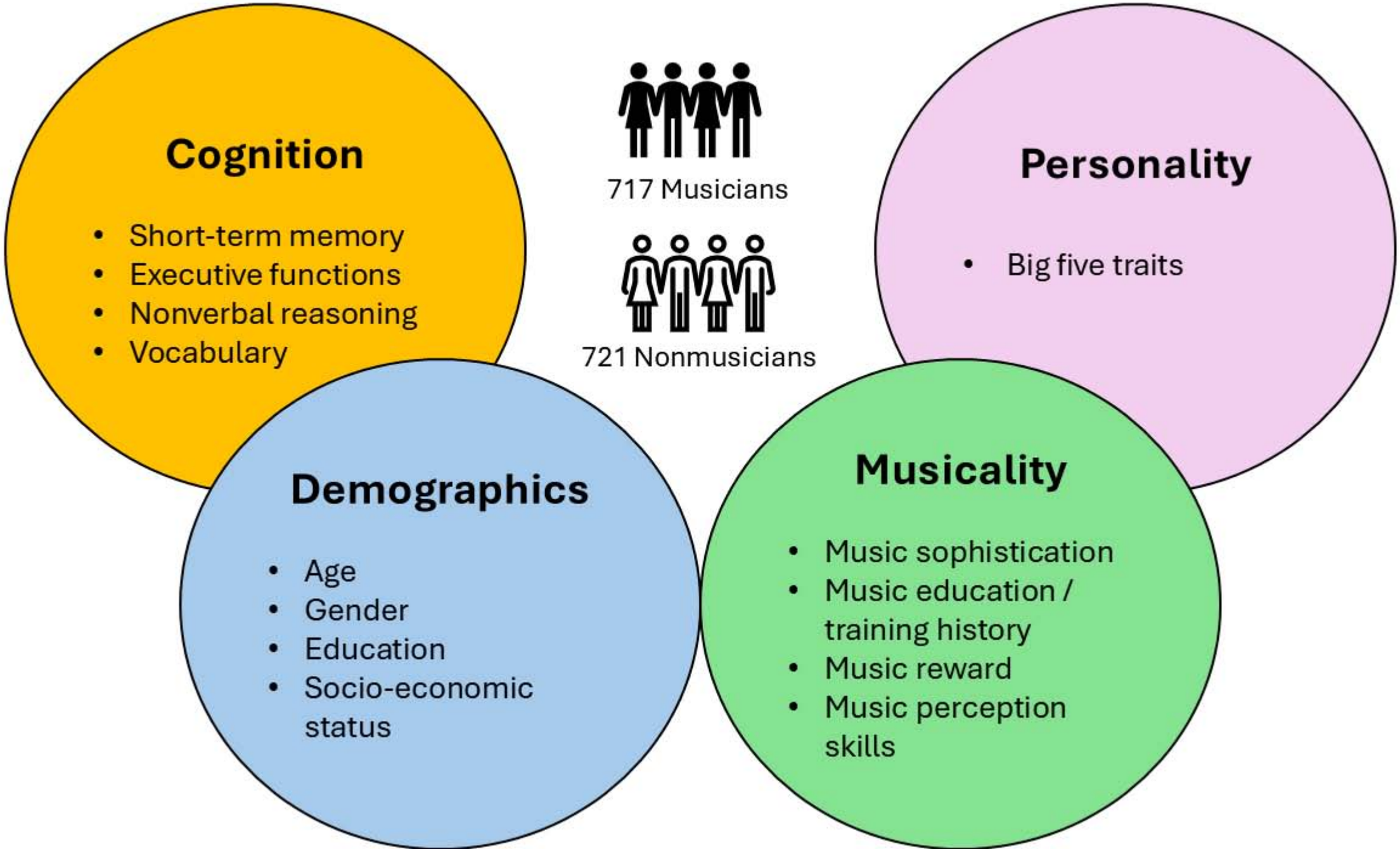

Participants (N)
160
120
80

OSF

```
Files
├── Data and analyses scripts
├── Digital materials
│   ├── jsPsych scripts
│   └── LimeSurvey Questionnaires
├── Publication in Advances in Methods...
└── Publication in Scientific Data
```

analysis.zip

```
multilab-musicians-memory
├── data
│   ├── clean
│   ├── fixing
│   ├── id
│   ├── raw-after-selection
│   ├── raw-before-fixing
│   └── raw-before-selection
├── files
├── plots
├── objects
├── R
├── renv
├── reports
├── scripts
│   └── create-figure-1
└── tables
```


Cognition
• Short-term memory
• Executive functions
• Nonverbal reasoning
• Vocabulary
717 Musicians
721 Nonmusicians
Personality
• Big five traits
Demographics
• Age
• Gender
• Education
• Socio-economic status
Musicality
• Music sophistication
• Music education / training history
• Music reward
• Music perception skills