Best Practices in Collecting and Encoding Behavioral Data for Analysis in fMRI Studies of Multilingualism and Musicianship

Hannah Folks, Caroline Gamard, Hope Morales, Devika Naphade, Krishna Sinha, Hanna Stern, Lena Yannella

The need to improve background and behavioral data collection for subjects participating in fMRI studies of bilingualism and multilingualism have long been noted (Andrews 2014, Birdsong 2006, de Bot 2009, and others). The current study of bi/multilingualism and musicianship includes interviews with 78 subjects and examines existing approaches that have been developed specifically for studies of bilingualism, critiques the problems with self-reporting, expands the scope to include multilingualism and musicianship, and proposes a new approach that strives to be both empirically and ecologically valid in combining interviews with proficiency testing. The Common European Framework of References for Languages (CEFR) is an internationally recognized framework for assessment of the world's languages with testing instruments across five modalities and six levels. Evaluation of different interview techniques and questions are informed by best practices in sociolinguistic and anthropological ethnographic and behavioral data collection. This multidisciplinary approach synthesizes qualitative and quantitative data points and introduces strategies to minimize bias in subject selection, interviews, testing and analysis. Our conclusions demonstrate the importance of non-essentialist variables and evidence-based empirical data in framing questions for bi/multilinguals and musicians, and multifactor proficiency testing results.

Introduction

Neuroimaging studies with a focus on bi/multilingualism and musicianship are becoming more prevalent in the past 15 years and it is time to examine behavioral assessment measures that accompany these studies. The goal of the current work is to review previous approaches to behavioral data, discuss strengths and weaknesses, and suggest more robust assessment process of bi/multilingualism and musicianship that would allow researchers to draw stronger conclusions from the data and analyses, as well as improve the basis for comparison across studies. We also examine a new data set of 78 subjects and attempt to address two of the major weaknesses currently found in the field, including (1) the lack of evidence-based empirical proficiency testing and (2) a focus on essentialist variables that fails to take into account important research from the fields of theoretical linguistics, anthropology, sociolinguistics and ethnomusicology.

Currently, neuroimaging studies of bi/multilingualism rely on a spectrum of behavioral data for establishing baselines, including self-reporting of language abilities and use, age of language acquisition, quantification of the regular use of two languages, and cognitive assessment tools including the Simon task, the Mini Mental State Examination (MMSE) and the Montreal Cognitive Assessment in determining eligibility for neuroimaging studies (Bialystok et al., 2007, Van de Putte et al., 2018), Del Maschio et al., 2019; Abutalebi et al., 2015, Kousaie et al., 2015). Another common practice in the field is the use of several different tests to assess each modality of language processing separately (Sleve & Miyake, 2006), picture naming tasks

(Del Maschio et al., 2019), and assessments such as the animacy judgement task that measure the speed with which participants can switch between languages (Kousaie et al., 2015).

The majority of neuroimaging studies of bilingualism relies on a minimum age of acquisition of a second language as an indication of bilingualism (Abutalebi et al., 2015; Del Maschio et al., 2019; Kousaie et al., 2015). However, language acquisition is a complex process that is determined by numerous "continuous, not all-or-nothing" variables beyond age of acquisition (Birdsong, 2006). Birdsong's important meta-analysis on the interrelationship of age and second language acquisition raises serious concerns about earlier approaches to possible critical periods in second language acquisition and brings forward important empirical data from both behavioral studies and neuroimaging studies of L2 (Birdsong, 2006). Other research has identified the importance of how bilinguals use their languages and how this affects outcomes, including different modalities (such as reading, writing, translation), the contexts in which they are used, and the level of immersion as (Van de Putte et al., 2018; Pliatsikas et al., 2015).

Neuroimaging studies of music have similarly neglected to include measures of proficiency, choosing to rely on age of acquisition and self-reporting as indicators of musicianship. The exception is studies on absolute pitch, where extensive testing of absolute and relative pitch are fundamental to the studies (Van Hedger 2015, 2019).

In our examination of all of these questions, we attempted to undertake a more robust and comprehensive assessment of language and music experiences in behavioral data collection that will be able to represent a variety of complex factors in not only acquisition of languages and musicianship, but also lifelong activity in both across multiple modalities.

One promising development in the field are questionnaires within neuroimaging studies that take into account numerous important, non-essentialist variables such as education, different contexts in which languages are used, and the method by which the language was learned (Anderson et al., 2018). This approach can be augmented by asking broad, open-ended questions in a live, face-to-face interview process.

Another recommendation is the inclusion of an internationally-recognized proficiency battery, such as the Common European Framework of Reference for Languages (CEFR), to standardized measurements across studies and allow comparisons to be drawn internationally. The CEFR is widely trusted by experts. It takes into account four different modes of language and conducts tests in oral and written tasks, making it one of the most robust proficiency references available (Arnott et al., 2017). A similarly robust assessment of musicianship can be developed, and inclusion of such empirically-based tests will vastly improve the validity of the conclusions drawn from neuroimaging data.

A common practice for music research is to recruit conservatory students as well as music teachers in higher education and professionals (Oechslin et al., 2013; Fasano et al., 2020) because these groups have established musical proficiency. The Helsinki Inventory of Music and

Affective Behaviors (HIMAB) is a questionnaire that has also been used to identify study participants (Gold et al., 2013). It includes both open-ended and quantitative questions on musical training, emotional connections to music, the use of music in different contexts, and music consumption (Gold et al., 2013). Such inventories are powerful, especially when paired with a subject with known proficiency, as they provide a detailed picture of musical history; however, a more open-ended, live interview process is necessary to encompass the complexities of musicianship.

In our study, we sought to combine these two primary methods, a language and music background questionnaire with robust proficiency testing, to demonstrate how an interview process, when accompanied by empirically based measurements for fMRI studies of multilingualism and musicianship, is the best practice for data collection.

Defining Basic Terms

Multilingualism and musicianship are popular areas of study in the field of neuroscience due to their dramatic effects on neural structure (Van de Putte et al., 2018; Pliatsikas et al., 2015; Weiss and Bidelman 2015; Jäncke 2009) and potential to build cognitive reserve (Perani et al., 2017; Grundy et al., 2017; Bialystok et al., 2007, Andrews et al. 2021). Cognitive reserve refers to neural changes resulting from high-end cognitive processes that increase the brain's resilience to neural insult throughout the life cycle, particularly during aging and dementia pathology (Bialystok et al., 2007).

There is no general consensus throughout the community of practice on a single definition of "bilingualism," "multilingualism" or "musicianship." Different studies of languages and brain have emphasized different aspects of bi/multilingualism, primarily focusing on a lifetime of continuous use of two (or more) languages (Bialystok et al., 2007; Anderson et al., 2018). The definition of bi/multilingualism used in this paper is the use of two or more languages on a daily basis at a high functional level; this emphasizes the necessity of robust, regular language use for generating data on language in the brain from neuroimaging.

Similarly, musicianship has not been well-defined in studies examining cognitive reserve, and there are rarely clear criteria for designating musical proficiency. Our definition of musicianship is the acquired abilities and skills that come with being a continuously dedicated musician, such as the capability to "selectively attend to the fine acoustic details of sound sequences" (Schroeder et al., 2015) and corresponding fine motor actions. Definitions of musicianship used in neuroimaging research on cognitive reserve should include quantifiable information (years of playing, performance, practice, daily/weekly number of hours spent playing, etc.) as well as include criteria to demonstrate high musical proficiency for singers and instrumentalists (including soloists, orchestra, symphony, theatre, choral, etc.)

Multimodality and embodied cognition are two methodological approaches that are critical in neuroimaging studies of cognitive abilities, including language(s) and music. Multimodality asserts that sensory and motor brain regions are interactive and interconnected in higher cognitive functions (including networks for language(s) perception and production (Gallese & Lakoff, 2005)). Not only are language and music embedded in the sensory-motor system, but they are also embodied. Embodied cognition refers to the way that human languages produce meaningful outcomes of our perceptions; it is "structured by our constant encounter and interaction with the world via our bodies and brains" (Gallese & Lakoff, 2005). Embodied cognition approaches suggest that languages and musics have the ability to not only influence individual perceptions and understandings of the world, but also affect how peoples' identities emerge from situated social interactions (Bucholtz & Hall, 2004). We considered these important social and methodological factors when selecting and framing our interview questions.

Sociolinguistic Factors of the Interview Process

For more than half a century, sociolinguistics and ethnographic research have worked to critique and fine tune behavioral data collection methods, including interview processes. Research in the field provides effective guidelines and approaches to conducting ecologically-valid interviews while also examining the researcher-subject relationship and interactions, variables of power and status, and implicit bias in the interview process and interpretation (Wolfram & Schilling-Estes, 1995; Schilling-Ester, 2008; Wolfram, 2018).

One crucial aspect of the interview process that has been researched extensively is the emergence of identities in interaction. Identities include macro-level demographic categories and specific local cultural positions; they encompass both temporary and specific interactional stances and participant roles that arise during the interview process (Bucholtz & Hall, 2005, 591-593). Identities are not inherent properties of individuals or groups: they are not given *a priori*. Rather they emerge through situated social interactions with other identities (Bucholtz & Hall, 2004, 375-376). Past experiences, societal norms (Pavlenko, 2005) and life scripts (Bernsten & Rubin, 2004), which are situated within social interactions and emerge from memory and culture, influence opinions about oneself and one's language/music abilities and preferences. These opinions are frequently associated with strong emotions, so they play a role in shaping identities and the emergence of identities during the interview process. Researchers in neuroimaging must apply this extensive research on the emergence of identities during the interview process and in behavioral questionnaires to future studies to facilitate ecological validity and more robust contextualization of results.

Empirical Methods and Analysis

Self-Reporting and the Need for Interviews

Self-reporting refers to a common behavioral data collection practice in which study participants are allowed to evaluate their own abilities. However, research has indicated that individuals have limited accuracy in evaluating their own language abilities (Labov, 1966; Trudgill, 1972). This method does not provide the most robust or significant insights into participants' actual language and music abilities, and it is not an empirically valid practice (Labov, 1966; Trudgill, 1972; Gollan et al., 2012; Andrews, 2014). Furthermore, the wide variety of self-reporting measures used across studies has created noticeable inconsistencies in interpreting results (Paradis, 2004). Despite these shortcomings, the field of neuroscience and many neuroimaging studies of languages have continued to rely on self-reporting. Therefore, 21st century neuroimaging studies of the brain and language must now demand an empirical, ecologically valid method of measuring multilingualism and musicianship: proficiency tests. Standardized proficiency testing across modalities measures language and musical abilities in a consistent, ecologically valid manner.

In addition to (not in lieu of) standardized proficiency testing, interviews provide a way of examining language abilities in a natural, real-time manner (Andrews, 2014). Interviews provide accommodation for both the interviewer and interviewee as the interviewer can modify their questions accordingly to gauge each person's individual language and musical proficiencies, thereby overcoming the restrictions and gaps of the self-reporting method. While an interview can allow for self-reporting of important biographical information, any self-assessment of language and musical ability becomes difficult to navigate; not everyone is qualified to evaluate their own language and music performance. Because of this, interviews must be paired with objective proficiency measurements in order to gain tangible statistical support for any information that has been self-reported.

Evidence-Based Empirical Data

Neuroimaging studies must move toward the use of evidence-based empirical data (EBED) to assess and identify study participants. EBED corroborates the qualitative behavioral data frequently collected on study participants with robust quantitative data, giving researchers the most comprehensive picture of their participants.

In general, EBED methods normally include tests for statistical significance of the experimental results, a multivariate analysis of covariance (MANCOVA), or linear regression approaches to detect significant trends within the empirical data without confounding variables. For example, Andrews et al. (2013) conducted a longitudinal study in which multilingual participants were followed as they acquired Russian over a 16-month period. These subjects were all uniformly proficiency tested in the Russian language over this period and put through to multiple sets of fMRI scans as well (Andrews et al., 2013). Because the research team applied MANCOVA as their primary statistical analysis, they were able to demonstrate that the obtained proficiency measurements correlated significantly with the fMRI scan data (Andrews et al.,

2013). Thus, the use of EBED in the statistical analysis increased the validity of the study and the strength of the conclusions.

In the current study, standardized proficiency baselines were utilized to provide EBED. Interviewees must receive a passing score at the required CEFR proficiency level for the study in order to be eligible to participate in the fMRI study. Proficiency testing took place after an extensive interview process, which was necessary to assess whether participants were good candidates for the study.

Proficiency Testing and Other Methods of Assessment

To collect robust, empirically-valid data on study participants, performance and proficiency testing is required for both musicians and multilinguals in our study. Our research team chose to rely on proficiency testing, rather than other methods of assessment because proficiency tests measure a person's general mastery of a certain language by including various dimensions of assessment and distinct levels for these skills (i.e. reading, writing, listening comprehension, etc.).

Similarly for music, proficiency batteries that specifically test for a wide variety of musical skills such as (but not limited to) absolute pitch, melodic dictation, and auditory memory skills are reliable indicators of multimodal musical abilities (Van Hedger et al., 2019; Dooley & Deutsch, 2010). Musicianship proficiency can be most easily and clearly designated by higher education, such as a PhD in performance, but the most robust and ecologically-valid way of assessing high-end musicianship is evaluation in live or taped performance before a panel of established music experts.

In contrast to proficiency tests, an achievement test measures how much a person has learned with reference to a specific course of study (Davies, 2007). For example, the Advanced Placement (AP) tests offered by the College Board are one of the most popular sources of higher education achievement testing. These standardized assessments, offered for over 20 high school subjects, measure students' crystallized knowledge after a year of study (College Board AP, n.d.). The College Board offers tests in literature and language composition, as well as one on music theory. Such achievement tests are given after one or two years of study at the grade-level; they are not comprehensive and do not assess for general mastery of an entire subject. Therefore, they are not the most ecologically-valid manner of measuring linguistic or musical proficiency.

Finally, the third kind of assessment is the placement test, which is designed to place students at an appropriate level in an educational program or course and are frequently unique to a particular institution. For instance, the University of Wisconsin system has its own Spanish placement test (2011), and The Peabody Institute of The Johns Hopkins University has an assessment of ear-training, keyboard skills, and sight singing to place students in an appropriate Music Theory seminar within their Masters of Music program. Placement tests are not designed to specifically evaluate a student's linguistic or musical background; rather, the sole purpose for

the evaluation is to accurately place students into an appropriate course level as they enter a new and unfamiliar institution (Davies, 2007).

Considering all of this information, proficiency testing is the most reliable measurement of linguistic and musical abilities. Proficiency tests should be used to establish certain mastery standards that are expected from all participants, acting as a gate-keeping mechanism to determine study eligibility. Proficiency tests also control for different educational experiences because not everyone has taken the same kinds of tests or assessments. Proficiency tests also reliably measure a range of modalities, such as comprehension, production, lexical and discourse competence, and writing, more reliably than the myriad of placement and achievements tests that are available (Tremblay, 2011). Finally, asking all study participants to complete and reach a particular threshold in one common proficiency battery allows fMRI activations to be appropriately contextualized and understood (Andrews et al., 2013).

Our research team chose the Common European Framework of Reference of Languages (CEFR), as the proficiency battery for the ongoing study. The CEFR, which was developed by the Council of Europe for the European Union and is recognized across the globe and across world languages, consists of a five-exam battery that tests multiple modalities of language: oral/speaking skills, listening comprehension, reading, grammar/lexicon, and writing (Andrews, 2014). Furthermore, the CEFR has a detailed and objective framework that is easily translatable to other languages, so it can be used to assess multiple languages, allowing for cross-study comparisons to be drawn internationally (Council of Europe, 2020). Other preeminent tests such as the TOEFL and IELTS are limited in their applicability to multiple languages because they only assess English language skills, possibly excluding valuable study participants (Kunnan & Carr, 2017). Similar proficiency tests exist in other languages, such as the DELE exam to test for Spanish and the TORFL for Russian. However, the CEFR has proven to be a reliable framework for proficiency testing across world languages and its international recognition makes it an excellent candidate to be used as a baseline standard in a majority of studies going forward.

Duke University Bilingualism, Multilingualism, and Musicianship Study

Method Design and Procedures

Seventy-eight subjects participated in the study. Interviews were collected starting in 2017. Study participants had demonstrated records of high-end proficiency in more than one language and at least one instrument. The research team had previous empirical knowledge of each participant's behavior and linguistic and musical histories, and interviews were conducted to gather qualitative and quantitative behavioral data about the interviewee's linguistic and musical experiences to determine eligibility for an fMRI study.

We interviewed 78 subjects, ages 20-70, and coded 18 of the interviews[1]. All interviews were conducted in English, but we acknowledged that each interview could have been conducted in the interviewee's preferred language for their comfort. Each subject signed an informed consent form that outlined the purpose of our study and the subject's role in the study. The purpose of our study was stated to participants as follows: "Both speaking more than one language and having musical experience are known to have certain advantages. We expect that our study will help us better understand the effects that music and language have on healthy aging."

Interview Process and Data Collection

The purpose of our interview process was to move away from a focus on self-reported proficiency, age of acquisition, or years of linguistic or musical practice as the determining, essential variables of language and music acquisition to a broader understanding of the non-essentialist, "continuous, not all-or-nothing" variables that play significant roles in language proficiency (Birdsong, 2006). These variables include "motivation to pass for a native, motivation to acquire lexico-grammatical accuracy, psycho-social integration with the L2 culture, aptitude, learning styles and strategies," L1-L2 pairings, level of L2 practice and use, relative dominances of languages (Birdsong, 2006), and purposes for which the languages are used (Van de Putte et al., 2018). Previous language background questionnaires such as those used by Perani et al. (2017) and Anderson et al. (2018) also emphasize the importance of gathering data on participants' use of various languages in distinct contexts with different people, as well as the language of education and certifications or degrees earned in each language.

With all of these complex, dynamic variables in mind, we compiled a list of broad, open-ended questions designed to allow the interviewees to elaborate as much as possible on personally relevant experiences to capture the factors that most predominantly affect their musical and linguistic experiences. The interview questions can be found in Appendix 1: Music, Language and Aging (MLA) Interview Questions. While previous studies have presented participants with predetermined categories of the common contexts in which they may be likely to use a particular language (Anderson et al., 2018), we kept our questions as open-ended as possible to allow interviewees to communicate their unique, personal narratives without being boxed into categories that may not fit their experiences. We also did not ask participants to self-assess their own proficiency, which continues to be a common practice in data collection for fMRI studies (Anderson et al., 2018; Kousaie et al., 2015), choosing instead to gather proficiency data through empirical testing. However, we did allow self-reporting of standardized test scores and any awards or achievements that indicated linguistic or musical proficiency. Finally, we also included an important aspect of language acquisition that has long been ignored by behavioral data collection: emotional connections to language and music.

---

[1] Due to the expansive data collected from 78 subjects, the coding process performed was only for a subset of 18 of those subjects. For more information, see Appendix 2.

Interviews were conducted in person or via Facetime. Past studies have used written questionnaires (Anderson et al., 2018, Perani et al., 2017; Gold et al., 2013), but we chose to conduct in-person interviews due to powerful sociolinguistic research that points to the advantages of open-ended (Gill et. al., 2008), face-to-face interactions (Hymes, 1972) and the importance of social cues exhibited by the interviewee, such as voice, intonation, body language, and facial expressions (Opdenakker, 2006), which, in combination with eye contact, enable *mind sharing*[2] (Donald, 2001). Interviews have been shown to improve mutual understanding between those involved due to increased clarification, vocalized utterances suggesting interviewer interest, and the ability of the interviewer to check for sufficiency and relevance of answers (Irvine et al., 2012).

All subjects were asked the same set of questions (provided in Appendix II) pertaining to both their linguistic and musical abilities to maintain uniformity. Interviewees were given the option to review the set of questions just prior to conducting the interview to prepare themselves and begin to formulate answers.

All study participants were asked about their willingness to participate in a recorded interview, and willing participants were recorded. The researchers took detailed written notes throughout the interview process, regardless of whether or not the participant agreed to be recorded. Once all interviews were complete, all of the questions were categorized and coded to produce extractable data.

Interview data was condensed into seven categories as shown in Figure 1.

Figure 1

Category	Data collected
Geography and home environment	Places, regions, settings, exposure, accessibility
Specific languages and instruments	L1(s), L2/3(s), language of education, instrument(s), different music for different instruments, make and model of instrument

---

[2] Merlin Donald defines *mind sharing* as a specialized form of social intelligence that is more highly evolved in humans than in other primates. It is the human capacity and evolutionary skill to not only analyze one's own thoughts, but to form effective judgements about others' mental states as well. It allows the documentation of "the phenomenology of experience… and consciousness" by monitoring dimensions like linguistic "mindreading vernacular" ("'I think,' 'they know,' 'you feel,' 'he is sorry'…") and other signs ("patterns of movement, shifts in gaze or attitude, gestures, emotional words…") effortlessly and simultaneously.

Number of years with languages and music	Beginnings and ends, training (formal/informal, methods of instruction)
Intensity	Daily, weekly, monthly, etc., playing, languaging, speech communities and communities of practice
Emotional ties	Memories, favorite texts (linguistic and musical), comfort levels, families and friends, cultural institutions, religion, code-switching, motivation, talent, prestige, discrimination
Proficiency measures for languages and music	Testing (ACTFL, SAT, AP, IB, CEFR, ILR [mono- or bilingual tests], etc., national competitions, concerts, organizations [band, orchestra, symphony], performances, medals, etc.
Professional aspirations and career paths	World musics, instrumentation and languages (importance of cultural context)

A subset of 18 interviews was coded according to the categories above to provide an example of the data collection process and the breadth of data afforded by the interview process.

Results

The aggregated results from this subset are shown in Figures 2 and 3, and a larger table showing each individual's results is found in Appendix II. One of the greatest advantages of the interview process is the rich, informative detail that the individual narrative provides, and examples of this qualitative data are also provided in Appendix II.

As shown in the data tables, a wide variety of languages and musical instruments are represented among our interview subjects. Various types of linguistic and musical assessment were reported as measurements of proficiency. Subjects also described a diverse array of instruction methods and learning environments that are not captured by the broad categories listed in Figure 2, which is why the open-ended interview format is so much more ecologically valid than questionnaires, and not all of the participants who fit in a particular category listed below had similar experiences.

We also gathered extensive detailed data on participants' emotional connections and attitudes toward their languages and music. This data was too individualized to be categorized in

Figure 2. Broad categories related to emotional connections such as "language/music used in religious services" or "genre of music of most vivid music-related memory" could be drawn, but this would lead to the loss of the individual detail that makes each participants' narrative so valuable and would not yield any meaningful conclusions. In this manner, the open-ended interview format is much more powerful than questionnaires. An example of powerful individual narrative is provided below to demonstrate how making broad generalizations about emotional connections to languages and music would eliminate the most revealing data.

Figure 2

Languages	Arabic, ASL, Chinese, English, French, German, Hindi, Japanese, Latin, Mandarin, Moldovan, Russian, Spanish, Telugu
Format of instruction	Language spoken in home, primary instruction in school, extracurricular instruction in school, higher education instruction, in-country experience, short immersion course, private tutoring, exposure in workplace, exposure in dance classes, virtual instruction, online resources, self-taught
Years of training[3]	1.5 – lifelong
Testing	AP exams (Spanish, Latin, Japanese), SAT Subject Test (Spanish), National Latin Exam, Japanese-Language Proficiency Test
Number of participants with two languages	7
Number of participants with three languages	6

---

[3] Many difficulties arose in identifying the number of years a participant had spoken each language. For instance, participants were not asked their age, so we could not assign numbers to languages they had been speaking their entire lives. There is also uncertainty about the age at which a person begins to speak a language. Finally, since many participants took language classes in high school but did not continue with any formal education in college, it was unclear if they still spoke that language. For this reason, the number of years of education in each language was much more easily described by the participant and more accurately reflects reality, so it is included in Table 1, rather than the number of years of speaking each language. In many cases, participants discussed receiving linguistic training during elementary or middle school, and the researchers had to estimate a number of years, which posed a challenge to assigning a number of years of formal foreign language training to each participant. We estimated that elementary school lasted 5 years and middle school lasted 3 years. This was also true for the musicianship data.

Number of participants with four languages	5

Figure 3

Number of participants	18
Instruments	Baritone, drums, flute, guitar, harmonium, harp, piano, piccolo, saxophone, trumpet, viola, violin, voice
Format of training	Musical higher education, private lessons, university orchestra, school orchestra, school choir, marching band
Years of training	1-67

Narrative Data: Emotional Ties (an example)

The participant's most enjoyable musical moment is when she is on stage playing a concerto. In 1995 or 1996 the emperor of Japan visited China and was given a private concert by the Chinese government, at which the participant was a soloist. She played a traditional classic piece for the guzheng called *High Mountain on the Fuji River*. It is a very difficult piece to understand, and after the concert, the emperor spoke to her about the piece and told her what it had made him think about. She was very impressed by his opinions on the piece. In 2007, she performed as a soloist with the North Carolina Symphony, which was the first time she had ever been on a professional stage in the United States. After her performance, the audience kept applauding for three minutes, and she kept bowing for three minutes because she did not realize that they wanted an encore. Encores are part of American performances that she had never experienced in China. She was very embarrassed, and she still regrets this moment deeply. Her favorite musical text is called *Butterfly Lover*, a tragic love story about lovers who kill themselves because they cannot be together and are reunited after death as butterflies. This piece of music makes her very emotional because she is a romantic who believes in true love. Her family and society expected a lot of her while she was growing up, but when she plays this piece, she is allowed to be emotional and becomes another person, a "crazy Chinese woman." She is allowed to escape harsh expectations and become fully immersed in the character.

Discussion

The interview process is a crucial component in collecting behavioral data for fMRI studies of language, music, and the brain. Moving beyond age of acquisition and focusing on non-essentialist variables, interviews provide valuable insight into participants' histories and emotional connections with language and music. However, it is worth emphasizing that while self-reporting can and should be used to understand complex individual narratives, studies have shown that individuals are unable to correctly self-report their musical and language proficiencies (Sleve and Miyake, 2006; Delgado et al., 1999). Therefore, interviews alone should not be used to determine proficiency.

Furthermore, the interview process presents some limitations and potential biases. In this study, participants were easily accessible and known by the interviewers before the process. However, in an actual fMRI study, random sampling is advised. While interviewer-interviewee familiarity increases ecological validity by encouraging authenticity in participants, it also limits the range of identities represented in the participant pool. Additionally, although we did give participants the choice of participating in the interview process in another language, this possibility was limited by the languages spoken by interviewers, and none of the interview participants requested to be interviewed in another language. All interviews in this study were conducted in English, contributing to an Anglocentric bias since English was the L1 or L2 of all participants. Participants with an L1 other than English may have felt uncomfortable, undermining the reliability of the interview process. Another factor in the interview process that may have contributed to participant discomfort was recording the interview process. The process of recording is beneficial because it allows researchers to recall specific response and reaction details (Westera et al., 2017). However, recording has been shown to generate feelings of judgment, a reluctance to open up, and reduced authenticity in responding. Even when subjects "accept the recording as part of the study," recording has been shown to minimize subject engagement (Al-Yateem, 2012). Our familiarity with the participants helped minimize the negative effects of recording, but it is important to be mindful of the limitation, especially when interacting with a random sample of participants.

The most salient shortcoming of the interview process is that open-ended interview data must be accompanied by evidence-based empirical proficiency data. Participants reported a wide range of proficiency assessments, as shown in Figure 2, that make comparisons across participants difficult. Furthermore, each different proficiency test measures different modalities of language and music on different scales. Therefore, fMRI studies of language and brain must combine this interview process with an internationally-recognized proficiency test such as the Common European Framework of Reference for Languages (CEFR). Participants should only contribute to a study after meeting a predetermined baseline on such a proficiency test and then participating in an extensive interview process since cognitive changes are associated with

increased levels of linguistic and musical proficiency. Therefore, selecting and investing in fMRI candidates efficiently requires strategic behavioral data collection.

Multilingualism and musicianship are important areas of research given their potential role in building cognitive reserve. A lifetime of multilingualism has been shown to delay the onset of dementia symptoms by an average of 4.1 years (Bialystok et al., 2007). Most recently, Andrews et al. 2021 showed the potential impact of lifelong high-proficiency musicianship on preserving white matter integrity in aging.

It is of the utmost importance to use ecologically- valid, empirically based methods of data collection, and proficiency testing in order to draw meaningful conclusions from fMRI studies of brain and language(s). Future research within the fields of multilingualism and musicianship and their role in building cognitive reserve should continue to incorporate the data and methods suggested in this article, in order to deepen the context for interpretation of neuroimaging results.

Conclusions: A short summary

I. Neuroimaging studies of brain and language should combine detailed behavioral data with empirically-based proficiency testing when assessing multilingualism and musicianship. The use of an internationally-recognized proficiency framework such as the Common European Framework of Reference for Languages (CEFR) is critical for providing baselines that allow for robust interpretation of activations resulting from the functional scan data.

II. Standardized proficiency measurements must be accompanied by interview data that captures the vast range of dynamic, non-essentialist variables that affect acquisition and processing of music and language.

III. Design of the interview format may be enhanced by implementing specific features drawn from sociolinguistic research on behavioral data collection, including, where possible, broad, open-ended questions and face-to-face interactions in real time.

Our study has clarified the best practices in the field for behavioral data collection for neuroimaging studies of brain and language: an ecologically valid, open-ended interview process in combination with rigorous, empirically based proficiency testing.

References:

Abutalebi, J., Canini, M., Della Rosa, P., Green, D. & Weekes, B. (2015). The neuroprotective effects of bilingualism upon the inferior parietal lobule: A structural neuroimaging study in aging Chinese bilinguals. *Journal of Neurolinguistics 33,* 3-13.

Al-Yateem, N. (2012). The effect of interview recording on quality of data obtained: A methodological reflection. Nurse Researcher *19*(4), 31-35.

Anderson, J., Grundy, J., De Frutos, J., Barker, R., Grady, C. & Bialystok, E. (2017). Effects of bilingualism on white matter integrity in older adults. *Neuroimage 167*, 143-150.

Anderson, J., Mak, L., Chahi, A. & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research and Therapy 50*, 250-263.

Andrews, E., Eierud, C., Banks, D., Harshbarger, T., Michael, A., Rammell, C.S. (2021). Effects of Lifelong Musicianship on White Matter Integrity and Cognitive Brain Reserve. *Brain Sciences 11*(1), 67; **https://doi.org/10.3390/brainsci11010067**.

Andrews, E., Frigau, L., Voyvodic-Casabo, C., Voyvodic, J. & Wright, J. (2013). Multilingualism and fMRI: Longitudinal study of second language acquisition. *Brain Sciences 3*(2), 849-876.

Andrews, E. (2014). *Neuroscience and Multilingualism*. Cambridge University Press.

Arnott, S., Brogden, L., Faez, F., Péguret, M., Piccardo, E, Rehner, K., Taylor, S. & Wernicke, M. (2017). The Common European Framework of Reference (CEFR) in Canada: A research agenda. *The Canadian Journal of Applied Linguistics 20*(1), 31-54.

Berntsen, D. & Rubin, D. (2004). Cultural life scripts structure recall from autobiographical memory. *Memory & Cognition 32*(3), 427-442.

Bialystok, E., Craik, F. & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia 45,* 459-464.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning 56*(1)*,* 9-49.

Bucholtz, M. & Hall, K. (2004). Language and identity. In Alessandro Duranti (ed.), *A companion to linguistic anthropology* (369-94). Hoboken: Wiley-Blackwell Publishing.

Bucholtz, M. & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies 7*(4-5), 585-614.

College Board AP. (n.d.) *AP Program*. College Board. https://ap.collegeboard.org/

Council of Europe. (2020). *The CEFR Levels*. Common European Framework of Reference for Languages (CEFR). https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

Davies, A. (2007). Applied linguistics and language learning/teaching. In A. Davies & K. Mitchell (eds.), *Introduction to Applied Linguistics: From practice to theory* (2nd ed., 86-89). Edinburgh University Press.

Delgado, P., Guerrero, G., Goggin, J. & Ellis, B. (1999). Self-assessment of linguistic skills by bilingual Hispanics. *Hispanic Journal of Behavioral Sciences 21*(1), 31-46.

Del Maschio, N., Fedeli, D., Sulpizio, S., & Abutalebi, J. (2019). The relationship between bilingual experience and gyrification in adulthood: A cross-sectional surface-based morphometry study. *Brain and Language 198*, 104680.

Donald, M. (2001). Chapter 3: The Governor of Mental Life. In *A Mind So Rare: The Evolution of Human Consciousness* (46-91). New York: W. W. Norton & Company.

Dooley, K. & Deutsch, D. (2010). Absolute pitch correlates with high performance on musical dictation. *The Journal of the Acoustical Society of America 128*(2), 890-893.

Egner, T. & Gruzelier, J. (2003). Ecological validity of neurofeedback. *NeuroReport 14*(9), 1221-1224.

Fasano, M., Glerean, E., Gold, B., Sheng, D., Sams, M., Sheng, D., Vuust, P., Rauschecker, J. & Brattico, E. (2020). Inter-subject similarity of brain activity in expert musicians after multimodal learning: A behavioral and neuroimaging study on learning to play a piano sonata. *Neuroscience 441*, 102-116.

Gallese, V. & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology 22*(3/4), 455-479.

Gill, P., Stewart, K., Treasure, E. & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups, *British Dental Journal 204*(6), 291-295.

Gold, B., Frank, M., Bogert, B. & Brattico, E. (2013). Pleasurable music affects reinforcement learning according to the listener. *Frontiers in Psychology 4*(541).

Golde, T., DeKosky, S. & Galasko, D. (2018). Alzheimer's disease: The right drug, the right time. *Science 362*(6420), 1250-1251.

Gollan, T., Weissberger, G., Runnqvist, E., Montoya, R. & Cera, C. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition 15,* 594–615.

Grundy, J., Anderson, J. & Bialystok, E. (2017). Neural correlates of cognitive processing in monolinguals and bilinguals. *Annals of the New York Academy of Sciences 1396*(1), 183-201.

Hymes, D. (1972). Editorial introduction to "Language in Society." *Language in Society 1*(1), 1-14.

Irvine, A., Drew, P. & Sainsbury, R. (2013). 'Am I not answering your questions properly?' Clarification, adequacy and responsiveness in semi-structured telephone and face-to-face interviews. *Qualitative Research 13*(1), 87–106.

Jäncke, L. (2009). The plastic human brain. *Restorative Neurology and Neuroscience 27*(5), 521-538.

Johns Hopkins Peabody Institute. (n.d.). *Music theory placement tests.* Peabody Institute. https://peabody.jhu.edu/academics/instruments-areas-of-study/music-theory/mtplacement/

Kaiser, J. (2018). The Alzheimer's gamble: Can the National Institute on Aging turn a funding windfall into a treatment for the dreaded brain disease? *Science 361*(6405), 839-841.

Glossos
язык limbă
γλῶσσα tongue
Issue 15, Summer 2020
http://seelrc.org/glossos/
glossos@seelrc.org

Kousaie, S., Laliberte, C., Zunini, R. & Taler, V. (2015). A behavioral and electrophysiological investigation of the effect of bilingualism on lexical ambiguity resolution in young adults. *Frontiers in Human Neuroscience 9*(682).

Kunnan, A. & Carr, N. (2017). A comparability study between the General English Proficiency Test- Advanced and the Internet-Based Test of English as a Foreign Language. *Lang Test Asia 7*(17).

Labov, W. (1966). The effect of social mobility on linguistic behavior. *Sociological Inquiry 36*(2), 186-203.

Oechslin, M. S., Descloux, C., Croquelois, A., Chanal, J., Van De Ville, D., Lazeyras, F. & James, C. (2013). Hippocampal volume predicts fluid intelligence in musically trained people. *Hippocampus 23*(7), 552-558.

Opdennaker, R. (2006). Advantages and disadvantages of four interview techniques in qualitative research. *Forum: Qualitative Social Research 7*(4), Art. 11.

Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism.* John Benjamins Publishing Company.

Pavlenko, A. (2005): Chapter 7: Social Cognition: "I no longer wanted to speak German." In *Emotion and Multilingualism* (192-226). Cambridge University Press.

Perani, D., Farsad, M., Ballarini, T., Lubian, F., Malpetti, M., Fracchetti, A., Magnani, G., March, A. & Abutalebi, J. (2017). The impact of bilingualism on brain reserve and metabolic connectivity in Alzheimer's dementia. *Proceedings of the National Academy of Sciences of the United States of America 114*(7), 1690-1695.

Pliatsikas, C., Moschopoulou, E. & Saddy, J. (2015). The effects of bilingualism on the white matter structure of the brain. *PNAS 112*(5), 1334-1337.

Schilling-Estes, N. (2008). Interview with Walt Wolfram. *Journal of English Linguistics 36*(4), 354-371.

Schroeder, S., Marian, V., Shook, A. & Bartolotti, J. (2015). Bilingualism and musicianship enhance cognitive control. *Neural plasticity 2016,* 1-11.

Sleve, L. & Miyake, A. (2006). Individual differences in second language proficiency: Does musical ability matter? *Psychological Science 17*(8), 675-681.

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition 33*(3), 339-372.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society 1*(2), 179-195.

University of Wisconsin System Boards of Regents (2011). *Spanish placement testing.* Center for placement testing. https://testing.wisc.edu/centerpages/spanishtest.html

Van de Putte, E., De Baene, W., García-Pentón, L., Woumans, E., Dijkgraaf, A. & Duyck, W. (2018). Anatomical and functional changes in the brain after simultaneous interpreting training: A longitudinal study. *ScienceDirect 99,* 243-257.

Van Hedger, S., Heald, S. & Nusbaum, H. (2019). Absolute pitch can be learned by some adults. *PLoS ONE 14*(9), 1-25.

Weiss, M. & Bidelman, G. (2015). Listening to the brainstem: Musicianship enhances intelligibility of subcortical representations for speech. *The Journal of Neuroscience 35*(4), 1687-1691.

Westera, N., Powell, M. & Milne, B. (2017). Lost in the detail: Prosecutors' perceptions of the utility of video recorded police interviews as rape complainant evidence. *Australian & New Zealand Journal of Criminology 50*(2), 252–268.

Wolfram, W. & Schilling-Ester, N. (1995). Moribund Dialects and the Endangerment Canon: The case of the Ocracoke Brogue. *Language 71*(4), 696-721.

Wolfram, W. (2018). The (in)significance of facts in sociolinguistic engagement. *Language in Society 47*(3), 374-377.

Appendix I:

Figure 1: Music, Language and Aging (MLA) Interview Questions

I. Multilingualism

1. Briefly describe your linguistic history (what is your first language [or do you have more than one first language], when and where did you use these languages, which languages do you speak on a regular basis, how did you learn the languages that you know, the intensity with your language and speech communities/communities of practice (daily, weekly, monthly, etc.)).

2. Did you study any languages in school (elementary, middle, high, college)? What was your language of education? How many years of study ((in)formal training, methods of instruction)? Can you speak, read and write in these languages?

3. Would you consider yourself "fluent" in these languages? Were you ever formally tested in any of these languages (classroom, AP, proficiency, job-related, etc.)?

4. Are there any bi- or multilingual members of your family (including parents, grandparents, siblings, spouse, children, grandchildren)? What languages are used (were used) with particular family members? Do you have any emotional ties to your language (memories, favorite texts, comfort levels, families and friends, cultural institutions, religion, code-switching, motivation, talent)?

5. If you are bi- or multilingual, which languages are primary in your central relationships at home and at work? Do you enjoy these languages in different ways?

6. What do you see as some of the benefits of using two or more languages?

II. Musicianship

1. Please describe your musical experience (including playing one or more instruments, singing, performance experience, listening to music, attending concerts, participating in musical group activities, musical training and education, musical career, aspirations, favorite performers/composers/pieces, etc.).

2. What are the most enjoyable aspects of your musical life?

3. What are some of your most vivid music-related memories (memories, favorite works, comfort levels, families and friends, cultural institutions, religion, motivation, talent)? Do you have memories that also include specific musical moments (performing, listening, lyrics, instruments, etc.)?

4. Can listening to a piece of music take you to a different time and place? This could include a wide spectrum of emotions (positive, negative) as well. Could you share some examples?

5. Are there any musicians in your family or amongst your friends? How has their musicianship impacted you?

III. Multilingualism and Musicianship

1. Do you feel a connection between music and languages, world language (importance of cultural context)?

2. Where are you from and what is your home environment like (regions, settings, exposure, accessibility)?

3. What is your in-country experience like (multiple countries of residence, immigration, discrimination, prestige)?

4. Any other comments?

5. Additional information:

Appendix II:

Qualitative Interview Data

- Interview data on geography/home environment for Participant 1: The participant's mother was a classically trained flautist and semi-classically trained pianist, so the participant had a lot of exposure to and encouragement of musicianship growing up. Her mother was the reason she became interested in music.

- Interview data on number of years for Participant 2: Participant grew up speaking Spanish in the home because her parents were from Argentina. She started speaking English when she started kindergarten in the United States. She moved overseas to Korea and began speaking Korean in elementary school and her first year of middle school. She also continued speaking English in school in Korea. She became fluent in Korean because it was the primary language of instruction in schools. Then she moved to Japan where she spoke Japanese and English in schools for the remainder of middle school and the beginning of high school. She lost her Korean within two years because she no longer spoke it in Japan, but she became fluent in Japanese because it was the primary language of instruction in school. She moved to the United States for the remainder of high school, during which she spoke English, and is now enrolled in college where she continues to speak English. She can still speak Japanese conversationally and read two of the three alphabets.

- Interview data on intensity for Participant 5: Participant speaks English daily in her professional and casual interactions. She speaks English with her family, at work, and with most of her relations. She described her English usage as 80% of her daily language usage. She also speaks Spanish daily, primarily at her job as a speech therapist with many Spanish-speaking clients and casually with her grandparents.

- Interview data on intensity for Participant 7: Participant took private violin lessons for two hours twice a week. She performed two violin recitals each year. She practices singing weekly and sings in her church once a month.

- Interview data on emotional ties to languages for Participant 12: The participant speaks primarily Chinese with her family. Both she and her brother are bilingual, but her other family members do not speak English very well, so they primarily use Chinese in the family. She thinks that speaking Chinese and English allows her to connect with all kinds of people, and she has met a lot of friends who also love languages and found a sense of community with other bilinguals. China is such a large country and so many people internationally speak Chinese, so she loves speaking Chinese because it gives her the ability to communicate with so many people and associated advantages. She would love to learn other languages, but right now she is primarily focused on improving her Chinese. She would also like to learn Spanish because she believes that it is the most

practical to know in the United States. At university, (she has since graduated), she was a member of the selective-living group known as LangDorm, a community dedicated to the celebration of different languages, so she found community through knowing other languages. She also believes that knowing Chinese forces her to think differently than if she only knew English, giving the example of how numbers in Chinese are grouped by fours, rather than threes. She spent three months in Taiwan after graduating from college because she did not have time to study abroad during college. She said that in-country experience was the best way for her to really improve her Chinese since the classical Chinese classes she had taken in college were not advanced enough for her as a native speaker.

- Interview data on emotional ties to music: The participant used to get stage fright. She loved the musical group she played with at her university because of its great community. She has a specific memory of a concert where her music blew off her stand and the whole concert had to be stopped so she could retrieve it; this was an incredibly nerve-racking and scarring experience for her. Currently, she listens to music that is mostly in English, but she has gone through phases where she has listened to primarily Chinese music or K-pop. Before she had Spotify, she listened to a lot of Chinese music because it was on YouTube. However, now that she has Spotify, there is less availability of Chinese music and K-pop, so she listens to more American music.

- Interview data on professional aspirations/career paths for Participant 8: Participant is a current music major and is required to take many music classes at the university level. The participant also views knowing Spanish as extremely beneficial to their future career in medicine.