Title	A neurocognitive comparison of listening to academic lectures and natural sounds: Implications for test validity
Author(s)	Janice Lee Jia Yi, Vahid Aryadoust, Li Ying Ng and Stacy Foo
Source	*International Journal of Listening,* (2020)
Published by	Taylor & Francis (Routledge)

A Neurocognitive Study of Listening Comprehension under Test and Non-Test Conditions: Implications for Validity

Abstract

With the advent of new technologies, assessment research has adopted technology-based methods to investigate test validity. This study investigated the neurocognitive processes involved in an academic listening comprehension test, using a biometric technique called functional near-infrared spectroscopy (fNIRS). Sixteen right-handed university students completed two tasks: (1) a linguistic task where they listened to a mini-lecture (i.e., Listening condition) and answered questions (i.e., Questions condition) and (2) a non-linguistic task where they listened to a variety of natural sounds and animal vocalizations (i.e., Sounds condition). Hemodynamic activity in three left brain regions was measured: the inferior frontal gyrus (IFG), dorsomedial prefrontal cortex (dmPFC), and posterior middle temporal gyrus (pMTG). The Listening condition induced higher activity in the IFG and pMTG than the Sounds condition. Although not statistically significant, activity in the dmPFC was higher during the Listening condition than the Sounds conditions. The IFG was also significantly more active during the Listening condition compared with the Questions condition. Although significant gender differences were observed in listening comprehension test scores, there were no differences in brain activity (IFG, dmPFC, and pMTG) between males and female participants. The implications for test validity are discussed.

Keywords: listening comprehension assessment; functional near-infrared spectroscopy; gender.

A Neurocognitive Study of Listening Comprehension under Test and Non-Test Conditions: Implications for Validity

Listening tests are integral for assessing academic language competency (Rost, 2016; Taghizadeh & Saadatju, 2020). Research on listening assessment is dominated by subskill taxonomies and frameworks that provide accounts of listening abilities purportedly engaged by listening tests (Richards, 1983; Buck & Tatsuoka, 1998; Eom, 2008; Liao, 2007; Shin, 2005). Interest in these theories reflects a growing concern with the dominant approach to listening research: formulating listening as a list of static subskills (Richards, 1983), which are at best "conjectural" and anecdotal (Aryadoust, 2020). This issue has been examined via quantitative methods such as item response theory (IRT) (Buck & Tatsuoka, 1998; Sawaki et al., 2009) or introspection with non-reproducible results (Buck, 1994; Goh, 2008). Although these approaches have addressed some gaps, they were only modestly effective at highlighting the nature of listening comprehension (Buck, 2001). This is because listening comprehension is multifarious and consists of cognitive, neurological, and behavioral layers (Bodie & Worthington, 2017; Worthington & Bodie, 2020). Nevertheless, most previous publications on listening assessment are inclined towards addressing the behavioral level using listeners' test scores. Despite its popularity, this approach fails to reveal the underlying mechanisms that result in listeners' test-taking behavior. The shortcoming of the behavioral approach has been long recognized in the available literature. For example, Buck (2001) lamented that "determining the competencies that underlie performance on a set of test tasks is a complex and indirect process, and we have no way of knowing for certain which competencies are required by any particular task" (p. 106). While behavioral research has failed to discover these listening competencies, neurocognitive methods seem to hold promise in highlighting listening mechanisms under assessment, as will be discussed below (Aryadoust, Ng, Esposito, & Foo, 2020).

Another challenge confronting listening assessment research is the interaction between listening performance and external factors such as gender, which has been a concern in validation research since the 1960s (Brimer, 1969; Boyle, 1987; Harding, 2011; Payne & Lynn, 2010; Sawyer et al., 2014). Sawyer et al. (2014) showed that listening tasks functioned differently across male and female native and non-native English speakers. Additionally, Payne and Lynn (2010) found that females outperformed males in second language comprehension tasks and suggested that females had better processing capacity in their second language. Research has yet to identify possible reasons for gender-mediated test performances.

Language assessment research has revealed that listening activities often engage an array of cognitive processes (Wang & Treffers-Dallerb, 2019) — most importantly the ability to understand literal and inferential messages (Buck, 2001; Leonard, 2019). In the past few decades, cognitive neuroscientists have investigated the neural mechanisms that subserve these processes using advanced techniques for measuring brain activity such as functional magnetic resonance imaging (fMRI) and fNIRS. Although they offer a vigorous approach to study listening mechanisms under test-taking conditions, to the best of our knowledge, the application of cognitive neuroscience methods in listening assessment research is scarce.

In light of this gap in understanding, the primary objective of this study was to investigate the neurocognitive mechanisms involved in listening comprehension assessment. Secondly, we aim to explore how neurocognitive research can inform our understanding of the cognitive processes that underpin listening comprehension under assessment conditions versus listening to natural sounds. According to Kintsch (1998), the main cause of oral language performance differentiation lies in the deepest biological or neurological layer of comprehension (Romeo et al., 2018). Significant associations have been found between listening processes (both literal and inferential) and certain areas in the left hemisphere, notably the dorsomedial prefrontal cortex (dmPFC), inferior frontal gyrus (IFG), and posterior middle temporal gyrus (pMTG) (Friederici, 2011; Hoffman et al., 2012; Rogalsky et al., 2008). These associations are mediated by the gender of the listeners, suggesting that the brains of different gender groups are wired differently, perhaps due to cultural norms and expectations (Johnston et al., 2000; Witelson et al., 1995).

The above findings were made under non-test conditions, and there is currently little research that examines the neurocognitive mechanisms of listening during test-taking (Aryadoust et al., 2020). We address this gap with this study and shine some light on the listening processes of a group of adult listening test-takers using fNIRS. In the following sections, we present a survey of extant neurocognitive research on listening and review the effect of gender on the neural mechanisms of listening. We then describe the aims and results of this study, concluding with the implications for listening assessment.

The Brain in Listening

Listening involves both bottom-up (literal) and top-down (inferential) processes (Berl et al., 2010; Bookheimer, 2002; Buchweitz et al., 2009; Furuya, 2019; Mason & Just, 2006). Bottom-up listening refers to interpreting language by segmenting the sound stream into separate meaningful units and building meaning from phonemes and words to full utterances or larger conversations (Buck, 2001). In support of bottom-up processing, empirical studies show that listeners' vocabulary knowledge has a crucial role in their comprehension (Bian et al., 2019; Furuya, 2019; Ghorbani Nejad, & Farvardin, 2019). Conversely, top-down processing involves applying context and prior knowledge by tapping into schemas of experiential, pragmatic, cultural, and discourse knowledge to meaningfully interpret spoken content (Furuya, 2019; Kintsch, 1991, 1998; Vandergrift & Goh, 2014). Neurological evidence supports the auditory processing streams associated with these listening processes. Berl et al. (2010) presented evidence for a ‘comprehension cortex’ that comprises the frontal cortex and certain areas of the left cortex. Friederici (2011) suggested that projections from the sensory cortex to the motor cortex (i.e., dorsal pathway I) likely support the bottom-up process, whereas links from Broca’s area (i.e., within the IFG) to the temporal cortex (i.e., dorsal pathway II) likely support the top-down process. In contrast, the dorsal pathway II purportedly facilitates information integration by making predictions about incoming information (Friederici, 2011). The roles of these pathways in listening were supported by Karunanayaka et al.’s (2007) large-scale study that highlighted engagement of the prefrontal cortex (PFC), frontal gyrus, IFG, and subcortical areas such as the hippocampus during story comprehension. Similarly, Abutalebi and Green (2007) showed that the PFC is implicated in top-down processing of language, whereas the IFG and middle frontal gyrus are associated with bottom-up processing.

In line with Berl et al.’s (2010) ‘comprehension cortex’, the left hemisphere is critical for listening comprehension, particularly (1) the dorsomedial prefrontal cortex (dmPFC), (2) IFG, and (3) pMTG. The PFC is generally implicated in higher-level thinking, creativity, ideological thoughts, and beliefs (Dietrich & Kanso, 2010; Ellamil et al., 2012). In contrast, the dmPFC is associated with top-down processing and social interactions, such as (1) inferring people’s mental states and forming impressions (Wagner et al., 2016), (2) producing metaphorical, non-literal language, and (3) formulating creative figures of speech using semantic memory (Benedek et al., 2013; Friederici & Gierhan, 2013). A meta-analysis of 120 neuroimaging studies by Binder et al. (2009) further supported the integral role of the dmPFC in inferential comprehension. According to Kröger et al. (2012) and van Kesteren et al. (2010), the dmPFC is associated with retrieving prior knowledge from long-term memory, which is essential for making inferences during comprehension.

Hallam et al. (2018) suggested that the left IFG and pMTG co-facilitate the bottom-up and semantic analysis of auditory verbal stimuli. The left IFG houses Broca's area, which is involved in amodal sentence comprehension through bottom-up processing and verbal working memory (Buchweitz et al., 2009; Rogalsky et al., 2008). In another study, Rüschemeyer et al. (2005) showed that the left IFG is associated with the second language processing, while Tatsuno and Sakai (2005) showed that less activity in this region correlates with higher language proficiency. Thus, high activity in the IFG may be indicative of lower listening proficiency, as listening is not fully automated in individuals with lower proficiency. This prediction was further supported by Jeong et al.'s (2007) study that used auditory stimuli during a sentence comprehension task. They found that the left IFG was more active while listening and processing a language that participants were less proficient in, concluding that the increased activation was due to increased processing loads (Jeong et al., 2007). A meta-analysis by Indefrey (2006) found that, in general, increased activation of the left IFG is observed during word-production tasks involving a language that participants are less fluent in. A review by Perani and Abutalebi (2005) also reached the same conclusion. These studies all point to increased brain activity — at least in the IFG — for participants with lower proficiency in that language, possibly due to increased cognitive loads during language processing.

Finally, the dmPFC is involved in top-down processing during listening (Ferstl et al., 2008; Hoffman et al., 2012; Perfetti & Frishkoff, 2008). These patterns were substantiated by an fMRI study that involved listening to sentences with varying noise levels (Zekveld et al., 2006).

Zekveld et al. (2006) found that the reduced need for top-down processing during intelligible, low noise speech resulted in lower peak amplitudes in frontal brain regions, while the increased need for bottom-up processing during less intelligible, high noise speech resulted in higher amplitudes in temporal brain regions. Furthermore, the two types of processing correlated with language activation and logical processes involving language comprehension and listening skills (Zekveld et al., 2006). Separate studies by Davey et al. (2015) and Whitney et al. (2010) also found significant relationships between temporary impairment of the pMTG and semantic cognition. Taken together, these pieces of evidence suggest that the pMTG plays key roles in bottom-up or literal comprehension, which was further supported by Noonan et al.'s (2013) meta-analysis of 53 brain imaging studies.

In summary, the causal relationships between activity of the dmPFC and inference-making ability (top-down) in listening, alongside the roles of the left IFG and pMTG in semantic processing (bottom-up), indicate that these three brain regions are three main neural substrates involved in listening comprehension tests.

Gender Differences in Listening

The literature on gender differences in listening is vast and varied; however, research focusing on the neurocognitive mechanisms of these differences is scarce. The majority of research in this domain has focused on language lateralization in male and female brains. Early research suggested that females show bilateral activation during listening, while males show lateralized activation in the left brain (Johnston et al., 2000). Generally, males have a more diffuse language network in the left hemisphere.

Males and females also show different brain organization regarding language. Schlaepfer et al. (1995) found that females had 23.2% and 12.8% greater grey matter mass than males in the dorsolateral PFC and superior temporal cortex, respectively, which are both language-related areas. Relatedly, Phillips et al. (2000) reported greater bilateral activation in the temporal lobes in females than males during passive listening of simple narrations. In Baxter et al.'s (2003) semantic processing study involving word categorizations, despite the absence of gender differences in task accuracy, males showed significant left hemispheric activation — especially the IFG, superior temporal gyrus (STG), and cingulate regions — whereas females showed significant bilateral activation in the STG. This result suggests that, despite similar performances in task accuracy, the neurocognitive pathways involved in comprehension may differ for each gender. Habl (2018) attempted to examine the outcomes of structural brain differences, hypothesizing that males were better at learning grammatical rules while females were better at vocabulary learning. However, no statistically significant differences between genders were found in the test results for grammar or vocabulary. This finding suggests that any brain structural differences influencing comprehension are more complex than direct links to performance on certain tasks.

The Present Study

One principle of neuroimaging research is to contrast brain activation during different tasks, e.g., listening to speech vs. non-speech. In the present study, non-speech refers to natural sounds. We adopted the methodologies of Belin et al. (2002), Belin et al. (2000), and Lewis et al. (2004) who contrasted listening to language and non-language stimuli. Comparing listening to speech with listening to natural environmental sounds and animal vocalizations can offer insights as to which brain regions are activated specifically in response to human speech. Belin et al. (2002) found that listening to speech increased left hemispheric activation which is sensitive to voice as well as bilateral activation of the primary auditory cortex (i.e., temporal lobe) when compared with listening to non-speech vocalizations. In earlier work, Belin et al. (2000) observed bilateral activation of voice-selective brain regions along the upper banks of the superior temporal sulci (STS) when subjects listened passively to language compared with listening to animal and environmental sounds. Relative to processing simple tone patterns, Lewis et al. (2004) showed that semantic processing of spoken words increased activity in the middle temporal gyrus (MTG), angular gyrus, and inferior frontal cortex (IFC), which is located near the dmPFC.

To extend the aforementioned work (Belin et al., 2000; Belin et al., 2002; Lewis et al., 2004), the current study investigated the brain activity of test-takers while they were listening to longer discourse rather than words. Therefore, this experiment is more similar to actual listening under assessment conditions. Using the two aforementioned contrasts enables us to examine the validity of listening comprehension assessments by referencing the neurocognitive processes that are elicited during listening tests but not when listening to natural sounds. This is similar to discriminant or divergent validity where conceptually unrelated and dissimilar stimuli, such as a listening test and a driving test, are not significantly correlated (Campbell & Fiske, 1959). Strong discriminant validity is indicative that psychological traits are not of the same entity and thus sufficiently unique (Voorhees et al., 2015).

The research questions of the present study are as follows:

1. Are there differences in activity across the IFG, dmPFC, and pMTG when listening to language (Listening condition), answering questions (Questions condition), and listening to natural sounds (Sounds condition)?
2. Are there differences in activity across the IFG, dmPFC, and pMTG between males and females under test conditions, and are the differences in brain activity congruent with their test scores?

As discussed, gender can result in systematic differences in test scores that are irrelevant to the construct being assessed (John & Benet-Martínez, 2000). Thus, test scores across gender were compared to identify any relationships between the neurocognitive processes involved and listeners' test scores. Test scores were intended to represent listeners' comprehension ability and, accordingly, brain activity should correlate with test scores. Similarly, if differences in brain activity are observed across gender groups, then differences should also be observed in their test scores.

Methods

Participants

Eighteen students (Age: M = 22.3, SD = 1.8 years) from a major university in Singapore were recruited for this study. All participants had a minimum of 12 years of education and reported no history of neurological or psychological disorders. They were compensated $SGD10 for their time.

Listening task

The linguistic task was a listening comprehension test where participants had to listen to a mini-lecture and then answer comprehension tests presented on a laptop (HP Pavilion, Hewlett-Packard, CA, USA) using stimulus presentation software (SuperLab 5, Cedrus Corporation, CA, USA). The lecture concerned a topic in ecology and was delivered by an Australian male speaker. The psychometric quality of this listening test was previously examined using Rasch measurement (Aryadoust, 2012). The test items had a unidimensional structure, indicating that the test scores obtained were reflective of test-takers' general listening ability.

In this study, the lecture recording was presented in 10 segments. Each segment was 20s long with a 20s pause after each segment. While listening to the lecture, participants took notes on a piece of paper (i.e., Listening condition). During the 20s pauses, participants focused their gaze on a fixation cross presented on the laptop. These pauses would allow brain activity levels to return to baseline (Nishiyori, 2016). After the lecture was delivered, participants completed 10 fill-in-the-blank questions that were also presented on the laptop. Similar to the lecture presentation format, each question was presented individually for 20s. There was a 20s pause between questions during which participants focused their gaze on a fixation cross.

Non-linguistic task (Sounds condition)

The non-linguistic task (i.e., Sounds condition) required participants to listen to 10 sets of natural sounds (i.e., fire crackling, thunder and rain, jungle, forest evening, and lake) and animal vocalizations (i.e., chicken, falcon, sheep, goose, and loon). Using SuperLab 5 software (Cedrus Corporation, CA, USA), each set of sounds was presented for 20s, with a 20s pause between sets. During the pauses, participants focused their gaze on the fixation cross presented on the laptop.

Data collection

Participants completed the three tasks in a single one-hour testing session in a quiet room within the university. After informed consent was obtained, participants completed the Edinburgh inventory (Oldfield, 1971) to establish whether they were right-handed or left-handed. The Edinburgh inventory comprises 15 questions about handedness preferences for a variety of tasks such as writing and throwing (Oldfield, 1971). At the end of the questionnaire, a laterality index was automatically generated: Right-handedness was denoted by a positive index, while left-handedness was denoted by a negative index. Participants then completed the abovementioned linguistic and non-linguistic tasks. The order of these two tasks was counterbalanced.

To measure brain activity during the tasks, a portable fNIRS system (NIRSport device, NIRx Medical Technologies, LLC, MN, USA) with 16 optodes (i.e., 8 pairs of sources and detectors) was used. Each source contained one light-emitting diode that emits infrared light at 760nm and 850nm. The 16 optodes were assembled on the left cerebral hemisphere based on a modified topographical montage to measure brain activity in the IFG, dmPFC, and pMTG. Before the start of the linguistic or non-linguistic task, participants were fitted with an fNIRS cap connected to the NIRSport device (NIRx Medical Technologies LLC). Subsequently, an automatic calibration process was performed using the NIRStar 15-0 recording software package (NIRx Medical Technologies LLC) to determine the optimum amplification factors for each of the 20 channels formed by the 8 source-detector pairs. As SuperLab 5 software (Cedrus Corporation) and NIRStar 15-0 software (NIRx Medical Technologies LLC) are not gen-locked, an additional c-pod (Cedrus Corporation) was used to send event markers via USB through connections with the NIRSport device (NIRx Medical Technologies LLC)). Each event marker was pre-set within SuperLab 5 software (Cedrus Corporation, CA, USA) to mark the start of each segment for both the linguistic and non-linguistic tasks.

Data Processing

Dichotomous scoring was used to mark the listening performance of the participants, with scores of one and zero awarded for correct and incorrect responses (including blanks), respectively. For brain activity measurements, the raw data collected with NIRStar 15-0 software (NIRx Medical Technologies LLC) were processed using nirsLAB v201706 software (NIRx Medical Technologies LLC) before data analysis. Pre-processing involved several steps to reduce noise (artifacts or spikes and consecutive timepoints where the data were saturated). The data were interpolated using an automated process and then visually inspected to further reduce any remaining noise. Any artifacts or spikes identified by visual inspection were replaced using random signals obtained from the data. Subsequently, a bandpass filter with low- and high-frequency cutoffs at 0.01 Hz and 0.2 Hz, respectively, was used. The processed data were subjected to the Beer-Lambert law to compute the hemodynamic states.

An automated statistical parametric mapping procedure was performed within nirsLAB (NIRx Medical Technologies LLC) to transform wavelength data into numerical representations (beta value) of the hemoglobin levels for each channel during each condition. These beta-values were averaged based on their respective brain areas, resulting in an averaged beta value for the IFG, dmPFC, and pMTG for each condition. These averaged beta values were subjected to statistical analysis.

Data were missing for two participants. Thus, only data from 16 right-handed participants were analyzed (age: 22.4 ± 1.9 years; Edinburgh inventory laterality index: 78.8 ± 13.0; males: N = 7, females: N = 9).

Statistical Analysis

The Shapiro-Wilk test of normality was first performed to assess if the obtained β values and test scores violated the assumption of normality. To answer the first research question, non-parametric Friedman tests were used to analyze the data across the conditions (i.e., Listening, Questions, and Sounds) as the β values were not normally distributed (Table 1). Further between-condition analyses were conducted using Wilcoxon Signed Rank tests when the Friedman tests yielded statistically significant results. Bonferroni corrections were used to control for multiple pairwise comparison Type 1 error rates. For the second research question, due to the small sample sizes, Mann-Whitney *U* tests were used to analyze the test scores and β values at each of the three brain regions between males and females. Effect sizes were calculated for all comparisons across both research questions following the recommendations by Field (2013) where $r = \frac{Z}{\sqrt{N}}$. An absolute *r* value of 0.1 to 0.3 indicates a small effect size, 0.3 to 0.5 indicates a medium effect, and 0.5 or higher indicates a large effect (Field, 2013).

Table 1

Shapiro-Wilk Test of Normality for Brain Activity

Brain regions	Conditions	Statistic	*df*	*p* values
IFG	Listening	0.87	16	0.025*
	Questions	0.96	16	0.73
	Sounds	0.96	16	0.65
dmPFC	Listening	0.67	16	< 0.0001****
	Questions	0.78	16	0.002**
	Sounds	0.94	16	0.35
pMTG	Listening	0.97	16	0.76
	Questions	0.92	16	0.16
	Sounds	0.93	16	0.23

Abbreviations: dmPFC – dorsomedial prefrontal cortex, IFG – inferior frontal gyrus, and pMTG – posterior middle temporal gyrus.

* $p < 0.05$

** $p < 0.01$

**** $p < 0.0001$

Results

Research Question 1: Brain Activity across Conditions

The Friedman tests revealed statistically significant differences across all conditions (i.e., Listening, Questions, and Sounds) for the IFG ($\chi^2(2, N=16) = 9.88$, $p = 0.007$), dmPFC ($\chi^2(2, N=16) = 7.13$, $p = 0.028$), and pMTG ($\chi^2 (2, N=16) = 8.38$, $p = 0.015$). Further analyses using Wilcoxon Signed Rank tests were performed to investigate specific differences between conditions per brain region (Table 2). Visualizations of these differences are presented in the form of topographical maps in Figure 1. Notably, the IFG (Listening: $\beta = 11.88 \pm 11.07\ e^{-5}$; Questions: $\beta = 4.17 \pm 10.83\ e^{-5}$, $Z = -2.48$, $p = 0.013$, $r = -0.44$) and pMTG (Listening: $\beta = 6.02 \pm 9.50\ e^{-5}$; Questions: $\beta = -0.39 \pm 9.24$, $Z = -2.69$, $p = 0.007$, $r = -0.48$) were more active during the Listening condition than in the Sounds condition. Although the activity level of the dmPFC was higher during Listening ($\beta = 5.53 \pm 18.03\ e^{-5}$) versus Questions ($\beta = 0.95 \pm 8.37$), the difference was not statistically significant after Bonferroni correction, $Z = -2.12$, $p = 0.034$, $r = -0.38$. Among the three brain regions, only the IFG was more active during the Listening condition ($\beta = 11.88 \pm 11.07\ e^{-5}$) than the Questions condition ($\beta = 4.17 \pm 10.83\ e^{-5}$), $Z = 2.48$, $p = 0.015$, $r = -0.44$ (Table 2). Unlike the dmPFC and pMTG, the IFG was similarly active during the Questions ($\beta = 4.17 \pm 10.83\ e^{-5}$) and Sounds conditions ($\beta = 1.12 \pm 8.38\ e^{-5}$), $Z = -0.88$, $p = 0.38$ (Table 2).

Table 2

Wilcoxon Signed Rank Test Results

Brain regions	Condition 1	β Values (10^{-5})	Condition 2	β Values (10^{-5})	Z	p	*r*
IFG	Listening	11.88 ± 11.07	Questions	4.17 ± 10.83	-2.48	0.013**	-0.44^{+}
	Listening	11.88 ± 11.07	Sounds	1.12 ± 8.38	-3.15	0.002**	-0.56^{++}
	Questions	4.17 ± 10.83	Sounds	1.12 ± 8.38	-0.88	0.38	-0.16
dmPFC	Listening	5.53 ± 18.03	Questions	4.17 ± 10.75	-1.03	0.30	-0.18
	Listening	5.53 ± 8.03	Sounds	0.95 ± 8.37	-2.12	0.034^	-0.38^{+}
	Questions	4.17 ± 10.83	Sounds	0.95 ± 8.37	-1.45	0.15	-0.26
pMTG	Listening	6.02 ± 9.50	Questions	3.84 ± 7.51	-0.83	0.41	-0.15
	Listening	6.02 ± 9.50	Sounds	-0.39 ± 9.24	-2.69	0.007**	-0.48^{+}
	Questions	3.84 ± 7.51	Sounds	-0.39 ± 9.24	-2.07	0.039^	-0.38^{+}

All β Values are presented as mean ± standard deviation. Abbreviations: dmPFC – dorsomedial prefrontal cortex, IFG – inferior frontal gyrus, and pMTG – posterior middle temporal gyrus.

** significant after Bonferroni correction ($\alpha = 0.0167$)

Brain regions	Condition 1	β Values (10^{-5})	Condition 2	β Values (10^{-5})	Z	p	*r*

^ significant before Bonferroni correction

$^{+}$ medium effect size

$^{++}$ large effect size

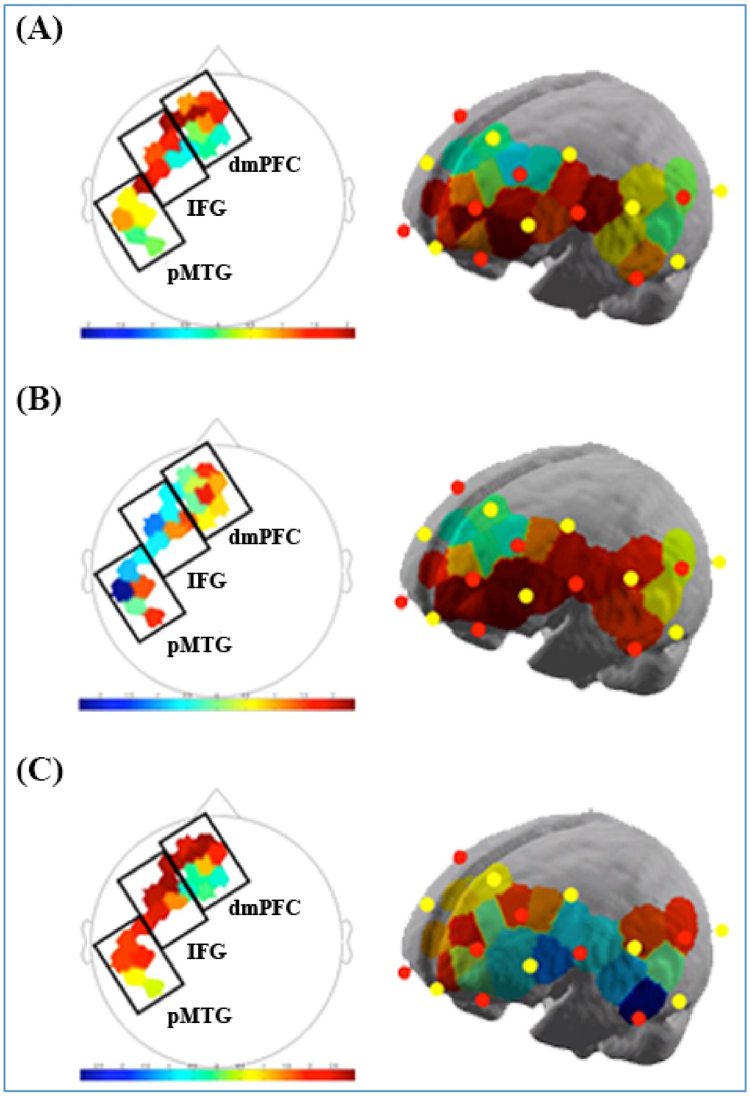

Figure 1. Two-dimensional topographical maps of oxyhemoglobin (OxyHb) levels in the left-brain hemisphere for the following conditions: (A) Listening vs. Questions, (B) Listening vs. Sounds, and (C) Listening across genders. Areas with colors towards the right end of the spectrum indicate significantly higher levels of OxyHb in the first condition than the second condition. In contrast, areas with colors towards the left end of the spectrum signify significantly higher OxyHb levels in the second condition than the first condition. Abbreviations: dmPFC –

dorsomedial prefrontal cortex, IFG – inferior frontal gyrus, and pMTG – posterior middle temporal gyrus.

Research Question 2: Gender Differences across Brain Regions

Mann-Whitney *U* tests were carried out to analyze gender differences in brain activity across each condition (i.e., Listening, Questions, and Sounds). None of the comparisons returned statistically significant gender differences across the IFG, dmPFC, or pMTG ($p > 0.05$) (Table 3).

Table 3

Mann-Whitney U Test Results for Gender Comparisons

Brain regions	Conditions	β Values for Females (10^{-5})	β Values for Males (10^{-5})	*Z*	*P*	*R*
IFG	Listening	10.03 ± 10.16	14.26 ± 12.53	-0.58	0.61	-0.10
	Questions	0.83 ± 8.41	8.47 ± 12.67	-1.64	0.12	-0.29
	Sounds	2.89 ± 6.06	-1.14 ± 10.77	-0.90	0.41	-0.16
dmPFC	Listening	1.12 ± 22.99	11.20 ± 6.39	-1.01	0.35	-0.18
	Questions	1.53 ± 13.57	7.56 ± 4.43	-1.01	0.35	-0.18
	Sounds	0.81 ± 9.61	1.14 ± 7.20	-0.05	1.00	-0.01
pMTG	Listening	7.22 ± 10.63	4.47 ± 8.36	-0.48	0.68	-0.08
	Questions	1.60 ± 5.08	6.73 ± 9.45	-1.54	0.14	-0.27
	Sounds	-0.18 ± 7.57	-0.66 ±11.70	-0.16	0.92	-0.03

Brain regions	Conditions	β Values for Females (10^{-5})	β Values for Males (10^{-5})	*Z*	*P*	*R*

Abbreviations: dmPFC – dorsomedial prefrontal cortex, IFG – inferior frontal gyrus, and pMTG – posterior middle temporal gyrus.

Test Scores

The average test score was 8.0 ± 1.5. Five participants scored below average (Score = 4, $N = 1$; Score = 6, $N = 1$; and Score = 7, $N = 3$), four participants had an average score of 8, and seven participants scored above average (Score = 9, $N = 5$; Score = 10, $N = 2$). The Mann-Whitney *U* test indicated that male participants (M = 8.9, SD = 1.1) outperformed female participants in the listening comprehension test (M = 7.3, SD = 1.6, $U = 12.50$, $p = 0.042$).

Discussion

This study investigated activity in three left hemisphere brain pathways associated with listening: the IFG, dmPFC, and pMTG. The study also aimed to identify differences in brain activity between males and females in the abovementioned brain areas during listening comprehension. The results for each research question are discussed accordingly.

Research Question 1: Brain Activity across Conditions

Firstly, the IFG and pMTG were significantly more active during the Listening condition than the Sounds condition. Although the Listening condition induced higher activity in the dmPFC than the Sounds condition (i.e., of medium effect), this difference did not reach statistical significance following the Bonferroni correction. Nevertheless, these findings are suggestive that the areas involved in speech recognition and language comprehension are more active when participants were listening to linguistic speech compared to non-linguistic sounds, including natural sounds and animal vocalizations. This finding aligns with Lewis et al. (2004) who found that the MTG and angular gyrus and IFC (i.e., within the proximity to the dmPFC) were more engaged during semantic processing of spoken words compared with processing of simple tone patterns. Additionally, our work supports the findings of Belin et al. (2000) who showed higher activity in the STG – an area near the IFG and pMTG – when processing speech sounds compared with non-speech vocal sounds such as moans and cries. The variation in brain areas identified may be due to equipment and task differences. In both of the abovementioned studies, the authors used fMRI to measure participants' brain activity while they listened to auditory stimuli with their eyes closed (Lewis et al., 2004; Belin et al., 2004). In the present study, fNIRS was used to measure participants' brain activity while they listened to auditory stimuli with their eyes fixated on a cross presented on the monitor. The spatial resolution of the fMRI equipment used by Lewis et al. (2004) and Belin et al. (2005) is higher than that of the fNIRS system in our study, and fMRI may thus provide more detailed imaging of the brain areas involved during listening. However, fNIRS helps maintain the ecological authenticity of the listening tasks in our study, which is an advantage over fMRI where participants must lie down and perform restricted tasks inside a fairly narrow chamber.

Our study also revealed that the IFG was significantly more active during the Listening than Questions conditions. As suggested by Karunanayaka et al. (2007), the IFG is involved in story comprehension. The IFG is also thought to be involved in making predictions about incoming information (Friederici, 2011). In our study, listening to the lecture involved story-based comprehension as well as prediction of what the speaker would say next. In contrast, answering questions does not seem to require these skills. Therefore, this finding provides support that the IFG is involved in processing language rather than in recalling information.

Research Question 2: Gender Differences across Brain Regions

The second research question addressed differences in brain activity between males and females while listening to speech under test conditions and listening to non-linguistic sounds. We found that activity in the IFG, dmPFC, and pMTG did not differ significantly between male and female participants, even though the former outperformed the latter in the listening comprehension test scores. We postulate that rather than dissimilarities in cortical brain activity, differences in test scores between male and female participants could be due to artifacts such as test methods and differences in linguistic backgrounds. Interestingly, our results contradicted previous research collectively suggesting that differences in listening performance between genders may be associated with differences in language processing at the neural level (Rossiter, 1972; Johnston et al., 2000). As we only measured left hemisphere brain activity, it was not possible to explore differences in brain lateralization between genders during the linguistic task (Listening and Questions condition).

More importantly, our findings exemplified the differences between listening as represented by absolute test scores and the neurocognitive processes that underlie listening. This difference warrants further investigation in future research. In particular, IRT methods such as differential item functioning (DIF) could be coupled with fNIRS to determine if the results of DIF analysis resonate with those of fNIRS. This method may allow investigation of test fairness at the neurocognitive level. Differences between test scores and neuroimaging methods when revealing differences between test-takers provide evidence that neuroimaging methods like fNIRS are suitable for validating language assessments.

Implications and Future Research Directions

Overall, our findings are important for the language assessment, as research often explores listening/language ability variation by resorting to quantitative modeling of variance in test scores through latent variable models like IRT, factor analysis, and the Rasch model. The observed difference between test scores and neurocognitive activity underscores the concern voiced by Mislevy (2009) about overreliance on test scores. Mislevy who pointed out that in test validation: "IRT characterizations of students and items […] are clearly simplifications, and they say nothing about the processes by which students answer items." Mislevy (2009, p. 3) raised two further questions that may be addressed only through neurocognitive research methods: "What is the nature of person parameters such as θ and η in latent variable models? Where do they reside?" Our study provided preliminary evidence that listening ability represented by test scores (analogous to θ in IRT) may be different from test-takers' actual neurocognitive processes during listening.

In addition, our study highlighted the brain regions in the left hemisphere associated with top-down and bottom-up processing during listening. Notably, our finding may be said to be inconsistent with the assumption of unidimensionality in latent variable models. This inconsistency is reminiscent of early debates on the suitability of latent variable models for language assessment where opponents asserted that the models were too stringent and did not consider all dimensions of language ability (Buck, 1994; Hamp-Lyonz, 1989), while proponents argued that psychometric and psychological dimensions were distinct and the former were constructed by analysts (McNamara, 1991, 1996). McNamara's (1991, 1996) argument suggested that psychometric dimensions did not have an ontological reality, even though the cause of variance in test data is assumed to be respondents' ability. We suggest that neurocognitive processes during test-taking could be the bridge between using purely psychometric or purely psychological conceptualizations of language proficiency (Kintsch, 1998). That is, listening construed as a psychometrically unidimensional construct (epistemology) seems different from the neurocognitive definition of listening (ontology) mediated in the aforementioned brain regions (Borsboom, 2005). Further research using neurocognitive methods is recommended to extend on the results of the presence study.

The incongruity between the neurocognitive processes of listeners and their test scores also warrants further research of more rigorous test development methods that are not only capable of differentiating high- and low-ability performers (i.e., via test scores), but also of establishing the neurocognitive processes of pre-specified brain networks. We propose that the "neurocognitive validity" of language assessments may be investigated at four levels:

> *Level 1:* Establishing the discriminant validity of the test by using relevant linguistic or non-linguistic auditory stimuli, such as natural sounds and vocalizations.
>
> *Level 2:* Establishing congruence between test-takers' scores and activation of specific brain pathways.
>
> *Level 3:* Collection of evidence to show that the potential causes of construct-irrelevant variance, such as gender, do not lead to differential brain activation.
>
> *Level 4:* The correlation between test-takers' neurocognitive processes under test and non-test conditions, which can be used to establish test authenticity.

This framework should be investigated and extended to shed light on the potential of neuroimaging techniques in language assessment research and validation.

Limitations

This study is limited by the small sample size and absence of information concerning the linguistic background of the gender groups (nine females and seven males). This small imbalance in sample sizes across gender groups could potentially influence the results, especially since the two lowest-scoring participants were females. Lastly, the lecture content in our listening comprehension assessment was in the field of ecology. Although none of the participants studied ecology, it is unknown as to whether their study courses affected the results.

Conclusion

To conclude, this study highlighted the potential of a neuroimaging technique (i.e., fNIRS) at validating language assessments by providing neurocognitive evidence supporting language processing under assessment conditions. The listening assessment used in this study engaged language-related regions of the brain. Furthermore, discriminant validity of the listening test used in this study was supported by differences in brain activation patterns during linguistic and non-linguistic listening. No significant gender differences were observed in left hemisphere activation measured across all three conditions. This implies that the listening comprehension assessment used did not statistically discriminate across genders.

Funding Statement

We wish to acknowledge the funding support for this project from Nanyang Technological University under the Undergraduate Research Experience on CAmpus (URECA) programme.

Ethical approval:

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964

Helsinki declaration and its later amendments or comparable ethical standards. IRB Number of the project is IRB-2018-02-011-02.

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(3), 242-275. https:/doi.org/10.1016/j.jneuroling.2006.10.003

Aryadoust, V., Ng, L. Y., Esposito, G., & Foo, S. (2020). A neurocognitive investigation of test methods and gender effects in a computerized listening comprehension test. *Computer Assisted Language Learning.* https://doi.org/10.1080/09588221.2020.1744667

Aryadoust, V. (2020). A review of comprehension subskills: A Scientometrics perspective. *System, 88*, 102180. https://doi.org/10.1016/j.system.2019.102180

Aryadoust, V. (2012). Differential item functioning in while-listening performance tests: The case of IELTS listening test. *International Journal of Listening, 26*(1), 40-60. https://doi.org/10.1080/10904018.2012.639649

Baxter, L., Saykin, A., Flashman, L., Johnson, S., Guerin, S., Babcock, D. & Wishart, H. (2003). Sex differences in semantic language processing: A functional MRI study. *Brain and Language*, *84*(2), 264-272. https:/doi.org/10.1016/s0093-934x(02)00549-7

Belin, P., Zatorre, R. J. & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17-26. https:/doi.org/10.1016/s0926-6410(01)00084-2

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P. & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309-312. https:/doi.org/10.1038/35002078

Benedek, M., Beaty, R., Jauk, E., Koschutnig, K., Fink, A., Silvia, P. J., … Neubauer, A. C. (2014). Creating metaphors: The neural basis of figurative language production. *NeuroImage*, *90*(100), 99-106. https:/doi.org/10.1016/j.neuroimage.2013.12.046

Berl, M. M., Duke, E. S., Mayo, J., Rosenberger, L. R., Moore, E. N., VanMeter, J., … Gaillard, W. D. (2010). Functional anatomy of listening and reading comprehension during development. *Brain and language*, *114*(2), 115–125. https:/doi.org/10.1016/j.bandl.2010.06.002

Bian, X., Cai, X., & Cai, D. (2019). The contributions of listening and reading vocabularies to listening comprehension of Chinese EFL students. *International Journal of Listening*. https:/doi.org/10.1080/10904018.2019.1623678

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12)*,* 2767–2796. https:/doi.org/10.1093/cercor/bhp055

Bodie, G., & Worthington, D. L. (2017). Measuring listening. In G. Bodie & D. L. Worthington (Eds.), *The sourcebook of listening research: Methodology and measures*. Wiley.

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience, 25*(1), 151-188. https:/doi.org/10.1146/annurev.neuro.25.112701.142946

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University Press.

Boyle, J. P. (1987). Sex differences in listening vocabulary. *Language Learning*, *37*(2), 273-284. https:/doi.org/10.1111/j.1467-1770.1987.tb00568.x

Brimer, M. A. (1969). Sex differences in listening comprehension. *Journal of Research & Development in Education, 3*(1), 72-79. https:/doi.org/10.1007/BF00287397

Buchweitz, A., Mason, R. A., Tomitch, L. M., & Just, M. A. (2009). Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. *Psychology & neuroscience*, *2*(2), 111–123. https:/doi.org/10.3922/j.psns.2009.2.003

Buck, G. (1994). The appropriacy of psychometric measurement models for testing second language listening comprehension. *Language Testing, 11*(2), 145-170. https:/doi.org/10.1177/026553229401100204

Buck, G. (2001). *Assessing listening*. Cambridge University Press.

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: examining attributes of a free response listening test. *Language Testing, 15*(2), 119-157. https:/doi.org/10.1177/026553229801500201

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105. https:/doi.org/10.1037/h0046016

Davey, J., Cornelissen, P. L., Thompson, H. E., Sonkusare, S., Hallam, G., Smallwood, J., & Jefferies, E. (2015). Automatic and Controlled Semantic Retrieval: TMS Reveals Distinct Contributions of Posterior Middle Temporal Gyrus and Angular Gyrus. *The Journal of Neuroscience, 35*(46), 15230-15239. https:/doi.org/10.1523/jneurosci.4705-14.2015

Dieler, A., Tupak, S. & Fallgatter, A. (2012). Functional near-infrared spectroscopy for the assessment of speech related tasks. *Brain and Language*, *121*(2), 90-109. https:/doi.org/10.1016/j.bandl.2011.03.005

Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin, 136*(5), 822-848. https:/doi.org/10.1037/a0019749

Ellamil, M., Dobson, C., Beeman, M., & Christoff, K. (2012). Evaluative and generative modes of thought during the creative process. *NeuroImage, 59*(2)*,* 1783–1794 https:/doi.org/10.1016/j.neuroimage.2011.08.008

Eom, M. (2008). Underlying factors of MELAB listening construct. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 6*(1)*,* 77–94.

Ferstl, E.C., Neumann, J., Bogler, C., & von Cramon DY. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping, 29*(5)*,* 581– 93. https:/doi.org/10.1002/hbm.20422

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th Ed.). Sage.

Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250-254. https:/doi.org/10.1016/j.conb.2012.10.002

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*(4), 1357-1392. https:/doi.org/10.1152/physrev.00006.2011

Furuya, A. (2019). How do listening comprehension processes differ by second language proficiency? Top-down and bottom-up perspectives*. International Journal of Listening.* https:/doi.org/10.1080/10904018.2019.1694411

Ghorbani Nejad, S., & Farvardin, M. T. (2019). Roles of general language proficiency, aural vocabulary, knowledge, and metacognitive awareness in L2 learners' listening comprehension. *International Journal of Listening.* https:/doi.org/10.1080/10904018.2019.1572510

Goh, C. C. M. (2008). Metacognitive instruction for second language listening development theory, practice and research implications. *RELC Journal, 39*(2), 188-213. https:/doi.org/10.1177/0033688208092184

Habl, C. (2018). Language aptitude and gender. In S. M. Reiterer (Ed.), *Exploring Language Aptitude: Views from Psychology, the Language Sciences, and Cognitive Neuroscience* (pp. 229-242). Springer International Publishing.

Hallam, G. P., Thompson, H. E., Hymers, M., Millman, R. E., Rodd, J. M., Lambon Ralph, M. A., … Jefferies, E. (2018). Task-based and resting-state fMRI reveal compensatory network changes following damage to left inferior frontal gyrus. *Cortex, 99*, 150-165. https:/doi.org/10.1016/j.cortex.2017.10.004

Hamp-Lyons, L. (1989). Applying the partial credit method of Rasch analysis: Language testing and accountability. *Language Testing, 6*(1), 109-118. https:/doi.org/10.1177/026553228900600109

Harding, L. (2011). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing, 29*(2), 163-180. https:/doi.org/10.1177/0265532211421161

Hoffman, P., Pobric, G., Drakesmith, M. & Lambon Ralph, M. A. (2012). Posterior middle temporal gyrus is involved in verbal and non-verbal semantic cognition: Evidence from rTMS. *Aphasiology*, *26*(9), 1119-1130. https:/doi.org/10.1080/02687038.2011.608838

Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning, 56*(s1), 279-304. https:/doi.org/10.1111/j.1467-9922.2006.00365.x

Jeong, H., Sugiura, M., Sassa, Y., Yokoyama, S., Horie, K., Sato, S., . . . Kawashima, R. (2007). Cross-linguistic influence on brain activation during second language processing: An fMRI study. *Bilingualism: Language and Cognition, 10*(2), 175-187. https:/doi.org/10.1017/S1366728907002921

John, O. P. & Benet-Martínez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 339-369). Cambridge University Press.

Johnston, M. K., Weaver, J. B., Watson, K. W. & Barker, L. B. (2000). Listening Styles: Biological or Psychological Differences? *International Journal of Listening*, *14*(1), 32-46. https:/doi.org/10.1080/10904018.2000.10499034

Karunanayaka, P. R., Holland, S. K., Schmithorst, V. J., Solodkin, A., Chen, E. E., Szaflarski, J. P., & Plante, E. (2007). Age-related connectivity changes in fMRI data from children listening to stories. *NeuroImage, 34*(1), 349-360. https:/doi.org/10.1016/j.neuroimage.2006.08.028

Kintsch, W. (1991). The role of knowledge in discourse comprehension: A construction-integration model. *Advances in Psychology*, *79*, 107-153. https:/doi.org/10.1016/s0166-4115(08)61551-4

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kröger, S., Rutter, B., Stark, R., Windmann, S., Hermann, C., & Abraham, A. (2012). Using a shoe as a plant pot: Neural correlates of passive conceptual expansion. *Brain Research, 1430,* 52–61. https://doi.org/10.1016/j.brainres.2011.10.031

Leonard, R. K. (2019). Examining the relationship between decoding and comprehension in L2 listening. *System, 87*. https://doi.org/10.1016/j.system.2019.102150

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R. & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex*, *14*(9), 1008-1021. https://doi.org/10.1093/cercor/bhh061

Liao, Y. (2007). Investigating the construct validity of the grammar and vocabulary section and the listening section of the ECCE: Lexico-grammatical ability as a predictor of L2 listening ability. *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, *5*, 37–78.

Mason, R. A., & Just, M. A. (2006). Chapter 19 - Neuroimaging Contributions to the Understanding of Discourse Processes. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics (Second Edition)* (pp. 765-799). Academic Press.

McNamara, T. F. (1991). Test dimensionality: IRT analysis of an ESP listening test1. *Language Testing, 8*(2), 139-159. https://doi.org/10.1177/026553229100800204

McNamara, T. F. (1996). *Measuring second language performance.* New York: Longman.

Mislevy, R. J. (2009). Validity from the perspective of model-based reasoning. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 83-110). Information Age Publishing.

Nishiyori, R. (2016). fNIRS: An emergent method to document functional cortical activity during infant movements. *Frontiers in Psychology*, *7,* 553. https://doi.org/10.3389/fpsyg.2016.00533

Noonan, K. A., Jefferies, E., Visser, M., & Ralph, M. A. L. (2013). Going beyond inferior prefrontal involvement in semantic control: Evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *Journal of Cognitive Neuroscience, 25*(11), 1824-1850. https://doi.org/10.1162/jocn_a_00442

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113. doi:10.1016/0028-3932(71)90067-4

Payne, T. W., & Lynn, R. (2010). Sex differences in second language comprehension. *Personality and Individual Differences*, *50*(3), 434-436. https://doi.org/10.1016/j.paid.2010.10.026

Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology, 15*(2), 202-206. https://doi.org/10.1016/j.conb.2005.03.007

Perfetti, C.M. & Frishkoff, G.A. (2008). The neural bases of text and discourse processing. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of Neuroscience of Language* (165–174). Elsevier.

Phillips, M. D., Lurito, J. T., Dzemidzic, M., Lowe, M. J., Wang, Y. & Mathews, V. P. (2000). Gender based differences in temporal lobe activation demonstrated using a novel passive listening paradigm. *NeuroImage*, *11*(5), S352. https://doi.org/10.1016/s1053-8119(00)91283-7

Richards, J. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, *17*(2), 219-240. https://doi.org/10.2307/3586651

Rogalsky, C., Matchin, W., & Hickok, G. (2008). Broca's area, sentence comprehension, and working memory: An fMRI study. *Frontiers in Human Neuroscience*, *2*(14). https://doi.org/10.3389/neuro.09.014.2008

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, *29*(5), 700–710. https://doi.org/10.1177/0956797617742725

Rossiter, C. M. (1972). Sex of the speaker, sex of the listener, and listening comprehension. *Journal of Communication*, *22*(1), 64-69. https://doi.org/10.1111/j.1460-2466.1972.tb00132.x

Rost, M. (2016). *Teaching and researching listening* (3rd Ed.). Longman.

Rüschemeyer, S.A., Fiebach, C. J., Kempe, V., & Friederici, A. D. (2005). Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human Brain Mapping, 25*(2), 266-286. https://doi.org/10.1002/hbm.20098

Sawaki, Y., Kim, H. J., & Gentile, C. (2009). Q-matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments. *Language Assessment Quarterly*, *6*(3), 190-209. https://doi.org/10.1080/15434300902801917

Sawyer, C. R., Gayle, K., Topa, A. & Powers, W. G. (2014). Listening fidelity among native and nonnative English-speaking undergraduates as a function of listening apprehension and gender. *Communication Research Reports*, *31*(1), 62-71. https://doi.org/10.1080/08824096.2013.844119

Schlaepfer, T., Harris, G. J., Tien, A. Y., Peng, L., Lee, S. & Pearlson, G. D. (1995). Structural differences in the cerebral cortex of healthy female and male subjects: A magnetic resonance imaging study. *Psychiatry Research*: *Neuroimaging*, *61*(3), 129-135. https://doi.org/10.1016/0925-4927(95)02634-a

Shin, S. K. (2005). Did they take the same test? Examinee language proficiency and the structure of language tests. *Language Testing*, *22*, 31-57. https://doi.org/10.1191/0265532205lt296oa

Swick, D., Ashley, V. & Turken, A. U. (2008). Left inferior frontal gyrus is critical for response inhibition. *BMC Neuroscience*, *9*(1), 102. https://doi.org/10.1186/1471-2202-9-102

Taghizadeh, M., & Saadatju, S. (2020). Engineering students' needs for listening scaffolding strategies and their perceptions of instructors' performance in an academic listening course. *Research in Science & Technological Education—first online*. https://doi.org/10.1080/02635143.2020.1764925

Tatsuno, Y., & Sakai, K. L. (2005). Language-related activations in the left prefrontal regions ae differentially modulated by age, proficiency, and task demands. *The Journal of Neuroscience, 25*(7), 1637-1644. https://doi.org/10.1523/jneurosci.3978-04.2005

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *The Journal of Neuroscience, 30*(47), 15888-15894. https://doi.org/10.1523/jneurosci.2674-10.2010

Vandergrift, L. & Goh, C. C. (2014). Teaching and learning second language listening: Metacognition in action. *Applied Linguistics*, *35*(2), 224-226. https://doi.org/10.1093/applin/amu002

Voorhees, C. M., Brady, M. K., Calantone, R. & Ramirez, E. (2015). Discriminant validity testing in marketing: An analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science*, *44*(1), 119-134. https://doi.org/10.1007/s11747-015-0455-4

Wagner, D. D., Kelley, W. M., Haxby, J. V. & Heatherton, T. F. (2016). The dorsal medial prefrontal cortex responds preferentially to social interactions during natural viewing. *The Journal of Neuroscience*, *36*(26), 6917-6925. https://doi.org/10.1523/jneurosci.4220-15.2016

Wang, Y., & Treffers-Dallerb, J. (2019). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System, 65,* 139-150. https://doi.org/10.1016/j.system.2016.12.013

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2010). The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cerebral Cortex, 21*(5), 1066-1075. https://doi.org/10.1093/cercor/bhq180

Witelson, S., Glezer, I., & Kigar, D. (1995). Women have greater density of neurons in posterior temporal cortex. *The Journal of Neuroscience*, *15*(5), 3418-3428. https://doi.org/10.1523/jneurosci.15-05-03418.1995

Worthington, D. L., & Bodie, G. (2020). *The handbook of listening*. Wiley.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M. & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *NeuroImage*, *32*(4), 1826-1836. https://doi.org/10.1016/j.neuroimage.2006.04.199