Repeating the listening text: effects on listener performance, metacognitive strategy use, and anxiety

Franz Holzknecht (University of Teacher Education in Special Needs, Zurich, Switzerland), Luke Harding (Lancaster University, United Kingdom)

Abstract

In second language listening assessment and pedagogy, practitioners hold different views on whether to repeat a listening text in contexts where inferences about listening ability are to be drawn from task performance. To address this issue, we investigated the effects of repeating the listening text (double play) on listener performance, listening strategies, test-taking strategies, test-taking anxiety, and listening anxiety. 306 Austrian secondary school students responded to four listening tasks drawn from the Austrian *Matura* exam and completed questionnaires measuring strategic behaviour and anxiety in a counter-balanced research design. Data were analysed using many-facet Rasch measurement (MFRM), factor analysis, and inferential statistics. Findings confirmed that double play led to higher levels of listener performance across two task types (multiple-choice items and note completion), however scores were higher in the single play condition compared to the first play of double play. Students also reported lower levels of anxiety, and the use of more listening strategies and fewer test-taking strategies in double play compared to single play with small effect sizes. We discuss the importance of balancing an empirically derived understanding of the effects of repeating the listening text with considerations of the purpose of an assessment, and the characteristics of the target-language use domain.

Introduction

Playing the listening text twice ('double play') is a traditional approach in the teaching of second language (L2) listening comprehension. Within language classrooms, the convention of double play is thought to confer several advantages for both teacher and learner. First, it provides learners the chance to adjust to characteristics of the recording, such as novel voices/accents and variable speech rates (Field, 2008; Hubbard, 2017). Second, it is thought to provide an opportunity for learners to achieve a deeper understanding of a listening passage because L2 listeners may "often miss the first parts of an aural text and … struggle to construct the context and the meaning for the rest of the message" (Vandergrift & Goh, 2012, p. 4). For

these reasons, practitioner-focused texts routinely recommend double, or multiple, plays of listening texts within instructional contexts (Field, 2008; Hulstijn, 2003; Jensen & Vinther, 2003; Vandergrift & Goh, 2012; Vandergrift & Tafaghodtari, 2010).

Within contexts of listening *assessment*, however, the approach to number of plays is more mixed. Double play is utilised in many high-stakes national L2 school leaving exams, for example in the German Abitur (across all German federal states), the Standardised Austrian Matriculation Examination (Matura), and the English exam in the French Baccalauréat, where some of the English listening texts are played up to three times. Internationally recognised language tests such as the Cambridge English Assessment suite also feature double play in all listening tasks across different levels (Field, 2013). Other international language tests (the British Council Aptis Test; the Duolingo English Test; the Oxford Test of English) give candidates choice over the number of times the listening text is played, or the number of plays varies between single play and double play across different tasks (e.g., Pearson Test of English [PTE]).

Several prominent international high-stakes tests, however, maintain a policy of playing listening texts only once. This is the approach taken by the Test of English for International Communication (TOEIC), the Test of English as a Foreign Language (TOEFL iBT), the International English Language Testing System (IELTS), the PTE Academic, the Occupational English Test (OET), and the General English Proficiency Test (GEPT).

Rationales for single play in listening assessment typically revolve around three main considerations: (1) practicality, (2) authenticity and (3) psychometric quality. On the first, in an assessment context, if listening texts are played only once, test providers can include a larger number of items in their tests, enhancing reliability and, potentially, construct coverage (Fortune, 2004; Green, 2017; Jones, 2011). This is a powerful consideration given the need to maximise measurement quality within the time constraints of operational testing. On the second consideration, a common argument for single play is that in many real-life situations people hear a listening text only once (e.g., an important announcement; a lecture), and thus listening assessments should replicate the real-world domain to draw valid inferences of future listening performance (Buck, 2001; Ruhm et al., 2016). Finally, single play might be implemented to enhance the psychometric quality of a listening test: to make a test more challenging, or to discriminate more effectively between candidates of different listening abilities, which might be a primary consideration depending on test purpose (Ruhm et al., 2016).

Proponents of double play in assessment contexts, on the other hand, base their arguments on three alternative considerations: (1) the limited capacity for single play listening tasks to provide sufficient information about context, (2) a view that the prevalence of real-life once only listening scenarios might be overstated, and (3) the desire to avoid construct-irrelevant variance. On the first point, some researchers have argued that in many real-world listening scenarios listeners will have some *a priori* knowledge of what a spoken text is going to be about based on previous encounters with the speaker, the physical setting of the conversation, and other contextual cues (Field, 2015). Listeners can typically access and use this contextual knowledge to aid comprehension. In this sense, double play can provide a means of modeling schema building within the relatively artificial context of a listening assessment task. Second, real-world listening contexts often involve opportunities for listeners to hear information repeated. For example, in face-to-face interaction interlocutors can often ask for clarification should they miss or mishear information, and repetition and paraphrase are well-understood phenomena of conversational discourse (Wong, 2000). In academic settings, course materials are increasingly offered either fully online or in a hybrid form including online and offline content (Sun & Chen, 2016), thus enabling students to play listening texts several times. On the third point, double play is sometimes seen as warranted on the grounds that playing listening texts only once in teaching and assessment scenarios may cause a level of anxiety among learners that exceeds the listening-related anxiety they would experience in real-word listening contexts, thus introducing a potential source of construct-irrelevant variance (Field, 2015; Winke & Lim, 2014).

Despite these conceptual debates, however, there has been little empirical research on the effects of playing a listening text once versus twice, and methodological limitations in the studies to date mean that our understanding of the effects of double versus single play remains limited. For language assessment practice, this represents an important site of enquiry as policy decisions at the design stage have the potential to influence measurement quality and construct representation. However, the importance of this topic extends from assessment into L2 learning and teaching contexts as well; washback from high-stakes exams creates a connection between language assessment design policies and approaches to classroom teaching (particularly in test preparation contexts) which might sustain either a double or single play approach without a sufficient evidence base. A deeper understanding of the effects of single versus double play in listening performance, which incorporates consideration of cognitive processes, strategic behaviour, and affective factors, would help to build theory around listening comprehension with a view to informing both assessment and pedagogical practice.

In this paper, we report on findings from part of a larger project designed to investigate the effects of double play (in comparison to single play) on listening performance, cognitive processing, metacognitive strategy use, and anxiety. The current study reports on the quantitative part of the project, which focussed on students' listening performance, metacognitive strategy use, and anxiety. The qualitative findings on cognitive processes are discussed elsewhere (see Holzknecht, 2019; Holzknecht, in preparation).

Literature review

Effects of double play on listening performance

Previous research on the question of repeating the listening text has tended to investigate the impact of repetition on listening task scores, item discrimination, or both. Two of these early studies reported no benefits of double play over single play (Brindley & Slatyer, 2002; Henning, 1991). For example, Henning (1991) found that double play tended to enhance task performance as indicated by lower item difficulty measures in the double play condition, however this did not reach statistical significance when mean difficulty was compared with single play. Henning's results also showed that double play did not have a positive effect on item discrimination, on item response validity (as indicated by Rasch measurement fit statistics), or on format construct validity (as indicated by a correlation matrix). Based on these findings, Henning suggested that there is no need to repeat the listening texts in the TOEFL test, a practice which still holds today. Similar findings were reported by Brindley and Slatyer (2002) in a study investigating the influence of multiple factors on listening task difficulty including speech rate, text type, input source, and item format, as well as single versus double play. Brindley and Slatyer found that speech rate and item format affected task difficulty, but double play did not have an observable effect.

Notable limitations of these studies concern the sample size and the tasks used. In Henning's (1991) study, 40 participants listened to the tasks in double play, and the tasks themselves were discrete-point items based on one-, two-, or three-sentence passages. As such, the study may have lacked sufficient power to determine the impact of repeating the listening text, and findings may not generalize adequately to longer listening passages. In Brindley and Slatyer's (2002) study, about 70 participants (the exact number is not mentioned by the authors) completed a double play task, providing only slightly better power compared with Henning.

In contrast to the findings discussed above, the majority of studies which have investigated the effects of double play in L2 listening pedagogy and assessment found that it was beneficial compared to single play in terms of (1) quantity and accuracy of recalled lexical items and propositions (Lund, 1991), (2) an increase in perceived listening performance by students (Kwon & Park, 2017), and (3) actual listening performance as indicated through test scores (Aryadoust, 2019; Berne, 1995; Chang & Read, 2006; Field, 2015; Iimura, 2007; Ruhm et al., 2016; Sakai, 2009; Sherman, 1997). Similar to the research above, however, notable limitations of these studies include small sample sizes, i.e. fewer than 50 participants in each condition (Aryadoust, 2019; Berne, 1995; Chang & Read, 2006; Iimura, 2007; Ruhm et al., 2016; Sakai, 2009; Sherman, 1997) or the use of listening tasks developed without field testing and/or prior statistical item analysis (Aryadoust, 2019; Berne, 1995; Chang & Read, 2006; Lund, 1991; Kwon & Park, 2017; Sherman, 1997). In addition, none of the studies adopted a fully counter-balanced research design where task ordering effects were considered.

The most comprehensive study of single play versus double play to date was conducted by Field (2015). In this investigation, Field first collected test score data from 73 participants taking two IELTS listening tasks (multiple-choice and gap fill). Before data collection, participants were told that they would hear the listening text only once, but after the first play they were informed that they could listen again and change their answers. Test scores increased by 1.03 on average (on a scale of 0 to 10) after the second play across all proficiency levels, with a mean standard deviation of 1.37. In the second part of the investigation, Field analysed stimulated recall protocols of 37 participants while they were solving the same two IELTS listening tasks. For this part of the study, participants were told from the outset that they would hear the text twice. Field's analyses revealed that it was only during the second play that many participants made use of higher order listening processes (e.g., constructing meaning for wider parts of the text as opposed to localised understanding). Participants also reported being less anxious in the double play condition.

Two limitations to Field's (2015) study design were that, (1) in the first (quantitative) part, listeners did not know they would be able to listen twice, and this might have led to different listening behaviours than if they knew they were experiencing a double-play condition; and (2) in the second part, listeners knew they would listen twice, but there was no comparison with a "pure" single play condition. Thus, building on Field's work, it is important to consider the nature of listeners' response processes in single versus double play conditions, comparing listeners who know *from the outset* that they will only have one chance to

comprehend a text (single play) and listeners who know that they will have a second chance after the first listening (double play).

Nevertheless, Field's study is important as it was the first to focus on students' response processes in relation to double play, moving beyond scores alone to provide insight into response processes and affective factors, both of which are relevant for understanding the nature of the operationalised listening construct (Hubley & Zumbo, 2017; Messick, 1995) For that reason, as well as investigating the impact of double play at the score level, we focused in this study on the impact of repeating the listening text on a range of metacognitive strategies in listening assessment, and on the key factor of anxiety.

Metacognitive strategies in L2 listening assessment

We classify metacognitive strategies in L2 listening assessment as comprising both *listening strategies* and *test-taking strategies*. Metacognitive *listening* strategies are conscious and goal-directed mental actions drawn on to aid comprehension (Vandergrift & Goh, 2012). They are particularly important for L2 learners whose comprehension is not yet fully automatized. Vandergrift and Goh (2012) propose 12 different metacognitive listening strategies: planning, focusing attention, monitoring, evaluation, inferencing, elaboration, prediction, contextualisation, reorganising, using linguistic resources, cooperation, and managing emotions. While there is some contention over the construct relevance of metacognitive strategy use in listening assessment (see Low & Aryadoust, 2021), researchers have argued that "any strategy a listener would have at his or her disposal 'in the wild' is part of that listener's listening proficiency, and should be reflected in scores" (Batty, 2021; see also Field, 2013). Following Field (2015), we hypothesise that the use of listening strategies could be impacted by the number of times the listening text is played. For example, listeners might focus their attention differently if they know from the outset whether a listening text is played once, twice, or multiple times. However, no study has yet investigated this in detail.

In contrast to metacognitive listening strategies, which play a major role in real-life listening, metacognitive test-taking strategies are specific to a test situation (Cohen, 2006). Cohen differentiates between *test-management* and *test-wiseness strategies*. Both types are applied by listeners to deal with the specific demands of test tasks, though test-management strategies are also reliant to some extent on language comprehension. For example, if students listen only for words which appear in the questions (a common test-management strategy), they still need to *listen*. Test-wiseness strategies, in contrast, are independent of language comprehension and solely informed by the test tasks; for example, when students guess an

answer before listening to the text. Nevertheless, both types of test-taking strategy relate to an overt orientation towards test taking, comprising behaviours specific to the test taking situation. It is not yet understood how single play and double play impact the use of test-taking strategies, though we would hypothesise that the more intense focus on "one-shot" listening in single play might promote more strategic test-taking behaviour.

Anxiety

Anxiety in relation to learning a foreign language has been defined as "the feeling of tension and apprehension specifically associated with second-language contexts" (MacIntyre & Gardner, 1994, p. 284). One form of language learning anxiety is L2 listening anxiety, which concerns negative feelings related to listening in a foreign language due to the unique features of spoken language, such as the need for real-time comprehension or lack of clarity (Vogely, 1998). Although research is still sparse, L2 listening anxiety appears to be negatively related to listening comprehension; that is, less anxious candidates in general perform better than more anxious candidates (see, for example, Révész & Brunfaut, 2013). Another form of language learning anxiety pertains to the test-taking process itself. Test-taking anxiety consists of "individuals' cognitive reactions to evaluative situations, or internal dialogue regarding evaluative situations, in the times prior to, during, and after evaluative tasks" (Cassady & Johnson, 2002, p. 272). It has been established through a large body of research that high levels of test-taking anxiety are generally related to a decline in test performance (for a review, see Winke & Lim, 2014).

It seems reasonable to hypothesise that double play could lower candidates' anxiety levels. If students know from the outset that they will hear the listening text a second time, they may feel less intimidated by having to listen in a foreign language. They may also be less worried and less stressed about taking the test and consequently more confident in their abilities. Field's (2015) research suggests that double play might indeed reduce anxiety – however, due to the research design Field was not able to directly compare the anxiety levels of candidates who experienced single play with candidates who experienced double play.

Methodology

Research questions

Based on the literature review, the following research questions were investigated:

RQ1. To what extent does listening performance differ across listening tasks completed in single play and double play (as indicated by differential task difficulty)?
   a. To what extent does listening performance differ across listening tasks completed in single play and in the first play of double play?

RQ2. To what extent does listeners' metacognitive strategy use (their use of listening strategies and test-taking strategies) differ between listening tasks completed in single play and double play?

RQ3. To what extent do listeners' anxiety levels differ between listening tasks completed in single play and double play?

Tasks

The listening tasks were taken from past live papers of the English listening section of the standardised Austrian matriculation examination (Matura), a professionally-developed high-stakes exam administered at the end of Austrian upper-secondary school (see Spöttl et al., 2018). Four tasks were used, all developed for CEFR level B2 (see Council of Europe, 2001). Two tasks were multiple-choice (MC) tasks, where students choose one correct answer from four options. The other two were note-form (NF) tasks, where students fill in gaps at the end of sentences with a maximum of four words. Both task types are commonly used in listening assessment and classroom comprehension activities and allow for comparison of results with other research in this area, notably Field (2015). The following seven criteria were considered when selecting the specific tasks from a wider pool of available Matura tasks:

1. The task was used in a live administration of the Matura, which guaranteed that it had passed all the quality control procedures.
2. The topics of the four tasks were suitably different to avoid potential overlap in terms of topical knowledge.
3. The tasks targeted a mix of standard British and American English to avoid overlap in terms of accent familiarity.
4. Tasks with the same format had the same number of items to allow for cross comparisons between item formats.
5. Both NF tasks were "fill in the blank at the end of sentences" format rather than "fill in the blank in the middle of sentences" or "answering questions" format to allow for cross comparisons (all three formats are developed for the Matura).
6. The tasks had similar task and item difficulty properties based on the field trial and standard setting results to allow for comparisons between tasks and item format.

7. The targeted listening behaviour within each task type was the same to allow for cross comparisons across task types.

Table 1 summarises the specific tasks chosen for the study. MC1 was an interview between a speaker with an American accent and a speaker with a British accent, MC2 was an interview between two speakers with a British accent, NF1 was a monologue of a speaker with a British accent, and NF2 was a radio program with three speakers with an American accent. All four listening texts were non-scripted authentic materials from real-world sources and included natural disfluencies. The speech rate of the different speakers in the four tasks was between 150 and 180 words per minute. As shown in table, the tasks were also similar in terms of audio file length and mean item difficulty (according to facility values (FVs) derived from field trials), with the notable exception of NF2, which had a shorter audio file and was also somewhat more difficult. NF2 had to be included due to the limited number of tasks available which matched the criteria above. Despite these differences, a group of ten expert judges placed all four tasks at CEFR B2 level during a formal standard setting procedure (the method through which assessment tasks are empirically assigned to proficiency levels to determine cut-off scores [see Council of Europe, 2009]). The two MC tasks were developed by the team of professional Matura item writers to test comprehension of main ideas and supporting details and the two NF tasks targeted comprehension of specific information and important details. All items are comparable in terms of the specific listening skills measured across tasks, as the items were designed through a process of "textmapping" (see Green, 2017, pp. 57-83 for details of this approach). The tasks and links to the audio files are included in the supplementary material (pp. 1-12).

Table 1: Summary of the tasks used in the study

Task ID	Task title	Accent*	Format	Number of items	Audio file length	Mean FV trial	Std. setting	Target**
MC1	Apted's film experiment	BE, AE	MC	6 + 1 example	4 min 01 sec	69%	B2	MISD
MC2	Useful plastic bottles	BE	MC	6 + 1 example	3 min 41 sec	69%	B2	MISD
NF1	Swan upping	BE	NF	9 + 1 example	3 min 40 sec	71%	B2	SIID
NF2	Lego master model builder	AE	NF	9 + 1 example	2 min 50 sec	43%	B2	SIID

*BE = standard British English, AE = standard American English

**MISD = understanding main ideas and supporting details, SIID = understanding specific information and important details.

Questionnaire

In addition to the listening tasks, the other main instrument of the study was a questionnaire targeting metacognitive strategies (listening strategies and test-taking strategies), and anxiety (test-taking anxiety and listening anxiety), designed to elicit data to inform RQ2 and RQ3. The questionnaire was constructed following the guidelines by Dörnyei and Taguchi (2009, pp. 127–128) and consisted of 25 statements to which participants had to indicate their level of agreement on a four-point Likert scale (disagree, partly disagree, partly agree, agree). Participants could also choose "I don't know". The questionnaire was administered in German; an English translation is included in the supplementary material (p. 13).

The items targeting listening strategies (14-19) were based on Vandergrift (1997) and had also been adapted by Winke and Lim (2014). The items targeting test-taking strategies (1-6) were based on Cohen and Upton (2007) and had been adapted by Winke and Lim (2014), and covered both test-management strategies (items 1, 2, 4, and 5) and test-wiseness strategies (items 3 and 6). The items targeting test-taking anxiety (7 to 13) were drawn from Winke and Lim (2014), who based their questionnaire on Cassady and Johnson (2002). The items targeting listening anxiety (20-25) were adapted from Elkhafaifi (2005) and were also used by Brunfaut and Révész (2015).

Several considerations guided the compilation of items for the questionnaire. The source questionnaires contained more items than could be administered in the study, so only those items which were considered most relevant for the purpose of the study were included. The decision on whether an item should be included in the study was taken by the authors based on

three factors. First, the wording of several items was phrased in very general terms, and it would have been difficult to relate these to the specific tasks the students had just performed in a single play or double play condition. Second, several items would not have been relevant to the tasks the students had just performed, and third, some items overlapped with other items and were therefore omitted.

Participants responded to the questionnaire twice: once after completing two tasks of the same format in a single play condition and again after completing the other two tasks of the other format in a double play condition (the rationale for relating the questionnaires to the task format is described in the research design section below). The other choices were to administer a questionnaire at the end of the testing process, or to administer a questionnaire after each of the four test tasks. We piloted all three options with 20 participants and found that participants preferred the design used in the current study as it allowed them to directly reflect on the tasks without the added distraction of adding two additional questionnaires. Following the findings from the same pilot study, all statements were phrased in past tense and detailed instructions were included to make it clear to the participants that they should relate their answers only to the two tasks they had just completed.

Participants

The target population for the Matura tasks used in the study are typically 17–19-year-old students in grade 8 of the Austrian academic upper secondary school system. However, according to the Austrian academic upper secondary curriculum, students should already have reached B2 at the start of grade 7. It was therefore decided to recruit students from grade 7 instead of grade 8 to take part in the study, as students in grade 8 might have already been familiarised with the tasks through test preparation activities (the tasks were in the public domain at the time of data collection).

Students were recruited via their class teachers, who were known to the first author through professional networks. In total, 306 students (197 female and 109 male) attending 16 different grade 7 classes took part in the study. The classes were spread across five academic upper secondary schools over three regions across Austria (Upper Austria, Styria, and Vorarlberg).

Most students were 16 or 17 years old (69.9% and 26.5% respectively, see Table 2), with a smaller number of students aged 18 (3.6%). German was the L1 of most of the participants (91.2%, as shown in Table 3). Thirteen percent of the students grew up bilingually, with 13 students (4.2%) having English as a second L1 (Table 4).

Table 2: Participants' age

Age	*N*	%
16	214	69.9
17	81	26.5
18	11	3.6
total	306	100.0

Table 3: Participants' L1

L1	*N*	%
German	279	91.2
Serbian	5	1.6
Turkish	4	1.3
French	1	0.3
Italian	1	0.3
other	8	2.6
missing	8	2.6
total	306	100.0

Table 4: Bilingual participants

Additional language	*N*	%
English	13	4.2
Turkish	6	2.0
French	4	1.3
Italian	3	1.0
Croatian	2	0.7
Hungarian	2	0.7
Serbian	2	0.7
Spanish	2	0.7
other	6	2.0
total	40	13.1

Research design and procedure

The 306 participants completed the tasks in a complex counter-balanced design. As shown in Table 5, participants attended one of 16 classes, each receiving the tasks in a different order. In Class 1 students first completed both MC tasks in a double play condition followed by the questionnaire, and then both NF tasks in a single play condition again followed by the questionnaire. Students were told by the exam invigilator and also through recorded

instructions that they would hear the recording once or twice. The same task type was used within each condition in order not to confound questionnaire responses with potential task type effects. For example, students might have reacted differently to a single play condition for MC tasks than to a single play condition for NF tasks in terms of listening strategies, test-taking strategies, test-taking anxiety, or listening anxiety. If the two different task types had been used within the same condition, such differences may have weakened the validity of the responses. To control for potential ordering effects, the test was administered in 16 different versions, which was the total number of all possible combinations.

Table 5: Research design

# of times heard	Class 1	Class 2	Class 3	Class 4
2	MC 1	MC 1	MC 2	MC 2
2	MC 2	MC 2	MC 1	MC 1
	Questionnaire	Questionnaire	Questionnaire	Questionnaire
1	NF 1	NF 2	NF 1	NF 2
1	NF 2	NF 1	NF 2	NF 1
	Questionnaire	Questionnaire	Questionnaire	Questionnaire
	Class 5	**Class 6**	**Class 7**	**Class 8**
2	NF 1	NF 1	NF 2	NF 2
2	NF 2	NF 2	NF 1	NF 1
	Questionnaire	Questionnaire	Questionnaire	Questionnaire
1	MC 1	MC 2	MC 1	MC 2
1	MC 2	MC 1	MC 2	MC 1
	Questionnaire	Questionnaire	Questionnaire	Questionnaire
	Class 9	**Class 10**	**Class 11**	**Class 12**
1	MC 1	MC 1	MC 2	MC 2
1	MC 2	MC 2	MC 1	MC 1
	Questionnaire	Questionnaire	Questionnaire	Questionnaire
2	NF 1	NF 2	NF 1	NF 2
2	NF 2	NF 1	NF 2	NF 1
	Questionnaire	Questionnaire	Questionnaire	Questionnaire
	Class 13	**Class 14**	**Class 15**	**Class 16**
1	NF 1	NF 1	NF 2	NF 2
1	NF 2	NF 2	NF 1	NF 1
	Questionnaire	Questionnaire	Questionnaire	Questionnaire
2	MC 1	MC 2	MC 1	MC 2
2	MC 2	MC 1	MC 2	MC 1
	Questionnaire	Questionnaire	Questionnaire	Questionnaire

Each participating class was assigned to take one of the 16 versions of the test, so all individuals within a class took the same tasks in the same order. Due to this research design, the participants were divided into two groups. All participants from group 1 took the MC tasks in double play and the NF tasks in single play and participants from group 2 took the MC tasks in single play and the NF tasks in double play. The tests were administered in pen-and-paper form following detailed test administration guidelines.

In addition to the test booklets, candidates were also given two pens in a different colour (blue and red). They were instructed to use the blue pen to record answers for the tasks in single play and for the first play of double play, and the red pen only to record answers (or changes

to answers) during the second play of double play. The instructions in the audio file also included this information. This was done to be able to compare the students' performance between the first and second play of double play and between single play and the first play of double play.

Analysis

To answer RQ1 we analysed the test data. Prior to data analysis the first author scored the 306 test booklets. All items in single play were scored dichotomously as either correct or incorrect. For the MC tasks, published keys were used. For the NF tasks, extended marking schemes were obtained from the Austrian Ministry of Education. These extended marking schemes represent a comprehensive set of answers refined during the nation-wide live administrations, where a group of experts collectively score thousands of individual answers as either correct or incorrect and establish a definitive marking guide (see Eberharter & Frötscher, 2012). The items completed in double play were scored twice according to the keys and the extended marking schemes: once for answers provided after the first listening (as indicated by the blue pen) and once for answers after the second listening (as indicated by the red pen).

After scoring, the test data was analysed in three stages. First, the reliability of the tests was calculated using Classical Test Theory (Cronbach's Alpha). Second, two bias analyses were conducted using Many-Facet Rasch Measurement (MFRM, see Linacre, 1994): one comparing listener performances across single play and double play conditions; the other comparing listener performance across single play and the first play of double play and single play. MFRM was useful as a between-participants research design was used – the two task formats were completed in different conditions by two groups of students – and the two groups may have differed in their average listening proficiency. The MFRM model, as an extension of the basic Rasch model, solves this problem by expressing student ability and task difficulty as a probabilistic function on the same latent variable (for details see Eckes, 2015, pp. 21–27). For both analyses, we specified a 4-facet model:

1. *Students.* The 306 students who participated in the study.
2. *Items.* The 30 items across the four listening tasks. These were grouped according to tasks.
3. *Tasks.* The four tasks (MC1, MC2, NF1, NF2). This facet was specified as a demographic (= dummy) facet, meaning that all the elements of the facet are anchored at 0 logits. This was done to avoid disjoint subsets in the data as the items were nested within the tasks.

4. *Conditions.* The conditions the students completed the items in. The elements of this facet differed between the two analyses. In the first analysis we included "single play" and "double play" and in the second analyses we included "first play" (of double play) and "single play". The *Conditions* facet was also specified as a dummy facet to bias/interaction with the *Tasks* facet (following guidance from Linacre, personal communication, January 29, 2018).

For RQ2 and RQ3, the questionnaire data was analysed through an exploratory factor analysis, a reliability analysis, and a subsequent test of statistical difference between the two conditions. We decided to join the data for the individual tasks in each condition to achieve a larger sample size and higher common factor variance without cross-loadings, following recommendations by Osborne and Costello (2005). The datasets for the two separate conditions were therefore responses based on both MC and NF tasks, across all participants (N=304, with 2 missing responses). For this reason, one questionnaire item was dropped prior to the analysis as it was only included for the MC tasks and not the NF tasks (statement 2 in the questionnaire for MC tasks). The remaining items were identical for the two tasks.

Principal axis factoring with Varimax rotation was chosen as extraction method. De Winter and Dodou suggest using principal axis factoring for data with "a relatively simple factor pattern" (2012, p. 708), which was the case as it was hypothesised that the factors would cluster according to the four sections of the questionnaire (test-taking strategies, listening strategies, test-taking anxiety, and listening anxiety). The analyses were run with both Varimax and Direct Oblimin rotation, which yielded essentially the same results. Only the results based on Varimax rotation are presented below.

Results

Test reliability

Test reliability was calculated using Cronbach's Alpha. For group 1 (MC tasks in double play and the NF tasks in single play) $\alpha = .82$, and for group 2 (MC tasks in single play and the NF tasks in double play $\alpha = .83$, so the overall reliability of the test was high (Pallant, 2007).

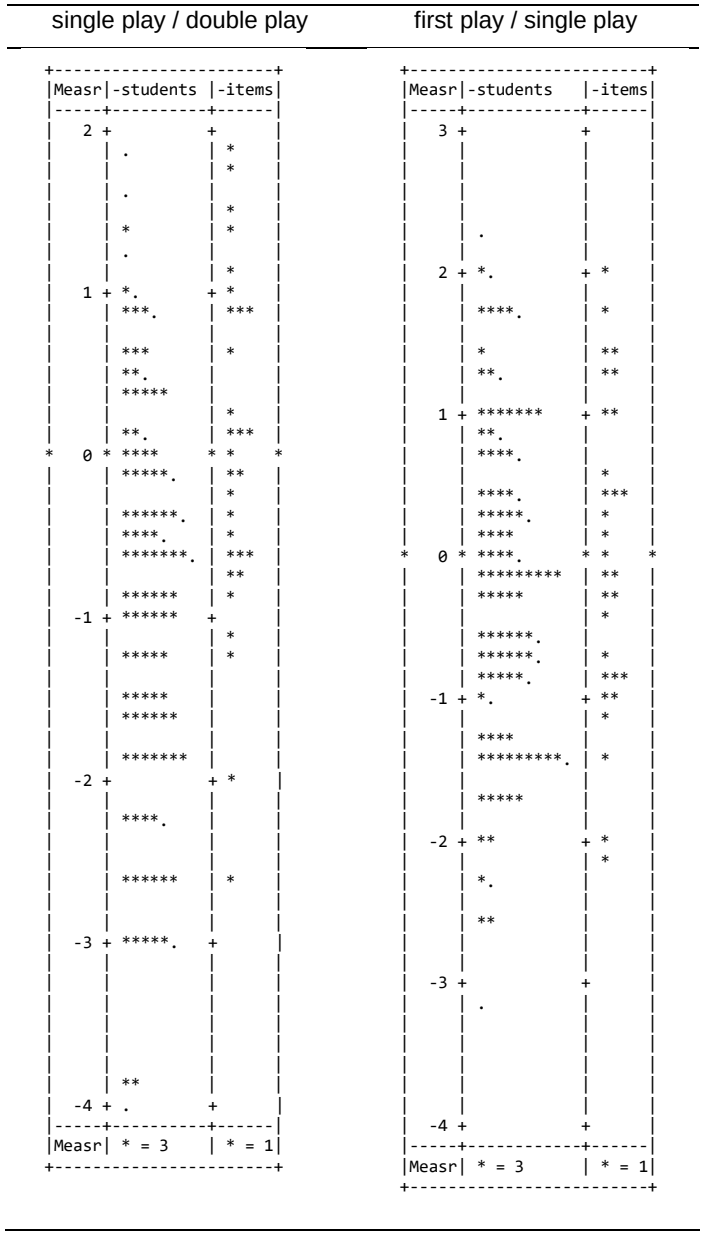
RQ1

MFRM summary statistics for the two analyses are shown in Table 6 and Wright maps in Figure 1. Infit MeanSquare (MS) values were close to the expected value of 1, indicating good model fit. Reliability of separation coefficients were all 0.82 or higher.

Table 6: Summary statistics of the two MFRM analyses (Tasks and Conditions were dummy facets and thus anchored at 0, so summary statistics for these two facets are not reported here).

	Single play / double play		first play / single play	
	Students	Items	Students	Items
N	306	30	306	30
Measures				
Mean	-0.95	0.00	-0.26	0.00
SD (pop.)	1.21	1.05	1.09	1.06
SE	0.49	0.15	0.45	0.14
RMSE (pop.)	0.51	0.15	0.45	0.14
Adjusted (True) *SD* (pop.)	1.10	1.04	0.99	1.05
Infit *MS*				
Mean	1.00	0.99	1.00	1.00
SD (pop.)	0.16	0.12	0.15	0.09
Outfit *MS*				
Mean	1.08	1.08	0.99	0.99
SD (pop.)	0.64	0.33	0.31	0.12
Homogeneity index (χ 2)				
df	305	29	305	29
p	0.00	0.00	0.00	0.00
Separation (pop.)	2.17	6.87	2.21	7.51
Reliability of separation (pop.)	0.82	0.98	0.83	0.98

Figure 1: Wright maps of the MFRM analyses (Tasks and Conditions were dummy facets and thus anchored at 0 logits, so they are not displayed in the Wright maps).



The MFRM bias analysis showed that the absolute measure of the four elements in the *Tasks* facet were higher in double play compared to the single play condition (Figure 2). In single play, absolute measures ranged from -0.12 to -0.54 logits across the four tasks (range = 0.42 logits), whereas in double play the range was smaller, spanning between 0.17 and 0.46 logits (range = 0.29 logits). All *t*-values in Figure 2 are larger than +/-2.00, indicating that the bias is significant (McNamara, 1996; McNamara et al., 2019, pp. 122–124), with small effect sizes (Cohen's $d = 2t / \sqrt{(df)}$).

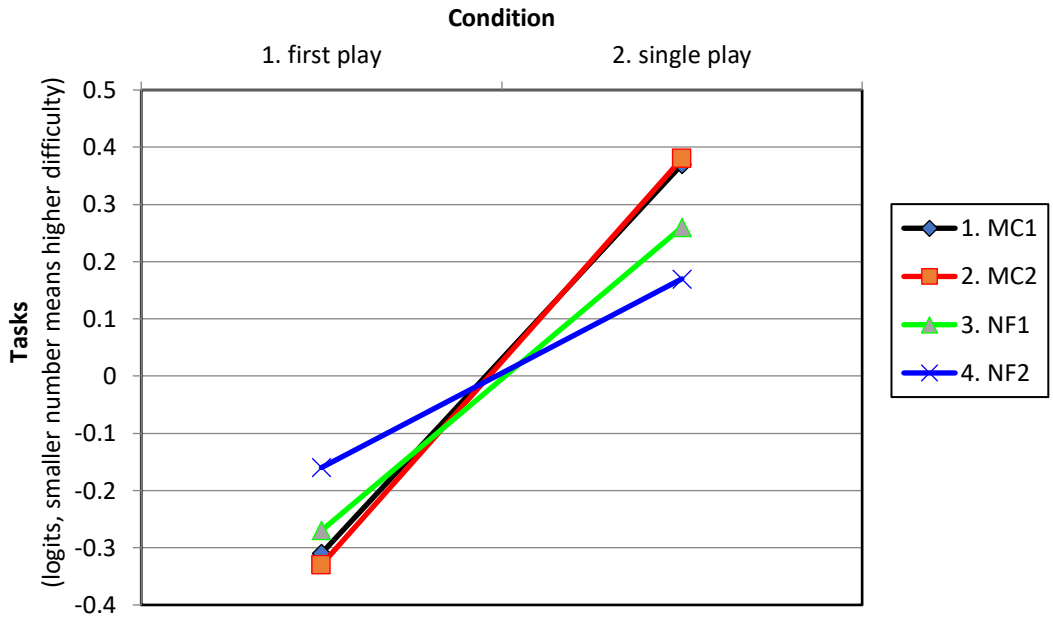
Figure 2: Facets bias analysis and associated t-values between single play and double play across the four tasks

	t-value	Cohen's *d*
MC1	-7.69	-0.36
MC2	-7.57	-0.35
NF1	-2.95	-0.11
NF2	-8.34	-0.32

A second bias analysis was conducted to detect potential differences in the absolute measure of the four tasks between the first play of the double play condition and the single play condition. As discussed above, the scores for the first play of double play were derived from the answers marked in blue pen. Whenever students had not selected an answer for a question, this was marked as zero.

The results show that there was a significant difference in absolute measure between the four tasks (Figure 3). The *t*-values for all possible pairs exceeded +/-2.00 and thus indicate statistical significance, with small effect sizes. The tasks had higher difficulty measures in the first play of double play compared to single play. For the MC tasks, only 0.7% of MC answers were left blank during single play, compared to 31.6% in the first play of double play, which may partly explain the higher scores. However, for the NF tasks, the number of non-responses was very similar between the two conditions (28.8% in single play and 28.2% in the first play of double play).

Figure 3: Facets bias analysis and associated t-values between the first play in double play and single play across the four tasks



	t-value	Cohen's *d*
MC1	-5.52	-0.26
MC2	-6.43	-0.30
NF1	-5.84	-0.22
NF2	-3.77	-0.14

To summarise, the MFRM results showed that students performed better on all four listening tasks in the double play condition compared with the single play condition, with small effect sizes. These findings are in line with previous studies which have found generally higher performance in a double play condition. Results from the second MFRM bias analysis demonstrated that students performed better after hearing listening texts in a condition where they *knew* that they could only listen once, compared with a condition where they expected to hear the text repeated. Analysis of non-responses, together with the results of the bias analysis, suggested that listening behaviour in the single play condition and in the first play of a double play condition for MC was not equivalent. These findings were less clear-cut for the NF task where effect sizes were smaller across both tasks, and where frequency of non-responses was relatively similar.

RQ2 and RQ3

As suggested by Osborne and Costello (2005, p. 3), a factor analysis was run on the questionnaire data for both conditions (single play and double play) to 1) identify the number of factors to be included in the final analysis by inspecting the scree plot each time and 2) detect outlier items which cross-loaded onto separate factors. The Kaiser-Meyer-Olkin (KMO) test for sampling adequacy and Bartlett's test for sphericity were performed for each separate analysis and were found to be adequate in each case (KMO ranged between 0.83 and 0.87 and Bartlett's test was significant at <0.001). For both conditions, the same three main factors were detected after inspection of the scree plots and five statements were identified and removed for the final analysis, as these statements each cross-loaded onto different factors (statements 1, 6, 14, 18, and 24).

The final analysis for both datasets was run with a fixed number of three factors. KMO was 0.87 for the single play condition and 0.86 for the double play condition and Bartlett's test was significant at <0.001 for both conditions. The total variance explained was 48.64 percent for the single play and 45.06 percent for the double play condition. For both conditions, the same items loaded mainly onto the same factors, except for item 3 and item 21, which loaded mainly onto a different factor in the double play condition. However, it was decided to keep these items in the analysis as in the single play condition they loaded mainly onto the same factor. As hypothesised, the identified factors corresponded to the pre-specified categories of the questionnaire. One factor relates to test-taking strategies (statements 3 to 5, whereby statement 3 was a test-wiseness strategy and statements 4 and 5 test-management strategies), one to listening strategies (statements 15 to 19), and one to anxiety (statements 7 to 13 on test-taking anxiety and statements 20 to 25 on listening anxiety loaded onto the same factor). The rotated factor matrix is included in the supplementary material (p. 14). Cronbach's alpha was calculated for each subscale: test-taking strategies, $\alpha = .70$; listening strategies, $\alpha = .75$, and anxiety, $\alpha = .93$, indicating that the individual factors were reliably measuring their respective constructs (see also Vogt, 2007).

A paired samples t-test was performed based on the means of the items within each factor to investigate differences between the single play and double play condition. The questionnaire included both positively as well as negatively formulated statements, so the data for the negatively formulated statements (statements 8 to 13 and 20 to 23) was reversed before calculating the mean to allow for cross-comparisons (see also Dörnyei & Taguchi, 2009, p. 90). As shown in Table 7, in the single play condition students reported relying slightly more

on test-taking strategies (p=.023) and slightly less on listening strategies (p<.001) (RQ2), and they were more anxious compared with the double play condition (p<.001) (RQ3). Effect sizes were small, though strongest for anxiety.

Table 7: Descriptive statistics, paired samples t-test, and effect sizes for the three factors of the questionnaire responses

Pairs	*N**	M**	SD	t	df	Sig. (2-tailed)	Cohen's d
test-taking strategies – single play	304	2.18	.65	2.29	303	0.023	0.13
test-taking strategies – double play	304	2.10	.64				
listening strategies – single play	303	1.91	.59	-3.93	302	<0.001	-0.23
listening strategies – double play	303	2.02	.61				
anxiety – single play	304	3.17	.61	6.15	303	<0.001	0.35
anxiety – double play	304	3.01	.67				

** Number of valid responses, total number of respondents = 306*

*** The mean is based on a four-point Likert scale where 1=disagree, 2=partly disagree, 3=partly agree, and 4=agree; the data for positively formulated anxiety statements (7, 24, 25) was reversed-coded to allow for cross-comparisons.*

Discussion

The findings of this study both confirm and extend previous research on the effects of repeating the listening text. First, the results demonstrate that listeners perform better in a double play condition, compared with single play, with a small but consistent bias effect found across all tasks. This finding agrees with the main share of previous research in this area, confirming the 'double play benefit' with a large sample of listeners. However, the current study also extends the knowledge base on the effects of double play in several ways.

First, the finding that performance on the *first* play of double play was significantly lower than performance on *single* play, coupled with the observation that many MC items were not answered during the first play of double play, suggests that learners may use a different type of listening behaviour in the first play of double play (for MC items). Knowledge that there will be an opportunity to listen more than once, therefore, emerges as a potentially important factor in shaping listening behaviour from the beginning of the task. This finding was less clear for NF items, however, indicating that task type may function as a mediating variable. Further qualitative research would be required to understand the nature of strategic behaviour during the task (see Holzknecht, in preparation).

Second, it was shown that listeners used listening strategies slightly more, test-taking strategies slightly less, and anxiety was lower in double play compared to single play (see also

Field, 2015). The findings for listening strategies and test-taker strategies matched the hypotheses expressed above – that use of listening strategies would differ across the two conditions, and that single play would encourage more strategic test-taking behaviour – though the findings from the questionnaire data revealed small (or even negligible) effects. The finding for anxiety, while still small, was stronger and confirmed the hypothesis based on previous qualitative research (Field, 2015) that double play would result in reduced levels of anxiety. This finding may be particularly important as previous research indicates a potential negative impact of anxiety on listening comprehension scores (see Holzknecht & Brunfaut, 2022).

For language assessment developers, the findings suggest that the decision to play a listening text once or twice in a listening assessment setting is not trivial. As well as differences at the score level, the results indicate there may be deeper construct implications as well – that the process of listening in both conditions, while related, is not entirely the same. For this reason, we argue that the decision to play a listening text once or twice should be guided by a clear understanding of the effects of double versus single play together with (1) a careful consideration of the purpose of the assessment (and its related practicality constraints) balanced with (2) a justification based on the prevalence (or not) of repeated listening in the target language use (TLU) domain. On the first point, different assessment purposes may routinely warrant single play listening based on overriding requirements for efficiency (e.g., placement assessments), or to model a specific degree of challenge (e.g., aptitude tests). Large-scale proficiency assessments may require items that discriminate candidates most effectively across multiple levels (taking an agnostic stance on the number of plays).

However, on the second point, for many assessments, decisions should be made with a clear understanding of the listening demands of the TLU domain. Some TLU domains for high-stakes assessment clearly warrant single play listening because precision is a key feature, such as aviation settings and many areas of health communication. Although possibilities exist for requesting clarification (e.g., 'readback' and 'hearback' sequences), it will often be important to draw inferences about what a listener can do in high-risk scenarios. In such contexts, the greater degree of challenge and higher levels of anxiety associated with single play may be construct-relevant as they would model the demands of the real-world listening environment. In other TLU domains, though, repeating listening texts may be a TLU domain feature. One example is academic admissions assessment; the opportunity to hear a listening text more than once appears to be increasing in the academic domain due to an increase in online learning environments (Sun & Chen, 2016) and double play may provide a more authentic condition that allows for accurate inferences about what a listener can achieve in future domains of

academic listening. In other assessments targeting more general domains of language use such as achievement tests (e.g., the Austrian Matura), classroom listening assessments, and other general proficiency assessments (such as the British Council Aptis test, for example), the findings of this study lend support to the double-play approach on the grounds that it allows listeners to demonstrate their understanding, reduces anxiety, and may encourage more listening-oriented and less test-oriented strategic behaviour. A fruitful approach in contexts where the choice based on a TLU domain analysis is not obvious could be to consider including both single play tasks and double play tasks in the same test, to ensure that a broad construct is captured.

Limitations and further research

This study had some limitations in terms of the research design, the nature of the sample and the selection of the test materials that should be acknowledged. First, the research design did not include conditions where listeners heard the same task type in both single and double play condition. As we explained above, the reason for this decision was to allow for the collection of relevant questionnaire responses after each listening test. As the Austrian Matura is a paper-based exam, we were also constrained by the need to use intact classes within each of the 16 versions. Future research using computer-delivered listening assessments could utilise designs in which tasks are randomly assigned to individual listeners in single or double play format, and where listeners experience the same task type in both conditions. Second, the population of listeners in this study was quite specific, and participants were all used to a double play convention, as double play is standard practice in the Austrian school system and the Matura exam. Although double play is also common in many educational settings around the globe (see Field, 2008; Hubbard, 2017), future research might explore the effects of double play with students who are more explicitly trained in responding to single play listening tasks . Finally, the Austrian Matura represents a specific approach to assessing listening in terms of task types and skills focus. We would encourage future research to explore the impact of double play across different task types in particular, and through different methods that may provide deeper insights into response processes than can be gained through questionnaires.

Bibliography

Aryadoust, V. (2019). Dynamics of item reading and answer changing in two hearings in a computerized while-listening performance test: An eye-tracking study. *Computer Assisted Language Learning*, 1–28. https://doi.org/10.1080/09588221.2019.1574267

Berne, J. E. (1995). How does varying pre-listening activities affect second language listening comprehension? *Hispania*, *78*(2), 316–329.

Brindley, G., & Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, *19*(4), 369–394. https://doi.org/10.1191/0265532202lt236oa

Brunfaut, T., & Révész, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly*, *49*(1), 141–168. https://doi.org/10.1002/tesq.168

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27*(2), 270–295.

Chang, A. C.-S., & Read, J. (2006). The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly*, *40*(2), 375–397.

Cohen, A. D., & Upton, T. A. (2007). `I want to go back to the text': Response strategies on the reading subtest of the new TOEFL®. *Language Testing*, *24*(2), 209–250. https://doi.org/10.1177/0265532207076364

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Council of Europe. (2009). *Manual for relating language examinations to the Common European Framework of Reference for Languages (CEFR)*. Language Policy Division.

de Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size.

Journal of Applied Statistics, *39*(4), 695–710. https://doi.org/10.1080/02664763.2011.610445

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Routledge.

Eberharter, K., & Frötscher, D. (2012). Quality control in marking open-ended listening and reading test items: Issues and practice. In D. Tsagari, S. Papadima-Sophocleous, & S. Ioannou-Georgiou (Eds.), *International experiences in language testing and assessment*. Peter Lang.

Eckes, T. (2015). *Introduction to Many Facet Rasch Measurement* (2nd ed.). Peter Lang.

Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal*, *89*(2), 206–220. https://doi.org/10.1111/j.1540-4781.2005.00275.x

Field, J. (2008). *Listening in the language classroom*. Cambridge University Press.

Field, J. (2013). Cognitive validity. In L. Taylor & A. Geranpayeh (Eds.), *Examining listening* (pp. 77–151). Cambridge University Press.

Field, J. (2015). *The effects of single and double play upon test outcomes and cognitive processing*. The British Council.

Fortune, A. J. (2004). *Testing listening comprehension in a foreign language – Does the number of times a text is heard affect performance ?* (Issue June). Unpublished MA dissertation: University of Lancaster.

Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, *34*(2), 165–182. https://doi.org/10.1016/j.system.2005.11.001

Green, R. (2017). *Designing listening tests: A practical approach*. Palgrave Macmillan. https://doi.org/10.1057/978-1-349-68771-8

Henning, G. (1991). *A study of the effects of variation of short-term memory load, reading response length, and processing hierarchy on TOEFL listening comprehension item performance (Educational Testing Service, RR 90-18).* Educational Testing Service.

Holzknecht, F., & Brunfaut, T. (2022). Individual difference factors for second language listening. In S. Li, P. Hiver, & M. Papi (Eds.), *The Routledge handbook of second language acquisition and individual differences* (pp. 331–346). Routledge. https://doi.org/10.4324/9781003270546-27

Holzknecht, F. (2019). *Double play in listening assessment*. Unpublished PhD dissertation, Lancaster University.

Holzknecht, F. (in preparation). Cognitive processing in single play and double play listening tasks.

Hubbard, P. (2017). Technologies for teaching and learning L2 listening. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 93–106). Wiley-Blackwell. https://doi.org/10.1002/9781118914069.ch7

Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning*, *16*(5), 413–425. https://doi.org/10.1076/call.16.5.413.29488

Iimura, H. (2007). The listening process: Effects of question types and repetition. *Language Education and Technology*, *44*, 75–85.

Jensen, E. D., & Vinther, T. (2003). Exact repetition as input enhancement in second language acquisition. *Language Learning*, *53*(3), 373–428. https://doi.org/10.1111/1467-9922.00230

Jones, G. (2011). *Research Summary: Once or twice? A critical review of current literature on the question how many times the audio recording should be played in listening comprehension testing items*. Pearson Education Limited.

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kwon, S. K., & Park, A. Y. (2017). The effect of double-play in L2 listening comprehension tests on test-takers' performance and performance appraisals. *English Language Teaching*, *29*(2), 27–49. https://doi.org/10.17936/pkelt.2017.29.2.2

Linacre, J. M. (1994). *Many-facet Rasch measurement* (2nd ed.). MESA Press.

Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, *75*(2), 196–204.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*(2), 283–305.

McNamara, T. (1996). *Measuring second language performance*. Longman.

McNamara, T., Knoch, U., & Fan, J. (2019). *Fairness, justice, and language assessment*. Oxford University Press.

Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749.

Osborne, J. W., & Costello, A. B. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, *10*(7), 1–9. https://doi.org/10.1.1.110.9154

Pallant, J. (2007). *SPSS survival manual* (3rd ed.). Open University Press.

Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, *35*(1), 31–65. https://doi.org/10.1017/S0272263112000678

Ruhm, R., Leitner-Jones, C., Kulmhofer, A., Kiefer, T., Mlakar, H., & Itzlinger-Bruneforth, U. (2016). Playing the recording once or twice: Effects on listening test performances.

International Journal of Listening, *30*(1–2), 67–83. https://doi.org/10.1080/10904018.2015.1104252

Sakai, H. (2009). Effect of repetition of exposure and proficiency level in L2 listening tests. *TESOL Quarterly*, *43*(2), 360–372.

Sherman, J. (1997). The effect of question preview in listening comprehension tests. *Language Testing*, *14*(2), 185–213. https://doi.org/10.1177/026553229701400204

Spöttl, C., Eberharter, K., Holzknecht, F., Kremmel, B., & Zehentner, M. (2018). Delivering reform in a high stakes context: From content-based assessment to communicative and competence-based assessment. In G. Sigott (Ed.), *Language testing in Austria: Taking stock (Sprachtesten in Österreich: Eine Bestandsaufnahme)* (pp. 219–240). Frankfurt: Peter Lang.

Sun, A., & Chen, X. (2016). Online education and its effective practice: A research review. *Journal of Information Technology Education: Research*, *15*(September 2015), 157–190. https://doi.org/10.28945/3502

Vandergrift, L. (1997). The comprehension strategies of second language (French) listeners: A descriptive study. *Foreign Language Annals*, *30*, 387–409.

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Meacognition in action*. Routledge. https://doi.org/10.4324/9780203843376

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning*, *60*(2), 470–497. https://doi.org/10.1111/j.1467-9922.2009.00559.x

Vogely, A. J. (1998). Listening comprehension anxiety: students' reported sources and solutions. *Foreign Language Annals*, *31*, 67–80. https://doi.org/http://dx.doi.org/10.1111/j.1944-9720.1998.tb01333.x

Vogt, W. (2007). *Quantitative research methods for professionals*. Pearson.

Winke, P., & Lim, H. (2014). The effects of testwiseness and test-taking anxiety on L2 listening test performance: A visual (eye-tracking) and attentional investigation. *IELTS Research Reports Online Series*, *3*, 1–30.

Wong, J. (2000). Repetition in conversation: A Look at 'first and second sayings'. *Research on Language and Social Interaction*, *33*, 407–424. https://doi.org/10.1207/S15327973RLSI3304_03