PREPARING LECTURE-STYLE TEXTS FOR UNIVERSITY LISTENING ASSESSMENTS

Nigel Dixon

ABSTRACT
Short lectures are an appropriate task for university listening assessments, but to produce quality tests, a number of issues should be considered at different stages of the test-preparation process. This article discusses the features of spoken language in lectures and outlines options for preparing texts which balance the need for authenticity with suitability for use as test material. Evidence for appropriate delivery speed for lectures is also considered. Finally, there is discussion of the advantages and drawbacks of different tasks that may be used to evaluate students' performance in lecture-based listening tests.

摘要
“（简）短讲座”很适合被用于大学听力测试。然而，要制作高质量的试题，应在考试准备的不同阶段对很多因素予以考虑。本文讨论了此类讲座的口语特点，并概述了在准备听力文本的过程中，如何兼顾此文本真实性与适用性。同时，本文也考虑了一些适合此类讲座的语速的例证。最后，文章讨论了在此类听力测试中，用于衡量学生听力水平的各类题型的利弊。

INTRODUCTION

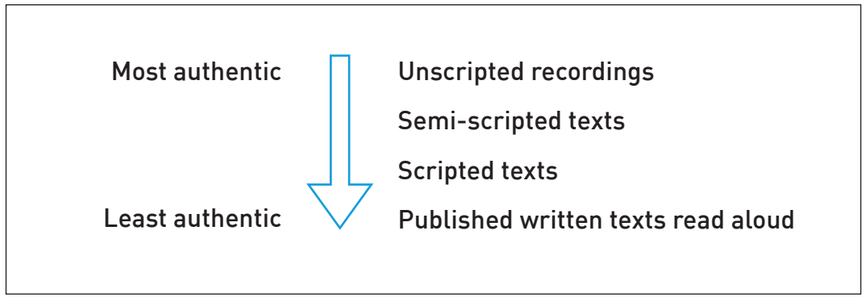
When preparing listening comprehension assessments in a university context, it makes sense to focus on the tasks that students need to perform in their academic studies. Although students will encounter interactional situations requiring them to listen and respond (Lynch, 2011), it is lectures that remain the main method of university teaching (Lee, 2009, p.42). It is generally accepted that listening to lectures places heavy demands on students' listening comprehension skills (Thompson, 2003, p.5), and thus, the ability to understand lectures is a key measure of students' second language competence (Flowerdew & Miller, 1992, p.60). Short lectures are therefore a logical choice of text for assessment of students' listening comprehension skills. However, for test writers, there are important considerations at different stages of test design. This paper will discuss issues related to text preparation and assessment tasks.

The Common European Framework of Reference of Languages (CEFR) B2 level is accepted as the minimum at which individuals can cope with university studies, while B1 level users may be admitted to pre-study programmes, such as foundation years prior to undergraduate courses. I will therefore make reference to the descriptors for B1 and B2 (Council of Europe, 2001) when discussing appropriate texts.

TEXT PREPARATION

There are several ways of generating lecture texts for listening exams, which can be placed on a continuum from most to least authentic (Figure 1). Rost (2005) argues in favour of using authentic samples of speech on the grounds of test validity: "To assess learners' listening ability, we need to focus on those aspects of proficiency and comprehension that are unique to listening" (Rost, 2005, p.170). In addition, as Flowerdew & Miller point out, features of spoken language in lectures also facilitate the task of comprehension (1997, p.34). These include "false starts, redundancies and repetitions", and short or incomplete clauses linked by pauses or simple conjunctions such as so and okay (1997, p.33). These features give listeners more time to process information, whilst the simpler grammatical structures are easier for them to parse in real time. By contrast, written articles commonly contain greater lexical density and use of embedded clauses than speech (Nesi, 2001), making them more difficult to understand if spoken, as well as sounding unnatural. This leads to the logical conclusion that reading an unaltered written text aloud as test input is unsatisfactory on the count of comprehensibility as well as authenticity.

On the other hand, the authentic option of using recordings of genuine lectures also has drawbacks. Firstly, such recordings may not be suitable for tests. Lectures publicly available on websites are generally too long to use in entirety within the time constraints of exams, and using short extracts creates unfair difficulties for listeners in orienting to the topic. In addition, public lectures are often extra-curricular events, with no intention that the audience should take notes or recall detailed information. They may also lack sufficient salient main ideas for test writers to use as the focus of comprehension ▷



▲
Figure 1: Lecture texts for listening assessments.

questions. However, some of these difficulties might be countered by editing or adding some scripted material to the transcripts of authentic lectures, and then re-recording them.

The third option of preparing scripted texts allows more control over text content, and skilled writers can reproduce at least some of the features of natural speech. This can be done by taking notes on written sources, making an outline, and then writing a final draft in spoken style. The writer can control the complexity of syntax, and may also add features to reduce difficulty, including selected repetition or restatement of main points (Flowerdew & Miller, 1996), and explicit signposting. The latter is necessary to support B2 level learners according to the CEFR descriptors (Council of Europe, 2001, p.8). Nonetheless, in my own experience, writing natural-sounding speech is very challenging, and even scripted materials in EAP textbooks have been criticized for lack of similarity to authentic lecture samples (Flowerdew & Miller, 1997; Thompson, 2003).
A final option is the semi-scripted lecture, in which a speaker improvises from an outline. This has the advantage of giving test-writers control over content while achieving features such as false starts and hesitations which are difficult to 'script'. However, although the technique is fairly well-established, examples in testing guidebooks such as Buck (2001) are limited to simple talks or dialogues on everyday topics rather than lectures, and commercial tests such as IELTS and TOEFL also still use scripted lectures. In a recent account, Clark (2014) describes the use of semi-scripted lectures in a college placement test, in which the volunteer speakers themselves selected topics and sources, and prepared outlines. The resulting lectures fulfilled the aim of creating natural, authentic-sounding speech, but five of the thirteen lectures were deemed unsuitable for the purpose for reasons of length, content, interest or suitability as a listening test (Clark, 2014, pp.9-10). I therefore suggest that given the relatively unproven nature of semi-scripted lectures, test-writers could consider trialling the technique with lower stakes classroom materials before attempting test recordings. For reasons of confidence, the speaker who records a semi-scripted lecture will ideally be experienced in giving presentations. Even then, the outline notes will need to be thorough, as speakers will probably lack the background knowledge to extemporize at length. Notes will likely be in bullet form, using indentation to distinguish key points and supporting examples and details. They should provide key words and phrases, but not complete sentences, in order to maximize the amount of natural spoken syntax. The test preparer may also indicate selected points which speakers should repeat or restate, and specify how much explicit signposting to use for bullet points. The latter two features may not be strictly consistent with natural lecture delivery in L1 settings. Thompson (2003, p.11), for example, found that explicit signposting is far less frequent in authentic lectures than in material typically recorded for EAP learners. Nonetheless, test-writers may consider signposting and some repetition necessary to achieve a degree of difficulty appropriate for L2 users at CEFR B1 or B2 levels. Finally, it may be useful to reassure speakers that natural hesitations are desirable and part of the 'authenticity' of the recording.

Even after taking the steps above, it seems unlikely that a perfectly usable version could be obtained at one attempt. In the article which first introduced the technique of semi-scripting material, Geddes & White therefore recommended recording two complete takes to begin with, and editing them together into the final product (1978, cited in Carr, 2011, p.86).

SPEED AND LENGTH

According to CEFR descriptors, students should be able to follow "standard spoken language" at B2 level, and "clearly articulated" speech at B1 level (Council of Europe, 2001, p.8). In other words, the listening input for the assessment should be at normal or near normal speed. In practice, this is not easy to define. To begin with, speaking speed may vary according to speech act and context. Table 1 compares rates of speech in four sets of data involving lectures and other recorded monologues, measured in words per minute (wpm). The rates for TED talks and radio monologues are at first sight noticeably faster than for lectures, but there is considerable variation in the range of speeds used by different speakers.

Speech event	No. of speakers	Mean wpm	Range	Source
TED talks	9	163	133-188	Dlugan (2012)
Scripted radio monologues	19	160.4	134-195	Tauroza & Allison (1990, p.98)
Lectures in the UK *	30	149.7	58-205	Nesi (2001, pp.207-208)
Lectures to non-native speakers**	22	141.7	102-199	Tauroza & Allison (1990, p.98)

Table 1: Comparison of different categories of monologue.

*averages for whole lectures, including periods of silence for reading calculations or writing
**based on samples of continuous delivery

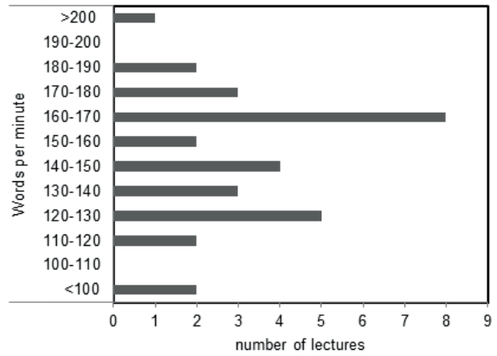


Figure 2: Distribution of BASE corpus lectures by speed (data from Nesi, 2001, pp.207-208)

Speed may be influenced by the purpose of the speaker. A breakdown of Nesi's data from the BASE corpus (2001) shows that lectures mainly range in speed from 110-190 wpm, with a large group falling between 160-170 wpm (Figure 2). However, Nesi notes that faster lectures were sometimes delivered by guest lecturers outside the main academic programme, or when the lecturer seemed not to expect notes to be taken. Nesi also notes examples of sharp changes of pace between sections of lectures containing anecdotes or background information, and sections where note-taking was expected. Although other lecturers in Nesi's sample still expected note-taking whilst speaking at a challenging pace, it would seem that 130-150 wpm could fall within the normal range for lectures where students are expected to identify and record key information.

Once the target speed for the examination lecture has been determined, this can be controlled both during and after the session when the lecture is formally recorded. The speaker can be given guidance about both overall speed and the length of pauses. Pauses should occur naturally at the ends of clauses or sentences, with longer pauses where there are changes of topic. The longer the pauses, the more time listeners have to process ideas. For scripted lectures, speed can also be adjusted by trialling an extract of 100 words or so, and calculating from this whether the speaker needs to go more slowly or quickly for the second take.

Lecture speed can also be adjusted by editing the recording using software such as Audacity®, which is freely downloadable1. The 'Tempo' function of Audacity® can be used to change the pace without distorting the pitch of speakers' voices, and pauses can also be lengthened by pasting in a quarter second of recorded silence. Buck claims tracks can thus be slowed by up to 20% without sounding unnatural (2001, p.186), but these are time-consuming measures which may be avoided by giving tactful feedback to speakers while recording.

The length of the lecture is a further consideration. This will depend on the number of passages included overall in the test and the required number of comprehension items. However, a test of candidates' ability to follow lectures in real life situations is likely to involve a relatively long passage compared with less demanding listening genres; Hughes considers ten minutes or more reasonable (2003, p.164).

TEST TASKS

Having created an appropriate text, test-writers need to decide on techniques and questions to measure the test-takers' ability to understand the lecture. As general rule, however, I would argue that candidates should be required to take notes during a lecture-based test rather than being shown questions before listening. H.D. Brown argues that assessment through note-taking has the advantage of authentically reproducing the experience of classroom lectures (2004, p.136). It is also a thorough measure of comprehension which allows candidates to demonstrate global understanding and ability to identify main arguments and key supporting information. By contrast, pre-questions may have the disadvantage of encouraging test-takers to listen selectively for details, which might be appropriate behaviour for administrative announcements, but not generally for lectures. The authenticity argument also suggests that the recording should be played only once, reproducing the experience of real life lectures. It should be acknowledged, however, that note- taking tasks will be unfamiliar to most students, so adequate classroom practice is essential to enable them to cope with the cognitive demands of simultaneously writing and listening (Carrell, 2007, p.45; Hughes, 2003, p.168).

Candidates' notes can be evaluated in three main ways: by scoring the raw notes, by setting tasks requiring reformulation or transfer of noted information, or through comprehension questions seen after hearing the lecture. There is evidence that raw notes reliably reflect candidates' ▷

listening competence. Song (2012) evaluated notes taken by candidates during an 8-minute lecture in comparison to their performance on short answer questions completed after listening. She found the raw notes were good indicators of competence, particularly when assessed in terms of candidates' ability to identify and record main topics (as opposed to minor details), and their ability to organize notes to show hierarchical relationships between ideas. Evaluating raw notes also allows candidates to be given credit for all key information they record, whereas comprehension questions may not cover every noteworthy point. In addition, if training in note-taking techniques has been a teaching focus, this could be factored into scoring. A test task developed by Kahn, for example, includes marks for visual lay-out and use of symbols and abbreviations as well as for accurate recording of main ideas (2002, cited in H.D. Brown, 2004, p.136).

The main drawbacks of assessing notes directly are increased marking time and reduced reliability in scoring (H.D. Brown, 2004, p.136ghes, 2003, p.168). Test-writers must prepare a full list of points to be identified in candidates' notes and the marks to be awarded, possibly including information meriting the award of partial credit. However, since it is virtually impossible to anticipate every acceptable wording of key points, marker subjectivity cannot entirely be avoided. It may not be a coincidence that the three examples of formal assessment of raw notes that I have encountered professionally were all scored by single markers to avoid inconsistency. Another argument against scoring notes directly is that note-taking methods and even the use of L1 are matters of individual freedom, and the comprehensibility of notes should not therefore be subject to external evaluation. However, if note-taking methods have been a focus of teaching, students should be aware of the need to record information clearly, and as Song points out, the seriousness of the test situation should motivate them to take notes relatively assiduously (2012, p74). In addition, students should be allowed to check and edit their notes before final submission, which would require no more time than completing alternative assessment tasks.

The second option, evaluating notes through a reformulation task, has the potential advantages of guiding students to present noted information in a clear format for a marker whilst also retaining the chance to gain credit for all information they have understood. Reformulation can take the form of an extensive writing task, such as a summary which can be scored according to inclusion of main points and/or supporting examples specified in the answer key. A more complex alternative is an integrated writing task requiring candidates to draw on both the lecture and some other input, typically a reading text. Examples include the University of Illinois English Placement Test (Cho, 2003) and the first writing task in the TOEFL iBT (ETS, 2015, pp.24-25). Extensive writing tasks are typically used to assess writing competence as well as ability to identify and use key information from the input. By contrast, candidates could also use their notes to complete an outline or a table, which could test identification of main ideas, but require only key words and phrases rather than a coherently written text. As with raw notes, reformulation tasks are relatively time-consuming and subjective to grade, and require preparation of a list of anticipated answers. This will be facilitated by piloting the test thoroughly with colleagues.

Thirdly, notes may be evaluated indirectly through comprehension questions which candidates see only after hearing the lecture and answer with reference to their notes. As Hughes argues, items can be scored more easily and reliably than raw notes (2003: p.168). However, when creating items, question focus must be considered with care, and Hughes emphasizes that questions must be "perfectly straightforward to anyone who has taken appropriate notes" (2003: p168). According to Weir (1993, cited by Rost, 2005, p.174), this is best achieved by basing the questions on a proficient set of notes taken while listening to the lecture, rather than on the full script. To do this, a rough recording can be made for a colleague to take notes from, with a second listening possibly permitted if the note-taker reports any losses of concentration which might occur without the 'authentic' pressure of being in a real examination. This would also allow incomplete phrases or abbreviations in the initial notes to be written more clearly for the test writer to refer to. Alternatively, the outline notes for a semi-scripted lecture could be used as a basis for question content, with revisions to items made after the actual recording if necessary.

The final decision concerns question style. Multiple Choice Questions (MCQs) are a popular method of assessing comprehension, and are easy to mark. However, they are beset by potential errors and traps for test-writers which result in unfair or unreliable items. In addition, they may inflate the grades of students who have better guessing techniques, or "test-wiseness" (Allan, 1992). Gap-filling items, in which sentences must be completed by one or two words, are easier to prepare. However, similarly to MCQs, there is a risk that answers may be deduced from contextual clues in the sentence, or simply from common sense. To help avoid this, a colleague can be asked to attempt the questions without knowledge of the text, to ensure that answers really are dependent on comprehension of the input. For a thorough discussion of pitfalls to avoid when preparing MCQs and gap-filling questions, including some helpful checklists for item-writers, see J.D Brown (2006).

CONCLUSION

Short lectures are an appropriate text type for academic listening assessments, but need to be prepared in such a way that they contain features of natural speech typical of this genre. Scripted, semi-scripted and genuine live lectures each have potential to achieve this, but test-preparers should take measures to counter the drawbacks of each of these approaches. Note-taking is recommended as the most authentic assessment task, provided candidates have received adequate training in this skill. Test-writers have three main options for evaluating candidates' performance, including scoring notes, different forms of transformation task, and post-listening comprehension items, but they should be aware of the positive and negative implications of each of these methods when deciding which to use in the test design. ○

REFERENCES

Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. Language Testing, 9(2), 101-119, doi:10.1177/026553229200900201

Brown, H.D. (2004). Language Assessment: Principles and classroom practices. White Plains, NY: Longman

Brown, J.D. (2006). Chapter 3: Developing good quality language test items. In J.D. Brown, Testing in Language Programs: A Comprehensive guide to English Language Assessment (pp. 41-65). Beijing, China: McGraw Hill / Higher Education Press

Buck, G. (2001). Assessing Listening. Cambridge, UK: Cambridge University Press.

Carr, N.T. (2011). Designing and analyzing language tests. Oxford, UK: Oxford University Press.

Carrell, P.L. (2007). Notetaking Strategies and their relationship to performance on Listening Comprehension and Communicative Assessment Tasks. ETS. Retrieved 7 April 2015 from http://www.ets.org/research/policy_research_reports/publications/report/2007/hsle

Cho, Y. (2003). Assessing writing: Are we bound by only one method? Assessing Writing, 8 (3), 165– 191, doi:10.1016/S1075-2935(03)00018-7

Clark, M. (2014). The use of semi-scripted speech in a listening placement test for university students. Papers in Language Testing and Assessment, 3(2), 1-26. Retrieved 8 April, 2015, from http://www.altaanz.org/uploads/5/9/0/8/5908292/clark.pdf

Council of Europe (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment: Structured overview of all CEFR scales. Retrieved 7 April, 2015 from: http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Key_reference/Overview_CEFRscales_EN.pdf

Dlugan, A. (2012). What is the average speaking rate? Retrieved 7 April, 2015 from http://sixminutes.dlugan.com/speaking-rate/

ETS (2015). TOEFL iBT® Test Questions (PDF). Retrieved 20 June, 2015 from http://www.ets.org/toefl/ibt/prepare/sample_questions

Flowerdew, J. & Miller, L. (1992). Student perceptions, problems and strategies in second language lecture comprehension. RELC Journal, 23(2), 60-80, doi:10.1177/003368829202300205

Flowerdew, J. & Miller, L. (1996). Lectures in a second language: Notes towards a cultural grammar. English for Specific Purposes, 15(2), 121-140, doi:10.1016/0889-4906(96)00001-4

Flowerdew, J & Miller, L. (1997). The teaching of academic listening comprehension and the question of authenticity. English for Specific Purposes 16(1), 27-46 doi:10.1016/S0889-4906(96)00030-0

Hughes, A. (2003). Testing for language teachers. (2nd Edition) Cambridge, UK: Cambridge University Press

Lee, J.J. (2009). Size matters: an exploratory comparison of small- and large-class university lecture introductions. English for Specific Purposes 28(1), 42-57 doi:10.1016/j.esp.2008.11.001

Lynch, T. (2011). Academic listening in the 21st century: Revising a decade of research. Journal of English for Academic Purposes, 10(2), 79-88, doi:10.1016/j.jeap.2011.03.001

Nesi, H. (2001). A corpus-based analysis of academic lectures across disciplines. In J. Cotterill, J. & A. Ife (Eds.), Language across boundaries (pp. 201-218). London, UK: Continuum.

Rost, M. (2005). Teaching and researching listening. Beijing, China: Foreign Language Teaching and Research Press.

Song, M-Y. (2012). Note-taking quality and performance on an L2 academic listening test. Language Testing, 29(1), 67-89, doi:10.1177/0265532211415379

Tauroza, S. & Allison, D. (1990). Speech rates in British English. Applied Linguistics, 11(1), 90-105. doi:10.1093/applin/11.1.90

Thompson, S.E. (2003). Text-structuring metadiscourse, intonation and the signalling of organization in academic lectures. Journal of English for Academic Purposes 2(1), 5-20, doi:10.1016/S1475-1585(02)00036-X

[1]Audacity® can be obtained from http://audacity.cn.uptodown.com/ (Chinese) or http://audacity.sourceforge.net/download/ (English)

Author Biography

Nigel Dixon is an EAP Tutor and former Examinations Officer at XJTLU. He has worked on test preparation and taught English in China, Indonesia, Turkey, Thailand and the UK, and has recently gained experience as a language test candidate in Chinese HSK exams.

Email:
nigel.dixon@xjtlu.edu.cn