

Listening Test Development Skills of IEP Instructors in the US:
Job Training Received and Desired

Valeria Bogorevich

Northern Arizona University

Abstract

The purpose of this study was to develop a Listening Test Development Skills Questionnaire (LTDSQ) enquiring about the on the job training that IEP instructors received, and if teachers have a desire for more on the job training. The current study focused on the ESL teachers who are employed at IEPs in the United States as well as on one particular skill of listening assessment for the classroom needs. In addition, it focused specifically on the process of test development and not on the issues of deciding what to test, test administration, washback or score interpretation. The results showed that only 33% (8 out of 24 people) received training for all outlined topics in all sub-areas of LTDQ. In terms of sub-areas, the average number of people who received training for Features of Passages was 74%, Item Writing 82%, Development and Analysis 81%. The numbers revealed that most of the participants evaluated the received training as fairly useful and that further training is fairly desired.

Keywords: IEP instructors, listening tests, on the job training, test development

Background

Tests are an important part of any teacher's job, they are the tools to collect evidence of students' learning (Gareis & Grant, 2015). There has been a lot of attention to the overall and language assessment literacy of school teachers recently (e.g. McGee & Colby, 2014; DeLuca & Klinger, 2010; Fulcher, 2012), but there has been less attention to assessment literacy of ESL/EFL (English as a Second/Foreign Language) teachers (e.g. Vogt & Tsagari, 2014). L2 (Second Language) English teachers can use standardized tests and tests from textbooks, but not always those types of assessment coincide with what was the focus on in the classroom, and teachers have to use their own tests. It is important for ESL teachers to be literate in test development to assess their students' knowledge accurately; therefore, a course in language assessment is a required one in most Master's programs in Teaching English as a Second Language. This fact gives us information that ESL teachers should be skilled in test development because of the presence of training in classes; however, it is important to sustain these skills by providing on the job opportunities for further professional development in this area. Likewise, courses on language assessment usually cover the fundamentals of assessment for each skill: Writing, Reading, Listening and Speaking, but most probably they do not cover test development for each skill in depth. It is important to know if ESL instructors receive any further in depth on the job training in testing of each skill.

The current study focused on the ESL teachers who are employed at IEPs (Intensive English Programs) in the United States as well as on one particular skill of listening assessment for the classroom needs. In addition, it focused specifically on the process of test development and not on the issues of deciding what to test, test administration, washback, or score interpretation. The purpose of this study was to develop a Listening Test Development Skills

Questionnaire (LTDSQ) enquiring about the on the job training that IEP instructors received, and if teachers have a desire for more on the job training.

Skills Needed for Listening Test Development

Assessment of listening skills is considered to be one of the most difficult tasks for teachers and test-developers because it is least developed and understood. Testing listening comprehension is far more time consuming and a more complicated process than testing reading comprehension (Buck, 2001). It can be inferred from this statement that IEP instructors would also consider listening assessment to be a complicated task. In this section an attempt is made to describe the skills that instructors should have in order to develop listening tests for the classroom needs.

To develop a listening test, two types of skills may be needed: General and specific to listening. The general test development skills involve skills that may be applied to different tests such as ability to apply the knowledge of test concepts, for example, test validity and reliability. The specific to listening test development skills involve the ability to apply knowledge about the listening construct to choose level-appropriate test listening passages. It was decided that the LTDSQ will engulf both types, but with more focus of the specific skills.

Research Questions

The literature reviewed served as a basis for creating statements for LTDSQ. The proposed questions for the future study using LTDSQ were answered:

1. How much training was received by the IEP instructors?
2. What is the perceived level of usefulness of the received training in each sub-area?
3. How much is further on the job training desired by IEP instructors?
4. How often are IEP teachers involved in test development?

5. Do IEP teachers perform peer review of their tests before test administration?
6. What is the level of test development support at IEPs?
7. How confident do teachers feel in listening test development?
8. To what extent do teachers do self-training on test development?
9. Is there a relationship between the confidence of IEP teachers in listening test development and their desire for further training, level of job support, the frequency of test-making, and peer reviews?
10. What is the origin of listening test development training?

Method

Participants

The sample consisted of current instructors and teaching assistants at a Southwestern IEP and volunteer participants at other American IEPs who used to work at the IEP mentioned above. The total number of complete responses was 24. Among the 24 respondents, 71% were female, which is representative of an average IEP. The age of the respondents ranged from 18 to 54 and was spread out this way: 4% were 18-24, 46% – 25-34, 37.5% – 35-44, and 12.5% – 45-54. In terms of degree, 54% either had a Master's degree in hand or were working on it, and 46% either had a PhD degree or were working on it. In terms of teaching experience, 23 out of 24 participants had experience of teaching all four skills: Listening, speaking, writing, and reading. Their experience of classroom teaching was also longer than two years. To provide further details, 25% had taught for 3 to 4 years, 21% for 5-6 years, 8% for 7-8 years, and 46% for 9 and more years. The current participants perceived themselves as confident in listening test development ($M = 4.21$) on a scale from 1 to 6. They described their frequency of test making as fairly often ($M = 3.67$) also on a scale from 1 to 6.

The measure of the study was a questionnaire developed by the researcher. The 19 statements in Part 1 and Part 2 covered three main areas of research interest, namely features of listening passages (items 1 through 8), item writing (items 9 through 13), and test analysis and development (items 14 through 19). These items constituted the three multi-item scales with 8, 5 and 6 items respectively (see Appendix). The survey was administered online through SurveyMonkey (<https://www.surveymonkey.com>). Both Cronbach's alphas for the Part 1 and Part 2 were high at .96 and .95 showing good psychometric properties. Cronbach's alphas for the multi-item scales ranged from .80 to .91 showing an acceptable level of reliability.

Results

How much training was received by the IEP instructors?

To answer this research question, a descriptive statistic table of all the items from Part 1 was used. It showed that only 33% (8 out of 24 people) received training for all outlined topics in all sub-areas. One person indicated no training for all outlined topics. In terms of sub-areas, the average number of people who received some training for Features of Passages was 17.75 (74%), Item Writing 19.60 (82%), Development and Analysis 19.33 (81%).

For Features of Passages, they were how to ensure level-appropriate grammatical difficulty of listening test passages (63%) and use of accented speech (58%). For Item Writing, only one topic had a low number of participants who indicated that they received training on that, which was the use of transcripts for item-writing (54%). For Development and Analysis, the lowest numbers were for topics of reliability (71%) and validity with 75%.

What is the level of perceived level of usefulness of the received training in each sub-area?

The numbers revealed that most of the participants evaluated the received training as fairly useful, around $M = 4.5$ on a 6-point scale (see Table 3). The differences of the usefulness

of training in each sub-area were not significantly different. The sub-area Features of Passages had the lowest mean of 4.37 and Item Writing had the highest mean of 4.75.

Table 3.

The average usefulness of overall training and training in each sub-area

	Minimum	Maximum	<i>M</i>	<i>SD</i>	Error
Overall	2.86	5.58	4.58	.82	.17
Features of Passages	2.75	5.50	4.37	.84	.18
Item Writing	2.80	6.00	4.75	.90	.19
Development and Analysis	3.00	6.00	4.67	.94	.20

**Note.* N = 23 as one person indicated no training in any topics

How much is further on the job training desired by IEP instructors?

The descriptive statistics showed that further training is fairly desired, around 4 on a 6-point scale (see Table 4). The differences in the desire for more training in each sub-area were not significantly different. The sub-area Item Writing had the lowest mean of 3.93 and Development and Analysis had the highest mean of 4.13.

Table 4.

The desire for further training overall and for each sub-area

	Minimum	Maximum	<i>M</i>	<i>SD</i>	Error
Overall	1.84	6.00	4.06	1.12	.23
Features of Passages	1.75	6.00	4.09	1.10	.23
Item Writing	1.40	6.00	3.93	1.30	.27
Development and Analysis	1.00	6.00	4.13	1.32	.27

**Note.* N = 24

In terms of individual topics, the highest desire was indicated for the ways of ensuring validity of listening tests ($M = 4.75$) and the use of different formats to test listening such as multiple-choice or summaries ($M = 4.54$). The lowest desire was indicated for the use of transcripts ($M = 3.46$) and writing detail questions ($M = 3.58$). It was interesting to see that the respondents indicated almost no training for the use of transcripts for item writing, and at the same time they did not express any desire to learn more about this topic.

Results of analysis of background questions

The results of six single-item questions on a 6-point semantic differential scale are provided in Table 5. The participants reported that they make listening tests quite often ($M = 3.67$). Peer reviews of listening tests before the administration are conducted fairly often ($M = 4.17$). The level of test development support at the participants’ job is also fairly high ($M = 4.67$). The participants’ confidence in developing listening tests was also fairly high ($M = 4.21$). The mean of the question asking about self-training was the lowest ($M = 3.04$) indicating the medium level of self-training done by the participants.

Table 5.

Six background questions

	Minimum	Maximum	<i>M</i>	<i>SD</i>	Error
How often	1	6	3.67	1.34	.27
Peer review	1	6	4.17	1.95	.40
Job support	1	6	4.67	1.71	.35
Confidence	2	6	4.21	1.14	.23
Self-training	1	6	3.04	1.55	.32

*Note. N = 24

Some questions from this section were used to answer the research questions asking if there is a relationship between the confidence in listening test development and desire for further training, level of job support, the frequency of test-making, and peer reviews. The correlations are provided in Table 6. The analysis revealed a significant negative correlation between the desire for further training and the confidence in listening test development. There were significant positive correlations between the confidence and the three other variables.

Table 6.

Correlations

	Desire	How often	Job support	Peer review
Confidence (r)	-.62**	.62**	.65**	.44*
Confidence (r ²)	.38	.38	.42	.19

*Correlation is significant at .05

**Correlation is significant at .01

The last research question was to provide the descriptions of the origin of training that was received by the participants. The most participants indicated receiving the training mostly on the job, partly on the job, and partly in classes (Figure 1). This finding corroborates the statement by Vogt and Tsagari (2014) that language teachers learn about language assessment on the job trying to compensate for the insufficient training in classes.



Figure 1. The origin of listening test development training.

Relevance to the PIE and Second Language Testing

The results of the pilot study can be used to make some conclusions about the sample of participants only and not the population of IEP instructors in the US.

According to the results of the study, there is a need for more training in the area of Features of Passages in general as fewer participants received training in it, especially on how to ensure level-appropriate grammatical difficulty of listening test passages and the use of accented speech. The results did not show the need of more training in the area of Development and Analysis because most of the participants received training in those areas. Nevertheless, the respondents indicated a desire to learn more this area, especially about the validity of listening tests. Probably, because the concept of validity is complicated and not quite clear; therefore, the participants indicated a desire for more training. In terms of the use of transcripts for items writing, the respondents indicated almost no training, but did not express any desire to learn more about this topic. It might mean that either the respondents learnt about this topic themselves or do not consider this topic important. Overall, the participants indicated a fairly high desire for more training, which means that they would like to learn more about listening test development

practices and issues in a form of further on the job training. This result goes along with the results in Vogt and Tsagari (2014) where foreign language teachers expressed the need in receiving further training in many different areas of language testing.

In addition, there is a negative relationship between the confidence and desire for more training. This might indicate that the instructors who have confidence in listening test development, have enough training. The positive relationship between the confidence and the frequency of test-making might indicate that teachers who make tests more often have more experience in test-making. The positive relationship between confidence and job support may show that the more job support is present, the more confident the teachers are in listening test development. The positive relationship between confidence and peer review practices might mean that more confident test-makers are more liable to follow the good practice of test development steps of subjecting their tests for peer review.

The indicated origins of training show that not many respondents learnt most of the information in classes. This should be considered by the program administrators so that more opportunities for hands-on experience can be provided during TESL training. In addition, the results showed that most of the training was received either mostly on the job or partly on the job indicating the importance of on the job experience and training. Taking into account the moderate desire for more on the job training, IEP's administrators should keep providing the opportunities for more training in listening test development.

References

- Buck, G. (2001). *Assessing listening*. Cambridge, UK: Cambridge University Press.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17, 419-438.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9, 113-132.
- Gareis, C., & Grant, L. W. (2015). *Teacher-made assessments: How to connect curriculum, instruction, and student learning*. Routledge.
- McGee, J., & Colby, S. (2014). Impact of an assessment course on teacher candidates' assessment literacy. *Action in Teacher Education*, 36, 522-532.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: findings of a European study. *Language Assessment Quarterly*, 11, 374-402.

Appendix

Listening Test Development Skills Questionnaire (LTDSQ)

Look at the skills for listening comprehension test development. Rate how **USEFUL** your current/past **ON THE JOB TRAINING** has been. Choose **N/A** if you have never had on the job training on any of these topics.

On the job training is **defined** as an opportunity to learn more about a specific skill for listening comprehension test development **at work** (e.g. professional development sessions, online webinars, meetings with assessment professionals, training sessions, etc.)

1. How to ensure level-appropriate <i>grammatical difficulty</i> of listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
2. How to ensure level-appropriate <i>vocabulary</i> of listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
3. How to ensure level-appropriate <i>length</i> of listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
4. How to ensure level-appropriate <i>speech rate</i> of listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
5. How to ensure level-appropriate <i>number of speakers</i> in listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
6. How to ensure <i>authenticity</i> of listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
7. The use of <i>accented speech</i> (e.g. Australian, Indian) in listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
8. The use of <i>videos</i> for listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
9. The use of <i>transcripts</i> for question writing.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
10. How to choose the <i>format</i> of listening test questions (e.g. multiple-choice, open-ended questions, summaries, etc.)						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
11. How to write <i>main idea</i> questions for listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
12. How to write <i>detail</i> questions for listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
13. How to write <i>inference</i> questions for listening test passages.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
14. How to identify <i>difficult/easy</i> test questions.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
15. How to calculate test <i>reliability</i> (consistency).						
Useless					Highly Useful	no job training on this

1	2	3	4	5	6	N/A
16. How to ensure test <i>validity</i> (how accurately it measures listening comprehension).						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
17. How to identify <i>ineffective distractors</i> (wrong answers in multiple-choice tests).						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
18. <i>Steps</i> of test development.						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
19. How to develop tests for different <i>purposes</i> (e.g. achievement, proficiency, placement).						
Useless					Highly Useful	no job training on this
1	2	3	4	5	6	N/A
20. Any other on the job training you have had on developing listening tests? Any suggestions for additional questions or deleting/ clarifying questions?						

Look at the skills for listening comprehension test development.

Rate your **DESIRE FOR MORE TRAINING** (e.g. professional development sessions, online webinars, meetings with assessment professionals, training sessions, etc.) on the following skills for listening test development.

1. How to ensure level-appropriate <i>grammatical difficulty</i> of listening test passages.						
Not Desired					Highly Desired	
1	2	3	4	5	6	
2. How to ensure level-appropriate <i>vocabulary</i> of listening test passages.						
Not Desired					Highly Desired	
1	2	3	4	5	6	
3. How to ensure level-appropriate <i>length</i> of listening test passages.						
Not Desired					Highly Desired	
1	2	3	4	5	6	
4. How to ensure level-appropriate <i>speech rate</i> of listening test passages.						
Not Desired					Highly Desired	
1	2	3	4	5	6	
5. How to ensure level-appropriate <i>number of speakers</i> in listening test passages.						
Not Desired					Highly Desired	
1	2	3	4	5	6	
6. How to ensure <i>authenticity</i> of listening test passages.						
Not Desired					Highly Desired	
1	2	3	4	5	6	
7. The use of <i>accented speech</i> (e.g. Australian, Indian) in listening test passages.						
Not Desired					Highly Desired	
1	2	3	4	5	6	
8. The use of <i>videos</i> for listening test passages.						
Not Desired					Highly Desired	
1	2	3	4	5	6	
9. The use of <i>transcripts</i> for question writing.						
Not Desired					Highly Desired	
1	2	3	4	5	6	
10. How to choose the <i>format</i> of listening test questions (e.g. multiple-choice, open-ended questions, summaries, etc.)						
Not Desired					Highly Desired	
1	2	3	4	5	6	
11. How to write <i>main idea</i> questions for listening test passages.						
Not Desired					Highly Desired	
1	2	3	4	5	6	

12. How to write <i>detail</i> questions for listening test passages. Not Desired 1 2 3 4 5 6 Highly Desired
13. How to write <i>inference</i> questions for listening test passages. Not Desired 1 2 3 4 5 6 Highly Desired
14. How to identify <i>difficult/easy</i> test questions. Not Desired 1 2 3 4 5 6 Highly Desired
15. How to calculate test <i>reliability</i> (consistency). Not Desired 1 2 3 4 5 6 Highly Desired
16. How to ensure test <i>validity</i> (how accurately it measures listening comprehension). Not Desired 1 2 3 4 5 6 Highly Desired
17. How to identify <i>ineffective distractors</i> (wrong answers in multiple-choice tests). Not Desired 1 2 3 4 5 6 Highly Desired
18. <i>Steps</i> of test development. Not Desired 1 2 3 4 5 6 Highly Desired
19. How to develop tests for different <i>purposes</i> (e.g. achievement, proficiency, placement). Not Desired 1 2 3 4 5 6 Highly Desired
20. Any other on the job training you would like to have on developing listening tests? Any suggestions for additional questions or deleting/ clarifying questions?

1. What is your age? 18-24 25-34 35-44 45-54 55 or more
2. What is your gender? Male Female
3. Choose all the degrees you hold OR studying for: Bachelor's Certificate Master's EdD PhD
4. Is your Intensive English Program is accredited by CEA? Yes No Don't Know I don't work for an Intensive English Program
5. How many years of classroom teaching experience do you have? 1-2 3-4 5-6 7-8 more than 8
6. What language skills have you taught?

Listening Speaking Reading Writing
7. How often do you make listening tests? Never Very Often 1 2 3 4 5 6
8. How often do your peers review the listening test you made before its administration? Never Always 1 2 3 4 5 6
9. What is the level of test development support at your job? No Support Plenty of support 1 2 3 4 5 6
10. How confident are you in developing listening tests? Not Confident Highly Confident 1 2 3 4 5 6
11. To what extent do you do self-training on test development? None Only Self-Training 1 2 3 4 5 6
12. Where did you receive your test development training? Mostly on the job Partly in the job Partly in classes Mostly in classes I haven't received any training Mostly self-training Partly self-straining
13. Any suggestions for additional questions or deleting/ clarifying questions?