

A Description and Evaluation of the TOEIC LPI

Kelly E. QUINN

Introduction

The TOEIC test was created by the Educational Testing Service (ETS), the same organization which makes the TOEFL and SAT tests. The TOEIC came from the Japanese Ministry of International Trade and Industry requests to the Educational Testing Service in the middle 1970's. It is "designed to measure the English-language listening comprehension and reading skills of individuals whose native language is not English. The TOEIC is used primarily by corporate clients, worldwide." (Wilson, 1989). Most examinees are in their mid-twenties to late forties, and working for a corporation. However, TOEIC test-takers have recently included many university graduating seniors, because corporations are requiring TOEIC scores for new employees more and more often. From its beginning nearly 20 years ago, the use of TOEIC has spread from Japan throughout Asia, and it is becoming more frequently used throughout Europe and South America.

In fact, since the TOEIC test was first administered in Japan, there has been a dramatic growth in the number of people taking the test. The number of people who took the test in 1995 was more than 560,000. More than 1,6000 corporations in a many different business fields use the TOEIC test.

The makers of the TOEIC claim that it shows the differences between low-beginner-to high-intermediate levels very well. Each company or organization has its own criteria for hiring, but the following scores can be considered typical for the employees described. Japanese companies that deal with English material on frequent basis, like the pharmaceutical industry or engineering firms, often consider a TOEIC score of 450 is acceptable for hiring practices. A TOEIC score of 600 is frequently considered the minimum acceptable for working overseas. Domestically-based engineers who have a TOEIC score of 500 are considered reasonably proficient in English. Reasonably proficient means able to read and write faxes and understands informal literature in his or her field in English. If the same engineer is being considered for a posting overseas, he or she must usually try for a TOEIC score of about 625. A domestically-based desk-worker with a TOEIC score of 600 is considered reasonably proficient in English. For the same desk-worker to go overseas, she or he must usually have a TOEIC score of 685.

In a "History and Status" report prepared by the TOEIC Steering Committee in 1996, TOEIC claims, "From the results of validity studies, based on the examinees scores, we can predict on the average how that person should be able to perform in English, and what range of competence that person has been able to attain in each of the language skills." However, in the "Language Proficiency Manual" published in 1996 and used to train interviewers, the writers claim, "While

some candidates may perform well in English classes or on standardized paper-and-pencil tests, these accomplishments do not guarantee adequate ability to speak English in a real life situation." Perhaps in order to reconcile this disparity, even within their own literature, TOEIC offers examinees who have scored 630 or higher on the written TOEIC test the opportunity to take a proficiency test, the Language Proficiency Interview (LPI), which directly measures their speaking ability. This test is a structured interview, lasting between 15 - 30 minutes. The purpose of the interview is to give the candidate the opportunity to demonstrate, in a realistic and natural conversational situation, the level at which he or she can speak English. The structure and method of test administration will be described in this paper and at the end the strong and weak points of the interview will be evaluated. Much of the material in the next section, that which deals with the history of the interview and its structure is taken from "Language Interview Manual".

Comparing Oral Interviews to Standardized Tests

It is possible to separate foreign language tests into the two broad categories of achievement tests and proficiency tests, according to the particular kinds of information that the instruments and procedures are intended to provide.

The purpose of language achievement testing is to determine students acquisition of various specified aspects of course content, eventually after training or instruction has been received. Achievement tests can range in scope from short quizzes to chapter tests to final examinations covering the content of a whole course. Because the background of material in an achievement test is clearly defined and limited to the content of a curriculum, it is possible to claim a perfect score. Those who have mastered the material to the same degree will receive the same score.

The distinguishing characteristic of all achievement tests is that they are based on, and respect, specified, preordained elements to which the student has been exposed in the course of the teaching program. A strict achievement test never contains questions on material that has not been covered. Results on achievement tests usually are expressed as percentages of correct answers.

The oral interview is not an achievement test, but a proficiency test. Language proficiency tests are designed to measure a candidate's ability in a language regardless of the type of training they may have had in that language. The purpose of language proficiency testing is to assess the candidate's ability to use the language effectively and appropriately in real-life situations. In contrast to achievement testing, in which the test material are based on the content of a specified curriculum, proficiency testing focuses on overall language competence without regard to the place, length of time, or manner in which that competence has been acquired.

Since a proficiency test does not cover a specified curriculum, it is not possible to anticipate what specific questions will be asked. An oral proficiency test such as the LPI will test everything an individual knows about how to use the language by sampling higher speech production on a variety of topics at different of levels. An individual can get a "perfect score" on the LPI oral proficiency test only by demonstrating speech production equivalent to that of a well-educated native speaker of the target language.

In a language proficiency test, candidates will always be asked questions for which they have not specifically prepared. The interviewer tries to systematically sample the candidate's language

and to elicit the highest level possible. Questions that are relatively difficult for the candidate are intentionally asked in order to find the candidate's upper limit.

Reliability and Validity in Oral Testing

Multiple choice tests grew in popularity in the United States after World War II, because of their high reliability and ease of administration. Reliability means score stability in a test/retest situation. Written essays or taped speech samples, were rejected on scientific grounds as unreliable. It was thought that human beings simply do not measure the same written or spoken sample as consistently as a computer can scan multiple choice answer sheets.

In developing or choosing a testing instrument, validity must be considered together with reliability. The oral interview was originally developed because the more reliable paper-and-pencil tests were not considered as valid as a direct measure of oral production. In this context a valid instrument means one that actually tests what it purports to test. The oral proficiency interview is regarded as having high face validity. Face validity, also a characteristic of the oral interview, concerns whether a test has the appearance of doing what it purports to. It is also important that the test seem valid to the candidate. A well structured oral proficiency interview assesses speaking ability in a real-life language context conversation.

Content validity refers to the extent to which the content of the test includes representative samples of the domain to be measured. The oral proficiency interview maintains content validity by including in each interview questions that test language functions and real life situations. L2 content validity should also be thought of as stability or consistency of content from one test to another the oral interview does not contain a series of specific questions; the topics and the questions asked vary from one interview to another. Well-trained interviewers, however, will administer interviews that can be thought of as parallel forms of the same testing procedure. Just as though the questions and the topics may differ, and indeed should differ, from test to test, the question topics remain the same. For example, an interviewer might ask a hypothetical question, such as, "If you could choose a new career, what profession would you like to have?" Rather than repeat the topic and risk test compromise, an interviewer could later use the same question type with a different topic: "If you were starting college now, what course of study would you undertake?" This strategy avoids test compromise while maintaining content validity.

If different interviews were not monitoring the same linguistic behaviors, content validity would indeed be weakened. However, workshops that train potential interviewers in elicitation techniques and provide them with a common understanding of the standards serve to strengthen the content validity as well as the reliability of the interview.

Structure of the Interview

Every oral interview follows the same general structure. The general structure guides the interviewer by directing higher attention to certain mandatory aspects of the test. An interview is generally conducted in four phases: Warm-Up, Level Check, Probes, and Wind-Down. Level Check and Probes make more time than the Warm-Up and the Wind-Down. At the very lowest level, the limitations of the candidate's language may be such that the four phases will be indistinguishable

from each other- At the very highest level, only a brief Warm-Up and Wind-Down will be necessary unless the candidate has not been speaking the language recently.

The Warm-Up

The Warm-Up consists of social amenities and simple conversation at a level that is easy for the candidate. (At the lowest levels this may not be possible.) There are three purposes to this phase of the interview: (1) to put the candidate at ease; (2) to reacquaint the candidate with the language if necessary; and (3) to give the interviewer a preliminary indication of the candidate's level.

For candidates, the main purpose of the Warm-Up is to put them at ease with the testing situation and to reintroduce them to the language. The length of the Warm-Up will depend on the circumstances. Candidates who have not spoken the language for some time may need to get back into it gradually, while with others they may immediately shift the conversation to a higher level. Interviewers should never skip this phase, but may shorten it considerably if the candidate does not seem to need it.

One quick way to begin the Warm-Up is for the interviewer to introduce himself or herself to the candidate in the target language—since introductions are usually learned early in foreign language classes it is easy for most candidates to respond, opening the way for further conversational exchanges.

For the interviewer, the Warm-Up serves the important function of giving a preliminary indication of the candidate's level. This preliminary indication must be confirmed, because many candidates answer questions at the level and in the style in which they are asked. The best approach is for an interviewer to assume that the preliminary indication checked in the next phase, the Level Check. In fact, the rest of the interview will be devoted to ascertaining whether or not this preliminary indication is accurate.

The Level Check

The purpose of the level check is to find the candidate's highest sustainable level of speaking performance. To find the level, the interviewer must test the breadth and depth of the candidate's ability in the language. How fluent is the candidate? How well does he/she pronounce the language? How accurate is the grammar? How wide is the vocabulary? How correct is the syntax? How native is the expression of ideas and concepts in the language?

Sometimes the level indication given by the Warm-Up is misleading, and the interviewer can begin the Level Check too low or too high. If the test begins at too low a level, the interviewer can simply raise the level of the questions and begin the Level Check over again. If the test begins at too high a level, the interviewer must bring the level down. Starting at too high a level is to be avoided, since bringing the level of an interview down is difficult to do without giving the candidate a sense of failure. In the Level Check interviewers should check a number of topics (both interest and non-interest areas) to see if the candidate can perform consistently at the level in question. Can the candidate accomplish the functions with suitable content and accuracy? When the candidate successfully passes the Level Check, his/her performance provides a bottom to the rating. The next phase aims at finding the ceiling.

The Probes

Intertwined with the Level Check phase is the Probes phase. The purpose of this phase is to make sure that the level the interviewer has been checking is the candidate's highest sustainable level. To probe, the interviewer should take the candidate above the previous level several times by introducing tasks and subject areas associated with the next higher level. If this phase is successful, every candidate should leave the testing room feeling that he/she has been tested to the limit of his/her ability.

The name of this phase is purposely in the plural, because there should be two or three probes. Probes should furnish clear examples of linguistic breakdown. "Sometimes the candidate actually tells the interviewer that the limit has been reached by saying, I don't know how to say that in your language," or "I know what I want to say but I can't say it." In other cases, a sharp drop in fluency, a sudden groping for words or a dramatic increase in grammatical errors gives evidence of the linguistic breakdown.

If the interviewer has carried out the Level Check at too low a level, the candidate may be able to respond to the Probes consistently well. If this happens, the interviewer must begin the process of Level Check and Probes over again at a higher level and continue until the ceiling of the candidate's proficiency is established.

While the Level Check gives evidence of what candidates can do, the Probes show what candidates cannot do. Without this phase of the interview, candidates may appear to be more proficient than they really are. The Probes allow an interviewer to explain why a candidate's speech is not at a higher level, providing diagnostic information with specific examples.

Experienced interviewers learn how to interweave the Level Check and Probes, so that the candidate is allowed to return to a level where performance can be sustained between higher-level questions.

The Wind-Down

The purpose of this phase is to leave candidates with a feeling of accomplishment after stretching their speaking ability to the limit. It is also the interviewer's last chance to check any aspect of the candidate's speaking ability that may still be incompletely assessed. Normally, the Wind-Down should return to the highest Level that the candidate was able to sustain during the interview. It may even be helpful, particularly at the lowest levels, to end the test by returning briefly to a topic discussed previously.

Description of the Levels

The LPI interview rates candidates from Level 0 to Level 5. Level 0 is the lowest possible score and represents zero proficiency in the target language. Level 0 candidates are unable to talk about any topics or subject areas and are unintelligible. Level 5 is the highest score and represents language ability equal to an educated adult native speaker.

Level 1 candidates can create with the language, ask and answer questions, and participate in short conversations. During the interview candidates at this level are often asked to ask the interviewer questions and to do a role play that involves an everyday survival situation. Survival

situation means, a common occurring daily activity that one would very likely encounter when traveling or living abroad. Common survival situations include, making a business appointment, asking for a camera to be repaired, visiting a dry cleaners, making hotel reservations. All of these situations require the candidate to give basic biographical information such as their name, address, a phone number where they can be reached, etc. It is important to note that role plays at level 1 include no complications. Time for appointments is always available, parts are never out of stock, all stains can be removed, and hotels are never completely booked, at level 1.

Level 2 candidates are able to participate in casual conversations. They can express facts, give instructions, describe, report and give narrative accounts of current, past, and future activities. During the interview candidates at this level are often asked by the interviewer to explain how to use a machine from their workplace — such as a copy or fax machine. Alternately, they can explain how to travel place to place, usually from the interview site to their home or they can explain how to cook a simple meal. It is important that the candidate know how to do what he is asked to do. Unless the candidate has said that he enjoys cooking, asking a fifty year old salaryman how to prepare curry rice may not be appropriate.

Candidates are often asked to describe either a room or a person. When describing a place, prepositions of location are important and details. The description of a room should not simply be an inventory of the furniture. When describing a person some interviewers feel that describing a person's character is acceptable, but basically the purpose of the task is physical description. Candidates are asked to explain a current event — something happening in the news, either in Japan or anywhere in the world. The topic can be anything, but candidates must be able to give all the background information and details to a listener who is hearing the story for the first time.

For present narration, candidates are usually asked to describe their daily routine. They should explain what they do from the time they get up until they go to bed at night. The past narration often describes a trip overseas or recent experience. This narration should not be just a simple itinerary, "We landed in LA and stayed two and then went to San Francisco for a week." The candidate must be able to give details and describe some experience in detail. For future narration, candidates are asked to explain their plans for the future, either in their careers or in their private lives. Oddly enough, most people are quite responsive to such simple prompts as, "Tell me about your plans for the future."

At Level 2, candidates are asked to do a role play of a survival situation similar to the ones done at Level 1. However, at Level 2 a complication occurs in the role play. Level 2 role-plays use the same situations as the role plays at Level 1, now however, a problem arises. For example, if they are making a business appointment, the person they want to meet may be unavailable at the requested time and the candidate has to negotiate a mutually acceptable time. Perhaps the camera which needed repair at Level 1 now has to be sent away for repair and will not be available for some days or the suit the candidate wanted cleaned will not be ready in time for his meeting or the hotel does not accept his credit card. The candidate has to listen to the cause of the problem and use his language skills to solve it. Level 2 role plays are much more true to life than those at Level 1. A common response for level 1 speakers faced with a Level 2 role play is to simply say "OK" and leave the situation. They make no effort to tackle and solve the problem linguistically.

Where Level 1 candidate's speech is accurate at the sentence level, level 2 candidates narrate. To put it another way, they speak in paragraphs. They are able to make several sentences about the same topic. Their speech is not error free and in fact can be very inconsistent, but their speech should be understandable to someone who is not used to dealing with foreigners.

Level 3 speakers are asked to speak about a wide range of concrete and abstract topics, hypothesize, defend an opinion, and do a role play for an unfamiliar situation. Where Level 2 speakers are asked to explain a current event, Level 3 candidates are expected to be able to give accurate information on a variety of topics -cultural differences between two countries, the state of the economy, the education system in their home country. Many speakers have expertise — vocabulary and knowledge — in one area, usually the field they work in. Expertise in one area is not enough to qualify as Level 3. Candidates must be able to give accurate information about topics outside their immediate life experience. Probes at Level 3 for concrete and abstract topics often flow naturally from the Level 2 current event probe. If the candidate describes a recent crime, for the Level 2 current event, the interviewer could probe to Level 3 by asking why crime appears to be increasing these days. This moves the linguistic target from a single concrete incident to a larger societal trend, an abstract topic. The candidate must speak concisely and accurately about the topic, giving examples or using statistics. Because the candidate must be able to speak on a wide range of topics, and this takes time, the interviewer must be very careful of the time.

When asking candidates to hypothesize, it is important that the question be something that the candidate has thought about beforehand or can reasonably be expected to have thought about. For example the question, "How would your life be different if you had not gone to university abroad?" is superior to the question "How would your life be different if you were Michael Jordan?" The question should be connected to the candidate's life. This means that the interviewer must listen to the candidate carefully and form a hypothetical question appropriate to his or her life.

For the substantiated opinion probe, interviewers often counter the candidate's explanation of an abstract topic. The interviewer offers an alternative explanation for the phenomenon. Candidates not at Level 3 will often simply roll over and accept the interviewer's counter-argument without trying to defend their own. Alternatively, the interviewer can offer some well known phenomenon and ask for explanation. An exchange might go something like this. The interviewer could ask the candidate whether or not English language education should be mandatory and if the candidate says yes and cites the role of English as an international language, its importance to business and commerce as reasons for the necessity of its being taught in schools, the interviewer could counter by saying that most students don't seem to reach a level of proficiency necessary to conduct business and in fact very few students will work directly with English speakers or need English themselves. A candidate not at Level 3 would probably give up the discussion, but a Level 3 speaker would defend his or her opinion and offer some new information to support it — perhaps the growing foreign investment in Japanese companies or the less tangible benefits that come from learning a language and foreign culture. Winning the debate is not important, and interviewers must be careful to not be abusive or argumentative. The purpose of the task is to see if the candidate can absorb new information and respond to it quickly and flexibly while maintaining or defending his or her own opinions.

Role plays for Level 3 involve responding to an unfamiliar situation. Something unplanned, unforeseeable has occurred and the candidate must explain the situation to the interviewer. Typical role plays include situations where the candidate has borrowed the interviewer's car and had an accident or the candidate's boss wants to know why the candidate has been missing so much work lately.

Level 3 speech is highly accurate. Errors are infrequent, but do occasionally occur. The candidate has a sophisticated, educated adult vocabulary and can use complicated tenses, conditionals and the perfect tenses, correctly. Level 3 candidates will still make errors, but their speech is much smoother and more fluent than a Level 2. This requirement of educated adult speech often disqualifies people who lived in English speaking countries when they were young, say elementary school, and then were educated in a non-English speaking country. Their fluency and grammar may be strong, but their world knowledge is difficult to express in English and their vocabulary is not developed enough to discuss the kind of topics a Level 3 speaker can handle.

Candidates at Level 4 must complete all of the Level 3 tasks and be able to adjust the register of their speech and be able to convince and persuade. Role plays are especially important at Level 4 and it is necessary to do several. Adjusting register basically means changing the way one speaks and people do it all the time. We talk differently to children than we do to our boss and we speak differently to good friends than we do when making a presentation before a group of strangers. To test candidates' ability to change register role plays are necessary. A typical role play asks the candidate to make a brief presentation at a city council meeting protesting the proposed construction of a new airport. Candidates, especially educated adults who use English in business or work are able to project their register up, but Level 4 candidates must also be able to lower their register. To test their ability to do this another typical role play asks the candidate to make a brief speech to young children on the dangers of smoking.

Level 4 candidates must demonstrate their ability to convince and again role plays are used. Typical role plays ask candidates to convince their boss to remove his son from a project where he is working with the candidate or the candidate has caused a traffic accident and must convince the driver of the other car to not call the police or his insurance agent.

Level 4 speech is highly accurate. Speech is nearly equivalent to an English native speaker. Speech is extensive, precise, and appropriate to every occasion. The candidate has an educated, sophisticated vocabulary. He uses idioms and native expressions accurately and appropriately. Errors are almost nonexistent. Only in the area of accent is listener aware that he is dealing with a non-native speaker. Henry Kissinger is always trotted out as an example of the Level 4 archetype. Henry Kissinger's persistent accent disqualify him from being a Level 5.

Level 5 speech is in all ways equal to an educated native speaker's. The interviewer is the same for a Level 4 and the only difference is that a Level 5 speaker has native pronunciation and accent. Because of this standard based on an educated native speaker, not all native speakers can be considered Level 5.

To qualify for a level of the interview, the candidate must complete all of the tasks. If a candidate fails one task, he or she cannot be considered at that level. If a candidate is able to complete two thirds of the tasks at the higher level he or she can be considered a plus level. For example,

candidates who can complete all of the Level 1 tasks plus complete two thirds of the Level 2 tasks can be considered 1+.

Evaluating the LPI

The TOEIC LPI is a highly accurate and reliable proficiency test. For teachers thinking of adapting it for use in class or helping students prepare for it there are some points of consideration. First, while the makers of the TOEIC written test claim that it shows the distinction between beginner and advanced students and further that the test is a valuable instrument for charting progress, the same cannot be said for the interview. The level of proficiency is quite different between the levels. Almost all candidates who qualify for the interview can complete the relatively basic tasks of Level 1, asking questions, providing simple biographical information and completing a survival role play, but the tasks of Level 2 seem disproportionately more difficult — particularly recounting a current event, giving instructions and describing. So great is this gap that it is quite possible that university English students will enter the language program at Level 1, study for four years and leave the program still at Level 1. This makes the interview unmotivating and unuseful as an indicator of progress.

Also, interviewers are trained to probe to a higher level and if the candidate fails in the task the interviewer should return to the previous level. For example, a candidate explaining a news item as a current task at Level 2 might be probed to Level 3 by the interviewer for an abstract topic task. If the interviewer returns to Level 2 for the next task he is indicating that he believes the candidate failed the probe. A second rater listening to the interviewer might not listen to the candidate's speech, but only note that the first interviewer returned to Level 2 so the candidate must have failed. The test's system of probes and tasks allows interviewers to give identical ratings to a candidate, but it does not guarantee that the interviewers are fairly evaluating the candidate's speech sample.

A related point is that as a proficiency test, the TOEIC LPI guarantees that candidates are able to perform the tasks at the Level they are evaluated at. The candidate might be able to do some tasks at a higher level, but because he or she cannot do them all he cannot be assigned the higher level. This system guarantees proficiency, but it should be remembered that it is possible that a candidate can likely perform some tasks at higher levels than the level guaranteed.

Conclusion

The TOEIC LPI is an accurate and reliable indicator of a candidate's English ability. However, the interview has drawbacks as an indicator of progress because of the large linguistic differences between the levels. This drawback could be overcome if the Level Checks at the lower levels were expanded so that the gap between the levels was not so large.

References

- ETS.(1993). TOEFL 1993-4 Bulletin of Information. Princeton, NJ: Educational Testing Service
- Gilfert, S. and Kim, V. (1990). TOEFL Strategies. Nagoya, Japan: Kawai JukuPress.
- Gilfert, S. and Kim, V. (1996). TOEIC Strategies, Tokyo: Macmillan JapanK.K.

- Heaton, J.B. (1990). *Classroom Testing*. London: Longman Group UK.
- Hinkel, E., Director of English Language and American Culture Program, Xavier University, Cincinnati Ohio, personal communication. Matsumoto, S and West, D. (1990).
- Wilson, K. M. (1989). *TOEIC Research Report #1*. Princeton, NJ: Educational Testing Service
- Woodford, P. (1982). "The Test of English for International Communication." In C Brumfit, (Ed) *English for international communication*. Oxford: Pergamon Press.