The Comparative Fallacy Reflected in L2 Proficiency Tests

Árpád Galaczi

Most language learners and language teachers around the world, if asked about their dream goal, would say something to the effect of: "(for my students) to speak/write/sound like a native speaker." While this is a natural and understandable striving – given that native speakers are the most proficient users of any language – it sets adult language learners up for ultimate failure. Even though in a handful of documented cases some adult L2 learners do achieve this goal, they are rather the exceptions than the norm and most post-puberty language learners stop short of native-like performance. As Cook (1999) puts it: "Whether or not one accepts that some L2 users can pass for native speakers, these passers form an extremely small percentage of L2 users. [They are] as typical of human beings as are Olympic high jumpers or opera singers" (p. 191). He goes on to make the point that L2 users should be viewed in their own right, as proficient users of highly complex linguistic systems (*interlanguages*) and not in terms of what they are not: "L2 users (should) be viewed as multicompetent language users rather than as deficient native speakers" (p. 185).

While clearly at fault for looking at native speakers as the benchmark, language teachers and language learners are not the only ones making this mistake: Many second language acquisition (SLA) researchers do the same and commit what Bley-Vroman (1983) calls the *comparative fallacy*. In Bley-Vroman's view, the picture we gain about the L2 learning process can be severely distorted if the target language is used as the ultimate norm when judging L2 performance. In his words, it is a serious mistake to rely on "a comparison of the target language in order to study the structure of the interlanguage" (p. 4).

The present paper will analyze a number of empirical SLA studies from the point of view of the comparative fallacy by looking at the L2 proficiency measures employed in them. These studies have been selected because they not only employ language proficiency tests, but also contain detailed descriptions of these test instruments – thereby providing a chance to search for instances of the comparative fallacy.

Harley and Hart's 1997 article examines the relationship between the starting age of intensive L2 exposure, "L2 outcomes" (performance), and specific components of the language aptitude of students enrolled in French immersion programs in Canada. The L2 proficiency tests the researchers employed were designed to measure different aspects of L2 knowledge, such as vocabulary, listening and reading ability, and oral and written production. The vocabulary recognition test consisted of a list of 66 real French words intermingled with 34 pseudo-words, with the students having to indicate how many words they recognize. As far as the exercise itself goes, there is no evidence of the comparative fallacy: Checking the ability to differentiate real words from nonsense words does not necessarily involve comparison with native-speaker standards. Where the comparative fallacy manifests itself is in the choice of vocabulary items: According to the authors, the real words were taken from a utility index of French words. This index is based on native-speaker standards that may not hold for L2 French speakers, especially if they acquired a significant part of their interlanguage in a classroom setting. While this is itself an empirical question, it is likely that native-speaker word usage and frequency differs

significantly from that of L2 learners'. As a consequence, a utility index based on French L2 learner corpora would have probably been in far better accordance with the subjects' L2 norms and would have offered a more realistic picture of the students' abilities.

A similar problem affects the writing task as well – at least partially. The first task is a cloze test consisting of three paragraphs from an authentic essay with words blanked out. Students had to fill in the missing words and then the complete texts were judged for "*nativeness"* by *native speakers* – needless to say using *native-speaker norms.* Only those words conforming to native-speaker standards were counted as correct and there was no partial credit given to non-native-like but justifiable word choices. The comparative fallacy makes its appearance again, not only through the use of TL norms when judging correctness of word choice, but also in the fact that an unadapted native-French written text was used as the basis of the cloze test.

In all fairness to the authors, however, it needs to be pointed out that even though instances of the comparative fallacy could be found in these two parts of the test, it is possible that they did not influence the overall findings of the study in a major way. The fact that words and texts based on L2 corpora would have given better results does not mean that the results of the employed test distorted reality beyond recognition. In all likelihood, similar results would have been yielded, and similar correlations between aptitude test scores and language proficiency scores would have been found, had the vocabulary recognition and the cloze test been based on L2 French word usage. Since all the subjects were given the same texts and same tests (proficiency and aptitude), the overall effect of the comparative fallacy on the group is probably minimal.

The second part of the written test in Harley and Hart's test battery, however, is in no danger of committing the comparative fallacy. It consists of written responses to open-ended questions, and these responses were scored based on criteria developed specifically for L2 French users, thereby doing justice to the students' L2 systems. Similarly, the listening comprehension texts the students had to listen to (and answer comprehension questions about) were tailored to their ability and status. Even though the texts sounded authentic, they were specifically designed for immersion students, piloted on similar student populations and changed based on their responses. Finally, the oral test seems to have successfully avoided the comparative fallacy as well: The sentence repetition tasks were scored in a way that students received credit not only for exact, native-like renditions but "semantically equivalent" non-native-like ones as well.

As with the above study, a study by Sparks, Javorsky, Patton, and Ganschow (1998) has language aptitude at its core and examines the links between foreign-language (FL) aptitude, FL performance, and native-language performance. The subjects were high school students enrolled in Spanish, German, and French classes, and a battery of different tests were administered to the subjects to measure their abilities and performance. Here we focus on the FL performance tests, as that is where insight into the comparative fallacy (its presence or absence) can be gained. The tests included measures of reading, writing, listening, and speaking ability in the respective foreign language. The tests in the different languages were exact equivalents to each other, with necessary minimal language-specific differences. For simplicity's sake, this paper will examine the Spanish FL tests, for they were thoroughly described in the study.

The reading comprehension test involved twenty multiple-choice questions *in English*: ten about a fictitious one-page letter, and ten about a brief original article from a Spanish magazine. The fact that the comprehension questions were in English, and that the texts varied in

difficulty in order to differentiate between more advanced and less advanced FL students, allow this part of the test to fulfill its goal without aiming too high or setting a much too difficult task for the test-takers. Unlike one of the subtests in the previous article, the present subtest couples an original Spanish text with an adapted one, thus avoiding the comparative fallacy of applying native-speaker standards to non-native speakers. Conducive to the same goal are the comprehension questions in the students' native language: They help the students focus on the content of the reading passages instead of perhaps getting bogged down in the grammar or structure of questions that may be beyond their level of proficiency. This is in tune with Cook's (1999) suggestion of acknowledging and using students' L1 in the FL classroom – a practice inspired by the multicompetence theory that supports the development of links between the different languages of a learner, rather than treating them as separate and isolated from each other.

A similar approach pertains to the listening and speaking test, which consisted of a 10-15 minute oral interview. The students not only received instructions in English prior to the interview, but were also given a conversation card in English to help them begin the conversation. As far as the scoring of the oral interview goes, not much is given away by the authors about it; therefore, there is not enough information to conclude the presence of the comparative fallacy or otherwise. Without providing a rubric or a description of the guidelines, the only piece of information the authors give about the scoring process is that it follows the guidelines laid down by the American Council on the Teaching of Foreign Languages (ACTFL). The same is true for the scoring of the written test. Therefore, to assess whether these two subtests are instances of the comparative fallacy or not, one needs to analyze the ACTFL proficiency guidelines (not easily available) from this point of view – a topic perhaps for a more comprehensive paper.

The unavailability of enough information about scoring procedures is far from being an isolated incident. Many empirical studies give only scant information about their measuring tools, hardly enough to give an idea about their reliability and validity, much less permit a more in-depth analysis, such as screening for the comparative fallacy. Perhaps the scientific community would benefit from additional publishing guidelines that request more information on testing instruments, as well as more research and awareness-raising about the comparative fallacy.

REFERENCES

Bley-Vroman, R. (1983). The comparative fallacy in interlanguage studies: the case of systematicity. *Language Learning*, *33,* 1-17.

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly, 33,* 185-209.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19,* 379-400.

Sparks, R.L., Javorsky, J., Patton, J., & Ganschow, L. (1998). Factors in the prediction of achievement and proficiency in a foreign language. *Applied Language Learning, 9*, 71-105.

Árpád Galaczi is a doctoral student in Applied Linguistics at Teachers College, Columbia University. His current research interests are fossilization in second language acquisition and foreign language learning aptitude. Correspondence should be sent to Árpád Galaczi, 40 Fennec Close, Cambridge CB1 9GG, United Kingdom. E-mail: ag486@columbia.edu.