*Compendium Study*

# Assessing English-Language Proficiency in All Four Language Domains: Is It Really Necessary?

*Donald E. Powers*

*September 2013*

Arguably, the answer to the question posed in the title of this paper is "Yes, very often it is highly desirable, but it is not always absolutely necessary." It is not always necessary because the performance of some tasks or jobs may require proficiency in only one of the four language domains of listening, reading, speaking, and writing, or in a subset of the four domains. Sometimes, therefore, assessing proficiency in a single area (speaking, for example) or in a pair of areas (speaking and listening, for example) may be sufficient, and the use of only a single, valid and reliable measure of the proficiency of interest (the *TOEIC*® Speaking test, for instance) may be completely justified. In other, possibly more common circumstances, however, a more comprehensive, four-skills assessment may be appropriate. For instance, when overall communicative competence (i.e., the ability to convey and receive information, exchange ideas, etc.) is of interest, as it often is for many jobs, it seems advisable to assess English-language skills in each of the four language domains. It is also advisable because the *TOEIC*® test can have a profound impact on how English is learned and taught around the world. Therefore, it is important to emphasize the testing of all four skills in order to have a salutary effect on teaching and learning practices.

## Objectives of This Paper

The primary purpose of this paper is to outline a rationale for a comprehensive, four-domain approach to assessing English-language proficiency. In doing so, the paper:

- discusses why a comprehensive approach is not always implemented

- considers the circumstances under which a comprehensive approach may be desirable and when it may be unnecessary

- outlines an argument for testing language proficiency in all four language domains

- buttresses the argument by selectively reviewing literature concerning the relationships among the four language domains

- demonstrates, by example, the utility of a comprehensive approach to assessing English-language proficiency

## Why *Not* Test in All Four Domains?

What are the reasons for *not* testing in all four domains? One reason might be interest in only a particular domain, as when evaluating the effectiveness of instruction for a single skill, for instance. An additional reason for testing selectively in fewer than four domains, instead of comprehensively, is efficiency: The relationships among the four domains are so strong, the argument goes, that an assessment of any single domain can serve as a reasonably good surrogate for an assessment in any of the others. In fact, early studies of the *TOEIC*® Listening and Reading test (e.g., Wilson, 1989) seem to have suggested this possibility. The strong correlation between scores on the TOEIC Listening and Reading test and performance on language proficiency interviews was viewed as evidence that TOEIC Listening and Reading scores could serve, at least *indirectly*, as moderately good proxies for *direct* assessments of speaking proficiency.

When speaking of proxies and surrogates, however, it is often difficult to determine (a) if the relationship between a measure and its proxy is strong enough or (b) if substituting one measure for another one makes sense logically. For instance, consider the relationship between health and obesity (too much body fat), which can be directly measured by such technologies as ultrasound or magnetic resonance imaging. Most often, however, body mass index (BMI) is used as a proxy for obesity, with BMI being a simple numeric function of both weight and height. The BMI is imprecise, however, because it fails to account for, among other factors, muscularity and size of body frame.

Just as obesity can be a factor in a person's overall health, oral proficiency (or language proficiency more generally) can be a factor in success on a job that requires the ability to speak English (or to use English more generally). To predict whether a person has the skills needed to function on the job, one could use scores from one TOEIC test as a proxy for another, in much the same way that the BMI is used as a proxy for obesity when predicting susceptibility to various diseases.

Although most analogies usually break down at some point, they can often be enlightening. But, whether the context is health or English-language proficiency, almost certainly, more persons will be misdiagnosed by a proxy than by a more direct measure of the construct of interest. The BMI may be an especially misleading index for some individuals—athletes and the infirm, for example. Likewise, using one TOEIC test as a proxy for another (or for a more comprehensive battery of measures) may misclassify some individuals disproportionately, for example those whose language training or experience using English is uneven across language domains. As discussed later in this paper, however, the use of both a direct measure and a related measure or proxy can, in combination, often provide more comprehensive and nuanced measurement (and thus more accurate prediction) than either one alone.

## Reasons for Not Using Proxies

Another argument against measuring proficiency in one domain of English-language proficiency as a proxy for proficiency in another is that the four domains are logically distinct. Moreover, although strongly correlated, they are also empirically distinguishable as, for example, Sawaki, Stricker, and Oranje (2009) have demonstrated for the TOEFL® test, and Liu and Costanzo (2013) have shown for the four TOEIC test measures. Furthermore, just as the relationship between height and weight may depend on various factors, so too may the relationship among measures of English proficiency in the four language domains. For instance, it may depend on the amount of formal training in the English-language and the degree of experience using English.

## An Argument for Testing in All Four Domains

The argument for testing English-language skills in all four domains is based on the following assertions:

- Language is the primary (and unique) vehicle by which human beings *communicate*, that is, share ideas, express intentions, convey information, and so forth.

- Effective communication is a complex endeavor, requiring at least some facility in each of the four language domains.

- To develop facility with language, one needs to encounter it, either through formal instruction or through informal experience and preferably through both.

- Testing in all four domains encourages teaching (and learning) in all four domains and therefore leads to better overall communicative competence.

## Relationships Among the Four Domains

The argument is based in part on the complex relationships among the four domains. At a broad level, pair-wise relations among the four domains have long been acknowledged as "complex relationships of mutual support" (Peregoy & Boyle, 2008, p.119). Both reading and listening are *receptive* (but not necessarily passive) activities that require the construction of meaning. The development of one, therefore, is likely to facilitate the development of the other, in part because some of the same processes underlie both skills. Both writing and speaking, on the other hand, are *productive* activities that involve the creation of a message for an audience. Similarly, then, their development is likely to co-occur.

Pair-wise relations also exist between speaking and listening, as both involve oral language, and between writing and reading, as both involve written language. Reading and speaking are linked in so far as reading can be an effective way to develop oral vocabulary. Listening and writing are related in that learning to listen may entail a focus on, for instance, rhythm and flow, processes that have counterparts in writing. Likewise, listening involves the same skills (prediction, hypothesizing, checking, revising, and generalizing) that are needed for reading and writing (Ronald & Roskelly, 1985).

All of the four domains are linked by virtue of five primary underlying linguistic elements— phonology (knowledge of sounds and how they are combined), morphology (the formation of words), semantics (knowledge of word combinations and limitations), syntax (grammar, or the rules that govern word order in sentences), and pragmatics (the way in which language is used in context, Baker & Baker, 2009). Given the same underlying processes, it is not surprising perhaps that some (e.g., Widdowson, 1998) have regarded the distinction among the four domains as being somewhat artificial.

## Why Teach All Four Skills?

Some experts (e.g., Peregoy & Boyle, 2008) have emphasized the desirability of language instruction that integrates processes from all four language domains. This assertion is based on the observation that good communicators move easily between written and oral modes, and so both modes are essential. In addition, the use of the same processes in each domain results not only in improvement in each domain separately, but also in overall development with respect to communication skills in general.

That the integration of multiple skills is regarded as the most effective method of developing language proficiency has been one of the major recent themes in second language teaching (Hinkel, 2006). As an example, by showing learners how to analyze text, teaching writing and reading together may enable learners to become both better writers and better readers. This assertion is supported by a significant body of research on the process of *reading to write* (Ruiz-Funes, 1999). Conversely, because of commonalities between reading and writing (for example, both involve composing or constructing meaning), learning to write may provide a means to develop reading skills (Zamel, 1992). Spack (1985), for example, described a composition course that focuses on the interrelationships among literature, reading, and writing. These instructional strategies are not surprising, perhaps, given the large number of processes that underlie both reading and writing ability (Shanahan, 2006).

Dependencies among the domains have also been a basis for designing courses on oral/aural communication. These dependencies have included the relationships among listening, speaking, and (as a subset of both) pronunciation, which have been characterized as "reciprocally interdependent oral language processes" (Murphy, 1991, p. 51). With respect to writing and speaking, partial support has been found for the hypotheses that (a) language learners who are able to use certain grammatical forms in their writing are also more capable of using them in their speaking and (b) learners who write extensively are more likely to exhibit improvements in both writing and speaking proficiency (Hubert, 2009). Dykstra-Pruim (2003) has also noted an underlying connection between speaking and writing.

Because listening and reading require many of the same processes (decoding, for example), developing listening skills is often a good strategy for developing reading skills (Roe & Ross, 2006). Conversely, reading exercises may be desirable not only for improving reading skills, but also for developing listening, speaking, and writing skills (Chen, Chen, & Sun, 2010). In an interview with Barbara Garner (2000), Hilferty pointed to the considerable research showing reciprocity between speaking and reading: Both depend on the same abilities required to process difficult materials, and, for some language learners, reading provides the primary input for learning a new language. In summary, the relations among the four domains have been demonstrated frequently and in numerous ways and in various combinations.

## Relation of Testing to Teaching

A number of language acquisition textbooks have emphasized the need for integrated teaching. Lee and VanPatten (2003), for example, stressed the importance of developing all four skills simultaneously, and the exercises in their book reflect this belief. Richards (2008) emphasized the role of teaching listening in an integrative approach. He advanced an alternative view of listening, focusing on its acquisition function (rather than the more traditional of listening as facilitating the understanding of spoken discourse). In this alternative acquisition view, listening can serve as a primary means of providing input to promote the development of other aspects of second language proficiency.

Richards (2008) also noted that the teaching of listening has received greater attention in recent years than in the past, asserting that … "if listening isn't tested, teachers won't teach it" (p. 1). This point applies with equal force to the reading, speaking, and writing domains. In order to reinforce this point, I refer the reader to discussions of washback and the consequences of testing. Messick (1996) defined *washback* as the extent to which a test influences learners and teachers to do things that they would not ordinarily do. Wall's (1998) discussion gave credence to the concept, and Choi (2008) provided a case study of the influence of one test in a particular context. In short, testing has the potential to change both teacher and learner behaviors, either for the good or for the not-so-good, depending on the nature of the test and the testing process.

## Utility of a Comprehensive Assessment for Assessing a Single Domain

Even when only a single domain (or fewer than four domains) is the primary concern, a comprehensive four-skills assessment may still be valuable in at least two ways. As Liu and Costanzo (2013) have shown, TOEIC Listening and Reading scores provide a relatively good indication of who is most likely to show improvement over time with respect to speaking and writing skills. When compared with their lower performing peers, TOEIC test takers with higher TOEIC Listening and Reading scores were much more likely to exhibit test score improvements over time on both the TOEIC Speaking and Writing tests. Furthermore, Powers, Yu, and Yan (2013) demonstrated that by considering multiple TOEIC test measures, it was possible to make meaningful distinctions among test takers with regard to their self-reported ability to perform a variety of everyday English-language tasks, even though such distinctions were not possible with scores from only a single TOEIC test. For example, among test takers having virtually identical *TOEIC®* Speaking scores, those exhibiting better performance on the TOEIC Listening and Reading tests were significantly more likely than their lower scoring peers to report being able to perform a variety of everyday *speaking* tasks, such as (a) making, changing, or canceling an appointment to see a person and (b) commenting on or reacting to someone's opinion during a discussion. The same was true for test takers having similar *TOEIC®* Writing scores: Those with higher TOEIC Listening and Reading scores were more likely to report being able to perform everyday *writing* tasks, such as (a) sending an email or letter to a public organization to request needed information and (b) preparing text and slides (in English) for a presentation at a professional conference. Thus, measurement in nontarget domains (e.g., listening and reading) can be useful for augmenting the prediction of both language development and everyday performance in other domains (e.g., speaking and writing).

## Summary

This paper can be summarized by reiterating the following points:

- Communicative competence is key. It is complex, involving all four language domains with many intricate relationships among its components.

- Developing communicative competence is probably best accomplished with an integrative approach.

- Testing can influence what is taught and what is learned, and the TOEIC tests can have a positive influence on teaching and learning.

- Each TOEIC test (i.e., Reading & Listening, Speaking, and Writing) can make a unique contribution to assessing English proficiency. As a result, a comprehensive approach to assessment has incremental value for predicting success in the workplace and for projecting future improvements in English-language proficiency.

As suggested earlier, proficiency in the four language domains is related in a variety of complementary, often subtle ways. As a result, although English-language skill in a single area can be measured accurately by a well-developed measure of that proficiency, an even more precise, nuanced estimate of the target proficiency can often be achieved by measuring proficiency in other language domains as well. We contend that usually (but not always), the additional effort and expense of a more comprehensive assessment is justified by the additional information that results.

# References

Baker, D., & Baker, S. (2009) *Second language acquisition*. Retrieved from http://www.education.com/reference/article/second-language-acquisition/

Chen, J.-M., Chen, M.-C., & Sun, Y. S. (2010). A novel approach for enhancing student reading comprehension and assisting teacher assessment of literacy. *Computers & Education*, *55*, 1367-1382.

Choi, I.-C. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, *25*, 39-62.

Dykstra-Pruim, P. (2003). Speaking, writing, and explicit-rule knowledge: Toward an understanding of how they interrelate. *Foreign Language Annals*, *36*, 66-76.

Garner, B. (2000, March). The relationship between reading and speaking skills: An interview with Ann Hilferty. *Focus on the basics: Connecting research and practice, 4*(A). Retrieved from the National Center for the Study of Adult Learning and Literacy. website: http://www.ncsall.net/index.html@id=328.html

Hinkel, E. (2006). Current perspectives on teaching the four skills. *TESOL Quarterly*, *40*, 109-131.

Hubert, M. D. (2009). The relationship between writing and speaking in the US university Spanish language classroom. *Dissertation Abstracts International Section A: Humanities and social sciences*, *69* (9-A), 3527.

Lee, J. F., & VanPatten, B. (2003). *Making communicative language teaching happen* (2nd edition). Boston, MA: McGraw-Hill.

Liu, J., & Costanzo, K. (2013). The relationship among TOEIC listening, reading, speaking, and writing skills. In D. Powers (Ed.), *The research foundation for the TOEIC tests: A compendium of studies* (Vol. II, 2.1-2.25). Princeton, NJ: Educational Testing Service.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*, 242-256.

Murphy, J. M. (1991). Oral communication in TESOL: Integrating speaking, listening, and pronunciation. *TESOL Quarterly*, *25*, 51-75.

Peregoy, S. F., & Boyle, O. F. (2008). Reading, writing, and learning in ESL: A resource book for teaching K-12 English learners. New York, NY: Pearson, Inc.

Powers, D. E., Yu, F., & Yan, F. (2013). The *TOEIC®* Listening, Reading, Speaking, and Writing tests: Evaluating their unique contribution to assessing English-language proficiency. In D.E. Powers (Ed.), *The research foundation for the TOEIC tests: A compendium of studies* (Vol. II, 3.1-3.14). Princeton, NJ: Educational Testing Service.

Richards, J. C. (2008). *Teaching listening and speaking: From theory to practice*. New York, NY: Cambridge University Press.

Roe, B. D., & Ross, E. P. (2006). *Integrating language arts through literature and thematic units*. New York, NY: Pearson Education, Inc.

Ronald, K., & Roskelly, H. (1985). *Listening as an art of composing*. Paper presented at the 36th Conference on College Composition and Communication. Retrieved from ERIC database (ED257094).

Ruiz-Funes, M. (1999). Writing, reading, and reading-to-write in a foreign language: A critical review. *Foreign Language Annals*, *32*, 514-526.

Shanahan, T. (2006). Relations among oral language, reading, and writing development. In D. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 171-186) New York, NY: Guilford.

Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing, 26*, 5-30.

Spack, R. (1985). Literature, reading, writing, and ESL: Bridging the gaps. *TESOL Quarterly*, *19*, 703-725.

Wall, D. (1998). Impact and washback in language testing. In C. M. Clapham & D. Corson (Eds), *Language testing and assessment. Encyclopedia of language and education* (Vol. 7, pp. 291-302). Dordrect, Netherlands: Kluwer Academic.

Widdowson, H. G. (1998). Skills, abilities, and contexts of reality. *Annual Review of Applied Linguistics*, *18*, 323-333.

Wilson, K. M. (1989). Enhancing the interpretation of a norm-referenced second-language test through criterion-referencing: A research assessment of experience in the TOEIC testing context (TOEIC Research Report No.1). Princeton, NJ: Educational Testing Service.

Zamel, V. (1992). Writing one's way into reading. *TESOL Quarterly*, *26*, 463-465.