Measuring the Power of Learning.®

Research Report
ETS RR–17-51

Articulating and Evaluating Validity Arguments for the *TOEIC*® Tests

Jonathan E. Schmidgall

December 2017

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Articulating and Evaluating Validity Arguments for the *TOEIC*® Tests

Jonathan E. Schmidgall
Educational Testing Service, Princeton, NJ

This report provides a brief overview of how the *TOEIC*® program has adopted an argument-based approach to validity in order to support the use of the TOEIC tests. This approach emphasizes the need to explicitly state claims about the measurement quality and intended use of a test and to support those claims with evidence. This report briefly summarizes how the assessment use argument (AUA) exemplifies this approach and was used to construct validity arguments for the TOEIC tests. After highlighting the practical applications of the validity arguments that were constructed, it highlights challenges associated with this work and proposes several extensions. Overall, this process demonstrates how TOEIC research takes a broad, critical, and rigorous approach to support the use of the TOEIC tests.

Keywords Test quality; validity; assessment use argument; *TOEIC*®

doi:10.1002/ets2.12182

How can you determine whether a test is suitable for the purpose for which it was designed? This fundamental question of validity has preoccupied test developers, researchers, and score users for decades. In the first *TOEIC*® *Compendium*, Powers (2010) provided a clear, accessible overview of validity that focused on two critical aspects: whether scores mean what they are supposed to mean and whether a test fulfills its designated purpose. Subsequently, consensus-based professional standards have come to embrace the view that test developers must convince stakeholders (i.e., anyone affected by the test) that the intended use of a test is appropriately supported or justified (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014; Educational Testing Service [ETS], 2015; Newton, 2012). This view is formalized in the argument-based approach to justifying test use.

The argument-based approach to justifying test use consists of a comprehensive set of claims made by the test developer. These claims are supported or undermined by evidence, which may include documentation from the test development process and ongoing research. Through an examination of the test developer's claims and the evidence to support them, various stakeholders may arrive at a global evaluation of whether the intended use of the test has been adequately justified. Different stakeholders may value different types of evidence; for example, teachers may be primarily concerned about evidence that the test has a positive impact on teaching and learning, whereas score users may be more concerned about the outcomes of decisions based on the test.

The purpose of this report is to provide an accessible introduction to the argument-based approach, its implementation for TOEIC tests, and the perceived benefits for stakeholders. I begin this report with a brief overview of the assessment use argument (AUA), a prominent argument-based approach to validation (Chapelle & Voss, 2014). Next, I detail the approach that has been used to articulate fully specified validation arguments for TOEIC tests. This approach incorporates evidence from a variety of sources, including test documentation, monitoring activities, and research. Finally, I provide an overview of the two primary ways in which the validation arguments are used: to influence the research agenda and to communicate with stakeholders. The report concludes with a brief discussion of the benefits of this approach, as well as several suggestions for extending it.

Corresponding author: J. E. Schmidgall, E-mail: jschmidgall@ets.org

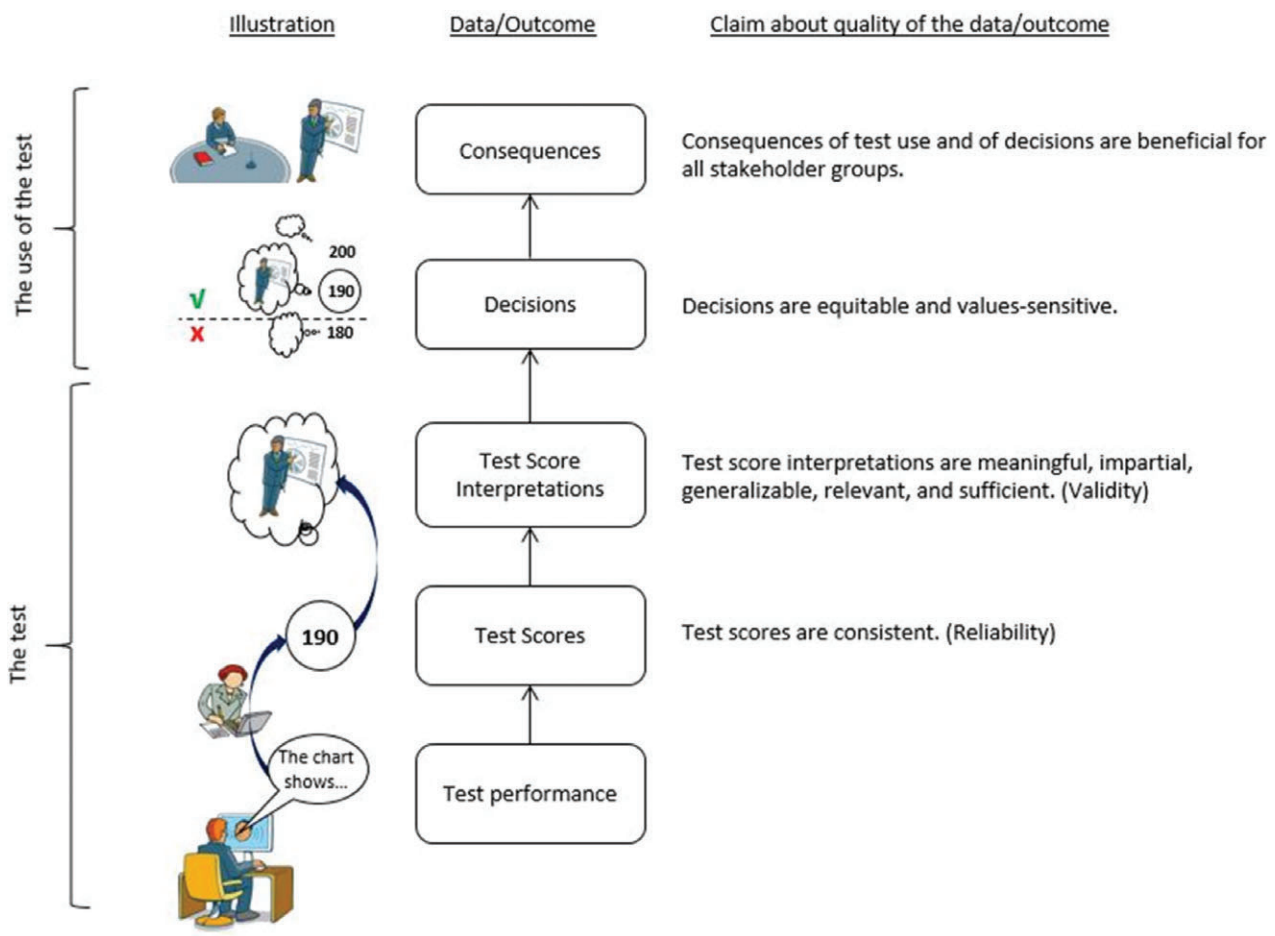

Figure 1 Data and claims in an assessment use argument.

The Assessment Use Argument

An AUA is "a conceptual framework for guiding the development and use of a particular language assessment, including the interpretations and uses we make on the basis of the assessment" (Bachman & Palmer, 2010, p. 99). The framework is structured as a hierarchical set of claims made by the test developer that specifies how test scores should be interpreted and used to make decisions. The AUA draws upon well-established conceptualizations of validity, including Messick's (1989) progressive matrix, and a formalized argument structure in which transformations of data represent inferences that support claims (Toulmin, 2003).

Although this approach may seem abstract, the use of argument structure requires test developers to make statements about the expected measurement quality and uses of a test explicitly and systematically. In the AUA framework, these statements take the form of four high-level claims. These four claims explicitly link test-taker performance to test scores, scores to interpretations about test-taker ability, interpretations about ability to decisions, and finally, decisions to the consequences that follow. Thus, the AUA framework takes the general form shown in Figure 1.

Each box in Figure 1 represents data related to a test or its use. Data are transformed into an outcome through an inference, represented by arrows. An outcome is expected to have particular qualities, as specified in the corresponding claim. The outcome also serves as data for the next step up the inferential ladder, linking test performances to consequences in the real world.

As seen in Figure 1, the foundational data for all subsequent inferences are the test performance. The illustration shows what this might look like for a computer-delivered speaking test: a test taker speaking in response to questions. The test taker's speaking responses are the test performance, which is transformed into a test score (e.g., 190) through a rating procedure. Figure 1 shows that a claim should be made about the qualities of test scores: They are consistent. Aspects of

the testing procedure that are unrelated to the test taker's ability (e.g., test forms, test administrations, and raters) should not unduly influence scores, and so, scores are expected to be consistent across these aspects.

Test scores are then transformed into interpretations about test takers' abilities. These interpretations should be meaningful, impartial, generalizable, and appropriate (relevant and sufficient) for the decisions to be made. In the case of our illustrative speaking test, a score of 190 may, for instance, imply that the test taker has a high level of speaking proficiency and would be expected to present familiar information orally with a high degree of comprehensibility.

Test score interpretations are typically used to make decisions. Consequently, the score interpretation is transformed into a specific decision category or contributes in some way to a broader decision-making process. As an outcome, decisions should be fair and sensitive to the values of the decision maker. In the illustrative example, the test taker's score of 190 exceeds the benchmark set by the decision maker, which contributes to a hiring decision. The benchmark set by the decision maker (190 or higher) is relatively high, which reflects the decision maker's need for a high level of speaking proficiency and the desire to minimize false-positive decision errors.

Ultimately, the use of a test and the decisions that ensue produce an outcome: real-world consequences. Consequences should be beneficial for stakeholder groups; otherwise, the effectiveness of the test (or decision-making process) may be in question. In the illustrative example, the hypothetical test taker who was hired is able to give effective oral presentations on the job in the real world, benefitting himself and his employer.

This approach provides a coherent way to relate the traditional measurement concepts of reliability and validity, and it treats validation as the act of providing evidence to support claims rather than a specific quality of a test (e.g., criterion-related validity, construct validity, etc.). This approach is also comprehensive in that it captures a wide range of desirable aspects of a test that have historically been collapsed into imprecisely defined categories (e.g., consequential validity, response validity, construct validity, content validity, etc.). Schmidgall and Choi (2011) reviewed language-testing research articles, categorizing their research questions using traditional categories of validity (e.g., construct validity) and claims and warrants in an AUA. They found that individual research questions in the publications reviewed could usually be linked to specific claims in an AUA and that traditional categories of validity may be scattered across levels in an AUA. This observation suggests that the use of argument-based approaches such as the AUA — as opposed to traditional categories of validity — may help clarify the implications of validity research.

Both decision makers and test developers share responsibility for justifying assessment use. Test developers are expected to provide evidence to support the claim that test scores are consistent and that scores may be used to make interpretations about test-taker abilities. Decision makers need evidence that decisions are values-sensitive and equitable and that consequences of decisions are beneficial. Unfortunately, decision makers may lack the expertise or resources needed to provide adequate backing for these claims, such as the ability to conduct a benchmarking or standard-setting study. Test developers often have this expertise but may not be aware of features of a specific decision-making context that may influence how decisions are made, such as the level of language proficiency required or the decision maker's tolerance for different types of decision errors. Consequently, an AUA may be enhanced through collaboration between decision makers and test developers. Decision makers can utilize a test developer's expertise to support benchmarking studies that promote values-sensitive and equitable decision-making, whereas test developers can receive feedback on the test's effectiveness.

The structure of an AUA provides a basis for a comprehensive justification of test use that links real-world concerns about decisions and their consequences with the traditional concerns of test developers: reliability and validity. As a comprehensive list of claims and evidence, an AUA can be used to identify weaknesses in the overall argument for test use and prioritize research or test development projects. For example, Wang, Choi, Schmidgall, and Bachman (2012) reviewed the Pearson Test of English Academic and, based on documents obtained from the test developer, produced a detailed AUA that explicitly specified claims regarding test use. The review generated a number of specific recommendations that could be used to inform research.

As a simple hierarchical set of claims, an AUA can be used as a communication tool that illustrates the key issues that determine important qualities of the usefulness of a test, including fairness, impact, reliability, and validity. The concerns of individuals and stakeholder groups vary, and one of the challenges for research is addressing these concerns in a coherent manner while enhancing the "assessment literacy" of stakeholders. For example, stakeholders may be concerned about the following issues:

- score consistency (How can you make sure that all raters follow the scoring guides?),
- interpretation of scores (When we calculate criterion-validity, who or what is the criterion?),

- decisions based on these interpretations (What are the cut scores in other institutions?),
- consequences of test use (How has TOEIC been helpful for job-seekers?), and
- test use that relates to a number of these issues (How can recruiters know that TOEIC scores meet the needs of the market?).

By delivering versions of an AUA oriented toward specific stakeholder groups, a test developer with a strong research program may be able to help stakeholders answer their questions and become more sophisticated consumers of assessment products.

Constructing Assessment Use Arguments for TOEIC Tests

The previous section discussed how the AUA can be used to specify four high-level claims about the measurement quality and intended use of a test. A fully specified AUA also contains a large number of warrants: statements made in support of each high-level claim. Bachman and Palmer (2010) elaborated a reasonably exhaustive list of potential warrants in their book-length description of the AUA. This general, idealized version of an AUA was designed to incorporate the accumulated knowledge of testing professionals as reflected in influential publications on validity and validation (e.g., AERA, APA, & NCME, 1999; Kane, 2006; Messick, 1989). When constructing an AUA, this is a logical place to start: adapting the generalized AUA to a specific context. Figure 2 illustrates the process the TOEIC research program used to create fully specified AUAs for TOEIC tests, beginning with the elaboration of idealized AUAs.

Step 1: Articulate Claims and Warrants

As shown in Figure 2, the first step in the process was to utilize the existing AUA framework as proposed by Bachman and Palmer (2010) to build idealized validation arguments for the TOEIC tests. This idealized version contained all of the claims (and supporting warrants) that test developers and score users might want to see supported. Essentially, it represented a best-case scenario for a test developer interested in adhering to best practices in measurement with multiple warrants supporting each of the four high-level claims summarized earlier (see Figure 1). Although testing occurs in the real world in which there are important trade-offs between reliability, validity, and practicality (i.e., cost and convenience) that must be carefully considered, the idealized AUA represents an aspirational set of claims and warrants for a test developer to aim to support.

One of the challenges of creating AUAs for TOEIC tests is the fact that the tests are used for multiple purposes. For example, TOEIC tests are intended to facilitate hiring, placement, promotion, and progress decisions (e.g., ETS, 2013, p. 27). Although an AUA should be articulated for each intended use of the test, we initially constructed AUAs aligned with one particular use or decision in mind: hiring. This use was chosen based on feedback from key stakeholder groups and is an example of a particularly high-stakes use of TOEIC tests.

Idealized AUAs were constructed for TOEIC Speaking, Writing, Listening and Reading, and Bridge tests, with adaptations to Bachman and Palmer's (2010) generalized AUA structure based on the particular design of each test and the intended use. For example, the TOEIC Listening and Reading tests do not use human raters, so warrants about inter- or intrarater consistency are not relevant to support the claim that TOEIC Listening and Reading scores are consistent. As another example, TOEIC tests are intended to be used with other criteria to facilitate hiring decisions (see ETS, 2013, p. 26), so a warrant that "TOEIC score interpretations provide sufficient information to facilitate hiring decisions" was not included in the AUAs constructed for this particular use.

Steps 2 and 3: Collect Evidence and Relate It to Claims and Warrants

The second step in the process illustrated in Figure 2 involved the collection and synthesis of evidence from the test design process, ongoing statistical and procedural monitoring, and research activity. Evidence from the test design process included documentation that was produced as part of the evidence-centered design process (see Hines, 2010; Schedl, 2010) and documentation that described test administration and scoring procedures. This included the initial justification for the definitions of the abilities measured by each test, as well as the item and test specifications. This documentation was synthesized and used to support a variety of warrants in the AUA, including warrants to support claims about the consistency of scores, plus the meaningfulness, impartiality, and generalizability of score interpretations.

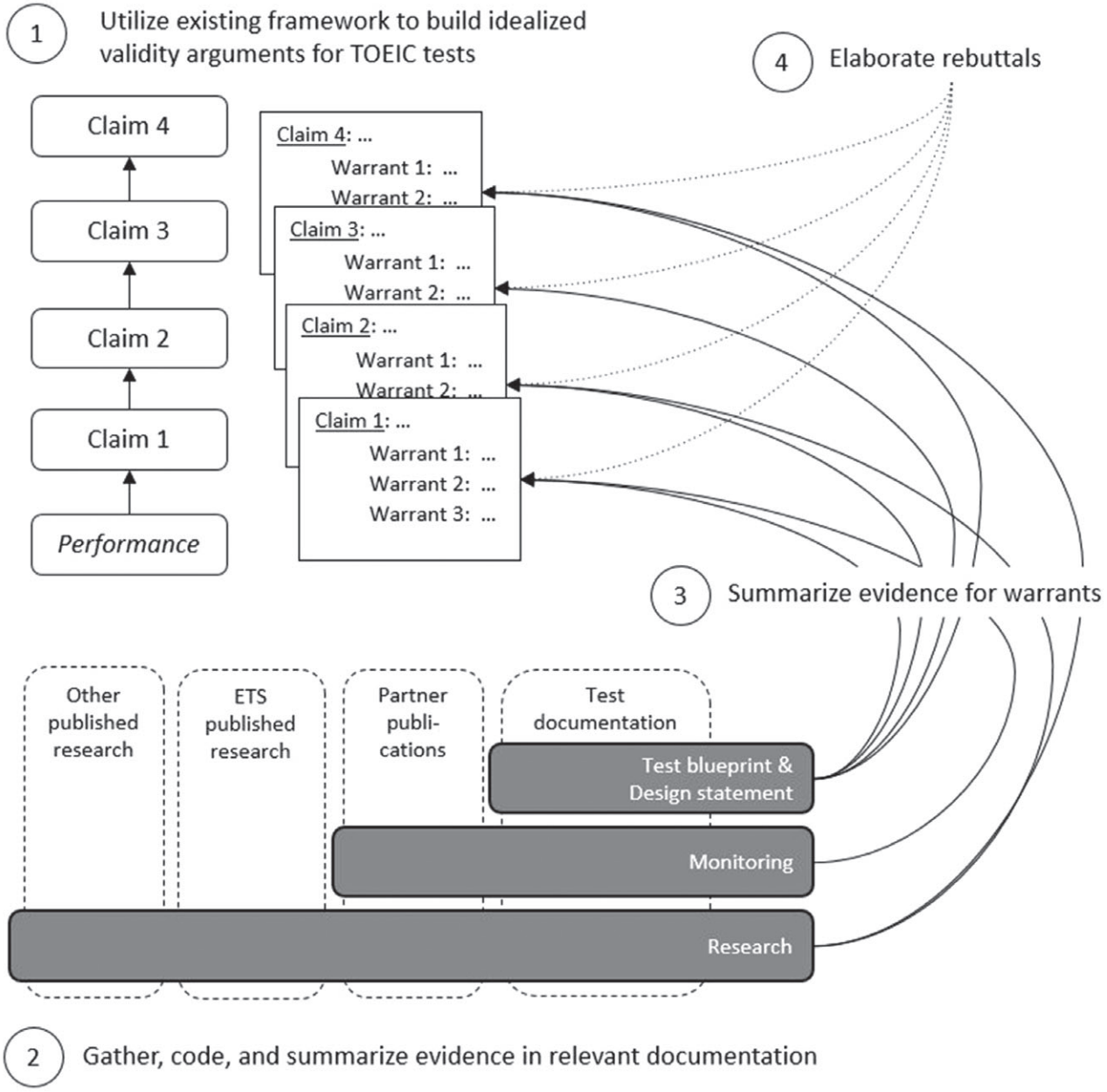

Figure 2 Overview of the process for creating TOEIC test assessment use arguments.

This documentation was produced entirely by the test developer (ETS), and much of it is confidential (e.g., item and test specifications).

The second type of evidence synthesized and summarized in each AUA derived from ongoing statistical and procedural monitoring. Most of this documentation was produced by ETS and includes statistical monitoring such as the stability of scores across test forms and administrations and potential changes in the demographic characteristics of the test-taker population. Procedural monitoring occurred as well and potentially included feedback provided to the test developer by test takers, score users, and local partners on test administration, security, the use of scores.

The final type of evidence included in each AUA derived from research and review articles published by ETS, its partners, trade journals, or individual researchers. The first two edited volumes in the *TOEIC Compendium* included more than 20 papers, each of which contributed evidence to support various warrants in TOEIC AUAs. Additional research and practitioner publications were identified periodically through manual searches of journals in language

assessment and keyword searches using Google Scholar. For example, periodic searches between June 2014 and June 2017 identified 113 publications that explicitly mentioned TOEIC tests, of which 76 were reviewed and coded for their relevance to TOEIC AUAs. Publications were excluded when their mention of TOEIC tests was cursory or without consequence for an AUA; for example, Lawn and Lawn (2015) mentioned TOEIC tests as an example of an English language assessment.

The 76 publications identified as relevant to TOEIC AUAs were reviewed and coded, and their findings or claims were incorporated as evidence (backing) or criticism (rebuttal) to a relevant warrant. Publications were coded based on the TOEIC tests to which they applied (Reading, Listening, Speaking, Writing, Bridge, unidentified), their substantive focus (reliability, validity, test use, test review), and local context (e.g., Japan, Korea). The vast majority of publications pertained to the TOEIC Listening (78%) and Reading (75%) tests; less than 10% pertained to the TOEIC Speaking, Writing, or Bridge tests. Publications varied in their substantive focal points, although many focused on issues pertaining to test use (51%) and validity (41%). Very few publications focused on reliability or score consistency (5%), which is not surprising; ideally, this quality of test scores should be examined under operational conditions and, thus, is primarily the responsibility of the test developer (ETS) and its local partners. The publications included in the review varied from unpublished graduate student papers to publications in international peer-reviewed journals, and around 13% of the publications were TOEIC test reviews. Most of the publications were published by researchers or practitioners in Japan (70%). Other local contexts included Korea (14%), Taiwan (8%), and China, Costa Rica, Indonesia, Thailand, and Vietnam (each less than 5%).

Step 4: Elaborate Rebuttals and Evaluate the Overall Plausibility of Assessment Use Argument

The final step in the process was to critically examine the existing evidence for each warrant and evaluate the overall plausibility of each AUA. Prior to this critical exercise, some potential rebuttals had already been documented based on the review of research and practitioner publications. For example, in a small-scale study of the impact of TOEIC Reading and Listening test use at a business school in Thailand, Apichatronajanakul (2011) found evidence of both positive and negative washback. Evidence of negative washback constituted a potential rebuttal to the warrant that the consequences of using TOEIC Listening and Reading tests would be beneficial to test takers, which could potentially undermine the overall claim that the consequences of TOEIC Listening and Reading test use are beneficial. When evidence for a particular warrant was mixed or mostly lacking, it underscored the need to consider the seriousness of existing or potential rebuttals and their impact on a broader claim about the measurement quality or use of the test.

Uses of Validity Arguments for TOEIC Tests

The fully specified AUAs reflect a broad consideration of evidence to support uses of TOEIC tests and have been developed with two primary applications in mind. First, they have been used to help inform a research agenda for the TOEIC program. Research is critical for supporting claims about the measurement quality and intended use of tests, but all test developers have limited resources. Based on critical evaluations of fully specified TOEIC AUAs, one area of research that the TOEIC program has pursued over the last several years has been focused on the uses of TOEIC tests and their potential impact on various stakeholder groups; several chapters in the third edition of the *TOEIC Compendium* address these focal issues (e.g., Hsieh, 2017; Oliveri & Tannenbaum, 2017).

The other application of the fully specified AUAs has been the creation of simplified versions aligned with the needs of different stakeholders. Although the wording of claims and warrants in an AUA are designed to be accessible to nonexperts, reviewing and evaluating a fully specified AUA requires a significant investment on the part of the reader. Simplified versions allow the fully specified AUA to be condensed and adapted with a particular readership in mind. For example, all test programs at ETS are periodically audited to ensure their compliance with professional standards. This compliance includes adhering to the *ETS Standards for Quality and Fairness* (2015). Given that test auditors are familiar with these standards, evidence presented in the simplified AUA is directly related to various ETS standards.

An extremely condensed version of the AUA is presented on the TOEIC research website (https://www.ets.org/toeic/research). Here, only the high-level claims (illustrated in Figure 1) are presented under the assumption that many readers will have an extremely limited assessment literacy. By focusing on the four fundamental, high-level claims, TOEIC research intends to communicate the key elements of an argument for test use to a broad audience. The four descriptive categories corresponding to the four high-level AUA claims on the TOEIC research website are shown in Figure 3.

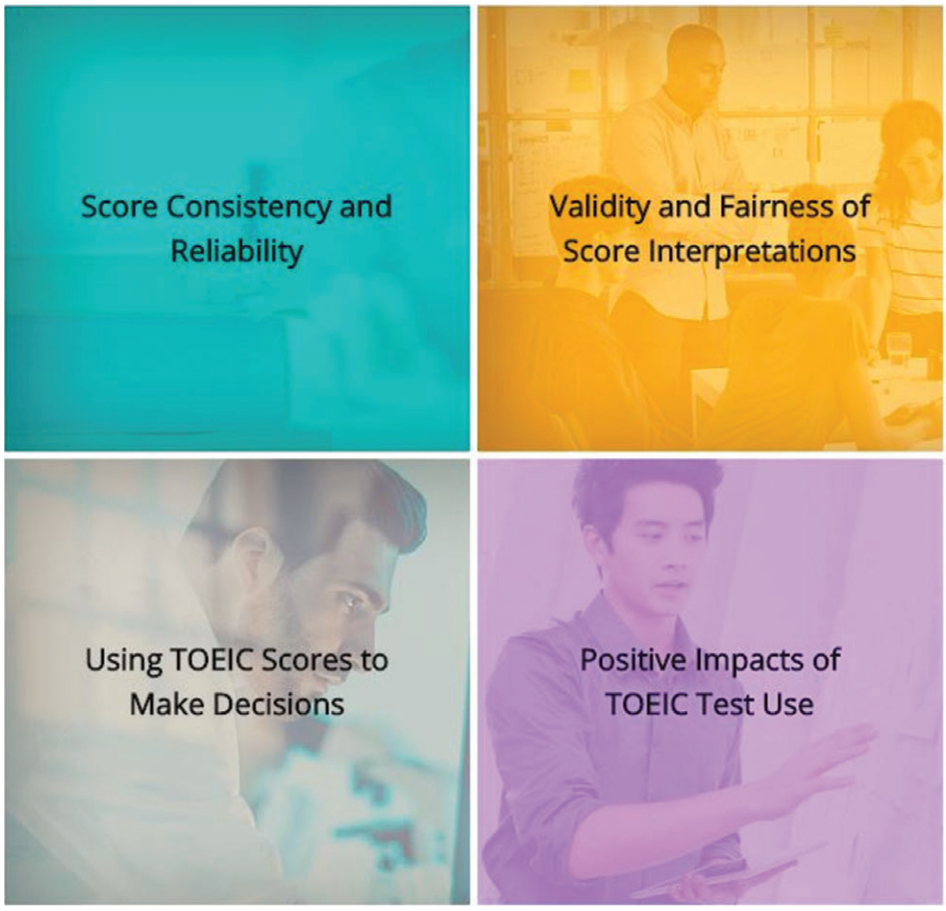

Figure 3 Website categories corresponding to the four high-level assessment use argument claims.

The four boxes shown in Figure 3 roughly correspond to each of the four overall claims in an AUA. If a website user places the mouse on one of the boxes, text appears to summarize the corresponding claim. For example, the following text corresponds to the score consistency and reliability category: "TOEIC scores are consistent and reliable, and are not improperly influenced by factors unrelated to language ability."

Website users who are interested in more information may click on one of the four categories or explanatory text (e.g., "TOEIC scores are consistent and reliable, and are not improperly influenced by factors unrelated to language ability") to read a brief and accessible summary of the types of warrants that support the overall claim (e.g., scores are consistent across test items, test forms, test administrations, raters) and review some of the evidence that is available to support claims and warrants. Figure 4 below illustrates how this has been implemented for the score consistency and reliability category.

As illustrated in Figure 4, each category includes a restatement of the overall claim, an outline of warrants that support the claim, and a list of relevant research evidence. An executive summary is provided for each research paper. This summary includes the purpose of the study, the evidence it produced, and the implications of the evidence for relevant claims. A link is provided to an electronic copy of the research publication for those that are interested.

Discussion

This paper provided a rationale for the argument-based approach to validation and an overview of the AUA, one such approach. It highlighted how this approach has been implemented in a novel way to produce fully specified AUAs for the TOEIC tests, which are then applied to guide TOEIC test research and disseminate the argument for TOEIC test use to various stakeholders. One of the purposes of providing simplified AUAs is to increase the assessment literacy of different stakeholder groups, including test takers and score users. The current design of the TOEIC research website

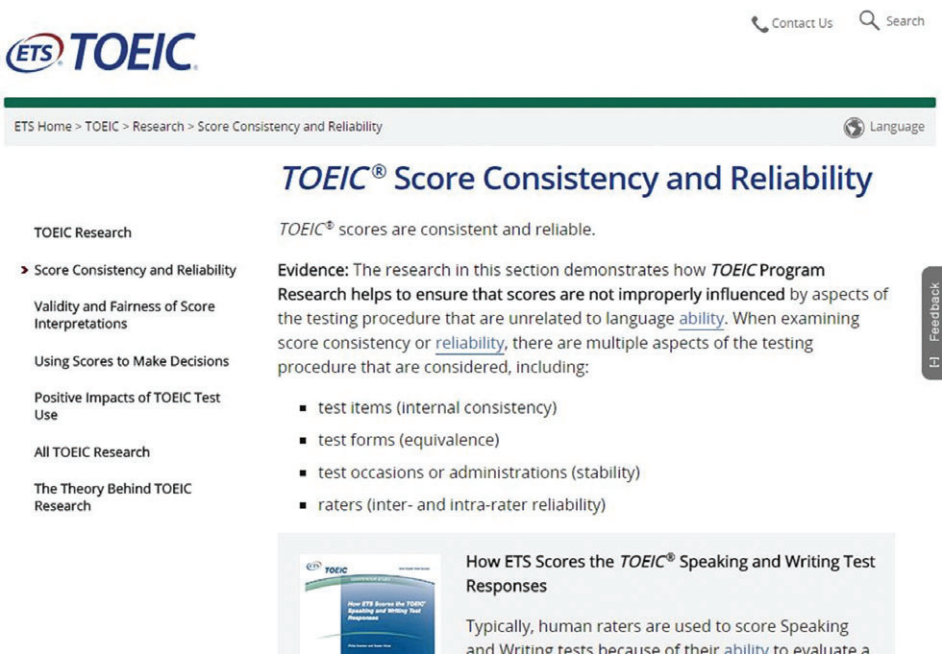

Figure 4 Descriptive text and information on the website to support the claim that scores are consistent and reliable.

reflects this intention. In the future, researchers at ETS hope to report on the effectiveness of the simplified AUA for promoting assessment literacy.

For tests that have multiple uses—such as the TOEIC tests—one of the challenges of using an argument-based approach to justifying test use is that it may require a number of individual AUAs. This implies a lot of documentation that could be difficult to create, maintain, and adapt for the purpose of communicating with different stakeholders. However, there is a potential solution to this challenge: a theory of action.

A theory of action is a logical model of how components of a test (e.g., test scores) can facilitate actions (i.e., decisions) that have intermediate and long-term outcomes (i.e., consequences). As exemplified by Bennett (2010), it includes a visualization that functions as a high-level summary of all of the intended uses of an assessment and their expected consequences. In a single figure, it could provide an accessible summary of the supported uses of a TOEIC test and the expected consequences of test use. Such a figure would also indicate the relationship between test components (e.g., scores), decisions, and consequences. Supporting documentation is expected to summarize the evidence to support each hypothesized relationship in the logic model, including potential rebuttals. Thus, the future publication of a theory of action for TOEIC tests may be a beneficial tool to communicate claims and supporting evidence about TOEIC test use in an accessible manner.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Apichatronajanakul, P. (2011). The washback effects of the TOEIC examination on the teachers and students of a Thai business school. *Language Testing in Asia, 1*(1), 62–75.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford, UK: Oxford University Press.

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement, 8*, 70–91.

Chapelle, C. A., & Voss, E. (2014). Evaluation of language tests through validation research. In A. Kunnan (Ed.), *The companion to language assessment*. New York, NY: Wiley.

Educational Testing Service (2013). *TOEIC user guide: Speaking and writing*. Princeton, NJ: Educational Testing Service.

Educational Testing Service (2015). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service.

Hines, S. (2010). Evidence-centered design: The TOEIC Speaking and Writing tests. In D. Powers (Ed.), *TOEIC compendium* (1st ed., pp. 7.1 – 7.31). Princeton, NJ: Educational Testing Service.

Hsieh, C.-N. (2017). *The case of Taiwan: Perceptions of college students about the use of* TOEIC® *tests to graduate* (Research Report No. RR-17-45. Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12179

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.). New York, NY: American Council on Education and Praeger.

Lawn, M. J., & Lawn, E. (2015). Increasing English communicative competence through online English conversation blended e-learning. *International Journal of Information and Education Technology, 5*(2), 105 – 112.

Messick, S. J. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13 – 103). New York, NY: Macmillan.

Newton, P. E. (2012). Clarifying the consensus definition of validity. *Measurement: Interdisciplinary Research and Perspectives, 10*(1 – 2), 1 – 29.

Oliveri, M. E., & Tannenbaum, R. J. (2017). *Insights into the use of TOEIC® scores to inform human resource management decisions* (Research Report No. RR-17-48). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12177

Powers, D. E. (2010). Validity: What does it mean for the TOEIC tests? In D. Powers (Ed.), *TOEIC compendium* (1st ed., pp. 1.1 – 1.11). Princeton, NJ: Educational Testing Service.

Schedl, M. (2010). Background and goals of the TOEIC Listening and Reading test redesign project. In D. Powers (Ed.), *TOEIC compendium* (1st ed., pp. 2.1 – 2.18). Princeton, NJ: Educational Testing Service.

Schmidgall, J. E., & Choi, I. K. (2011, May). *Frameworks for validity: A comparison of traditional and argument-based approaches for reviewing research*. Paper presented at the 14th annual conference of the Southern California Association for Language Assessment Researchers, Los Angeles, CA.

Toulmin, S. E. (2003). *The uses of argument* (updated ed.). Cambridge, UK: Cambridge University Press.

Wang, H., Choi, I., Schmidgall, J., & Bachman, L. F. (2012). Review of Pearson test of English academic: Building an assessment use argument. *Language Testing, 29*(4), 603 – 619.

Suggested Citation

Schmidgall, J. E. (2017). *Articulating and evaluating validity arguments for the* TOEIC® *tests* (Research Report No. RR-17-51). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12182

Action Editor: Donald Powers

Reviewers: Ikkyu Choi and Veronika Timpe-Laughlin