



The Feasibility of Testing Academic Literacy in EAP Reading Comprehension Tests

Esmat Babaii

Kharazmi University, Tehran, Iran

Description of the Problem

ESP and its celebrated offspring EAP (English for Academic Purposes) are now an indispensable part of the English language teaching enterprise in almost every major academic center around the globe. A brief review of the EAP literature reveals considerable developments in the areas of materials preparation and teaching methodologies (Charles & Pecorari, 2016; Guardado & Light, 2020; Hyland, 2006). These developments and modifications in syllabi and teaching procedures, which were informed by and in line with recent re-conceptualizations in the field, call for a new type of evaluation scheme. In'nami et al. (2016), for instance, acknowledge a need for constructing and validating an English academic test battery to measure a construct they call “academic English proficiency required for learning and researching at universities” (p. 2). In their article, nevertheless, they do not explicitly present a set of *a priori* assumptions and arguments about the characteristics of such a test and they only provide *a posteriori* statistical examination of the Test of English for Academic Purposes (TEAP) they have constructed for Japanese university students. Hence, examining TEAP in terms of what Weir (2005) refers to as “*a priori* validity evidence” as theory-based specifications “before the test event” (p. 17) is not possible here. Nevertheless, their efforts in recognizing and dealing with EAP testing as an issue, *per se*, can be considered a step forward.

The evaluation component of the puzzle of EAP curriculum, at least in some contexts such as Iranian higher education, does not seem to fit in as it has hardly been able to adjust to the new developments in EAP. It is disappointing to see how testing specialists have remained negligent to the advancements in EAP scholarship and tend to continue using traditional subject-specific text comprehension, and to a lesser extent, production, as the most frequent method for testing language proficiency of EAP students. Of course, some EAP instructors are now benefiting from more innovative and involving, albeit informal, assessment techniques in their classes (see Cochran, 2004, for instance); however, this is NOT the *status quo* for the majority of large-scale EAP tests currently in use. The literature on EAP testing is replete with articles and book chapters challenging or supporting the contribution of ‘topical subject-specific knowledge’ to test takers’ test performance. For more than three decades, EAP testing has hardly stepped beyond tests of ‘specialized CONTENT’ as comfortable, familiar operationalization of EAP knowledge (cf. Alderson & Urquhart, 1985a & b; Alderson, 1988; Ansary & Babaii, 2002; Clapham, 1996; Davies, 2001; Krekeler, 2006, Tan, 1990 among many others). In fact, in contrast to non-specific general language tests where content/background knowledge tends to be “viewed as a confounding factor, masking “true” language ability” (Douglas, 2013, p. 369, emphasis in the original), selecting discipline-



relevant text for EAP reading tests has been a common practice to ensure face validity, at least. To the best of my knowledge, the condition of EAP testing has not improved drastically over these years. Even in the most recent works of EAP that I could access, EAP evaluation was discussed only in terms of some general points and suggestions (for example Charles & Pecorari, 2016; Douglas, 2013; Guardado & Light, 2020). The undisturbed, continued use of the conventional tests, nonetheless, should not imply that EAP testing is in its ideal shape. To further the discussion, I will try to tackle the issue by focusing on EAP reading comprehension tests and the abilities they are expected to measure.

Hyland (2005, 2006) candidly mentions achieving ‘academic literacy’ as the top item on the EAP agenda—“empower[ing] learners by initiating them into the ways of making meanings that are valued in their target courses and disciplines” (Hyland, 2006, p. 31), an ability which is no longer a non-native learner’s issue: English native speakers are also among the audience of this new approach to EAP pedagogy. According to Hyland (2006, p. 31) “students and academics alike are judged by their control of the discourses of their discipline” and it seems unlikely that the ‘control of disciplinary discourses’ Hyland (2006) is talking about can be assessed by traditional tests. I believe that evaluating the end product of EAP courses ONLY through measuring the comprehension of some subject-specific texts is falling short of catching up with the recent developments witnessed in EAP scholarship. Such tests could merely touch a small portion of the ‘academic literacy’ we expect the learners to achieve. The negative washback of this limited type of tests—to be more specific, narrowing the curriculum (Messick, 1996)—is far too obvious.

It should be emphasized at the outset that my argument does not have much to do with the long lasting debate over a wide-angled vs. narrow-angled approach to EAP (Basturkmen, 2003; Williams, 1978) and we can leave the question of ‘degree of specificity’, i.e., selecting discipline-specific or general academic content, to the test constructor—the choice that is relevant to his/her approach to EAP. The problem diagnosed here simply points to construct under-representation (cf. Messick, 1989, 1996) of the current EAP tests ending up in the negative washback of such tests for the curriculum. To elaborate on this argument, it should first be clarified what is supposed to be measured by an EAP reading test: (a) the specialized content knowledge of a given discipline or (b) the academic discourse through which the ideas, concepts and findings are presented and discussed in the mentioned discipline? My position is that measuring (a) cannot and should not be our purpose in language testing. We, as language testing practitioners, should be concerned with examining whether reading comprehension EAP tests can tap the process of ‘unfolding the meaning’ in abstract and complex academic texts which is characterized by certain textual features such as nominalization, grammatical metaphor, high lexical density and high degree of formality and impersonalization (cf. Charles & Pecorari, 2016; Martin, 1996, 2005).

In their introductory book to EAP, Charles and Pecorari (2016) offer several useful, discourse-informed procedures for teaching academic oral and written genres to EAP students, admitting that academic texts need unpacking and elaboration but, unfortunately, they remain relatively silent when it comes to testing EAP reading comprehension. The point is that if we admit that teaching academic texts is different from teaching general reading (Charles & Pecorari, 2016, p. 131) why should we not try to treat testing EAP reading differently as well? If we design and implement an EAP curriculum informed by recent developments in academic literacy and genre-based pedagogy but we continue using traditional reading comprehension tests which do not align with the curriculum’s objectives, gradually those objectives which we fail to measure in our tests will be de-emphasized, and then, put aside by the teachers and the students. Consequently, under-representation of academic literacy could lead to narrowing the curriculum, leaving out some important aspects of academic literacy and discourse knowledge that are not included in the tests.

To illustrate this point, I will try to analyze a typical, traditional TOEFL-like test commonly encountered in the Iranian EAP context (Sample Test 1). This EAP reading comprehension test consists of a passage related to the field of civil engineering followed by not-so-sophisticated multiple-choice questions that, nine out of ten, can be tackled without a full understanding of the passage by most Iranian students of civil engineering (as I have observed in several academic settings). To me, this is probably not a good candidate for testing academic literacy.

Sample Test (1)

There are two main types of concrete dams: arch dams and gravity dams. Arch dams are tall, curved shells of concrete that can be as little as 3 meters thick. Their arch shape gives them great strength. Large gravity dams are also made of concrete, but it is their vast weight that prevents them from bursting.

The largest dams are embankment dams, which are made by piling up a huge barrier of earth and rock. A core of clay or concrete in the middle keeps water from seeping through the dam. The side is covered with stone to protect it from water. Rogunsky Dam in the Soviet Union is the world's highest dam. It is 325 meters high. Hoover Dam, one of the world's highest concrete dams, measures 221 meters in height. It is an arch dam that spans the Colorado River and supplies water for irrigation and electricity to California, Arizona, and Nevada.

1. The “great strength” referred to in line 3 is a result of
 - A. size
 - B. weight
 - C. shape
 - D. concrete

2. It can be inferred from the passage that gravity dams differ from arch dams in that gravity dams are
 - A. thicker than arch dams
 - B. not as sturdy as arch dams
 - C. not as attractive as arch dams
 - D. made of a different material than arch dams

3. According to the passage, the core of clay in the center of an embankment dam serves which purpose?
 - A. To support the structure
 - B. To hold the side together
 - C. To form the shape of the dam
 - D. To prevent water seepage

4. According to the passage, how tall is the Hoover Dam?
 - A. 325 meters
 - B. 185 kilometers
 - C. 221 meters
 - D. 3 meters

5. According to the passage, the water supplied from Hoover Dam is used for
 - A. strength and support
 - B. irrigation and electricity
 - C. protection and irrigation
 - D. electricity and support

Taken from ‘*Entrance Examination preparation course book for students of civil engineering*’, compiled by M. Parna (2006, p. 279-280)

A noteworthy feature of the passages found in most subject-specific EAP reading modules, as pointed out by Krekeler (2006, p. 104), is that although “they clearly deal with a relevant academic subject [...] they explain key terminology and concepts and do not assume previous knowledge of the field.” As a consequence, they enjoy a low degree of technicality and the EAP test taker is usually so familiar with the content that he/she can answer the questions without reading the passage. This comment specifically applies to Items 1, 2 and 3 in Sample Test (1) as the concepts are covered by most students of engineering at their early stages of BS programs. This, *per se*, leads to low interactiveness of the test as it fails to

involve the test taker's cognitive and linguistic abilities (Bachman & Palmer, 1996). In other words, a typical BS student of engineering can answer these items correctly without activating the linguistic resources needed for text comprehension because the technical background knowledge compensates for possible incompetence in language ability. So, one may wonder what aspect of reading ability is supposed to be tapped by these items: items 1-3 are about the differences between a gravity and an arch dam: too rudimentary for a senior student of civil engineering; item 4 can be answered by spotting 'Hoover dam' and its corresponding measure of length; and item 5 is only checking common sense.

The rationale for this type of test probably lies in their face validity and apparent authenticity (Fulcher, 1999). Such tests, adopting what Bachman (1990) calls a 'Real-Life' approach to the authenticity of assessment, attempt to copy or simulate the test takers' Target Language Use (TLU) domain in their design. Accordingly, the learners' comprehension and production of certain subject-specific texts (mostly written) are considered as indicators of their ability to handle academic tasks in English. Interestingly, the simulation is not perfect, as it only tries to copy the 'input' as is, with no attention to characterizing the EAP task, defining it in terms of its constituents and the discursive processes underlying the text's decoding and encoding. This approach towards testing EAP can be criticized in two respects:

1. It puts undue emphasis on the role of background knowledge at the expense of other linguistic, discursive and cognitive capacities needed to handle EAP tasks effectively. As such, the construct validity of these tests is under question due to 'construct under-representation'. This will, in turn, lead to a well-known instance of negative washback called 'narrowing the curriculum' which means that EAP teachers and students will be encouraged to deal with those educational objectives that are examined in the tests and downplay or ignore those which are left out.
2. It obscures the main objective of EAP programs, which is *equipping the learners with academic literacy not teaching them specialized content*. In fact, we tend to forget that EAP students' mastery over content knowledge is to be checked in their specialized courses and by content teachers not in EAP courses through English tests. EAP testing, rather than indulgence in subject-matter, should concern itself with the link between discourse and knowledge construction in academic texts and try to check whether EAP students have mastery over utilizing this ability.

Speaking about advanced academic literacy, Tardy (2005) contends that in later stages of literacy at universities "students need to learn ways of thinking about, interacting with, and constructing knowledge of disciplinary communities and content" (p. 326). She considers these as requirements for disciplinary community membership. In line with her argument, I believe that this level of literacy requires a deep understanding of rhetorical and discursive resources through which scientific ideas and practices are disseminated and/or challenged. Even though academic literacy has proved to be a complicated construct to define (Tardy, 2005), EAP testing will not benefit from neglecting the notion and denying its necessity. Offering field-specific, context-based definitions of academic literacy may seem a modest achievement in comparison to grand theory-building endeavors, but it can potentially lead to constructing more evidentially and consequentially valid EAP tests. In what follows, far from any claim for finding panacea, I intend to make some suggestions to pave the way for more defensible and useful EAP tests. This report argues that EAP testing can, and in fact, should benefit from the current works in academic literacy as well as discourse analysis of professional genres used in academia.

An Alternative Approach to EAP Testing

The hard fact about EAP context is that EAP students will need to understand certain written genres like textbooks, research articles, book reviews, dissertations, and comprehend some oral genres such as lectures, seminars, dissertation defense sessions, while they are also expected to produce a range of genres including oral presentations, case study reports, laboratory reports, essays, reaction papers and

project proposals. Genre analysis research has explored and documented the macro-structures and stylistic features of most of these academic genres, revealing cross-disciplinary as well as cross-cultural similarities and differences (a long list of sources, too many to mention here, can be found by browsing applied linguistics journals, especially *Journal of English for Academic Purposes* and *Journal of English for Specific Purposes*). In what follows, I attempt to present a blueprint for constructing EAP tests which is informed by the current research into the nature of academic genres and activities:

1. A detailed and in-depth analysis of current and future experiences of EAP students in a given discipline should be provided. This would lead to a thorough description of discursive activities expected to be handled by a successful candidate (see Kuo, 1993, as a good example of identifying the needed skills for listening comprehension of science lectures). Here, instead of prescribing a test simply consisting of simulated activities, attempts should be made to explore and deconstruct the activities to find their salient characteristics in terms of rhetorical functions, logical and/or temporal order, stylistic features, register, and linguistic as well as textual peculiarities. The process, perhaps needless to say, would not rely on armchair reflection but involve triangulation of data obtained through different techniques including observation of current and target language situations, inquiring subject experts as well as students about students' current and target language-related needs, as well as content analysis of the texts used in the program.

2. Specifications of each test task in terms of measurable behavioral objectives should be provided. General descriptions such as 'The student can complete a lab report' or 'The student can write an argumentative essay' are to be avoided since they are too broad, and would encourage subjective interpretations. The following examples, as components of different academic genres, can better lend themselves to explicit and objective measurement (see Table 1).

TABLE 1

Examples of Behavioral Objectives

-
- (i) The student can re-state the information displayed by a line graph.
 - (ii) The student can paraphrase lexically-dense, nominalized sentences using simple sentences.
 - (iii) The student can recognize and highlight the persuasive expressions/clauses in the paragraph.
 - (iv) The student is able to summarize the main elements of argumentation in the essay.
 - (v) The student is able to arrange the clauses/sentences according to a logical order.
 - (vi) The student can follow the cohesive devices.
 - (vii) The student can process and unpack long, embedded sentences.
 - (ix) The student can identify cues for new ideas, topic shift, and concluding remarks.
 - (x) The student can recognize organizational pattern (e.g., general to specific, cause-effect, compare and contrast)
-

3. The criteria for the successful completion of each task, and the weight assigned to it as compared to other tasks within the context of the genre it belongs to should be determined by a panel of EAP practitioners together with subject specialists. For instance, the panel may decide to assign more importance to 'the ability to understand visual information displayed in graphs and figures' in science and engineering textbooks than in humanities texts due to higher incidence of this type of information display in the former materials.

4. The specifications and criteria outlined in the previous stages should be translated to test tasks constructed and piloted to a sample of EAP students similar to the target audience. The piloting phase will help the test constructors to examine the feasibility and usefulness of the test for the target test takers and make the necessary modifications to improve the test quality for further use.

In Sample Test (2), I attempt to assess certain linguistic and textual features such as knowledge of connectives, text-patterning, paragraph macro-organization and inferencing. I assume that it would be unlikely to provide correct responses to the questions without reading the passage carefully and utilize one's linguistic and textual knowledge. If this assumption is correct, then the test would enjoy a high

degree of interactiveness and can well represent the complexity of an EAP reading task. As such, it would have positive washback for EAP teaching and learning as well.

Sample Test 2

Water-generated power is clean and renewable, **and** early industrial communities were sited along rivers to use waterpower. However, the huge dams required to generate electricity from water (hydroelectric energy) can displace people from their homes, devastate vast areas through flooding, and decimate fish populations that need open rivers to spawn. A recent development by Northeastern University engineering professor emeritus Alexander Gorlov represents a new solution to the problem of how to harness hydropower without building a dam. Gorlov's helical turbine design is potentially suitable for use in most of the world's rivers, as well as in ocean channels. They have been installed in test installations in Brazil and in Vinalhaven, Maine.*

1. Which of the following words could replace 'and' in linking the two parts of the first sentence?
(Because, that's why, for instance, in conclusion)
2. The second sentence starts with 'however', because the author wants to make a 'contrast' to the positive features of water-produced energy. Name at least two of the disadvantages of constructing huge dams.
3. Using one of these connectors: (as a result, nevertheless, in fact, in sum), link the first sentence to the following sentence
[There are countries that are economically dependent on waterpower]
4. In the next paragraph, the author would mostly probably
 - a. describe Dr. Gorlov's invention and how it works
 - b. talk about other sources of energy like wind
 - c. detail the disadvantages of constructing dams
5. What is the function of "They have been installed in test installations in Brazil and in Vinalhaven, Maine" in the last sentence?
(evidence, a supporting statement or an example?)
6. Underline the clauses that can make the "generalization—examples" pattern with this sentence:
[Constructing dams to use hydropower can damage the environment.]

* Passage taken from *Building systems for interior designers* by C. Binggeli (2003, p. 163)

Concluding Remarks

This paper was an attempt to suggest a different approach towards testing academic literacy. Far from any claim for perfection, the sample test was provided to demonstrate the feasibility of constructing more interactive EAP tests. I hope that this paper would encourage serious discussions about the nature of academic literacy, leading to lucid characterization of the academic tasks that should be included in EAP test design. It should also be pointed out that the mastery over academic genres can be, and are being, checked by examining the students' final products or the portfolios made during the course. With no intention to question the value of formative techniques like portfolio assessment, it is my conviction that description of TLU must find its way to EAP test specification as well. New research projects may be launched to examine the following relevant issues: how EAP content teachers, EAP TEFL instructors and students would perceive the type of items suggested in Sample Test 2; whether Sample Test 2 can really function better than Sample Test 1 in terms of authenticity, interactiveness and positive washback,

etc. The results of this line of research could help revise the specifications, and consequently, the content of future of EAP tests.

The Author

Esmat Babaii is Professor of Applied Linguistics at Kharazmi University, Iran, where she teaches research methods, language assessment and discourse analysis to graduate students. She has published articles and book chapters dealing with issues in Language assessment, Systemic Functional Linguistics, test-taking processes, and critical approaches to the study of culture and language. Her most recent work is a co-authored article about the treatment of L1 pragmatic cultural schema in pragmatic assessment done by non-native teachers, published in *TESL-EJ* (May, 2021).

Department of Foreign Languages,
Kharazmi University
No. 43, Mofatteh St., Tehran 15719-14911, Iran
email: babai@khu.ac.ir

References

- Alderson, C. (1988). Testing English for specific purposes: How specific can we get? In A. Hughes (Ed.), *Testing English for university study* (pp. 16-28). Modern English Publications and British Council.
- Alderson, C., & Urquhart, A. (1985a). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing*, 2, 192-204.
- Alderson, C., & Urquhart, A. (1985b). This test is unfair: I'm not an economist. In P. Hauptman, R. LeBlanc, & M. Bingham-Wesche (Eds.), *Second language performance testing* (pp. 25-45). University of Ottawa Press.
- Ansary, H., & Babaii, E. (2002). Topic variable in narrow-scope EAP short-context reading tests. *RELC Journal*, 33(1), 55-74.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford University Press.
- Basturkmen, H. (2003). Specificity and ESP course design. *RELC Journal*, 34(1), 48-63.
- Binggeli, C. (2003). *Building systems for interior designers*. John Wiley & Sons, Inc.
- Charles, M., & Pecorari, D. (2016). *Introducing English for academic purposes*. Routledge
- Clapham, C. (1996). *The development of IELTS: A study of the effect of background knowledge on reading comprehension (Studies in Language Testing)* (Vol. 4). Cambridge University Press.
- Cochran, M. (2004). "Cynical politics" versus "scientific truth": Communication crisis at Brookhaven National Lab. *IEEE Transactions on Professional Communication*, 47, 256-267.
- Davies, A. (2001). The logic of testing languages for specific purposes. *Language Testing*, 18, 131-147.
- Douglas, D. (2013). ESP and assessment. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 367-384). Wiley-Blackwell.
- Fulcher, G. (1999). Assessment in English for academic purposes: Putting content validity in its place. *Applied Linguistics*, 20(2), 221-236.
- Guardado, M., & Light, J. (2020). *Curriculum development in English for academic purposes*. Palgrave Macmillan.
- Hyland, K. (2005). Perspectives on EAP: An interview with Ken Hyland. *ELT Journal*, 59, 57-64.
- Hyland, K. (2006). *English for academic purposes: An advanced resource book*. Routledge.
- In'nami, Y., Koizumi, R., & Nakamura, K. (2016). Factor structure of the Test of English for Academic Purposes (TEAP®) test in relation to the TOEFL iBT® test. *Language Testing in Asia*, 6(3), 1-23.

- Krekeler, C. (2006). Language for special academic purposes (LSAP) testing: The effect of background knowledge revisited. *Language Testing*, 23(1), 99-130.
- Kuo, C. (1993). Teaching lecture comprehension to non-native science students. *IEEE Transactions on Professional Communication*, 36, 70-74.
- Martin, J. (1996). Waves of abstraction: Organizing exposition. *The Journal of TESOL France*, 3, 87-104.
- Martin, J. (2005). Discourses of science. In J. Martin & R. Veel (Eds.), *Reading science* (pp. 3-14). Routledge.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 13-103). Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
- Parna, M. (2006). *Entrance examination preparation course book for students of civil engineering*. Azade Publication.
- Tan, S. (1990). The role of prior knowledge and language proficiency as predictors of reading comprehension among undergraduates. In J. de Jong & D. Stevenson (Eds.), *Individualizing the assessment of language abilities* (pp. 214-224). Multilingual Matters.
- Tardy, C. (2005). "It's like a story": Rhetorical knowledge development in advanced academic literacy. *Journal of English for Academic Purposes*, 4, 325-338.
- Weir, C. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.
- Williams, R. (1978). EST: Is it on the right track? *MALS Journal*, Summer, 23-28.

(Received October 16, 2021; Revised November 20, 2021; Accepted December 18, 2021)