

Article

Evaluating Perceptions towards the Consequential Validity of Integrated Language Proficiency Assessment

Asli Lidice Gokturk Saglam ^{1,*}  and Dina Tsagari ²¹ Department of Language and Literature Studies, University of South Eastern Norway, 3184 Borre, Norway² Department of Primary and Secondary Teacher Education, Oslo Metropolitan University, 0167 Oslo, Norway; dints@oslomet.no

* Correspondence: asli.saglam@usn.no

Abstract: This research study explores teacher and student perceptions to verify consequential validity and the potential washback effect of a locally developed university-level English language proficiency test which consists of reading and listening-to-writing assessment tasks. The integrated language proficiency test is used upon completion of the English language preparatory program in the Turkish context to determine learners' access to further English medium academic courses in their departments. To examine whether this source-based proficiency test has achieved its intended outcomes, 39 freshman students and 19 university instructors, who offered courses in various departments, were surveyed through questionnaires. Interviews were conducted with the instructors to gauge their perspectives about the validity of the integrated proficiency test-based decisions (whether students pass or fail) over time in terms of the language competency and academic skills of their learners. Quantitative and qualitative data analysis also revealed evidence both for positive and negative issues concerning the consequential validity of the test. Findings may help educators to reach a better understanding of the construct of integrated language assessment tasks in EAP contexts and the consequences of their use in achieving the intended and unintended curricular goals.

Keywords: consequential validity; integrated language assessment; test consequences; teacher perceptions; student perceptions; washback; EAP



Citation: Gokturk Saglam, Asli Lidice, and Dina Tsagari. 2022. Evaluating Perceptions towards the Consequential Validity of Integrated Language Proficiency Assessment. *Languages* 7: 65. <https://doi.org/10.3390/languages7010065>

Academic Editors: Juana M. Licerias and Raquel Fernández Fuertes

Received: 23 November 2021

Accepted: 24 February 2022

Published: 9 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The assessment of English for Academic Purposes (EAP) integrated assessment (integration of writing, listening, and reading texts) has become more prominent in recent years. This is so because external academic texts (sources) provide support for content, act as a repository for language, improve validity, and bring about positive washback (Weigle and Parker 2012; Cumming et al. 2005; Cumming 2006). Research studies have also highlighted the importance of academic tasks. These play a critical role in academic success as they are commonly based on using external resources and integrating reading-writing skills (Hale et al. 1996; Rosenfeld et al. 2001). Another benefit for integrated assessment is that text-based information provides test-takers with content and ideas, minimizing the impact of topic familiarity, creativity, and life experiences (Weigle 2004). In addition, source texts provide test-takers with rhetorical structures to model vocabulary and grammar (Leki and Carson 1997). In fact, writing an essay solely based on background knowledge of an unseen topic is not regarded as authentic (Cumming et al. 2000). By eliciting discourse synthesis through organizing, selecting, and connecting (Spivey 1984), integrated tasks relate to the reading-to-write and listening-to-write processes in the target language using situation and lead to more appropriate assessment in academic writing (Plakans 2009).

Students need to be a part of the academic literacy and academic conversation by responding to external sources and constructing their own responses based on source-based information (Hamp-Lyons and Kroll 1996). Recent research explored student practice and

ability for writing from sources within the scope of EAP (Cumming et al. 2016; Wette 2017, 2018) “because of their crucial importance in higher education for demonstrating the acquisition of new knowledge in course papers and examinations and for establishing identities within academic discourse communities internationally” (Cumming et al. 2018, p. 2). In their study Cumming et al. (2016) focused on a synthesis of recent research on writing from sources for academic purposes. Based on the analysis of empirical evidence from contexts of first and second language education, researchers concluded that: (1) students experience difficulties, but develop certain strategies to deal with, the complex processes of writing from sources; (2) prior knowledge and experience influence students’ performance in writing from sources; (3) differences may appear between L1 and L2 students in their understanding and uses of sources in writing; (4) performance in tasks that involve writing from sources varies by task conditions and types of texts written and read; and (5) instruction can help students improve their uses of sources in their writing.

Student writers evidently need support in developing proficiency in the cognitively challenging task of writing using sources in their undergraduate years through practice and formal teaching and learning (Mansourizadeh and Ahmad 2011; Thompson et al. 2013; Wette 2017). This will help them shift from knowledge-telling to knowledge transformation (Bereiter and Scardamalia 1987) in the academic writing process. This transformation requires comprehension of propositions by other authors, synthesis, and acknowledgement of connections between multiple sources, integration of borrowed information into student writer’s own ideas as well as good knowledge of grammar and vocabulary in second language—L2 (Currie 1998; Storch 2009). In addition, Windsor and Park (2014) stress that L2 reading to write tasks in (online) higher education contexts foster deep learning because expert student writers go beyond reproducing content knowledge through synthesizing contextually appropriate information from external texts into their own work. Instead, source-based writing motivates learners to “create new knowledge by interacting objective procedural and declarative knowledge with more contextually subjective and informal tacit knowledge” (p. 96). Finally, the authors also argue that instruction should also go beyond “teaching procedural and declarative reading and writing skills and content knowledge” (p. 96).

In tandem with the above discussions in the field of language assessment, it is often hypothesized that when there is a curricular alignment within a language program between what is taught and what is tested, washback is going to be strong to be strong (Tsagari and Cheng 2016). The integrated proficiency test examined in this study functions in a local context and aims for such alignment is aimed for. The test consists of reading-to-writing and listening-to-writing tasks that focus on the same topics that are involved in the course content. Skills tested in the test also replicate the target language use (TLU) domain. Consequently, it differs from other standardized high-stakes tests because it particularly aims at distinguishing test-takers who can use English for academic purposes in university classrooms. Therefore, investigating its impact is of vital importance in safeguarding the positive washback and consequential validity of the test. Consequential validity is defined as the potential social impact of test interpretation and use (Messick 1989). According to Messick (1996) washback is an essential part of construct validity which is framed under consequential validity. Washback is seen as an inherent quality of any kind of assessment, especially when the results are used for important decisions (Cheng 2013; Tsagari 2009). Advocates of Messick’s views (e.g., Bachman and Palmer 2010; Kane 2013) also concurred that the effect of a test on learning and teaching is an integral aspect of its validity. For Messick, washback is associated with the consequential aspect of construct validity. The current study mainly focuses on the consequential aspect of validity with particular focus on washback. Washback is seen as an instance of test validity and conceptualized under the consequential aspect because this aspect entails both educational and social consequences. Messick commented that the consequential aspect of validity involved evidence and rationale for evaluating the intended and unintended consequences of interpretation and use of scores “ . . . in both the short- and long-term, especially those

associated with bias in scoring and interpretation, with unfairness in test-use, and with positive and negative washback effects on teaching and learning" (Messick 1996, p. 251). The concept of validity was expanded by taking aspects such as the effects of assessment on teaching and learning and the consequences of how assessment information is used into consideration (Messick 1989, 1995) to reach a better, more in depth, and complete understanding of how a testing program functions (Kane 2013).

This study is an attempt to explore aspects of consequential validity of an integrated proficiency test used in a Turkish university setting by drawing on perceptions of different stakeholders using an integrated language proficiency test. It is commonly agreed that researching effects of a test may have connotations for educational administration, materials development, teacher training, and resourcing, as well as test development and revision (Abbas and Thaheem 2018; Barnes 2017; Gokturk Saglam 2018; Gokturk Saglam and Farhady 2019; Hawkey 2009; Spratt 2005). Green (2007, p. 30) states: "It is important to gain ecologically grounded understandings of how a test operates within an educational context, rather than (or in addition to) seeking to isolate the effects of testing in experimental fashion". Pursuing ecologically grounded understandings may result in critical analysis of the alignment between test data and instruction. Thus, a deliberate focus on test consequences gains prominence, as it reveals whether language and skills manifested and described as objectives in the curriculum are acquired due to instructional practices. Furthermore, examining consequential validity of the test and test related decisions over time acts as a confirmatory study of the potential washback of the integrated language proficiency test used in the current Turkish EAP context at the tertiary level. This may provide valuable information about how this integrated proficiency test operates in the local context it is used in and shed light on how to make the best use of integrated tests to engineer positive washback in similar contexts. In the current study, the following research questions were addressed:

(1) What do teachers report about test consequences based on their evaluation of the students' English language skills and academic performance?

(2) What do freshman students report about the test consequences based on their self-evaluation of their proficiency levels in English language skills and academic performance?

As mastery of language and real-life academic skills (e.g., reading-to-write and listening-to-write tasks) are critical for academic success in English Medium Instruction (EMI) contexts, the findings of the study will contribute to the growing literature regarding consequential validity of integrated tests, which is relatively underexplored. Therefore, this study sets a useful agenda for inquiry and aims at reaching an understanding of integrated assessment and the viability of test-based decisions resulting from an integrated English proficiency test.

2. Materials and Methods

2.1. Participants and Setting

The present study was conducted at the preparatory English program (PEP) of an English medium Turkish university. The program teaches English for general and academic purposes across different language competency levels to aid students to cope with their undergraduate academic studies. Therefore, the overarching aim of the PEP is developing students' English skills and academic performance. At the end of instruction, PEP utilizes an integrated language proficiency test, which consists of theme-based reading and listening-to-write tasks, to determine whether test-takers can access their further academic studies beyond PEP. Test-takers are required to read 4 texts with different lengths on the same topic and listen to a note-taking listening task. In the writing section of the test, students are given back their notes and allowed to integrate information from oral (listening) and written (exam readings) external texts. The cut-off score is 65 out of 100. The teaching and learning curriculum of PEP is also closely aligned to discourse synthesis across various texts. Students had 20 contact hours a week for 16 weeks in the PEP program. This program catered for university students' academic and linguistic skills using a variety

of themes, such as environment, culture, and globalization as content. Course objectives aimed to improve language and academic skills. The course was built on an integrated skills approach and used a variety of integrated tasks in which students were asked to use information from reading and listening texts into their written and spoken output. To this end, course content included instruction in academic writing conventions (such as summarizing, responding to, and paraphrasing information from external texts in the course content) as well as cross textual reading.

The formative (midterm level assessment test used during the PEP program) and summative assessment (level achievement test used at the end of the PEP program) were administered before the proficiency test. These assessment procedures also adopted an integrated approach in determining the achievement of course objectives. They were administered in a systematic way. PEP instructors were required to oversee standardized invigilation guidelines. Also, before grading, instructors attended standardization sessions following tests that are used for formative and summative purposes in which norms of successful levels are discussed and exemplified through discussion of actual student outcomes.

Participants included 39 freshman students and 19 university instructors who were teaching regular courses in English in various departments. To examine the viability of test decisions over time, a questionnaire (Appendix A) was administered to freshman students who had been previously placed in upper-intermediate level in PEP and received instruction for 16 weeks. This group of students were targeted because, upon taking the language proficiency test and their completion of PEP, they gained access to their departmental courses within the same academic year, and this was convenient to trace their progress within the longitudinal design of the study. The students were aged between 19 and 24 and they had different majors from architecture, international relations, industrial engineering, international business and trade, hotel management, business administration, and gastronomy. The group was recruited on the basis of convenience sampling procedures.

University teachers had a mixed profile of teaching experience ranging from 7 years to 20 years. All held PhD degrees. Although they were teaching at different departments (see Table 1), they offered courses using English as the medium of instruction. A total of 19 teachers participated in interviews and 17 of them responded to the questionnaire. These teachers were selected based on theoretical and stratified sampling. They are faculty staff members who teach first year students when students pass the language proficiency that is administered at the end of PEP.

Table 1. Informants of teacher interviews on consequential validity.

Departments	N
Psychology	2
International Relations	2
Under-graduate English	2
Architecture	4
Mathematics	2
Aviation	1
Hotel Management and Tourism	2
Engineering	3
Business Administration	1

2.2. Data Collection Instruments

To investigate the research questions regarding consequential validity of the integrated proficiency test and viability of test-based decisions over time, semi-structured, one-on-one interviews were conducted with academic members of staff who offered departmental courses at the undergraduate level. In addition, teacher and student questionnaires surveyed conceptions of different stakeholders to investigate test consequences. [Beran and Rokosh \(2009\)](#) claim that “for a measure to be highly useful, it must provide the type of information required to be used for its intended purpose” (p. 499). Research studies on

consequential validity often focus on revealing the relationship between students' English language skills and academic performance (Huang et al. 2018) and emphasize the importance of exploring stakeholders' perspectives to improve reliability and validity of the tests (Haertel 2013; Michaelides 2014; Tsagari 2009, 2014). Kane (2013) argues that stakeholders who have personal experience of the test consequences should be included during exploration of consequential validity. Raising awareness of the complexity and versatility of student perceptions about assessment, Michaelides (2014) argues that "empirical evidence collected via interviews with test-takers can be invaluable in building the validity argument and informing about consequential aspects of examination programmes" (p. 438).

In relation to the research questions, the current data collection instruments focus on the targeted objectives enclosed in the proficiency test and the test-takers academic performance based on teacher and student perspectives. To examine different aspects of consequential validity both questionnaires and interviews required participants to evaluate English language skills and academic performance of successful proficiency test takers in their departmental courses, compare the overlap between targeted objectives and test constructs with observed student performance, elaborate on personal experience of test consequences, and report observations related to the test consequences based on academic performance.

2.3. Teacher Questionnaire and Interviews

In addition, the same instructors (N = 17) were surveyed through a questionnaire to elicit their conceptions about how effective the decisions made on the proficiency test scores were in identifying the language competency required for academic study within the current university program. The aim of the research was explained to the participants, and they were asked for their consent through a cover letter that accompanied the questionnaire. The first part of the questionnaire was to elicit their background information regarding their major, work experience, gender, highest qualification achieved, and the courses they delivered at the university. The second section surveyed their perceptions regarding how well PEP prepared the students in English for Academic Purposes (EAP). They were required to reflect their points of view by responding to 12 items on a 4-point Likert scale ranging from 'not prepared' (1) to 'well-prepared' (4). Also, after piloting the questionnaire through a read-aloud protocol with one of the university instructors, necessary modifications were made, and ambiguous items were reworded. Cronbach's alpha reliability coefficient for internal consistency of the items for the 12 items on the scale was $\alpha = 0.92$. In the third section, in response to an open-ended item, participants were asked to write their comments about the strengths and weaknesses of their students in terms of their language skills. Finally, in the fourth section, participants were asked about their suggestions to improve the PEP.

The questionnaires were e-mailed to 35 university instructors of mainstream courses and 17 completed the questionnaire. The questionnaire was administered towards the end of the spring semester when PEP graduates were attending classes for nearly 5 months.

To follow up on the questionnaire results, one-on-one interviews were conducted with 19 university instructors who offered undergraduate courses. The interviews were recorded, transcribed, and coded with a focus on meaning, condensing the meaning, and interpreting (Bogdan and Biklen 1998). University instructors were asked to make comments about the English language competency and academic skills of their students in meeting the expectations of the academic demands of the departments. Some interviews were conducted in English and some in Turkish, based on the preference of the participants, and lasted around 35 min. Interviews conducted in Turkish were translated into English.

2.4. Student Questionnaire

A perception questionnaire was given to 39 freshman students who had taken the language proficiency test after their instruction at PEP. The purpose of the questionnaire was to survey the freshman students' perceptions of the effectiveness of the TRACE-based

decisions in identifying the language competency and academic skills required for academic study within their department. Before administration, it was piloted with 22 freshman students and necessary changes were incorporated to ensure clarity of the items.

The questionnaire surveyed (1) information as to their department; (2) their perceptions regarding the extent to which PEP courses prepared them in English and academic skills for their departments; (3) their perceived strengths and weaknesses in terms of language proficiency; and (4) suggestions to improve instruction in PEP. Accompanied with a cover letter and a consent form, the questionnaire gathered information about student perceptions as to how well they think the PEP prepared them regarding their English language ability and academic skills training. This questionnaire had 12 items on a 4 points-Likert scale ranging from 'not prepared' (1) to 'well-prepared' (4). Participants were instructed to check the appropriate box on the questionnaire that reflected their point of view. Cronbach's alpha reliability coefficient for this part was $\alpha = 0.88$. In addition, students were required to respond to an open-ended question and self-evaluate their strengths and weaknesses in terms of their English language competency. The final item of the questionnaire was intended to elicit their suggestions to improve PEP to prepare students for the use of English for Academic Purposes at mainstream departmental university courses.

Students were given questionnaires towards the end of the spring semester after having some exposure to and experience in departmental courses and relevant academic demands. Consequently, their observations and experience over time, after 4 months, were expected to contribute to exploring the research questions about viability of test decisions made on the proficiency test scores.

3. Results

Perception questionnaires and interviews required teachers to comment on the correspondence between the exit criteria of PEP (the language proficiency test) and students' actual academic performance at various departments. Teacher and student perceptions were surveyed through the questionnaires about students' achievement in English and academic skills in their further academic studies. The thematic analysis and frequency counts of interviews and teacher/student questionnaire data mapped out the perceived strengths and weaknesses of the students in their English language proficiency and academic skills, revealing a variety of issues.

- **How well PEP graduates were prepared**

Data analysis of the teacher and student responses reported through questionnaires revealed discrepancies between how well students and their instructors think the preparatory program supported the students for their further academic studies in a range of skills. Table 2 outlines a summary of teacher perceptions regarding how well PEP prepared their students whereas Table 3 presents student perceptions.

Results of teacher perceptions provided negative evidence for the validity of the test consequences. Teachers claimed that even though their students had passed the English language proficiency test, they were not (well) prepared in most English and academic skills. It is important to note here that the language proficiency test did not have a speaking section, and this may have affected the consequential validity of the test as the most negative dispositions were associated with students' speaking skills such as discussing ideas and expressing opinions clearly and accurately in their speech (82%) and asking questions (76%). Thus, for many teachers, students had not developed their skills in many target domains even though they passed the proficiency test in PEP.

Table 2. Summary of teacher perceptions towards how well PEP prepares students (N = 17).

	Not Prepared		Fairly Prepared		Prepared		Well Prepared		SD
	f	(%)	f	(%)	f	(%)	f	(%)	
Reading academic texts and understanding the main ideas	3	(18)	9	(53)	5	(29)	0	(0)	0.83
Taking reading notes	1	(6)	13	(77)	2	(12)	0	(0)	0.44
Understanding lectures	1	(6)	8	(47)	7	(41)	0	(0)	0.62
Taking listening notes	3	(18)	8	(47)	4	(24)	1	(6)	0.83
Writing an organized essay	3	(18)	9	(53)	3	(18)	0	(0)	0.65
Discussing ideas and expressing opinions clearly and accurately in their speech	9	(53)	5	(29)	2	(12)	1	(6)	0.92
Asking questions in class	8	(47)	5	(29)	3	(18)	1	(6)	0.95
Using a range of vocabulary appropriately	5	(29)	11	(65)	1	(6)	0	(0)	0.56
Using a range of grammatical structures in their written and spoken work	7	(41)	9	(53)	1	(6)	0	(0)	0.61
Using different sources (notes, summaries, etc.) to support ideas in their written and spoken work	10	(59)	4	(24)	2	(12)	0	(0)	0.73
Giving feedback to peers	7	(41)	7	(41)	2	(12)	0	(0)	0.70
Revising own written work based on given feedback	5	(29)	7	(41)	3	(18)	0	(0)	0.74

Table 3. Summary of student perceptions (N = 39).

	Not Prepared		Fairly Prepared		Prepared		Well Prepared		SD
	f	(%)	f	(%)	f	(%)	f	(%)	
Reading academic texts and understanding the main ideas	1	(3)	7	(18)	20	(51)	10	(26)	0.75
Taking reading notes	2	(5)	11	(28)	18	(46)	7	(18)	0.81
Understanding lectures	1	(3)	9	(23)	20	(51)	9	(23)	0.76
Taking listening notes	1	(3)	12	(31)	12	(31)	14	(36)	0.89
Writing an organized essay	3	(8)	6	(15)	13	(33)	17	(44)	0.95
Discussing ideas and expressing opinions clearly and accurately in their speech	2	(5)	14	(36)	18	(46)	5	(13)	0.77
Asking questions in class	2	(5)	10	(26)	19	(49)	8	(21)	0.81
Using a range of vocabulary appropriately	4	(10)	12	(31)	16	(41)	7	(18)	0.90
Using a range of grammatical structures in their written and spoken work	4	(10)	11	(28)	18	(46)	6	(15)	0.87
Using different sources (notes, summaries, etc.) to support ideas in their written and spoken work	5	(13)	14	(36)	15	(39)	5	(13)	0.88
Giving feedback to peers	3	(8)	15	(39)	12	(31)	9	(23)	0.92
Revising own written work based on given feedback	3	(8)	8	(21)	13	(33)	14	(36)	0.96

Contrary to the teacher perspective, PEP graduates perceived themselves prepared in many English and academic skills including reading academic texts and understanding the main ideas, using a range of grammatical structures in written and spoken work, revising

own written work based on given feedback, understanding lectures, taking listening notes, and writing a well-organized essay. Overall, it can be inferred that most of the students held positive perceptions in terms of their competency in most English sub-skills in a stark contrast to their teachers' opinion.

However, some students claimed that the program did not prepare them sufficiently in some certain skills such as discussing ideas and expressing opinions clearly and accurately in their speech and asking questions. Except for these negative conceptions (indicating insufficient competency in these speaking skills), student perceptions did not tend to resemble teacher opinions. This difference may stem from learners' low level of assessment literacy and evaluative capacities. Also, student response displayed diversity regarding how well the PEP assisted them in using different sources such as notes of external text-based information and summaries to support ideas in their written and spoken work. Analyzing information from different sources and integrating these into one's work was a major construct in the proficiency test and relevant skills were targeted in the instructional design. However, there seems to be a contradiction between student and teacher evaluation with respect to this issue, as teachers expressed rather negative perceptions and indicated that students were not prepared and needed further practice. Therefore, these findings imply the need to raise student awareness about assessment literacy as well as raising awareness on different levels of performance and descriptors that define these levels.

- **Identified strengths and weaknesses in students' academic performance**

During the interviews the instructors evaluated their students' strengths and weaknesses in general, not of those students that had passed the language preparatory program. However, some commented that they observed differences between PEP graduates and others who were exempt from the program. One of the respondents commented that there were some students in the departmental courses who managed to complete the PEP program and pass the proficiency test despite their low level of language competency.

When I got a class in the very beginning of the semester, I asked how many students came from PEP. About twenty out of thirty raised hands. Out of those twenty, five or six will be very good, very well-equipped. Although they are still, especially in speaking they would be very shy and very insufficient. Let's say five out of twenty would be equipped in terms of writing and can do the work in discipline. About ten will not be up to standard, really. So, they struggle. About five, they shouldn't be there at all. And I am trying to be realistic. I mean knowing the context, knowing the educational background and the possibilities what could be done in PEP in a certain amount of time, feasibility . . . I'm taking in all those factors and I'm trying to give a kind of realistic and generous answer; Thirty per cent of prep school graduates shouldn't be there. They are not ready. They are effectively, really, still in intermediate or even pre-intermediate level, in some cases. And somehow, they managed to slip through the net.

Comparing students who attended PEP and who were exempt, another instructor shared the following observation: "They have a great difficulty in self-expression and talking in English. If they come from prep school, they have great difficulty in speaking as well as understanding English but if they come from a good high school, they do not have problems in speaking". The comment highlighted the insufficiency of speaking skills of students who come from PEP. Consequently, these views imply negative consequential validity regarding the potential implications of the decision on the validity of the study.

In addition to speaking, mentioned above, teachers outlined certain weaknesses in students' language and skills, summarized in Table 4 below.

According to the teachers, speaking was prioritized as a domain in which students required further practice. This finding was also reflected in the teacher questionnaire results where the majority (77%) argued that student competency was inadequate in terms of speaking ("the weakest point of an average student") even though this skill was conceived as the most useful skill for academic success. Lack of motivation and self-confidence was often remarked as a major attribute of insufficient speaking skills of the students as reflected

in the following comment: *“They feel so insecure when they speak in English. I assume it is due to the lack of confidence in speaking a foreign language”*. Therefore, teachers suggested that students needed more instruction in speaking in English.

Table 4. Teacher perceptions of students’ weaknesses in language and academic skills.

Students’ Weaknesses in Language and Academic Skills	N	%
Reading	7	37
Listening	4	21
Writing	8	42
Speaking	15	79
Grammar	9	47
Vocabulary	10	53
Understanding exam questions	6	32
Academic Skills (discourse synthesis, knowledge of citation, evaluation of online sources)	3	16

According to the teachers, writing skills also proved to be both daunting and difficult for the students with respect to content generation, self-expression, and citing skills as reflected in this statement: *“I think they are not really good at academic writing. Yes, there are some examples where there are a lot of spelling problems, grammar problems, mistakes, but I have seen a lot of papers that had good grammar and good spelling yet not very good at communicating what they have in mind”*. Questionnaire findings also supported this perspective. It was highlighted by most teachers (71%) that students were good at organizing their essays with regard to making an outline, integrating supporting ideas, writing a thesis statement, and having an overall organization in their written outcome. However, some (35%) expressed that essay organization was also perceived as a weak area, especially in ensuring the flow of ideas and generating content. Coherence between ideas was framed as an ‘inability’ for the students when they attempted to convey their thoughts and build up their own arguments in a properly organized academic structure. According to some instructors (30%), another negative disposition in writing skill was associated with inadequate citation skills pertaining to inaccurate paraphrasing, quoting, and making use of citation mechanics, such as the use of APA. These instructors held a rather negative impression towards students’ inaccurate and inadequate citation practices when they borrowed information from external texts into their own writing.

Low level of language competency was deemed as one of the weak areas of students. Some teachers (47%) stated that students’ low level of grammar and vocabulary knowledge intervened in their understanding. Some teachers (32%) claimed they had difficulty while marking students’ written tasks as they could not decide whether to take quality of language or content into consideration. This was defined as the “big dilemma”. Consequently, some of the participants concurred they compromised and ignored the language and focused on the content. Questionnaire findings also reflected this issue as most of the teachers (59%) noted the inefficiency of using accurate grammar and lack of adequate vocabulary knowledge (29%) as a major handicap. It was claimed that *“students sometimes stock phrases and collocations that are wrong. They complete their work with a limited number of words: Therefore, written assignments generally look so simple and lack depth of adequate discussion”*.

Instructors’ evaluations of their students’ reading and listening skills in English indicated both positive and negative conceptions. In terms of listening skills, few instructors commented that their students were confident in note taking (24%), finding the main and the supporting idea(s) (6%), inferring attitude and purpose (6%), and identifying signal words (6%). Some teachers pinpointed negative perceptions related to students’ difficulty to understand lectures (35%), lack of motivation to listen to long lectures (12%), and difficulty to understand class discussions (6%). However, it was the contention of most of the teachers (60%) that students lacked a good level in variety of reading skills in inferencing–identifying tone and purpose, drawing conclusions through critical thinking,

analysis, and synthesis of main ideas, coping with comprehension of long texts, and finding main ideas. This was inferred as negative validity evidence as the integrated proficiency exam and the aligned instruction of PEP placed emphasis on these skills.

- **Students’ performance in their departmental courses**

When teachers considered their students’ language and academic performance beyond PEP, they expressed negative perceptions. Data analysis led to emerging themes (as summarized in Table 5) and prioritized certain weaknesses in students’ English skills and academic performance including inadequate speaking skills, effect of students’ educational background on their achievement, and low motivation for reading. The criticism towards inefficient student performance regarding academic tasks which required an integrated approach (reading-to-writing) towards classroom tasks was also reflected in the emerging themes of the interviews.

Table 5. Emerging themes of the interviews with department teachers.

Themes	N	%
Inadequate speaking skills of the students	15	79
Effect of educational background on student achievement		
Exam orientedness	3	16
Mismatch between prior learning and university culture	4	21
Low motivation for reading	8	42

Most of the teachers (79%) conveyed inadequate speaking skills of their students in English.

Unfortunately, most of the students are unable to follow the class because of the language problem. And they are unable to ask questions in foreign language. And that affects the course very bad and negatively. We talk, we show, we discuss, we explain, and we expect students to interact with us to join to the class to contribute to the class, ask the questions, discuss the concepts with us. But they prefer to stay silent and just watch. Then I feel, and most of us feel like, we’re just lecturing in front of a wall. That’s a big concern (T1).

The comment highlights the insufficiency of speaking skills which may be due to an unintended test consequence, as speaking is not tested on the proficiency test. There may be more explicit student focus on other skills in comparison to the speaking skill which is not tested.

Another common theme concerned the effect of educational background on students’ achievement. Some teachers (37%) commented that there was a mismatch between previous educational culture, which relied on exam-oriented approach to learning and rote memorization, and university culture which emphasized critical thinking and (re)constructing knowledge by synthesizing information from various sources. Some teachers claimed that students do not place emphasis on the evaluation of their performance as they are product oriented.

There is a mismatch between students’ educational background and the skills required at the university. Schools are busy with teaching students how to solve a multiple-choice question without having the knowledge. Students focus on the correct answer rather than why that’s the correct answer. Often why is never asked. Thinking, evaluating, criticizing is a mind-set and most students do not seem to have that. Here at the university, they need to be formatted and it is very difficult (T2).

Although the integrated language proficiency assessment test in this study set out to provide students with positive washback and acquisition of the ‘mindset’ described by the comment above, the analysis of conceptions that teachers had about the proficiency assessment indicated some negative test consequences. In other words, the proficiency test and the aligned curriculum/instruction might not be efficient in impacting the students’ attitudes towards learning (in terms of “formatting the students” as rephrased by the

teacher comment above). Teachers tended to believe that educational background of students fostered an exam-oriented approach to learning. One of the teachers concurred: *“Students are very much accustomed to multiple choice items, and they prefer responding to this format”* (T3).

In addition, teachers are inclined to think that their students have difficulty in borrowing information from texts and integrating these into their own work (discourse synthesis) as reflected in the comment below.

The ability to synthesize ideas is really, widely-challenging. Sometimes we’re not sure whether it is language issue or whether it is a critical thinking issue. I mean, synthesizing is putting ideas together. For example, seeing, detecting the patterns, similarities, new connections between them, there is a critical thinking skill. We’re not sure, and the students are not generally very good at it (T4).

In addition, response of the teachers indicated that students “resist reading”. This resistance was at times associated with an exam-oriented approach that students have towards learning. In other words, it was suggested that students were inclined to respond to a certain type of exam questions which would involve multiple choice format. A teacher explained: *“I was really surprised to see that if there is a question which is more than 5–6 lines in the exam, they ignore the question and don’t do it. They don’t even consider putting in the effort to read it. They don’t read the instructions to an assignment. They ask and want me to explain. They run away when they see a reading text”* (T5).

The teacher’s response above seems to imply that, even though reading is tested in the proficiency test through multiple texts in different genres and it is linked to the writing skill to reflect real-life language use domain. This does not seem to exert a powerful effect and a positive washback for the students to pay deliberate attention to improve their reading skills. One of the teachers pointed out: *“They don’t seem to have understood the logic and purpose behind reading skills. They try to memorize and therefore their affective filters are up. They are really anxious”* (T6). Consequently, teachers stressed that there was a gap between language and skills required for academic success and actual student performance.

However, student perceptions were not in agreement with teacher opinions. Students tended to comment that they felt competent in writing skills, especially writing an academic essay (46%). However, they did not tend to mention any of the criticisms raised by their teachers such as lack of mastery in ensuring flow of ideas, building an argument based on expanded justifications, and integrating information from external texts into one’s own oral and written work. Students seemed to feel themselves confident in most of the English and academic skills except for the speaking. Showing similarity to overall teacher evaluations, some students considered themselves weak in terms of speaking fluently.

- **Suggestions for improvement**

Suggestions of university teachers and freshman students to improve the PEP converged on some certain concepts, including a deliberate focus on enhancing speaking and writing skills. Overall, improving speaking skills was the most voiced suggestion by the participating teachers. They stressed the students’ need to practice speaking more to gain confidence. One of the respondents noted: *“I believe in the rigor of the Preparatory English Program; however, faculty members share a common belief that students have a big problem in speaking. I would strongly suggest the English teachers to encourage more speaking in their classes and to test students’ speaking ability maybe with a different method”*. Here, it is also important to focus on the suggested idea of “testing speaking ability” to improve this skill as it signifies the reliance on testing as a lever for change in how teachers teach and how students learn. This resonates with the findings of [Huang et al. \(2018\)](#) who argue that lack of speaking tests leads to undesirable consequences. They remark that, although there is high demand for communication skills in both speaking and writing, the majority of the English language programs in higher education prioritize academic writing, as speaking tests are more labor-intensive in terms of administration and scoring.

Another teacher suggested: *“Some students can pass the proficiency exam despite low writing skills as they get higher grades from the other parts of the exam”*. It can be inferred that some teachers believe students can be successful on the proficiency tests due to being test-smart, mastering sections that involved responding to reading and listening sections of the test. Therefore, taking assessment-driven measures, such as focusing on and prioritizing the source-based writing through reading-to-writing and listening-to-writing tasks, is seen as an effective way of maintaining higher language competency level. Furthermore, teacher response indicated a higher focus on writing skills through deliberate teaching of grammar as well as citing information through accurate and conceptually appropriate summarizing, paraphrasing, and (in)direct quotation. It was often argued that students would highly benefit from working on skills such as summarizing, paraphrasing, and basic citation in APA style. One comment concurred: *“students should learn how to summarize articles/videos and write response paragraphs. They have difficulty in summarizing and reflecting on sources in terms of how these contribute to their own arguments both in writing and speaking”*. Therefore, it was argued that a more deliberate effort towards teaching critical thinking skills through integration of source-based information into students’ oral and written outcome was a necessity for the instruction in PEP.

Freshman students offered a variety of measures, which resonated with the teacher suggestions for improving the PEP. These included having more practice in grammar and vocabulary, especially focusing more on speaking skills. Like teachers, students also highlighted that it was necessary to add a speaking component to the proficiency exam, commenting that if it is tested then they would pay more attention to this skill. It was remarked that learning about purposes of basic citation skills as well as mechanical application of citation such as using the APA would be useful for effective learning. Some students also mentioned that learning vocabulary related to their academic discipline would prove to be useful. In addition, there were some comments which highlighted the concept of different teaching methodology between teachers in the PEP by stating: *“PEP needs to self-check about the teachers and the application of the plan (means the curriculum). Plan and the approach to education is okay but some problems happen on the stage”*.

4. Discussion

One of the primary objectives of English preparatory programs in higher education is to prepare their students for the language skills and academic demands of their future studies. This study mainly aimed at outlining how teacher and student perceptions can inform the validation process of an integrated English language proficiency test beyond given/achieved scores. This consequential validation study found evidence of both positive and negative washback of the integrated English proficiency test. Positive washback is regarded as related to consequential validity, whereas negative washback is associated with lack of validity (Ferman 2004).

Teacher conceptions of the skills that are required for academic success in higher education elicited through this study overlap with the construct and targeted skills of an integrated assessment in a university EAP program. Most instructors pointed out that cross textual reading skills and synthesis of information from diverse texts were elemental for academic success. Consequently, learners were expected to integrate information from external texts into one’s own oral and written work to build an argument.

University mainstream teachers confirmed that students who engaged in their academic studies beyond PEP have encountered an array of difficulties in their English and academic skills. Teachers expressed doubts about the effectiveness of the test-based decisions in the EAP program in terms of identifying the language competency and skills required for academic study at the tertiary level. One of the main teacher criticisms was geared towards students’ weak speaking skills that was deemed as hindering their academic success. Student perceptions were also concerted with teacher views. Furthermore, both teachers and students suggested to place more deliberate focus on speaking in instruction and make speaking as a part of the proficiency test. Therefore, the findings of this

study point to lack of consequential validity resulting in a narrowing of the curriculum to tested skills which eventually hinders learning.

Another area of student performance that received teacher criticism was using sources effectively. Findings in this study resonate with previous research which concluded that knowing about source selection, integrating information from external texts into academic writing, maintaining contextual appropriacy, and mastering technical accuracy in citation practices (e.g., use of APA) pose considerable challenges for L2 students (Thompson et al. 2013). Teachers indicated that their students lacked adequate proficiency in academic writing using sources. They attributed difficulties that hindered their students' performance to low levels of linguistic competency, lack of reading motivation and their exam-oriented background, claiming that they needed a new mindset into deep learning. These findings agree with the conclusions of prior studies which reported ongoing challenges that undergraduate student writers face on the way to achieving proficiency in a complicated academic literacy (Pecorari and Petrić 2014; Wingate 2015; Wette 2017). Therefore, as suggested by the participant teachers of the study, the instruction should include more practice and guidance in integrated assessment. To illustrate instruction into source-based academic writing can entail a deliberate focus on raising awareness on the functions of citation (Hirvela and Du 2013) to help students understand the role of summaries, quotes, and paraphrases (Shi 2008). Therefore, instruction may focus on functions of citations as well as cross-textual reading skills to help them improve their language and academic skills. Understanding the construct of integrated assignment and drawing assessment and instruction closer may bring forth positive consequential validity. According to Wette (2017), instructional support should be extended throughout undergraduate years to provide students with gradual support in gaining proficiency in this challenging new literacy that is necessary in the higher education. As experience in source-based writing plays a significant role on student performance she argues priorities must be set in writing courses because "while novices may be capable of paraphrasing single ideas from individual texts, experienced writers are able to synthesize and comprehend connections between multiple sources, and to use the writing process to transform current knowledge conceptually and linguistically as well as to advance their own thinking" (p. 47).

This study found a discrepancy between teacher and student conceptions regarding consequences of the test. Students tended to report a positive impression of the test. They were confident that the PEP program prepared them for the English and academic skills that were required at their majors, whereas teachers held a rather negative impression about the students' competency and performance. It is often argued that student perceptions about examinations reflect their knowledge of assessment literacy (Taylor 2009). Students seem to disregard the rationale behind integrated assessment and neglect deliberate strategies for life-long learning. Therefore, fostering assessment literacy of the students is crucial for raising awareness on self-assessment of their competency in English and academic skills as well as determining further learning objectives.

5. Conclusions

Exploring consequential validity may establish a means for ongoing dialogue between different stakeholders involved in a testing program. Unanticipated consequences of the test should be taken into consideration as a part of the validation process in a systemic and broader point of view. This broad systemic validation process may guide educators in a deliberate and concerted effort to cater for the real-life needs of the parties who are affected by the test consequences. Furthermore, focusing on consequential validity during the instructional design of assessment procedures during the validation process (Reckase 1998) and "motivating test developers to assume responsibility for more aspects of test usage" (Iliescu and Greiff 2021) can be a means of resolving unwanted, unintended test consequences. Therefore, effects of assessment in terms of how it influences instructional and learning processes (Tiekstra et al. 2016) should be critically considered in the development of integrated assessment procedures in future. Highlighting the importance of

consequential validity and devising a wholistic and systemic perspective towards test consequences during the instructional design and validation of the test/assessment procedures may resolve unwanted/unintended test impacts.

The first research question set out to explore how teachers viewed the test consequences based on their evaluation of the students' English language skills and academic performance. Their perceptions seemed to cast doubt over the effectiveness of the decisions made by an integrated English proficiency test used in an EAP program in identifying the language competency and skills required for academic study at the tertiary level. They remarked that the speaking skills of their students were inefficient. This may be an unwanted/unintended test impact due to narrowing of the curriculum to the tested skills. On the other hand, findings of the second research question, which investigated perceptions of the freshman students about the test consequences based on their self-evaluation of their proficiency levels in English language skills and academic performance, presented a stark contrast as students tended to hold positive views.

These findings bring about implications for materials development, teacher training, and enhancing student assessment literacy alongside instructional design. Integrated assessment should be embedded more efficiently into the curriculum. Student outcome can include reading/listening into speaking instead of excessively reading and listening for writing. Teachers confirmed that speaking skills constitute an important part of academic success. Therefore, it could be integrated into formative, summative, and proficiency assessment procedures. Course materials can also be designed to reinforce improvement of all skills and student peer/self-evaluation to help learners assess their progress and identify further learning goals. Focusing on purposes of academic citing practices and strategy training (e.g., reading for main ideas, cross textual reading, taking reading notes . . . etc.) may also raise student awareness of source-based writing and speaking.

Despite these implications, this study has several limitations. Although we were able to elicit teacher opinions through different lines of data collection, we could gather learner views only through a questionnaire due to time constraints and lack of voluntary student participation. In addition, when evaluating the quality of an integrated assessment and exploring its consequential validity, resorting to perceptions of different stakeholders should hold a more prominent place to design better assessment. Thus, further research studies can integrate multiple lines of data from students. Consequently, a broader understanding of the extent to which integrated assessment impacts the process of instruction and learning can be reached. However, perceptions tend to be affected by assessment literacy knowledge (Tsagari 2020). Therefore, future research could make use of actual student performance (e.g., written reports, oral presentations, etc.) to scrutinize consequential validity beyond a test. Also, studies can target other stakeholder perceptions to extend the scope of the validation process and use larger samples of teachers/students for generalizability purposes.

There are various implications of this study for instructional designers (e.g., test developers and curriculum advisors) and teachers. Investigating consequential validity of a test may cast light upon the unintended test consequences and provide instructional designers with insights into unintended negative impact in the validation process. In this vein, intended positive test consequences can be confirmed and enhanced, whereas unintended consequences can be minimized. Another implication concerns improvement of instruction. Good test consequences (even if unintended) should be identified and used, while negative consequences should be retaliated as much as possible (Taleporos 1998).

Author Contributions: Conceptualization, A.L.G.S.; methodology, A.L.G.S.; formal analysis, A.L.G.S.; investigation, A.L.G.S.; writing—original draft preparation, A.L.G.S.; writing—review and editing, A.L.G.S. and D.T.; visualization, A.L.G.S. and D.T.; supervision, D.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Ozyegin University (27 August 2013).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We extend our sincere thanks to the participants of this study.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Consequential Validity Teacher Questionnaire

At the end of the Preparatory School Program in English language, our students take the English language proficiency test. Test results are assumed to be linked to the use of English for Academic Purposes at mainstream departmental university courses. We would appreciate your answers to the following questions.

Background Questions

What is your major (graduate and postgraduate) (TESOL, Literature, EFL, Applied Linguistics, Education . . . etc.)?

What is your gender?

Years of work experience

1–5 6–10 11–15 16–20 More than 20 years

What is the highest qualification you achieved?

Which course(s) do you teach?

Perceptions about validity of test-decisions

1. Do you think students who completed Preparatory School Program are well-prepared regarding their English language ability and academic skills training? Check the degree in the table below that reflects your point of view.

	Not Prepared	Fairly Prepared	Prepared	Well Prepared
1. Reading academic texts and understanding the main ideas				
2. Taking reading notes				
3. Understanding lectures				
4. Taking listening notes				
5. Writing an organised essay				
6. Discussing ideas and expressing opinions clearly and accurately in their speech				
7. Asking questions				
8. Using a range of vocabulary appropriately				
9. Using a range of grammatical structures in their written and spoken work				
10. Using different sources (notes, summaries etc.) to support ideas in their written and spoken work				
11. Giving feedback to peers				
12. Revising own written work based on given feedback				
If you have further comments about any other English language ability and academic skills, please write them here:				

2. Where do you see the strengths and weaknesses of the students who completed Preparatory School Program? Please comment under relevant headings in the table below.

	Strengths of the Students Who Completed Preparatory School Program	Weaknesses of the Students Who Completed Preparatory School Program
In Writing		
In Reading		
In Listening		
In Speaking		
In using Grammar and Vocabulary		
If you have any further comments about any other use of English for Academic Purposes at mainstream departmental university courses, please write them here.		

3. Do you have any suggestions for improving the Preparatory School Program?

Appendix A.2. Consequential Validity Student Questionnaire

At the end of the Preparatory School Program in English language, our students take the English proficiency test. Test results are assumed to be linked to the use of English for Academic Purposes at mainstream departmental university courses. We would appreciate your answers to the following questions.

1. What is your department?
2. Where do you see the strengths and weaknesses of the students who completed Preparatory School Program? Please comment under relevant headings in the table below.

	Strengths of the Students Who Completed Preparatory School Program	Weaknesses of the Students Who Completed Preparatory School Program
In Writing		
In Reading		
In Listening		
In Speaking		
In using Grammar and Vocabulary		
If you have any further comments about any other use of English for Academic Purposes at mainstream departmental university courses, please write them here.		

3. Do you have any suggestions to improve to Preparatory School Program to prepare you better for the use of English for Academic Purposes at mainstream departmental university courses?

Appendix A.3. Interview Guide

Introduction

Key points of the study; purpose, confidentiality, media and timing

1. Correspondence between exit criteria of the English proficiency test and expected academic skills at the departments
Could you share your observations regarding student achievement in:
 - In Writing
 - In Reading
 - In Listening
 - In Speaking
 - In use of Grammar and Vocabulary
 - In academic skills
2. What skills are required at the department?
3. Do you think that students who complete Prep Program can cope with the academic demands of your department

4. What needs to be done in prep program to better equip these students for their departments?
5. Other Views
6. Round up and Thanks

References

- Abbas, Asiba, and Sahiba Sarwar Thaheem. 2018. Washback Impact on Teachers' Instruction Resulting from Students' Apathy. *Research on Humanities and Social Sciences* 8: 45–54.
- Bachman, Lyle F., and Adrian S. Palmer. 2010. *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford: Oxford University Press.
- Barnes, Melissa. 2017. Washback: Exploring what constitutes "good" teaching practices. *Journal of English for Academic Purposes* 10: 1–12. [\[CrossRef\]](#)
- Beran, Tanya N., and Jennifer L. Rokosh. 2009. Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science* 37: 171–84. [\[CrossRef\]](#)
- Bereiter, Carl, and Marlene Scardamalia. 1987. *The Psychology of Written Composition*. Hillsdale: Erlbaum.
- Bogdan, Robert, and Sari Knopp Biklen. 1998. *Qualitative Research for Education: An Introduction to Theory and Methods*. Boston: Allyn and Bacon.
- Cheng, Lying. 2013. *Consequences, Impact, and Washback*. The Companion to Language Assessment. Edited by Antony John Kunnan. New York: John Wiley & Sons Inc., pp. 1130–47.
- Cumming, Alister, Robert Kantor, Donald Powers, Terry Santos, and Carol Taylor. 2000. *TOEFL 2000 writing framework*. Princeton: Educational Testing Service.
- Cumming, Alister. 2006. *Analysis of Discourse Features and Verification of Scoring Levels for Independent and Integrated Tasks for the New TOEFL*. TOEFL Monograph No. MS-30 Rm 05-13. Princeton: ETS.
- Cumming, Alister, Robert Kantor, Kyoko Baba, Usman Erdosy, Keanre Eouanzoui, and James Mark. 2005. Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing* 10: 5–43. [\[CrossRef\]](#)
- Cumming, Alister, Conttia Lai, and Hyeyoon Cho. 2016. Students' writing from sources for academic purposes: A synthesis of recent research. *Journal of English for Academic Purposes* 23: 47–58. [\[CrossRef\]](#)
- Cumming, Alister, Luxin Yangc, Chenhui Qiuc, Lian Zhange, Xiaoling Jid, Junju Wange, Ying Wange, Ju Zhanf, Fengjuan Zhangf, Chunyan Xuf, and et al. 2018. Students' practices and abilities for writing from sources in English at universities in China. *Journal of Second Language Writing* 39: 1–15. [\[CrossRef\]](#)
- Currie, Pat. 1998. Staying out of trouble: Apparent plagiarism and academic survival. *Journal of Second Language Writing* 7: 1–18. [\[CrossRef\]](#)
- Ferman, Irit. 2004. The washback of an EFL national oral matriculation test to teaching and learning. In *Washback in Language Testing: Research Contexts and Methods*. Edited by Liying Cheng, Yoshinori Watanabe and Andy Curtis. Mahwah: Lawrence Erlbaum Associates, pp. 3–18.
- Green, Anthony. 2007. *IELTS Washback in Context: Preparation for Academic Writing in Higher Education*. Cambridge: Cambridge University Press.
- Gokturk Saglam, Asli Lidice. 2018. Can exams change how and what teachers teach? Investigating the washback effect of a university English language proficiency test in the Turkish Context. *Eurasian Journal of Applied Linguistics* 4: 155–176. [\[CrossRef\]](#)
- Gokturk Saglam, Asli Lidice, and Hossein Farhady. 2019. Can Exams Change How and What Learners Learn? Investigating the Washback Effect of a University English Language Proficiency Test in the Turkish Context. *Advances in Language and Literary Studies* 10: 177–187. [\[CrossRef\]](#)
- Haertel, Edward. 2013. How Is Testing Supposed to Improve Schooling? *Measurement: Interdisciplinary Research and Perspectives* 11: 1–18. [\[CrossRef\]](#)
- Hale, Gordon, Carol Taylor, Brent Bridgeman, Joan Carson, Barbara Kroll, and Robert Kantor. 1996. *A Study of Writing Tasks Assigned in Academic Degree Programs*. Princeton: Educational Testing Service. [\[CrossRef\]](#)
- Hamp-Lyons, Liz, and Barbara Kroll. 1996. Issues in ESL writing assessment: An overview. *College ESL* 6: 52–72.
- Hawkey, Roger. 2009. A study of the Cambridge Proficiency in English (CPE) exam washback on textbooks in the context of Cambridge ESOL exam validation. In *Language Testing Matters: Investigating the Wider Social and Educational Impact of Assessment, Proceedings of the ALTE Cambridge Conference, April 2008*. Edited by Lynda Taylor and Cyril Weir. Cambridge: Cambridge University Press, pp. 326–43.
- Hirvela, Alan, and Qian Du. 2013. "Why am I paraphrasing?": Undergraduate ESL writers' engagement with source-based academic writing and reading. *Journal of English for Academic Purposes* 12: 87–98. [\[CrossRef\]](#)
- Huang, Heng-Tsung Danny, Shao-Ting Alan Hung, and Lia Plakans. 2018. Topical knowledge in L2 speaking assessment: Comparing independent and integrated speaking test tasks. *Language Testing* 35: 27–49. [\[CrossRef\]](#)
- Iliescu, Dragos, and Samuel Greiff. 2021. On consequential validity. *European Journal of Psychological Assessment* 37: 163–66. [\[CrossRef\]](#)
- Kane, Michael. 2013. The Argument-Based Approach to Validation. *School Psychology Review* 42: 448–57. [\[CrossRef\]](#)
- Leki, Ilona, and Joan Carson. 1997. Completely different worlds: EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly* 31: 36–69. [\[CrossRef\]](#)

- Mansourizadeh, Kobra, and Ummul K. Ahmad. 2011. Citation practices among non-native expert and novice scientific writers. *Journal of English for Academic Purposes* 10: 152–161. [CrossRef]
- Messick, Samuel. 1989. Validity. In *Educational Measurement*, 3rd ed. Edited by Robert L. Linn. New York: Macmillan, pp. 13–103.
- Messick, Samuel. 1995. Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist* 50: 741–49. [CrossRef]
- Messick, Samuel. 1996. Validity and washback in language testing. *Language Testing* 13: 241–56. [CrossRef]
- Michaelides, Michalis P. 2014. Validity considerations ensuing from examinees' perceptions about high-stakes national examinations in Cyprus. *Assessment in Education: Principles, Policy & Practice* 21: 427–41. [CrossRef]
- Pecorari, Diane, and Bojana Petrić. 2014. Plagiarism in second-language writing. *Language Teaching* 47: 269–302. [CrossRef]
- Plakans, Lia. 2009. Discourse synthesis in integrated second language writing assessment. *Language Testing* 26: 561–87. [CrossRef]
- Reckase, Mark D. 1998. Consequential validity from the test developer's perspective. *Educational Measurement: Issues and Practice* 17: 13–16. [CrossRef]
- Rosenfeld, Michael, Susanna Leung, and Philip K. Oltman. 2001. *The Reading, Writing, Speaking, and Listening Tasks Important for Academic Success at the Undergraduate and Graduate Levels*. TOEFL Monograph Series 21; Princeton: Educational Testing Service.
- Shi, Ling. 2008. Textual appropriation and citing behaviors of university undergraduates. *Applied Linguistics* 31: 1–24. [CrossRef]
- Spivey, Nancy N. 1984. *Discourse Synthesis: Constructing Texts in Reading and Writing*. Outstanding Dissertation Monograph. Newark: International Reading Association.
- Spratt, Mary. 2005. Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research* 9: 5–29. [CrossRef]
- Storch, Neomi. 2009. The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing* 18: 103–18. [CrossRef]
- Taleporos, Elizabeth. 1998. Consequential Validity: A Practitioner's Perspective. *Educational Measurement: Issues and Practice* 17: 20–23. [CrossRef]
- Taylor, Lynda. 2009. Developing assessment literacy. *Annual Review of Applied Linguistics* 29: 21–36. [CrossRef]
- Thompson, Celia, Janne Morton, and Neomy Storch. 2013. Where from, who, why and how? A study of the use of sources by first year L2 university students. *Journal of English for Academic Purposes* 12: 99–109. [CrossRef]
- Tiekstra, Marlous, Alexander Minnaert, and Marco G. P. Hessels. 2016. A review scrutinizing the consequential validity of dynamic assessment. *Educational Psychology* 1: 112–37. [CrossRef]
- Tsagari, Dina. 2009. *The Complexity of Test Washback: An Empirical Study*. Frankfurt am Main: Peter Lang GmbH.
- Tsagari, Dina. 2014. 'Investigating the face validity of Cambridge English First in the Cypriot context'. *Research Notes* 57: 23–31. Available online: <http://www.cambridgeenglish.org/images/177881-research-notes-57-document.pdf> (accessed on 18 February 2022).
- Tsagari, Dina, and Lying Cheng. 2016. 'Washback, Impact and Consequences'. In *Encyclopedia of Language and Education*, 3rd ed. Edited by Elana Shohamy and Nancy H. Hornberger. Cham: Springer, vol. 7, pp. 1–13. [CrossRef]
- Tsagari, Dina. 2020. Language Assessment Literacy: Concepts, challenges and prospects. In *Perspectives on Language Assessment Literacy: Challenges for Improved Student Learning*. Edited by Hidri Sahbi. New York: Routledge, pp. 13–33.
- Weigle, Sara Cushing. 2004. Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing* 9: 27–55. [CrossRef]
- Weigle, Sara Cushing, and Keisha Parker. 2012. Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing* 21: 118–33. [CrossRef]
- Wette, Rosemary. 2017. Source text use by undergraduate post-novice L2 writers in disciplinary assignments: Progress and ongoing challenges. *Journal of Second Language Writing* 37: 46–58. [CrossRef]
- Wette, Rosemary. 2018. Source-based writing in a health sciences essay: Year 1 students' perceptions, abilities, and strategies. *Journal of English for Academic Purposes* 36: 61–75. [CrossRef]
- Windsor, Angela, and Sang-Soon Park. 2014. Designing L2 reading to write tasks in online higher education contexts. *Journal of English for Academic Purposes* 14: 95–105. [CrossRef]
- Wingate, Ursula. 2015. *Academic Literacy and Student Diversity: The Case for Inclusive Practice*. Bristol: Multilingual Matters.