

RESEARCH

Open Access



Writing assessment literacy and its impact on the learning of writing: A netnography focusing on Duolingo English Test examinees

Chengyuan Yu¹ and Wandong Xu^{2,3*}

*Correspondence:
20031501r@connect.polyu.hk

¹ College of Communication and Information, Kent State University, Kent, OH, USA

² Department of Chinese and Bilingual Studies, Faculty of Humanities, The Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hung Hom, Hong Kong SAR, China

³ School of Foreign Languages, Anhui Agricultural University, 130 Changjiang West Road, Shushan District, Hefei 230036, Anhui, China

Abstract

Language assessment literacy has emerged as an important area of research within the field of language testing and assessment, garnering increasing scholarly attention. However, the existing literature on language assessment literacy primarily focuses on teachers and administrators, while students, who sit at the heart of any assessment, are somewhat neglected. Consequently, our understanding of student language assessment literacy and its impact on learning remains limited. Moreover, previous language assessment literacy research has been predominantly situated in classroom assessment contexts, and relatively little scholarly attention has been directed to large-scale testing contexts. To address these gaps, this study investigated the role of language assessment literacy in students' learning for the writing section of the Duolingo English Test (DET) through a netnographic approach. Twenty-three online videos posted by test takers were analyzed with reference to the existing conceptualizations to examine learners' language assessment literacy of the DET writing section. Highlighting learners' voices, we propose a new model relating writing assessment literacy to learning which has the potential to develop the learner-centered approach in language assessment literacy research. It elucidates the internal relationships among different dimensions of students' language assessment literacy and their impacts on the learning of writing. We, therefore, discussed the findings of this study to argue for the importance of the transparency of assessment and the opportunities to learn provided by large-scale assessments, and to call for teachers' attention to students' language assessment literacy and understanding of the writing construct.

Keywords: Assessment literacy, Learning of writing, Duolingo English Test (DET), Netnography

Introduction

Language assessment literacy (LAL) has been broadly defined as “knowledge about language assessment principles and its sociocultural-political-ethical consequences, the stakeholders' skills to design and implement theoretically sound language assessment, and their abilities to interpret or share assessment results with other stakeholders” (Lee & Butler, 2020, p. 1099). It has been an important research topic during the past two decades due to its impact on learning, teaching, and assessment (Abrar-ul-Hassan &

Nassaji, 2024; Hamp-Lyons, 2016; Vogt et al., 2024). Despite the disagreement in the definition of LAL (Coombs & DeLuca, 2022; Lee & Butler, 2020), it is believed to be crucial for all stakeholders who engage in assessment-related actions (Fulcher, 2012; Levi & Inbar-Lourie, 2019), including delivering/interpreting/using assessment results. An increasing number of studies have investigated the LAL of preservice and/or in-service teachers (e.g., Baker & Riches, 2018; Fulcher, 2012; Kelly et al., 2020; Levi & Inbar-Lourie, 2019; Sun & Zhang, 2022; Vogt & Tsagari, 2014), university admission officers (Deygers & Malone, 2019), employers (Pan & Roever, 2016), and policymakers (Pill & Harding, 2013). However, many scholars have argued that more empirical investigations are warranted to deepen our understanding of the LAL of learners (e.g., Xu et al., 2023). Firstly, learners' LAL research is important in many learner-centered contexts of language teaching, as the interconnected relationships among instruction, learning, and assessment in language education and curriculum development converge on learners. Secondly, learners' LAL is closely associated with their self-regulated learning which materialize in learning outcomes (Smith et al., 2013; Torshizi & Bahraman, 2019). Specifically, learners' understanding of assessment purposes influenced the content of their learning (Sato & Ikeda, 2015), and assessment-literate learners were expected to engage actively and gain a greater agency of learning (Vlanti, 2012). The linkage among instruction, learning, and assessment is widely observed in previous studies concerning learners' views and understanding of assessment (e.g., Knoch & Elder, 2010; Lee, 2008; Sato & Ikeda, 2015; Xie & Andrews, 2012; Xie, 2013, 2015), although they were less discussed theoretically in the scholarship of LAL.

Furthermore, current research has been mainly situated in classroom assessment contexts to conceptualize LAL (e.g., Butler et al., 2021; Lee, 2017), to survey stakeholders' LAL (e.g., Lam, 2019), or to examine the effectiveness of LAL training (e.g., Deeley & Bovill, 2017). Only a few studies have been conducted to investigate the role of LAL in learning, while most of them were conducted to examine teachers' LAL and its indirect effect on students' achievements (e.g., Mellati & Khademi, 2018). Relatively few studies have focused on learners' assessment experience to conceptualize their LAL and its effect on learning outcomes especially in a high-stakes testing context. There is an apparent need for researchers to articulate how test takers perceive and comprehend assessment and the role of these perceptions in their learning from their own voices. These efforts hold significant promise in reshaping the traditional assessment structure by acknowledging learners' active involvement and making informed decisions about current assessment practices.

To fill the abovementioned gaps, this study investigated the role of LAL in students' learning towards the writing section of a large-scale high-stakes language test, namely DET, an at-home English as a second language (L2) test. By conducting a netnography of 23 vloggers in an online video sharing community, this study could offer a better understanding of the effects of writing AL (WAL) on the learning of L2 writing and further shed light on the relationship between assessment and learning. Focusing on writing, this study could enhance our understanding of LAL for writing or WAL, a construct attracting increasing scholarly attention because of the importance of writing in language learning, teaching, and assessment (Rad & Alipour, 2024; Weng, 2023). Situated in large-scale testing contexts, this study could add new insights to the current discourse of

LAL that has been dominated by classroom assessment. Under the influence of COVID-19 global pandemic, at-home online language tests are gaining increasing popularity as a “new norm,” which it warrants continued attention. Here we would like to mention that learner, student, and test taker are used interchangeably in this study. We understand that these constructs are different in meaning, but in the present study, they refer to the same population, who are university students and potential test takers learning English to prepare for DET.

Literature review

LAL research has been influenced by research of AL. LAL has also borrowed conceptualizations from AL. Therefore, this section starts with a review of the concept of AL. AL is a multidimensional construct encompassing stable knowledge of assessment theories and contexts, practical skills in assessment development and usage, and fundamental principles guiding the proper use of assessment and its consequences (Brindley, 2001; Davies, 2008; Inbar-Lourie, 2008; Lee & Butler, 2020; Taylor, 2013). The sociocultural-political-ethical nature of AL indicates that the understanding of language assessment is not independent of contexts where they are used, for example, an international English proficiency for university admission. As a response, Fulcher (2012) advocated an inductive approach to AL research which is situated in specific cases of assessments.

While few studies have been conducted to investigate learners' AL, it has been found that even young learners, including primary school students, may already possess substantial AL in different dimensions (Butler et al., 2021). Butler and her colleagues (2021) found that these young learners were able to reflect on and articulate their assessment experiences even without formal assessment training. They had formed a clear understanding that assessment should be learning-oriented and motivation-driven, and focused on language use rather than linguistic forms. Surprisingly, they were also aware of the impact of construct-irrelevant issues such as anxiety and time allocation. Additionally, learners were willing to share their needs, interests, and test-taking strategies to facilitate assessment development. Their studies suggest that we should hear learners' voice, further develop their AL, and not take our assumptions for granted.

A few studies have investigated learners' engagement with and understanding of assessment as well as its effects on learning. Sato and Ikeda (2015) explored the extent to which students' understanding of the ability measured by high-stakes test items was similar to test developers' intention. They found that many learners lacked the knowledge of the intended constructs underlying the test under investigation, resulting in ineffective learning as they devoted significant time and efforts to prepare for skills or abilities that were not de facto measured or less important. Conversely, test takers with greater knowledge of the test demonstrated higher levels of confidence and self-regulation in test preparation (Xie & Andrews, 2012). Generally, learners are more concerned with final scores than the assessment purpose and procedures, potentially impacting their learning autonomy in a negative way (Vlanti, 2012). Smith et al. (2013) found that AL interventions were conducive to the development of evaluative judgement in self- and peer-assessment, thus enhancing learning outcomes. While these studies enhanced our understanding of AL, they did not specify how different dimensions of AL influence actual learning performance separately and jointly.

In the field of large-scale language testing, it has been found that test takers are able to form their views of the test, including its assessed knowledge and skills, difficulty, validity, and reliability (Messick, 1982). Such views, often referred as test takers' perceptions of the test (Gu et al., 2017; Yu & Zhao, 2021; Xie, 2015), have been investigated in a limited number of studies, which result in a mixed picture of their effects on learning. For example, Lee (2008) and Knoch and Elder (2010) found that test takers' preferences for test task types and perceptions of time arrangement had no effect on their learning outcomes. In contrast, Qi (2007) and Xu and Wu (2012) found that test takers perceived aspects like good handwriting, formatting, essay length, and accurate language use to be more important than official rating criteria and invested more time and efforts in practising these aspects. Qi (2007) argued that it might be a byproduct of teachers' instruction. Nonetheless, more research is needed to frame learners' AL and articulate its different dimensions in relation to their learning.

Theoretically, research has conceptualized student AL, but not specifically for language assessment. For example, Chan and Luo (2021) proposed four dimensions underlying learners' AL in holistic competency development: *knowledge*, *attitude*, *action*, and *critique*. In alignment with the majority of AL studies, *knowledge* dimension serves as the threshold entailing the understanding of why and how the assessment is administered. The *attitude* dimension highlights learners' self-appraisal of assessment value, regulation of emotion-laden assessment experience, and willingness to engage in assessment practices. By *action*, learners are supposed to develop strategies for the completion of different assessment tasks, reflect intentionally, and evaluate and uptake assessment and feedback to seek further improvement. The *critique* dimension refers to learners' awareness of their right to question, analyze, engage in dialogues, and work to improve the assessment mechanism. This framework has been used in WAL research and proved to be suitable (Rad & Alipour, 2024; Xu et al., 2023). Therefore, the present study used it as the theoretical framework.

Duolingo English Test and its writing section

The Duolingo English Test (DET) is an online language assessment developed and managed by the Duolingo Language Learning program (<https://www.duolingo.com/>). Unlike traditional academic English proficiency tests, DET aims to assess daily English proficiency or "real-world language ability" (<https://testcenter.duolingo.com/faq>) across four subskills: reading, writing, listening, and speaking (Ye, 2014). It has been gaining popularity as a large-scale and high-stakes test, particularly after the outbreak of the Covid-19 pandemic (Wagner, 2020; Wagner & Kunnan, 2015). DET is designed as an at-home computer-adaptive test in which the difficulty of subsequent items is determined by test takers' performance in previous items. Test takers can take the test at their convenience on any reliable Internet-connected computer with a supported browser, front-facing camera, microphone, and speaker, in a quiet and well-lit room. In the writing section of DET, test takers are presented with four extended writing tasks, three of which begin with a picture prompt and one with a written prompt (We studied the 2021 version in this study). These tasks require test takers "to describe, recount, or make an argument" (LaFlair & Settles, 2019, p. 10). Due to its adaptive

nature, different test takers may encounter prompts of varying difficulty in the writing section. It is also worth noting that there is an additional unscored writing task, which is not reflected in the test data and not addressed in the present study. The scoring of all DET sections, including the writing one, is done automatically. Test takers receive a single holistic score and separate scores of subskills on a scale of 10–160.

As mentioned above, this study investigated the role of learners' AL in learning specifically in the testing context of high-stakes DET writing section. It aims to address the following two research questions:

- (1) What are DET test takers' knowledge, attitudes, actions, and critiques regarding the DET writing section?
- (2) How does DET test takers' writing assessment literacy (knowledge, attitudes, actions, and critiques) affect their learning of English language writing?

Methods

Data collection

In response to Fulcher's (2012) call for an inductive approach in LAL research, this study adopted a qualitative design to study learners' first-hand perspectives and experiences. Different from traditional qualitative approaches that draw on interviews, field notes, or journals, this study collected data from online platforms, a methodology increasingly embraced by researchers in applied linguistics in general (e.g., Kessler et al., 2021; Kulavuz-Onal, 2015) and language assessment specifically (e.g., Kim, 2017; Yu & Zhao, 2021). Researchers argue that data collected from online platforms are naturalistic and unobtrusive, thus truthfully reflecting frontline stakeholders' voices.

In this study, the authors conducted searches on Bilibili, the most influential Chinese video-sharing platform, with keywords such as "DET" and "DET writing." The returned results were ordered regarding the times of being viewed. To ensure that the dataset captured meaningful interactions reflective of viewers' WAL, videos with either less than five comments or less than five bullet chats were screened out. The final dataset included 23 videos that introduced DET writing and/or shared test preparation and learning experiences, together with viewers' comments and bullet chats. As these video data were interactive, they can reflect test takers' "abilities to interpret or share assessment results with other stakeholders," which is an important aspect of language assessment literacy (Lee & Butler, 2020, p. 1099). Table 1 displays detailed information of the selected videos. The number of bulletin chats ranged from 0 to 184; the number of comments ranged from 6 to 582; the length ranged from 4:32 to 52:34. As mentioned by these test takers, their scores ranged from 110 to 145, indicating a diversity of performance levels. The length of the videos and the number of comments and bulletin chats demonstrated the richness of insights from the vloggers.

Data analysis

The data analysis in this study involved two distinct but complementary approaches. To address RQ1, Chan and Luo's (2021) framework (including knowledge, attitude, action, and critique) served as the initial theoretical coding scheme for the deductive identification and characterization of assessment literacy in the selected videos. The data analysis

Table 1 Details of video data

Video no	Test takers' score	No. of bulletin chats	No. of comments	Length
Video 1	135	184	196	34:17
Video 2	110	16	127	9:39
Video 3	125	27	126	13:18
Video 4	125	113	582	7:53
Video 5	125	43	155	10:20
Video 6	130	38	197	20:25
Video 7	140	5	36	24:18
Video 8	145	0	6	17:22
Video 9	135	28	85	15:00
Video 10	125	20	70	35:46
Video 11	NM	15	73	22:03
Video 12	115	3	22	9:19
Video 13	115	16	38	9:27
Video 14	125	9	9	6:50
Video 15	NM	13	12	7:29
Video 16	NM	2	7	5:22
Video 17	NM	39	7	24:38
Video 18	NM	67	49	27:17
Video 19	NM	20	31	52:34
Video 20	NM	43	34	9:11
Video21	NM	24	12	7:00
Video 22	NM	2	16	4:32
Video 23	NM	45	7	4:46

NM means not mentioned

for RQ2 is completely inductive, featuring a process of open, axial, and selective coding (Corbin & Strauss, 1990), and helps to explore the role of AL in test takers' learning of writing.

The collected videos were subject to meticulous examination and analysis by the two researchers in this study. They were firstly transcribed verbatim and cross-checked by the researchers to ensure accuracy. This process allowed the researchers to deeply engage with the video data and develop a basic understanding of its content. Subsequently, the video transcripts were segmented into different meaning chunks that are of interest and relevance to the research questions. Then Chan and Luo's framework (2021) was referenced to code data segments deductively in terms of different dimensions of AL during test preparation, taking, and reflection stages. The bullet chats posted by the viewers, who constituted DET candidature, were treated as interactions that revealed test takers' agreement or disagreement with the vloggers' perspectives on the understanding, preparation, and practices regarding the DET writing section. The synchronous bullet chats were analyzed in conjunction with the corresponding meaning chunk from the video within the same time frame. For instance, when a vlogger mentioned the importance of rich vocabulary, several audiences requested recommendations on vocabulary books and lists in their synchronous bullet comments. These bullet chats were analyzed in line with the *knowledge* dimension of AL. It should be noted that although the analysis of videos and bullet chats did not follow a multimodal approach explicitly,

the visual components of the videos aided in contextualizing the researchers' analysis. Additionally, the asynchronous comments accompanying the videos were exported and consolidated into a single document for analysis. This strand of data was analyzed also in an inductive way (Corbin & Strauss, 1990) and triangulated the analysis of videos and their embedded bullet chats. We also drew a diagram to better demonstrate the intricacy of multiple dimensions underlying AL and their role in the learning of writing based on results obtained in this study. It was initially created by the first author and underwent rounds of discussion among the two researchers and another colleague to improve its representativeness and accuracy. As the textual data used in this study are in Chinese, we translated the excerpts presented in this paper for international audience. The translation was prepared by the second author and checked by the first author who is a certified English-Chinese translator by Chinese government.

For researcher positionality, we are English as a foreign language speakers with professional proficiency and study writing and language testing and assessment. We received education that emphasized exams and scores and believe in the importance and usefulness of test preparation. Our prior knowledge in this research field and educational experience can influence our analysis of the data in this study. We cannot eliminate the impact of our prior knowledge and experience, but we discussed our preconception during the data analysis through constant discussion. Here, we would like to remind the readers to interpret the findings of this study within the context and keep it in mind that they should not generalize the findings of this study to other contexts.

Results

Informed by Chan and Luo's (2021) model of AL, the analysis of the 23 videos led to four dimensions of test takers' AL in relation to the DET writing section: (1) knowledge, (2) attitude, (3) action, and (4) critique. Following are the findings of this study, which are structured according to the four dimensions of Chan and Luo's model.

Knowledge

Most test takers articulated their different understanding of the purpose, process, and evaluative criteria regarding the DET in general and the writing subtest specifically in the videos.

As for the assessment purpose, most vloggers reached a consensus on the appearance of DET as a timely alternative to the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL) during the Covid-19 pandemic when test centers of the two influential English tests were temporarily shut down in China. Due to its high-stakes nature as the gatekeeper for admission to international universities, a utilitarian view of assessment gained popularity among these vloggers. It motivated them to invest more effort and get well-prepared for the sake of assessment scores since they did not want to lose the opportunity to get an authoritative language proof for their university application:

There are thousands of universities and institutions around the world accepting DET scores in the admission process. It is a good choice for students who have a basic command of English and an urgent need for a proficiency certification in this language. (Video 8)

The test is not as easy as we thought. If you want to get a high score, it is necessary to get fully prepared and wholly engaged in different practices. You can only take the test twice within one month, so seize the opportunity firmly. (Video 5)

In addition to realizing DET as a test for admission purposes, some vloggers (e.g., Videos 1, 6, and 9) also held DET as a comprehensive measurement of speaking, writing, reading, listening, and vocabulary skills which provides an incentive for them to focus on their English learning and get all-rounded development. As one vlogger mentioned in Video 12:

DET requires more effort and time to learn English if I am strongly determined to gain an ideal score on this test. It also offers me an opportunity to develop my English proficiency, especially my writing skills. Without these test preparation activities, I wouldn't put in that much effort and time to have regular learning. (Video12)

Inconsistent with the side effects of assessment in Chan and Luo's study (2021), some vloggers mentioned the beneficial effects of assessment on their learning. They recommended a basic understanding of the assessment before test taking and affirmed the value of assessment in promoting learning and enhancing learning autonomy and persistence.

I strongly recommend the introduction video posted by Mr. Pan (one prominent vlogger on the Bilibili platform) which provides very useful information about the nature, structure, and question format of the test. It aids you in arranging suitable learning activities and better interpreting your test results. The score achieved for the first time is also a good reference for you to find out which aspects you are not qualified for and make much progress. (Video 6)

Once I knew the test well, I created a general learning plan. I find different materials for each type of assessment task and set a daily goal to achieve. There is no shortcut to success in English speaking and writing, and accurate guidance and continuous practice are the key. (Video 2)

I completed two writing sample tasks every day and I kept doing it for almost 35 days. During the test preparation process, I do feel that persistence is the best teacher who guides you to find the right way to write more. (Video 14)

Viewers hardly posted their comments and requests for explanations when the vloggers delivered their understanding of the purpose of assessment and its impacts on their learning of English in general and writing skills specifically, suggesting that viewers were less critical of the assessment.

When approaching the assessment process, three features of DET were salient: (1) computer delivered, (2) adaptive, and (3) the use of automated scoring for writing. Some vloggers and their viewers emphasized that the potential candidates should be familiar with the computer-delivered writing tests and possess adequate typing speed. For example, one viewer mentioned: "If you could type fast, you will have a big advantage" (Video7_Comment). This perception forced them to practice typing speed while preparing for DET writing:

It is very necessary to practice typing since DET writing tasks require composing an essay within a strict time limit. Generating ideas, selecting appropriate words and phrases, translating and organizing what you have thought about into a text, all these processes pose high levels of pressure on your mind as well as your hands. (Video6_Comment)

Apart from mentioning the computer-delivered format of assessment, test takers were aware of the adaptive nature of DET:

DET is essentially adaptive in which the question difficulty is aligned with the test takers' proficiency. The difficulty of the following items is determined by your performance on the foregoing ones. Adaptive testing offers students items at the right difficulty and produces more accurate results than conventional tests. (Video8)

In addition, some vloggers expressed their understanding of the assessing and scoring of writing skills in the DET and commented on its influence on L2 learning performance and writing skill development. Specifically, they believed that automated scoring mainly evaluated lexical and syntactic diversity and was not capable of assessing coherence and cohesion. Such a belief made them memorize and “fill” advanced vocabularies into complex template sentences that they had learned and practised when preparing for the DET writing section.

DET writing responses are scored automatically with statistical machine learning. Machines are not able to make judgements on writing logic, so more attention should be directed to other aspects of text quality like lexical sophistication and diversity. Test takers will not be penalized for chaotic structures. (Video10)
However, the writing section of the DET is not designed to measure coherence and cohesion due to automated scoring. So, there is no need to pay attention to it. The key is to fill in advanced vocabularies into the template sentences I have prepared. (Video12)

Some viewers of these videos asked for lists of advanced vocabulary and complex sentences in comments and bullet chats with requests like “Could you please share the word lists?”. One viewer even sent such a bullet chat: “I am ready to note down. It's super useful!” (e.g., Videos 3, 12, 20) when the vloggers were presenting template sentences. All these comments and bullet chats revealed the tacit agreement on taking advantage of assessment features in test preparation and learning. An official intervention is needed to help students learn and analyze the scoring system of assessment, thus applying assessment appropriately and accurately for actual learning development rather than a higher score.

Some vloggers were also aware of the evaluative criteria for the DET writing section and summarized key components of the criteria with their personal insights injected in such as length, lexical and syntactic complexity, accuracy, and relevance to the topic:

Length is an important criterion for the writing section of DET. If you want to achieve a score beyond 120, you need to type more than 85 words in the Writing Sample task. In this way, I kept practicing my typewriting and tried to improve its speed. (Video3)

I think the most important part of the DET is to test our vocabulary base because it supports different assessment tasks. (Video4)

From the natural acquisition, we should prioritize grammatical accuracy over grammatical complexity in writing. On the contrary, we must enhance the complexity of sentences in practice at first and then do repeated revisions to ensure accuracy when facing an exam-focused situation. (Video20)

Such beliefs were consistent with those held by the viewers. For example, as the viewers mentioned, “The sophistication of vocabulary and sentences is the most important” (Video20_BulletChat), and “Use ‘first and foremost’ rather than ‘first,’ and the machine will give you a higher score” (Video23_BulletChat). Also worth noting is that length was held by most vloggers as an important criterion of writing, although it is not specified in DET official documents. Besides, almost all vloggers in this study took advantage of the limitations of automated scoring by focusing on lexical and grammatical elements rather than on content and neglecting the organization of writing deliberately in their writing. A deliberate emphasis on length but disregard for coherence and cohesion jointly suggests test takers’ inadequate grasp of the evaluative criteria of the DET writing section.

To summarize, test takers’ knowledge of the purpose, process, and evaluative criteria influenced specific learning strategies in their test preparation process of DET writing (i.e., *knowledge of test purpose, process, and evaluative criteria* → *develop learning strategies*).

Attitude

The vloggers affirmed the value of the DET as an authoritative proof of their general English proficiency and writing in specific. By critically comparing the DET with IELTS and TOEFL, a large majority of the vloggers preferred this test as a reliable alternative to the latter two influential assessment programs. It also suggested that test takers’ knowledge of the test purpose could be associated with their attitudes towards the test (*knowledge of test purpose* → *general attitude*). While some of them knew that the DET was easily accessible as a result of its online format and shorter sitting time, they still insisted that having a serious attitude was necessary. Keeping a serious attitude towards assessment, they managed to set learning goals, came up with various learning strategies, accessed multiple sources of materials, and expended more time and effort. The findings indicate that the attitude towards DET was associated with the test takers’ investment in learning (*attitude* → *investment in learning towards the test*). The positive attitudes not only influenced their test preparation activities, but also their emotional experience along with the assessment process. A few vloggers recalled their experiences of managing anxieties and other negative emotions:

Never feel nervous when encountering a difficult topic in writing tasks. Although you don’t understand what the question means or have no background knowledge relevant to this topic, you can write down the template sentences at first and then add content bravely according to your understanding of the question even if it is incorrect. Keep writing down! (Video17)

I tried to re-adjust my heart quickly through deep breaths when doing sample tests. (Video13)

Not only were the pressures caused by the test situation, but also by unsatisfactory results that test takers received from the mock test system. As one viewer reported, “I am even more anxious after I received a low score in the mock test than before I began to have test preparation” (Video3_Comment). However, a few vloggers could address the pressure by reasonably interpreting these low scores in the sample mock tests. For example, one vlogger thought that “it is not necessary to panic when you receive a low score in the mock test as it may have just selected a bunch of items that you are not good at. From a different angle, you can identify weaknesses to overcome and find ways to improve” (Video6_Transcript). In this way, knowledge of the testing process (e.g., the formation process of test papers) has close connections to the attitude dimension of assessment literacy, which is positively associated with test takers’ engagement with the assessment and contributes to better learning outcomes (*knowledge of test process* → *emotion regulation* → *engagement with assessment* → *learning outcomes*).

Although some vloggers and viewers were motivated to practice writing during the test preparation process, they lacked the long-term motivation to practice different aspects of their writing skills. Most of them merely set a temporary goal for their English learning and writing skills, which was, to achieve the ideal score that met the university admissions requirement (*general attitude* → *long-term learning*). Long-term motivation to practice different aspects of their L2 proficiency and writing skills was not strongly determined. It can be explained by the fact that these students preferred to regard the Duolingo English Test as a tool to enter a university in English-speaking countries, as captured in the following quote:

The holistic and subtest scores attained in the DET helped me get an offer for further education at a foreign university. It means a temporary end for this stage of English learning. (Video3)

The abovementioned short-term and instrumental motivation can also be gleaned from the comments and bullet chats, as many viewers mentioned the required overall or section score, for example: “For the university I would like to apply, all section scores should be above 100” (Video5_BulletChat).

Action

Consistent with Chan and Luo (2021), this study also identified three types of assessment literate actions: the ability to (1) develop strategies for different assessment tasks, (2) reflect intentionally, and (3) engage with feedback. Firstly, most vloggers were able to develop strategies for different assessment tasks and critically examine the effectiveness of these strategies. Video viewers also actively participated in sharing and posting numerous requests for a wide range of learning materials, including word lists, writing samples, and tutoring videos to aid their learning and preparation for the DET writing section, as shown by the frequent comments and bullet chats “I want word lists!” and “I want templates!”. Vloggers recommended and viewers requested writing templates. It seems likely that writing for the DET tasks were regarded as a mechanical practice to win higher scores instead of meaningful communication.

Why could I produce a text of more than 100 words for the writing task? It is because

I used a long template. It is composed of different types of sentences with high lexical and syntactic complexity. I don't think the essay template may lead to a low score in Duolingo as it is automatically scored. More importantly, it helps me organize my ideas and meet the length requirements under time pressure. (Video10)

Most strategies were executed by test takers to improve scores rather than improving their actual writing ability, as manifested in their preparation. This pattern of strategy use was determined by knowledge of the evaluative criteria of DET writing and general attitudes towards the DET (*knowledge of evaluative criteria & general attitude* → *developing strategies for learning*). With the knowledge of evaluative criteria in mind, vloggers developed corresponding strategies, which occupied a large part of their test-preparation procedures:

We need to ensure grammatical accuracy which means no errors exist in the sentences we write, so feedback and revisions by ourselves or from expert writers are necessary...I need to use many beyond-B2-level words in CEFR vocabulary lists... But we should still write our response in relevance to the given topic. (Video20)

It was also found that most test takers reflected intentionally on their previous test-taking experience to improve task performance and writing ability (*reflection* → *developing strategies for learning* → *long-term learning*). As one vlogger said, the topic categorization that she used in previous writing assessments (the IELTS writing task) had an impact on the DET writing task, as well as future writing practices.

Like IELTS writing tasks, topics of the DET writing tasks can be categorized in line with some themes including social problems and history. Learning to summarize these topics is an important strategy to enhance writing performance. When assigned a task about digital devices, we should pick up their associations with online learning and the potential side effects of technology. (Video2)

Critique

Critique can be identified in two aspects: (1) critically examining the DET writing section and the writing scores and (2) engaging in critical dialogues to improve the assessment. For the first aspect, test takers examined DET and its writing section by referring to their experiences of taking other similar English language tests:

IELTS, TOEFL, and DET are different in writing tests. The DET writing task requires students to produce a response in more than 50 words with a time limit of 5 minutes. I think it is hard to have a careful measurement of writing skills in such simple tasks. At least, the logic aspect of writing is not examined since the text is composed of only 50 words, equal to a few sentences. (Video10)

The DET is an adaptive test, so the difficulty of questions varies significantly. The difficulty of the open-ended writing tasks is determined by my performance on the foregoing objective items, and I am wondering whether they can measure my writing competence consistently. I undertake a few free practice tests and the score often falls between 125 and 150, which is aligned with what I have gained in IELTS. (Video1)

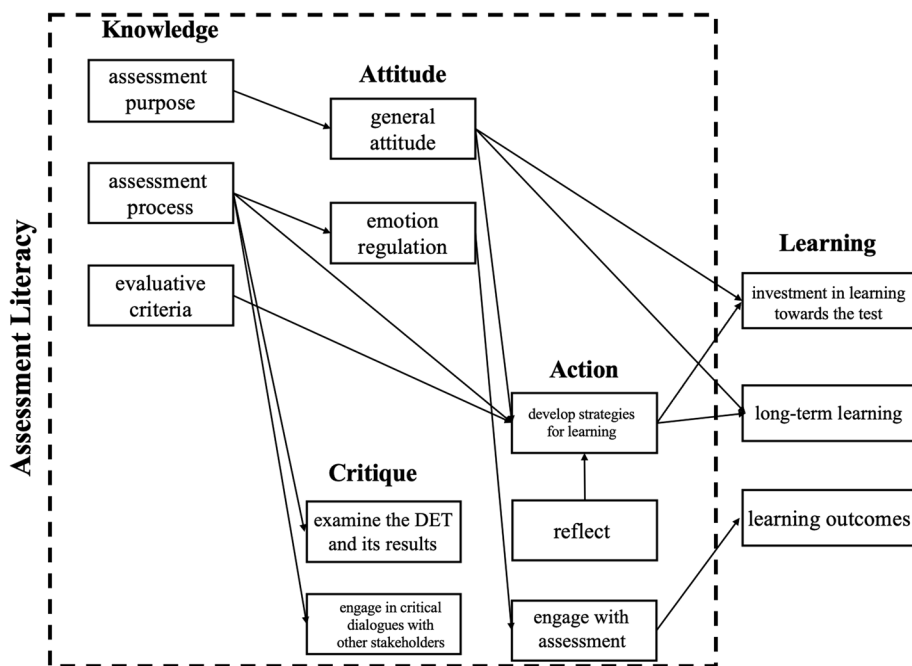


Fig. 1 The relational model of assessment literacy and learning

The second aspect of critique is expanded by engagement in critical dialogues of how the DET writing section should be improved:

Human raters in IELTS and TOEFL pose a high demand for the hidden logic behind our texts. Machine scoring is advantageous for its high reliability compared to human raters. However, current artificial intelligence is so limited to imitation that only surface features of text quality including grammar accuracy, sentence complexity and vocabulary richness can be carefully evaluated. It cannot think independently and probe into the complexity of writing. Maybe the combination of human raters and machine scoring would be better to evaluate and grade students' writing skills. (Video19)

It suggests that knowledge of the test process and evaluative criteria helps to form the critiques of DET writing (*knowledge of test process & knowledge of evaluative criteria* → *critiques*). However, their online critiques might mainly be an outlet for complaints. Their critiques were not found in close associations with their learning in the current dataset. It can be explained by the fact that these test takers have no access to the initial design, development, and validation procedure of DET. Due to its high-stakes nature as a gatekeeper of future admissions, they could only become accustomed to the test, rather than take any actions to improve it.

From the findings presented above, relationships among different dimensions of AL and learning have been identified (see Fig. 1).

This model describes the four dimensions of AL and learning. The knowledge dimension consists of knowledge of (1) assessment purpose, (2) assessment process, and (3) evaluative criteria. The attitude dimension consists of (1) general attitude and (2) emotion regulation. The critique dimension comprises (1) examining the test and its results

and (2) engaging in critical dialogues with other stakeholders. The action dimension comprises (1) developing strategies for learning, (2) reflecting, and (3) engaging with assessment. The learning dimension consists of (1) investment in learning towards the test, (2) long-term learning, and (3) learning outcomes. The knowledge of assessment purpose influences test takers' general attitude towards the tests. The knowledge of assessment process influences both subdimensions of critique (i.e., examining the test and its results and engagement in critical dialogues with other stakeholders), emotion regulation, and actions to develop strategies for learning. The knowledge of evaluative criteria influences actions to develop strategies for learning. The general attitude influences actions to develop strategies for learning. The general attitude towards the test can also directly influence learning in the form of investment in learning towards the test and long-term learning. Emotion regulation influences engagement with assessment. Within the action dimension, reflecting influences the development of strategies for learning, which influences investment in learning and long-term learning. The engagement with assessment influences learning outcomes. Mediation effects can also be observed in this model, and they are summarized as below:

- (1) Knowledge of assessment purpose → general attitude → action to develop strategies for learning/investment in learning towards the test/long-term learning.
- (2) Knowledge of assessment process → emotion regulation → engagement with assessment → learning outcomes.
- (3) Knowledge of assessment process → action to develop strategies for learning → investment in learning towards the test/long-term learning.
- (4) Knowledge of evaluative criteria → action to develop strategies for learning → investment in learning towards the test/long-term learning.
- (5) Reflecting → action to develop strategies for learning → investment in learning towards the test/long-term learning.

Discussion and conclusion

The present study analyzed online videos posted by test takers of the DET. It confirms that AL is a multi-level and multidimensional construct (e.g., Chan & Luo, 2021; Davies, 2008; Smith et al., 2013), and different AL dimensions are involved in complex interactions. This present study not only demonstrates that Chan and Luo's (2021) four-dimensional model is interpretable and applicable in the context of large-scale, high-stakes testing, but also extends their model and provides a nuanced, contextualized account of test takers' AL with the same four dimensions: (1) knowledge (of assessment purpose, process, and evaluative criteria); (2) attitude (i.e., general attitude and emotion regulation); (3) action (developing strategies for learning, reflecting intentionally, and engaging with the assessment); and (4) critiques (critically examining the test and its results and engaging in critical dialogues with other stakeholders).

Consistent with previous research, test takers in this study were able to form their own understanding and knowledge of the multiple aspects of assessment (Knoch & Elder, 2010; Lee, 2008; Messick, 1982; Xie, 2015; Yu & Zhao, 2021). As shown in this study, their knowledge of assessment served as an important indicator of AL. It was also found that the knowledge dimension could exert indirect effects on learning via their direct

effects on the other three AL dimensions (i.e., attitude, action, and critique), which helps elucidate the mechanism of how knowledge and understanding of assessment influence learning (Sato & Ikeda, 2015; Vlanti, 2012; Xie & Andrews, 2012). Specifically, knowledge of the assessment purpose is associated with test takers' general attitude towards the test, which in turn influences how they take actions to regulate their learning and engage in assessment. The strategies they developed for learning not only have a positive relationship with their task performance but also partly influence their investment in learning for the test and long-term writing development. Test takers' knowledge of the assessment process (i.e., the delivery, administration, and scoring procedures) can activate their emotion regulation that indirectly influences their learning outcomes through the mediation of engagement with assessment. Students' knowledge of the assessment process directly influences their actions such as developing strategies for learning, which further contributes to their long-term learning. Furthermore, the two aspects of the critique dimension were mostly about the assessment process, suggesting a close relationship between knowledge of the assessment process and critiques. This perhaps means that the face validity of a test is most related to its assessment process. The knowledge of evaluative criteria is directly associated with test takers' ability to develop strategies for learning and is indirectly associated with their investment in learning towards the test and long-term learning. In addition to the intricate connections between dimensions, different aspects in the same dimension may correlate with each other, for example, test takers' ability to reflect can influence their ability to develop strategies for learning. All these suggest that AL training should be done holistically.

Like previous research, this study also found that learners may have a partial understanding of the measured construct or overemphasis on some construct-irrelevant factors in their learning or test preparation (Qi, 2007; Xu & Wu, 2012). According to Qi (2007), for the writing section of traditional paper–pencil, large-scale, and high-stakes tests, test takers can put more efforts into improving handwriting, formatting, essay length, and the accuracy of language use. In the online at-home DET writing test, as revealed in the findings, learners also paid much attention to essay length and tried to use advanced vocabulary and sophisticated sentence structures in their writing to achieve a high score. It suggests that in the context of large-scale, high-stakes testing like DET, learners' motivation features an instrumental orientation (i.e., caring about scores) instead of an integrative one (i.e., achieving further writing development). Therefore, they focused on the easier and more concrete aspects (e.g., surface-level linguistic complexity and length) that can be improved in a short period of time rather than more complex ones like content, coherence, and cohesion. Such findings seem to be contrary to existing studies situated in the context of classroom assessment where final scores tend to be less important, echoing the view that AL is sociocultural and dependent on specific contexts (Lee & Butler, 2020). Teachers may need to direct students' attention to these more important aspects to facilitate real learning.

Our finding also points to the potential problems of current writing testing methods, particularly the employment of the promising but problematic automated scoring. When facing heavy test pressure, learners tend to “play with” and please the testing system. In this sense, test developers need to hear the voice of test takers to identify the problems that may arise in the actual use of an assessment program (e.g., learning/preparing for

the test), and to examine the negative washback on students' long-term learning. Test developers also need to shoulder the responsibility of increasing the transparency of test materials, particularly scoring criteria, to provide opportunities for test takers to learn to be assessment literate (Kunnan, 2018). When developing an automated scoring system, coherence, cohesion, logic, and content should be integrated to characterize the human's real perception of writing assessment.

In this study, learners expressed their critiques of the DET writing section and its result and engaged in critical dialogues with other test takers. Additionally, assessment-literate learners can see good points of the test, hold positive attitudes, and develop corresponding strategies to facilitate the long-term development of writing. It indicates that elements of large-scale, high-stakes writing tests can also be utilized to facilitate learning (Yu, 2023). Teachers are recommended to pay attention to the improvement of students' writing assessment literacy and guide students to understand the construct of writing in authentic settings as a complement to the measured construct in the writing tests.

Participation in assessments can be emotional/affectional which requires conscious regulation (Cheng et al., 2014; Liu & Yu, 2021). Therefore, it is of great value for teachers to cater to students' affect, including motivation and anxiety in assessment tasks. By doing so, the use of writing assessments could be more beneficial to students.

Finally, this study found that students could actively engage in online informal learning through watching educational videos. Their comments and bullet comments are not only data for this study but also showcase their willingness to contribute to the online learning community. However, their engagement indicates a problematic understanding of language learning, especially the writing construct. In this sense, teachers play a key role in the whole process if they can offer formal classroom instruction targeting to provide fundamental knowledge concerning these issues.

While a model integrating AL and learning in the context of writing tests has been developed and supported by first-hand learner data, findings in this study should be interpreted with caution for its exploratory nature and small sample size. Furthermore, the data in this study can only reflect the views of a certain group of population with particular experience. Researchers can construct learners' writing assessment literacy questionnaire with reference to the model proposed in this study and thus examine its validity with a large population across different contexts.

Abbreviations

DET	Duolingo English Test
AL	Assessment literacy
LAL	Language assessment literacy
IELTS	The International English Language Testing System
TOEFL	Test of English as a Foreign Language

Acknowledgements

We would like to thank the journal editor and the anonymous reviewers for their time and effort in processing our manuscript.

Authors' contributions

The authors were involved in data collection, data analysis, writing, and revising the drafts. They read and approved the submitted manuscript.

Authors' information

Chengyuan Yu researches the testing and assessment of writing, academic writing, and digital and information literacy. His publications have appeared in *Language Testing*, *Asia Pacific Education Review*, *International Journal of Qualitative*

Studies in Education, Asian-Pacific Journal of Second and Foreign Language Education, and *International Conference Proceedings of Association of Language Testers in Europe*. Email: cyu15@kent.edu.

Wandong Xu has defended her doctoral dissertation at the Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, and is a newly employed lecturer at the School of Foreign Languages, Anhui Agricultural University. Her major research interests include language assessment and cognition, reading and writing development, cross-linguistic transfer, metacognition and self-regulated learning, argumentation, and higher-level thinking skills. Wandong's work has been published in international journals like *System*, *Metacognition and Learning*, *Innovations in Education and Teaching International*, *Sustainability*, and some Chinese journals. E-mail: 20031501r@connect.polyu.hk.

Funding

We received no funding for this study.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 27 May 2024 Accepted: 28 June 2024

Published online: 08 July 2024

References

- Abbar-ul-Hassan, S., & Nassaji, H. (2024). Rescoping language assessment literacy: An expanded perspective. *System*, 120, 103195. <https://doi.org/10.1016/j.system.2023.103195>
- Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 35(4), 557–581. <https://doi.org/10.1177/026553221771673>
- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126–136). Cambridge University Press.
- Butler, Y. G., Peng, X., & Lee, J. (2021). Young learners' voices: Towards a learner-centered approach to understanding language assessment literacy. *Language Testing*, 38(3), 429–455. <https://doi.org/10.1177/026553221992274>
- Chan, C. K. Y., & Luo, J. (2021). A four-dimensional conceptual framework for student assessment literacy in holistic competency development. *Assessment & Evaluation in Higher Education*, 46(3), 451–466. <https://doi.org/10.1080/02602938.2020.1777388>
- Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET, and GEPT. *TESOL Quarterly*, 48(2), 300–330. <https://doi.org/10.1002/tesq.105>
- Coombs, A., & DeLuca, C. (2022). Mapping the constellation of assessment discourses: A scoping review study on assessment competence, literacy, capability, and identity. *Educational Assessment, Evaluation and Accountability*, 34, 279–301. <https://doi.org/10.1007/s11092-022-09389-9>
- Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13, 3–21.
- Crusan, D., Plakans, L., & Gebriel, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43–56. <https://doi.org/10.1016/j.asw.2016.03.001>
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347. <https://doi.org/10.1177/0265532208090156>
- Deeley, S. J., & Bovill, C. (2017). Staff student partnership in assessment: Enhancing assessment literacy through democratic practices. *Assessment & Evaluation in Higher Education*, 42(3), 463–477. <https://doi.org/10.1080/02602938.2015.1126551>
- DeLuca, C., Lapointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/s11092-015-9233-6>
- Deygers, B., & Malone, M. (2019). Language assessment literacy in university admission policies, or the dialogue that isn't. *Language Testing*, 36(3), 347–368. <https://doi.org/10.1177/0265532219826390>
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Gu, X., Hong Y., Yu, C., & Sarkar, T. (2017). A comparative study on the washback of CET-6, IELTS and TOEFL iBT writing tests: Evidence from Chinese test-takers' perspectives. In *ALTE Conference 2017 Proceedings* (pp. 160–165). Bologna: Association of Language Testers in Europe. <https://www.alte.org/resources/Documents/ALTE%202017%20Proceedings%20FINAL.pdf>
- Hamp-Lyons, L. (2016). Purposes of assessment. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 13–27). De Gruyter.
- Hyland, K. (2013). Writing in the university: Education, knowledge and reputation. *Language Teaching*, 46(1), 53–70. <https://doi.org/10.1017/S0261444811000036>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402. <https://doi.org/10.1177/0265532208090158>
- Kelly, M. P., Feistman, R., Dodge, E., St Rose, A., & Littenberg-Tobias, J. (2020). Exploring the dimensionality of self-perceived performance assessment literacy (PAL). *Educational Assessment, Evaluation and Accountability*, 32(4), 499–517. <https://doi.org/10.1007/s11092-020-09343-7>

- Kessler, M., De Costa, P., Isbell, D. R., & Gajasinghe, K. (2021). Conducting a netnography in second language acquisition research. *Language Learning*, 71(4), 1122–1148. <https://doi.org/10.1111/lang.12456>
- Kim, E. Y. J. (2017). The TOEFL iBT writing: Korean students' perceptions of the TOEFL iBT writing test. *Assessing Writing*, 33, 1–11. <https://doi.org/10.1016/j.asw.2017.02.001>
- Knoch, U., & Elder, C. (2010). Validity and fairness implications of varying time conditions on a diagnostic test of academic English writing proficiency. *System*, 38(1), 63–74. <https://doi.org/10.1016/j.system.2009.12.006>
- Kulavuz-Onal, D. (2015). Using netnography to explore the culture of online language teaching communities. *CALICO Journal*, 32(3), 426–448. <https://doi.org/10.1558/cj.v32i3.26636>
- Kunnan, A. J. (2018). *Evaluating language assessments*. Routledge.
- LaFlair, G., & Settles, B. (2019). Duolingo English test: Technical manual. Retrieved April 28, 2020, from <https://s3.amazonaws.com/duolingo-papers/other/Duolingo%20English%20Test%20-%20Technical%20Manual%202019.pdf>
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78–89. <https://doi.org/10.1016/j.system.2019.01.006>
- Lee, H. K. (2008). The relationship between writers' perceptions and their performance on a field-specific writing test. *Assessing Writing*, 13(2), 93–110. <https://doi.org/10.1016/j.asw.2008.08.002>
- Lee, I. (2017). *Classroom assessment and feedback in L2 school contexts*. Springer.
- Lee, J., & Butler, Y. G. (2020). Reconceptualizing language assessment literacy: Where are language learners? *TESOL Quarterly*, 54(4), 1098–1111. <https://doi.org/10.1002/tesq.576>
- Levi, T., & Inbar-Lourie, O. (2019). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168–182. <https://doi.org/10.1080/15434303.2019.1692347>
- Liu, X., & Yu, J. (2021). Relationships between learning motivations and practices as influenced by a high-stakes language test: The mechanism of washback on learning. *Studies in Educational Evaluation*, 68, 100967. <https://doi.org/10.1016/j.stueduc.2020.100967>
- Malone, M. (2017, October). *Unpacking language assessment literacy: Differentiating needs of stakeholder groups*. Paper presented at East Coast Organization of Language Testers, Washington, DC.
- Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education*, 43(6), 1–18. <https://doi.org/10.14221/ajte.2018v43n6.1>
- Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice. *Educational Psychologist*, 17(2), 67–91. <https://doi.org/10.1080/00461528209529246>
- Pan, Y., & Roever, C. (2016). Consequences of test use: A case study of employers' voice on the social impact of English certification exit requirements in Taiwan. *Language Testing in Asia*, 6(6). <https://doi.org/10.1186/s40468-016-0029-5>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381–402. <https://doi.org/10.1177/0265532213480337>
- Qi, L. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education: Principles, Policy & Practice*, 14(1), 51–74. <https://doi.org/10.1080/09695940701272856>
- Rad, H. S., & Alipour, R. (2024). Unlocking writing success: Building assessment literacy for students and teachers through effective interventions. *Assessing Writing*, 59, 100804. <https://doi.org/10.1016/j.asw.2023.100804>
- Sato, T., & Ikeda, N. (2015). Test-taker perception of what test items measure: A potential impact of face validity on student learning. *Language Testing in Asia*, 5(10). <https://doi.org/10.1186/s40468-015-0019-z>
- Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students' assessment literacy. *Assessment & Evaluation in Higher Education*, 38(1), 44–60. <https://doi.org/10.1080/02602938.2011.598636>
- Sun, H., & Zhang, J. (2022). Assessment literacy of college EFL teachers in China: Status quo and mediating factors. *Studies in Educational Evaluation*, 74, 101157. <https://doi.org/10.1016/j.stueduc.2022.101157>
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403–412. <https://doi.org/10.1177/0265532213480338>
- Torshizi, M. D., & Bahraman, M. (2019). I explain, therefore I learn: Improving students' assessment literacy and deep learning by teaching. *Studies in Educational Evaluation*, 61, 66–73. <https://doi.org/10.1016/j.stueduc.2019.03.002>
- Vlanti, S. (2012). Assessment practices in the English language classroom of Greek junior high school. *Research Papers in Language Teaching and Learning*, 3(1), 92–122.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11, 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- Vogt, K., Böhn, H., & Tsagari, D. (2024). Language assessment literacy. *Language Teaching, First View*. <https://doi.org/10.1017/S0261444824000090>
- Wagner, E. (2020). Duolingo English Test, Revised Version July 2019. *Language Assessment Quarterly*, 17(3), 300–315. <https://doi.org/10.1080/15434303.2020.1771343>
- Wagner, E., & Kunnan, A. (2015). Review of the Duolingo English test. *Language Assessment Quarterly*, 12(3), 320–331. <https://doi.org/10.1080/15434303.2015.1061530>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- Weng, F. (2023). EFL teachers' writing assessment literacy: Surveying teachers' knowledge, beliefs, and practises in China. *Porta Linguarum*, 40, 57–74. <https://doi.org/10.30827/portalin.vi40.23812>
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10(2), 196–218. <https://doi.org/10.1080/15434303.2012.721423>
- Xie, Q. (2015). "I must impress the raters!" An investigation of Chinese test-takers' strategies to manage rater impressions. *Assessing Writing*, 25, 22–37. <https://doi.org/10.1016/j.asw.2015.05.001>
- Xie, Q., & Andrews, S. (2012). Do test design and uses influence test preparation: Testing a model of washback with structural equation modeling. *Language Testing*, 30(1), 49–70. <https://doi.org/10.1177/0265532212442634>

- Xu, J., Zheng, Y., & Braund, H. (2023). Voices from L2 learners across different languages: Development and validation of a student writing assessment literacy scale. *Journal of Second Language Writing*, 60, 100993. <https://doi.org/10.1016/j.jslw.2023.100993>
- Xu, Y., & Wu, Z. (2012). Test-taking strategies for a high-stakes writing test: An exploratory study of 12 Chinese EFL learners. *Assessing Writing* 17(3), 174–190. <https://doi.org/10.1016/j.asw.2012.03.001>
- Ye, F. (2014). Validity, reliability, and concordance of the Duolingo English Test. Retrieved from <https://s3.amazonaws.com/duolingo-certifications-data/CorrelationStudy.pdf>
- Yu, C. (2023). Testing issues in writing. In H. Mohebbi, & Y. Wang (Eds.), *Insights into teaching and learning writing: A practical guide for early-career teachers* (pp. 56–70). Castledown. <https://doi.org/10.29140/9781914291159-5>
- Yu, C., & Zhao, C. G. (2021). A "netnographic" study of test impact from the test-takers' perspective: The case of a translation test. In *Collated Papers for the ALTE 7th International Conference* (pp. 63–66). Madrid: Association of Language Testers in Europe. <https://www.alte.org/resources/Documents/ALTE%207th%20International%20Conference%20Madrid%20June%202021.pdf#page=70>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.