English Teaching, Vol. 78, No. 1, Spring 2023, pp. 125-144
DOI: https://doi.org/10.15858/engtea.78.1.202303.125
http://journal.kate.or.kr

L2 Students' Views on Writing Tools: Investigating Domain Definition Within Argument-Based Validation Framework

Elizabeth Lee*

Lee, Elizabeth. (2023). L2 students' views on writing tools: Investigating domain definition within argument-based validation framework. *English Teaching*, *78*(1), 125-144.

In many high-stakes testing situations, test-takers are not allowed to draw on external writing resources while writing, a practice observed more frequently in classroom settings. This may pose problems with the representativeness of test tasks and score interpretations. This study investigates the domain definition of one particular test known as the English Placement Writing Test within an argument-based validation framework. Focusing on the domain definition inference, the following rebuttal was evaluated: Certain essential contextual factors in the academic writing domain are not modeled in the test tasks. To do so, lower- and intermediate-level ESL students (n=92) who previously took the test were surveyed and interviewed regarding their uses of computer-based and face-to-face human-assisted writing tools. Results showed that students at both levels were statistically similar in their attitudes toward and uptakes of such tools while writing. The difference in availability of external writing tools between the target and test domain may point to issues with task authenticity of the test.

Key words: writing tools, feedback, attitudes, academic writing

*Author: Elizabeth Lee, Visiting Professor, Institute of Liberal Education, Incheon National University;
119 Academy-ro, Bldg. 12-201, Yeonsu-gu, Incheon-si, 22009, Korea; Email: edl@inu.ac.kr
Received 31 December 2022; Reviewed 15 January 2023; Accepted 17 March 2023

1. INTRODUCTION

When international students are admitted to an English-medium university, they are expected to be equipped with the basics of academic language skills (Andrade, Evans, & Hartshorn, 2015). To ensure this expectation, students' TOEFL iBT and IELTS scores as well as placement test scores are used to make admission and placement decisions, and numerous studies support the use of these high-stakes tests as they can accurately and appropriately measure students' language skills and thereby inform decision-makers whether students are prepared to handle academic language demands at the university level (Chapelle, Enright, & Jamieson, 2008). However, being proficient in academic writing goes beyond meeting certain lexical, syntactic, and discourse standards. Good academic writers are keenly aware of the rhetorical situation in which they produce the text (Ferris, 2009). They also actively plan, monitor, and evaluate not just what they are writing about but also how they go about completing their writing goals (Ferris, 2009). This often involves drawing on external writing resources such as computer-based and human-assisted writing tools (Andrade et al., 2015). In other words, good academic writing does not happen in a vacuum but rather occurs in an open communicative space with multiple "readers"—be that with an automated written corrective feedback (AWCF) tool such as *Grammarly* or with a graduate-level writing tutor attending to a student's writing needs.

Due to their product-oriented output, however, test scores drawn from high-stakes proficiency or placement tests do not inform decision-makers how capable international students are at accessing and exploiting external writing resources as part of their regular writing process. This is not ideal, given that faculty expect students—ESL or not—to make good use of technology and other writing resources to complete various academic writing tasks, ranging from basic 5-paragraph essays to graduation theses. Although testing learners' linguistic knowledge is an essential feature of language test tasks, if the task is missing a critical component reflective of the target domain (in this case, the drawing of external writing resources), the information that one could garner from a test score is severely limited (Dimova, Yan, & Ginther, 2020).

Because high-stakes tests such as placement tests play an important role in many universities that admit and place international students, this warrants an investigation into test task representativeness, particularly one that generates more participation from test-takers. Motivated by the dearth of research on this subject, this study therefore seeks to investigate the domain definition of one particular test known as the English Placement Writing Test (EPWT) within an argument-based validation framework (Kane, 2006). The findings underscore the importance of developing authentic test tasks that go beyond testing language knowledge and skills.

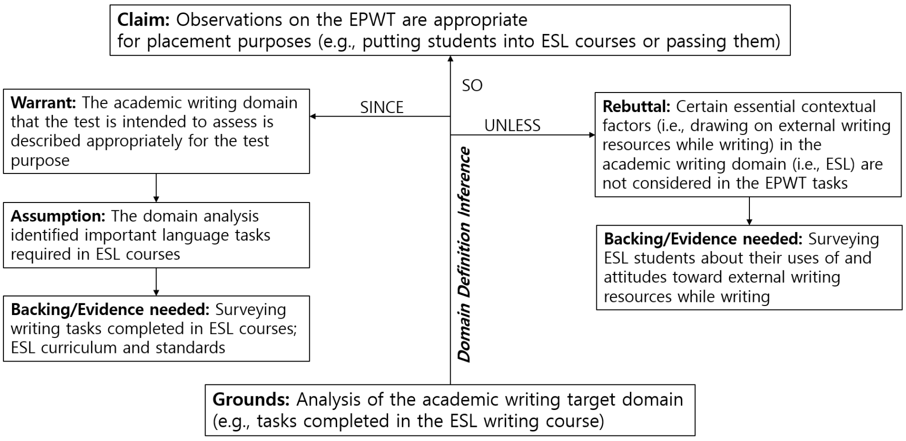
2. REVIEW OF THE LITERATURE

2.1. An Argument-Based Approach to Validity

Validity, which is "the degree to which evidence and theory support the interpretations of test scores for proposed uses of test scores" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 1), is an important aspect of language testing. Without strong validity, it would be difficult to justify to stakeholders the uses of test scores for their intended purposes. Although there are many ways to go about establishing validity, in recent years, using an argument-based validity framework such as the one proposed by Kane (2006) has been drawing strong traction among language testers and test researchers. An argument-based validation is essentially a research program (Chapelle, 2020) which requires researchers to accomplish two phases: The first is to establish an interpretive/use argument (IUA) framework in which the intended interpretations, uses, potential rebuttals, and the evidence needed to support each claim is explicitly stated. The claims are stated in an orderly chain-like structure using inferences (see Figure 1). An inference refers to the "if-then" argument structure (modeled after Toulmin's model of argument) where a conclusion/claim is made from the grounds/premise (Kane, 2006), and it is also a process in which a researcher goes about drawing conclusions based on the gathered validity evidence (Chapelle, 2020). Although there is not a one-size-fits-all approach, generally, an IUA may include the following inferences: *domain definition*, *evaluation*, *generalization*, *explanation*, *extrapolation*, *utilization*, and *consequence implication* (for a more detailed discussion on each inference, see Kane, 2006 and Chapelle, 2020). More will be said about structuring arguments within an IUA framework in the following paragraph. In the second phase, researchers have to evaluate the claims stated in the IUA and then establish a validity argument. According to Kane (2022), researchers would take on a more "confirmationist approach" (p. 56) in the first phase, and in the later phase, a more "critical stance" (p. 56), since the goal of developing an IUA is to justify test score interpretation and uses, whereas the goal of establishing a validity argument is to test the credibility of such claims. In addition to its pragmatic advantages, the argument-based validation offers "rhetorical devices" that facilitate researchers to state their claims (Chapelle, 2020).

FIGURE 1

The Domain Definition Inference, Adapted from Li (2015)


Claim: Observations on the EPWT are appropriate for placement purposes (e.g., putting students into ESL courses or passing them)
SO
SINCE
UNLESS
Warrant: The academic writing domain that the test is intended to assess is described appropriately for the test purpose
Assumption: The domain analysis identified important language tasks required in ESL courses
Backing/Evidence needed: Surveying writing tasks completed in ESL courses; ESL curriculum and standards
Rebuttal: Certain essential contextual factors (i.e., drawing on external writing resources while writing) in the academic writing domain (i.e., ESL) are not considered in the EPWT tasks
Backing/Evidence needed: Surveying ESL students about their uses of and attitudes toward external writing resources while writing
Domain Definition Inference
Grounds: Analysis of the academic writing target domain (e.g., tasks completed in the ESL writing course)

A single test score may have several interpretations and uses, all of which have to be stated and validated separately. In this paper, we will investigate the domain definition inference of the EPWT. An IUA for this placement test has been previously developed by Li (2015) and will be adopted in the present study with some modifications made. Specifically, Li's IUA closely adapted Chapelle et al.'s (2008) TOEFL IUA and consequently some of the wordings appeared a little too generic. For this reason, statements were updated to specify the local context and modeled after Chapelle's (2020) rhetorical device. As seen in Figure 1, the domain definition inference is where one establishes a claim about the appropriateness of the observed test performance from the grounds, the target domain. According to Chapelle (2020), from a constructivist-realist view, a domain refers to both an abstract area of social importance (e.g., academic writing) as well as an area established by a social community through its artifacts (e.g., the academic writing curriculum set by an English department at one university). A target domain refers to non-testing performances taking place in such a domain (e.g., academic essays written by students for non-testing purposes). Within the domain definition inference, the goal is to observe the extent to which test performances (e.g., EPWT score) are reflective of the performances accomplished in the domain of interest (e.g., essays written in ESL courses), which often involves (but not limited to) identifying key knowledge, skills, and abilities as well as tasks performed in the target domain.

To establish the claim from the grounds, however, one must elaborate further and this is accomplished through warrants and assumptions. Warrants are statements that link the grounds/data to the claim/conclusion, which need to be validated. For example, the following warrant, "the construct domain that the test is intended to assess is described appropriately for the test purpose," allows one to infer that observations of EPWT performances are appropriate (claim) from the observations of performances in ESL courses (grounds), provided proper evidence is gathered. Assumptions further expand on the warrant and identify the specific evidence that would be gathered. In the same example, an underlying assumption of the aforementioned warrant would be, "the domain analysis identified important language tasks required in ESL courses." The evidence that would have to be gathered would come from a domain analysis. A domain analysis typically involves examining the tasks and task characteristics of the target domain, the curriculum and standards, and test specifications, which are conducted by a panel of content experts (e.g., instructors) (Chapelle, 2020; Chapelle et al., 2008). Chapelle (2020) underscores the importance of collecting and analyzing both quantitative and qualitative types of data that address research questions which are motivated by a need to evaluate assertions made within one's IUA.

While the IUA and the development stage of the validity framework take on a confirmationist approach, this does not exclude researchers from stating and investigating potential rebuttals. Rebuttals are statements that could modify, weaken, or dismiss the claims that are stated in one's IUA. They act as the flipside of warrants (Chapelle, 2020), and similar to assumptions, researchers have to specify the evidence that would be needed to evaluate and validate the rebuttals. Although rebuttals are not desirable for those trying to defend test use, finding only positive evidence for the stated assumptions and warrants is not enough to establish strong credibility; evidence of absence of rebuttals is also needed to ensure that the claims remain legitimate. In our model, the rebuttal "certain essential contextual factors (i.e., drawing of external writing resources while writing) in ESL courses are not modeled in the EPWT tasks," requires backing that certain essential contextual factors that students do in the ESL writing course are in fact missing from the test task. Test tasks that fail to integrate essential contextual factors are likely to generate test scores that may not be wholly relevant to the context in which score-based decisions have to be made. In this study, the researcher acts as an outside evaluator (i.e., someone who did not participate in the test development process) and investigates the rebuttal associated with the domain definition inference. The backing used in this study draws on test-takers' perspectives, both quantitatively and qualitatively, which have not been adequately addressed in peer-reviewed studies investigating the domain definition inference.

2.2. L2 Students' Uptakes of Computer-Based and Face-to-Face Human-Assisted Writing Tools

Consulting outside sources while writing, such as digital assistance writing tools or feedback from a writing expert, is a common practice observed among many ESL/EFL students in academic writing settings. Due to a wide variety of feedback tools and writing resources made available nowadays, we limit computer-based writing tools (CBWT) and face-to-face human-assisted writing tools (FFHAWT) to those that ESL/EFL learners in higher education are reported to commonly use or are exposed to: CBWT include AWCF (Koltovskaia, 2020; Ranalli & Yamashita, 2022) such as Grammarly, Microsoft Word's spelling and grammar checker (MS-NLP) (Ranalli & Yamashita, 2022), online dictionaries (ODs) (O'Neill, 2019), online translators (OTs) (Cancino & Panes, 2021; O'Neill, 2019), and search engines (SEs) (O'Neill, 2019). FFHAWT include writing center tutors (Williams & Severino, 2004), course instructors (Hawe & Dixon, 2014; McMartin-Miller, 2014), and peers (Tigchelaar & Polio, 2017).

Numerous studies looking into CBWT and FFHAWT have investigated students' motivation for and responses to these writing resources. When it comes to CBWT, L2 learners are more likely to select a tool that provides clear, specific, and accurate feedback (or the desired output); is highly accessible; is perceived to be trustworthy; and is least cognitively demanding. Ranalli (2018) reported that ESL students were more likely to accurately correct errors based on the feedback of an automated writing evaluation tool, known as *Criterion*, when it was specific and an assessment of the tool's accuracy was not required (in other words, reducing student's cognitive demands for having to evaluate the trustworthiness of the accuracy of feedback). In O'Neill's (2019) study, students used ODs and OTs, and to a lesser extent, SEs for graded and non-graded assignments; their perceptions of using ODs and OTs for writing (in Spanish or French) were largely positive. In particular, ODs were reported to be useful for looking up words or for confirming, whereas OTs were reported to help understand phrases and sentences. Interestingly, although OTs were deemed less trustworthy as a result of providing incorrect translations, students in O'Neill's study used OTs anyway due to their high accessibility and ease in lowering cognitive demands (that is, students would have to think less while using OT as opposed to without it). Although students' preferences for certain computer-based tools may not always yield accurate corrections (Koltovskaia, 2020; Ranalli, 2018; Ranalli, Link, & Chukharev-Hudilainen, 2017), studies have demonstrated that the uptake of a particular computer-based tool (so long as it fulfills the purpose(s) of the task at hand) can potentially lead to a higher writing quality/score than without the uptake of such a tool (e.g., Cancino & Panes, 2021).

Similar to technology use, students' responses to FFHAWT also depended on the quality of feedback received, accessibility to the helper, and the quality of affective-cognitive interaction involved. Previous studies showed that students' level and quality of engagement with instructor feedback (Hawe & Dixon, 2014; McMartin-Miller, 2014) and the approaches taken to providing corrective feedback to ESL writers (Ferris, 2014), can influence the extent to which L2 learners benefit from instructors. In other words, the more both parties (students and instructors) are willing and positively motivated, and the more the feedback is relevant and digestible to the learner's level, students were found to respond more positively and make appropriate corrections. With peers, learners' ability to trust and negotiate feedback provided by their peers (Tigchelaar & Polio, 2017), the amount of training that was offered prior to peer response, the quality of teacher involvement (Storch, 2017; Tigchelaar & Polio, 2017), and the mode of interaction (Storch, 2017), can impact how much learners can benefit from peer corrective feedback. Specifically, students who trusted their peers' ability to provide good feedback and were thoroughly trained in peer response (with teachers present), were more likely to perceive peer help as useful and make appropriate revisions and corrections on their papers.

Consulting with a writing center tutor is another resource that many L2 writers rely on. Research has shown that L2 learners tend to request more sentence-level corrective feedback compared to native speakers of English; and that they are more inclined to view tutors as editors rather than advisors (Williams & Severino, 2004). It has also been reported that writing center assistants are more direct with L2 learners because the latter find these interactions more productive and less cognitively taxing (Moussu & David, 2015; Williams & Severino, 2004). Furthermore, the aforementioned studies reported that L2 writers tend to gain the most benefit from consulting with tutors who have extensive experience working with L2 learners, a solid grasp of the subject, and strong interpersonal skills. Overall, drawing on FFHAWT can be highly advantageous so long as the helper has a clear understanding of and the ability to effectively deliver the writer's needs as well as a proper, trusting relationship being formed with the student.

As can be seen, L2 students use and are exposed to a variety of digital and human-assisted writing tools for academic and non-academic writing purposes in higher education. However, this reality may be overlooked, especially in cases where testing specific language knowledge and skills are prioritized over factors that have a considerable effect on students' writing process. To this end, this study explores ESL students' uses and attitudes toward CBWT and FFHAWT to examine the degree to which drawing of external writing resources is modeled in the EPWT test tasks, using the following research question:

What are the lower- and intermediate-level ESL students' uses of and attitudes toward CBWT and FFHAWT? And for what reasons do students at these two ESL writing levels use CBWT and FFHAWT?

The findings to research question will be used to evaluate the rebuttal and by extension, the claim made within the domain definition inference.

3. METHODOLOGY

3.1. Participants and Context of Study

Ninety-two ESL students who previously took EPWT at a large North American university and were subsequently placed into ESL writing courses participated in the study: 35 from two sections of the lower-level and 57 from four sections of the intermediate-level course. The sample included 60 male and 32 female students. Most spoke Chinese (41) as their first language, followed by Korean (15) and Arabic (9). Other self-reported languages that were represented include Vietnamese (5), French (3), English (2), Spanish (2), Telugu/Hindi (2), Malay (2), Nepali (2), Portuguese (1), Japanese (1), Marathi (1), Bahasa (1), Urdu (1), Kinyarwanda (1), Bangla (1), Runyankole (1), and Marvadi (1).

At the time of this study, the EPWT was a required placement test for all international students who received a TOEFL score (or equivalent) between 70 and 99. The EPWT consisted of two integrated reading-writing tasks: a 15-minute-long summary followed by a 30-minute-long argumentative essay, drawing ideas from two short reading passages and their personal experiences. The essays were typed in a learning management system where no computer-assisted writing tools were enabled. Two or three raters rated each set of essays; undergraduate students were placed into either the lower- ('B') or intermediate-level ('C') ESL writing course or were enrolled into a first-year composition (FYC) course ('Pass'). The courses are sequenced such that a student who is enrolled in a lower-level course in the first semester, moves on to the intermediate-level in the next semester, and then enrolls in an FYC thereafter. Generally, students in a lower-level course tended to struggle more with micro skills of writing and so the lower-level ESL course focused more on grammar and paragraph writing. On the other hand, students in an intermediate-level needed to develop better macro skills and so the intermediate-level ESL course focused on writing various types of essays for general academic purposes (e.g., compare/contrast, summary/response) and the writing process. FYC, on the other hand, emphasized writing essays that consider the rhetorical situation and prompt students to input critical thinking.

3.2. Data Collection and Analysis

3.2.1. Quantitative

The data comes from a larger research project that investigated the validity of the EPWT within an argument-based validation framework. An online questionnaire consisting of 7-point Likert scale closed-ended and open-ended questions and a semi-structured interview with a smaller sample of students were used to answer the research question. The Likert-scale goes from 1 (strongly disagree) to 7 (strongly agree), and it was adopted because seven-scale survey items were found to effectively distinguish respondents' perceptions (Finstad, 2010). For the purposes of the current study, only items that targeted students' uses and attitudes toward CBWT and FFHAWT were drawn (see Appendix A). Moreover, for ease of comprehension, CBWT was replaced with *technology* and FFHAWT was replaced with *outside help*, and the definitions of these words were clarified in the questionnaire and verbally by the researcher at the time of data collection. Before conducting the study, IRB was approved.

In weeks 3 and 4 of the semester, students in all six sections of the ESL writing courses completed the questionnaire, which was treated as part of an ongoing pedagogical activity. It was given to all students and only those who agreed to participate in the research were collected for data analysis. The researcher, who participated as a guest lecturer, walked around the room and addressed any questions students had while completing the questionnaire. This was to ensure that students were responding to the best of their ability.

After data screening, students' responses to the closed-ended questions were downloaded as a CSV file and were subsequently analyzed in Excel using the Real Statistics Resource Pack software (Release 7.6). Responses to negatively-worded statements were reversed prior to running the descriptive statistics and inferential statistics. Item analysis was conducted and Cronbach's alpha was α=0.91 and Pearson's correlation coefficients were between 0.66 and 0.89, which is considered reliable and converging. Because data were non-normally distributed, a Mann-Whitney U test was used to analyze lower- and intermediate-level students' responses to the closed-ended questions.

3.2.2. Qualitative

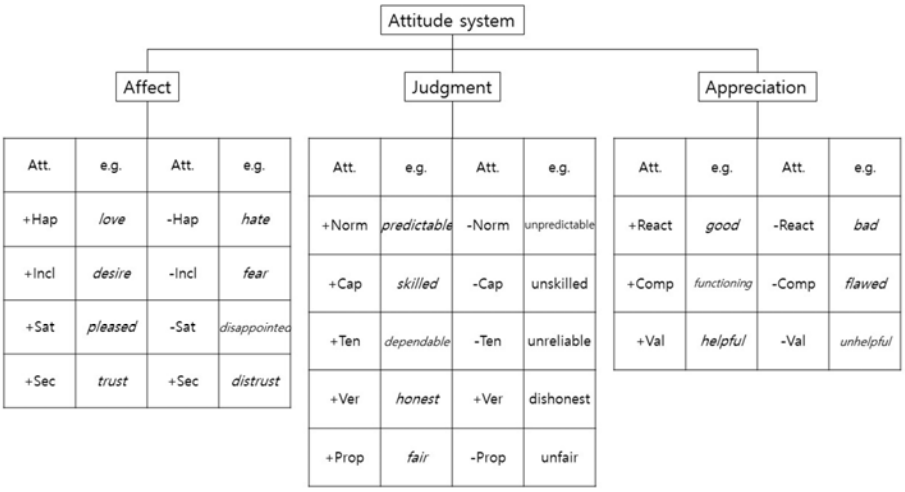
In Weeks 4 and 6, 26 of the 92 students (11 lower-level and 15 intermediate-level) participated in a follow-up one-on-one semi-structured interview. Participants who consented to research participation on the survey were individually e-mailed and only those who responded to partake in the interview were drawn. All interviews took place in a quiet office room with the researcher, and each session was audio-recorded in Audacity. Questions prompted students to share their views on CBWT and FFHAWT and explain their reasons for using such resources (see Appendix B). Transcripts of the interviews were typed manually by the researcher in Word and were coded following Saldaña's (2016) recommendation. To identify attitudes, Martin and White's (2005) Appraisal framework was adopted, which is briefly described next.

One of the ways in which attitudes can be analyzed is with the application of Martin and White's (2005) Appraisal theory. It argues that interpersonal meaning can be found in texts that are spoken or written by people. Interpersonal meaning includes what and how people express their feelings toward things and other people; and how speakers/writers position themselves, imagine their audience, align or disalign and interact with their audience (Martin & White, 2005). The Appraisal framework, which is based on the Appraisal theory, is described as a three-system network, made up of attitude, engagement, and graduation. Each system taps into a different aspect of the interpersonal meaning found in speakers/writers' evaluative language and language resources. In this study, we focus on the attitude system.

The attitude system is made up of three sub-systems known as affect, judgment, and appreciation (see Figure 2). Each sub-system focuses on a particular aspect of a speaker/writer's attitude. Affect is composed of language resources that have to do with positive and negative feelings and reactions of an individual. Affect-related resources are further categorized into happiness/unhappiness (e.g., love, hate), inclination/disinclination (e.g., desire, fear), satisfaction/dissatisfaction (e.g., pleased, disappointed), and security/insecurity (e.g., trust, distrust). Students, for example, could express affect by stating which mediums they find most comfortable working with (+security) while composing their academic essays.

FIGURE 2

Martin and White's (2005) Appraisal Framework, Adapted



Att.	e.g.	Att.	e.g.
+Hap	*love*	-Hap	*hate*
+Incl	*desire*	-Incl	*fear*
+Sat	*pleased*	-Sat	*disappointed*
+Sec	*trust*	+Sec	*distrust*

Att.	e.g.	Att.	e.g.
+Norm	*predictable*	-Norm	unpredictable
+Cap	*skilled*	-Cap	unskilled
+Ten	*dependable*	-Ten	unreliable
+Ver	*honest*	+Ver	dishonest
+Prop	*fair*	-Prop	unfair

Att.	e.g.	Att.	e.g.
+React	*good*	-React	*bad*
+Comp	*functioning*	-Comp	*flawed*
+Val	*helpful*	-Val	*unhelpful*

Judgment is a sub-system of resources related to positive and negative evaluations of people and their behaviors, as perceived by the speaker/writer. Similar to affect, judgment resources are grouped into categories of social esteem (how well an individual follows social norms) and social sanction (how ethically one behaves). Positive and negative normality (e.g., predictable, unpredictable), capacity (e.g., skilled, unskilled), and tenacity (e.g., dependable, unreliable) constitute social esteem. Positive and negative veracity (e.g., honest, dishonest) and propriety (e.g., fair, unfair) form social sanction. If a student seeks help from a writing tutor due to the tutor's expertise in academic writing, the student is making an assessment of the capacity (+capacity) of the tutor.

Whereas judgment has to do with evaluations of people and their behaviors, appreciation has to do with evaluations of physical and abstract objects. Once again, language resources associated with appreciation can be further grouped into positive and negative reaction (e.g., good, bad), composition (e.g., functioning, flawed), and valuation (e.g., helpful, unhelpful). For example, an online grammar checker that enables a student to quickly see the errors in his/her draft and thereby helps the student avoid making similar mistakes in the future, is perceived to have a positive valuation (+valuation).

The Appraisal framework was adopted to identify the attitudes found in the interview and the open-ended question. In the first phase of coding, transcripts and responses to the open-ended question were printed and read line-by-line to identify students' attitudes following Martin and White's (2005) framework. In this study, we focused only on students' attitudes related to CBWT and FFHAWT. A codebook was developed at the same time to ensure consistent coding and analysis across all transcripts (see Lee, 2020). In the second phase of coding, all coded parts of the transcripts were double-checked by the researcher and were categorized in terms of reasons for using CBWT and FFHAWT. A matrix was developed in Excel to tally and visually organize the kinds of attitudes that students expressed for CBWT and FFHAWT (see Table 3).

To prepare the second coder analysis, 20% of the transcripts that were stratified sampled were prepared in Excel. Each line of the transcript (minus the researcher's lines) was logged in each row of the Excel sheet and a second coder, who was also an expert in Appraisal analysis, was asked to identify the attitudes. Brief contextual information was provided where it was necessary, and an hour-long training took place before the second coder analysis. An inter-rater reliability of k=0.87 was reached, which is considered excellent in agreement. Disagreements were reviewed together and subsequent changes were made to the codebook and the rest of the data analysis.

After all data collection and analysis were completed, results from the questionnaire and interview were triangulated to answer the research question.

4. RESULTS

4.1. Students' Uses and Attitudes toward CBWT and FFHAWT

Table 1 summarizes the uses and attitudes that students at the lower- and intermediate-level ESL writing courses expressed toward CBWT and FFHAWT, as found in the questionnaire responses. On average, students considered one or both types of writing tools useful for improving their writing quality, grades, and communication with peers and instructors. Mean ratings for the first seven statements ranged between 4.80 and 5.14 at the lower-level and between 5.18 and 5.70 at the intermediate-level, suggesting that students on average *somewhat agreed* to *agreed* that drawing on external writing tools was useful for academic writing purposes.

TABLE 1

Students' Uses and Attitudes Toward CBWT and FFHAWT

Items (Reversed)	Lower-Level		Intermediate-Level	
	M	*SD*	*M*	*SD*
1. I use technology (e.g., Grammarly) and/or outside help (e.g., writing center) when I write my essays	4.94	0.24	5.35	0.18
2. Using technology and/or outside help improves my essay grades in the ESL writing course	5.14	0.23	5.70	0.16
3. Using technology and/or outside help improves my essays	5.00	0.21	5.53	0.18
4. Using technology and/or outside help improves my written communication with instructors/professors	4.89	0.25	5.42	0.18
5. Using technology and/or outside help improves my written communicate with peers/classmates	4.80	0.20	5.18	0.21
6. Using technology and/or outside help improves my academic writing in English	4.86	0.21	5.32	0.20
7. Using technology and/or outside help solves my academic writing problems in English	4.91	0.23	5.33	0.18
8. How often do you use the following when writing essays? - Technology such as Grammarly, online dictionaries, or spellcheckers	4.23[a]	0.26	4.77	0.22
9. How often do you use the following when writing essays? - Outside help such as writing center tutors, peers, or teachers	2.74	0.21	3.25	0.19

[a] Ratings for statements 8 and 9 are on a 7-point Likert scale with 1 (*Never (0%)*), 2 (*Rarely, about 10% of the time*), 3 (*Occasionally, about 30% of the time*), 4 (*Sometimes, about 50% of the time*), 5 (*Usually, about 70% of the time*), 6 (*Frequently, about 90% of the time*) and 7 (*Every time (100%)*)

However, when it came to rating frequencies of uses, on average, students at both levels reported using technology about 50% of the time while writing, whereas they reported using outside help only 10% to 30% of the time. A Mann-Whitney U test was run to determine if there were differences in ratings across students at the two levels (see Table 2). Except for item no. 3, the differences between the median ratings were statistically similar for lower- and intermediate-level students. In other words, no significant differences were found in terms of how students used and perceived external writing resources. On other hand, for item no. 3, "using technology and/or outside help improves my essays," the difference between the median rating for lower-level (5.0) and intermediate-level students (6.0) was statistically significant, *U*=745.00, *z*=2.03,*p*=0.04, *r*=0.21. However, the effect size was 0.21, suggesting that this was a significant but small effect.

TABLE 2

Output of Mann-Whitney U Test

Items	Mann-Whitney U	Wilcoxon W	z	*p*
1	821.00	1451.00	1.42	0.16
2	759.00	1389.00	1.97	0.06
3	745.00	1375.00	2.03	0.04*
4	784.50	1414.50	1.76	0.09
5	809.00	1439.00	1.55	0.13
6	781.50	1411.50	1.77	0.08
7	830.00	1460.00	1.38	0.18
8	809.50	1439.50	1.53	0.13
9	802.50	1432.50	0.13	0.12

* p<.05

4.2. Reasons for (Not) Using CBWT and FFHAWT

While we were able to gather the fact that students from both course levels generally viewed CBWT and FFHAWT in a positive light, the closed-ended questions did not prompt students to give separate ratings on CBWT and FFHAWT. This was due to the sheer size of the original questionnaire from which the data for the present study was collected: It had prompted students to respond to 66 close-ended questions related to academic writing in a single setting. Due to practical constraints, the researcher was not able to further expand or administer a separate questionnaire. Instead, the researcher opted to investigate specific attitudes related to CBWT and FFHAWT qualitatively, through responses to the open-ended question and semi-structured interviews.

As shown in Table 3, students' attitudes and reasons for using CBWT and FFHAWT somewhat varied. In total, 129 unique attitudes were captured from the questionnaire, and 125 unique attitudes were identified from the interview. This is due to some respondents expressing more than one attitude toward one or more external writing resources. Furthermore, higher percentages of positive attitudes related to FFHAWT and negative attitudes toward both types of tools were captured in the interview compared to the questionnaire, and this may be due to students having had more time to jog their memories and elaborate on their experiences and attitudes during the interview.

Overall, the most used external writing tools among students were related to CBWT: Grammarly and other AWCF (43/92 and 15/26), online dictionaries (9/92 and 12/26), search engines (4/92 and 7/26), online translators (3/92 and 5/26), and MS-NLP (4/92 and 1/26). Twenty-eight out of ninety-two respondents and one interviewee did not specify the form of CBWT that they used. Of the 92 questionnaire respondents, 32 lower-level and 51 intermediate-level students reported that they used one or more of the aforementioned tools because they were reliable (+security), helpful (+valuation), good for revising and editing (+reaction), and accessible and easy to use (+composition). Similarly, 10 lower-level and 15 intermediate-level interviewees stated the same reasons for using CBWT as their primary writing resource when writing their essays.

TABLE 3

Reasons for (Not) Using CBWT and FFHAWT

+/- Tool	Reason (Attitudes)	Lower-Level				Intermediate-Level			
		Open-Ended		Interview		Open-Ended		Interview	
		N	%	*n*	%	*n*	%	*n*	%
+CBWT	Easy, accessible, and quick turnaround (+Sec, +Val, +React, +Comp)	32/35	91.4	10/11	90.9	51/57	89.5	15/15	100.0
+FFHAWT	Meeting one's writing goals when does with the "right" people (+Sec, +Val, +React, +Comp)	4/35	11.4	9/11	81.8	13/57	22.8	11/15	73.3
-CBWT	Unaware of its affordance; the user is insecure; something about it is inconvenient or unhelpful (-Sec, -Val, -React, -Comp)	1/35	2.9	0/11	0.0	5/57	8.8	3/15	20.0
-FFHAWT		2/35	5.7	3/11	27.3	1/57	1.8	7/15	46.7
Relying primarily on one's self (rarely uses CBWT or FFHAWT)	There is no need to draw on external writing tools (-Val)	1/35	2.9	0/11	0.0	3/57	5.3	0/15	0.0

Drawing on FFHAWT was not as frequently self-reported as using CBWT but was nonetheless present. The most frequently utilized forms of FFHAWT include peers (16/26 and 2/92), instructors (only reported by 13 out of 26 interviewees), writing center tutors (5/26 and 4/92), regular tutors (2/26 and 2/92), and family (only reported by 3 out of 26 interviewees). Ten out of ninety-two questionnaire respondents did not specify the type of FFHAWT they relied on. Students used FFHAWT at various writing stages: from brainstorming ideas (prewriting), organizing ideas into an essay (composing), to revising and editing. Students also sought FFHAWT for assignments that they perceived to be high stakes (e.g., an assignment contributing to a large proportion of the final grade) and making "errors" was deemed less forgiving. Feedback that helped students meet their writing goals was considered useful (+valuation), positive (+reaction), and comprehensible (+composition). In the interview, students mentioned that it was important to seek feedback from the "right" person (+security, +capacity), who was often perceived to be an expert in writing, an evaluator, and/or a proficient speaker of English.

Interviewees also reported that ESL instructors (as well as other course instructors) encouraged students to draw on CBWT and FFHAWT while writing. Grammarly, for example, was a tool that was often introduced in the lower-level ESL course and was used for revising and editing their ESL writing assignments (as reported by 15 out of 26 students). Students were expected to copy and paste parts of their essays to correct grammar, spelling, and mechanics errors before handing in their final drafts to their instructors. They also reported being encouraged to visit writing center tutors by their instructors (ESL or elsewhere) for idea generation and content organization among other reasons (5 of 26 students). It is clear that drawing on external writing resources, whether it be CBWT or FFHAWT, is an integral part of the writing process for a great deal number of ESL students in the ESL writing domain.

However, not everyone who used these tools viewed them with complete positivity. Among students who expressed more critical viewpoints, some leading reasons include the fact that CBWT (6/92 and 3/26) and FFHAWT (3/92 and 10/26) were time-consuming and/or inaccessible (-composition); gave confusing (-composition), incorrect (-reaction), and/or unhelpful feedback (-valuation); and that using one or more of the tools had somehow triggered a lack of confidence in the user (-security). In addition, 4 out of 92 students (and none from the interviewees) explicitly reported that they rarely used CBWT or FFHAWT and instead relied primarily on themselves to complete their writing assignments. For these individuals, they did not see a need to draw on CBWT or FFHAWT (-valuation) to complete writing assignments done for ESL and other courses. Although negative attitudes toward CBWT and FFHAWT were expressed in much smaller proportions, these concerns can partially explain the relatively low usage of FFHAWT as well as CBWT among the lower-level and intermediate-level ESL students (see Table 1).

5. DISCUSSION AND CONCLUSION

In this paper, an evaluation of the rebuttal within the domain definition inference is made based on the gathered findings of the current study. To briefly recap the findings of this research, the majority of ESL students' attitudes were positive toward both CBWT and FFHAWT, and students at both course levels were statistically similar in their viewpoints. The majority of questionnaire respondents and interviewees used CBWT and/or FFHAWT because these tools were perceived to be accessible, reliable, useful, and appropriate for revising and editing purposes. This echoes the findings of previous research where ESL/EFL learners were found to respond better to online writing tools that generate clear, specific, and accurate feedback (Ranalli, 2018); reduce cognitive demands (O'Neill, 2019; Ranalli, 2018); are highly accessible (O'Neill, 2019); and are perceived to be trustworthy (O'Neill, 2019; Ranalli, 2018). The findings also support students' preferences for forms of FFHAWT that generate positive, trustworthy interaction (Hawe & Dixon, 2014; McMartin-Miller, 2014; Storch, 2017; Tigchelaar & Polio, 2017); provide feedback that is relevant and comprehensible (Ferris, 2014); and lower cognitive demands (Moussu & David, 2015; Williams & Severino, 2004).

This study also found that students were expected and encouraged to use one or more of these tools by their instructors to complete assignments written for their ESL courses. In addition, a handful of students expressed more negative attitudes toward these tools, perceiving them to be time-consuming and inaccessible; their feedback to be confusing, incorrect, and unhelpful; and the interaction triggering insecurity in the user. This would partially explain the rather low average usage of these tools.

Overall, the results of the present research point to evidence that the EPWT tasks did not fully consider the fact that the majority of ESL students draw on CBWT and/or FFHAWT while writing, which is an essential component of ESL students' writing process. Although two reading passages were made available in the placement test, as with most integrated reading-writing tests, this feature is mainly to assess students' ability to draw on outside sources (Weigle & Parker, 2012) rather than to be used as a form of self-selected CBWT. This leads to reasonable support for the rebuttal stated within the domain definition inference, which in turn calls for a careful review of the claim, that is, the extent to which observations on the EPWT are in fact appropriate for placement purposes.

This study collected and analyzed both qualitative and quantitative data to address research question motivated by a need to evaluate the rebuttal as well as the strength of the claim; however, the sampling population was limited to ESL students, and other types of evidence should be gathered. For example, it would be useful to survey and interview ESL instructors and even expand the sampling population to those enrolled in FYC, since placement decisions also impact international students who were directly enrolled in FYC.

Cognitive-behavioral studies that explore the process in which students write essays (e.g., Koltovskaia, 2020) in the test domain versus the target domain would also be useful, as this would show the degree to which drawing on writing resources is actually employed by students.

Nevertheless, the findings in the current study raise some concerns regarding the authenticity of the EPWT tasks. A writing test that limits students' use of external writing resources may increase test security, but it might not reflect how students actually write for their courses and how such writing is judged by their instructors (Dimova et al., 2020). This has the potential to misplace certain students into a course that is either too easy or too challenging for them. Furthermore, the goals of each writing course (i.e., lower-level vs. intermediate-level vs. FYC) are different, and so it would be important for test developers to develop test tasks and a rubric that clearly link and measure these different learning goals: "[a]s a test developer of a local language test, you should seek out opportunities to ground the test within the local language program by linking it to instructional goals, learning outcomes, and instructional practices, as well as by involving different stakeholders (instructors, students, parents, administrators) in the testing process" (Dimova et al., 2020, p. 22). If instructors expect students to regularly draw on external writing resources such as CBWT and FFHAWT to produce quality writing, it would be important for test tasks to model this aspect more authentically without compromising test security.

Placement tests are commonly observed in many higher education settings where there exist large numbers of international students. Although the stakes of a placement test may not be deemed as consequential as that of a TOEFL or IELTS test, requiring completion of one or two remedial courses could cost added tuition and lengthen time to graduation. Therefore, it is crucial that test developers and decision-makers take a closer look at not only the linguistic components but also the authenticity of test tasks and ensure strong validity evidence not only of assumptions and warrants but also of absence of rebuttals.

Applicable level: Tertiary

REFERENCES

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

Andrade, M. S., Evans, N. W., & Hartshorn, K., J. (2015). Perceptions and realities of ESL students in higher education: An overview of institutional practices. In N. W. Evans, N. J. Anderson & W. G. Egginginton (Eds.), *ESL readers and writers in higher education: Understanding challenges, providing support* (pp. 18-36). New York: Routledge.

Cancino, M., & Panes, J. (2021). The impact of Google translate on L2 writing quality measures: Evidence from Chilean EFL high school learners. *System, 98*, 102464

Chapelle, C. A. (2020). *Argument-based validation in testing and assessment.* Los Angeles: Sage.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the test of English as a foreign language.* New York: Routledge.

Dimova, S., Yan, X., Ginther, A. (2020). *Local language testing: Design, implementation, and development*. London: Routledge.

Ferris, D. R. (2009). *Teaching college writing to diverse student populations*. Ann Arbor, MI: The University of Michigan Press.

Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing, 19*, 6-23.

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers, 22*(5), 323-327.

Hawe, E. M., & Dixon, H. R. (2014). Building students' evaluative and productive expertise in the writing classroom. *Assessing Writing, 19*, 66-79.

Kane, M. T. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 17-64). Westport, CT: Praeger.

Kane, M. T. (2022). Articulating a validity argument. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed.) (pp. 53-69). London: Routledge.

Koltovskaia, S. (2020). Students engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 44*, 100450

Lee, E. (2020). *An evaluation of the English placement writing test using students' self-assessments and instructors' judgments.* Unpublished doctoral dissertation, Iowa State University, Ames, IA, USA.

Li, Z. (2015). *An argument-based validation study of the English placement test (EPT): Focusing on the inferences of extrapolation and ramifications*. Unpublished doctoral dissertation, Iowa State University, Ames, IA, USA.

Martin, J. R., & White, P. R. R. (2005). *Language of evaluation: Appraisal in English*. New York: Palgrave Macmillan.

McMartin-Miller, C. (2014). How much feedback is enough?: Instructor practices and student attitudes toward error treatment in second language writing. *Assessing Writing, 19*, 24-35.

Moussu, L., & David, N. (2015). Writing centers: Finding a center for ESL writers. In N. W. Evans, N. J. Anderson & W. G. Eggington (Eds.). *ESL readers and writers in higher education: Understanding challenges, providing Support* (pp. 49-63). New York: Routledge.

O'Neill, E. (2019). Online translator, dictionary, and search engine use among L2 students. *CALL-EJ, 20*(1), 154-177.

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning, 31*(7), 653-674.

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology, 37*(1), 8-25.

Ranalli, J., & Yamashita, T. (2022). Automated written corrective feedback: Error-correction performance and timing of delivery. *Language Learning & Technology, 26*(1), 1-25.

Saldaña, J. (2016). *The coding manual for qualitative researchers*. London: Sage.

Storch, N. (2017). Peer corrective feedback in computer-mediated collaborative writing. In H. Nassaji & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (pp. 66-79). New York: Routledge.

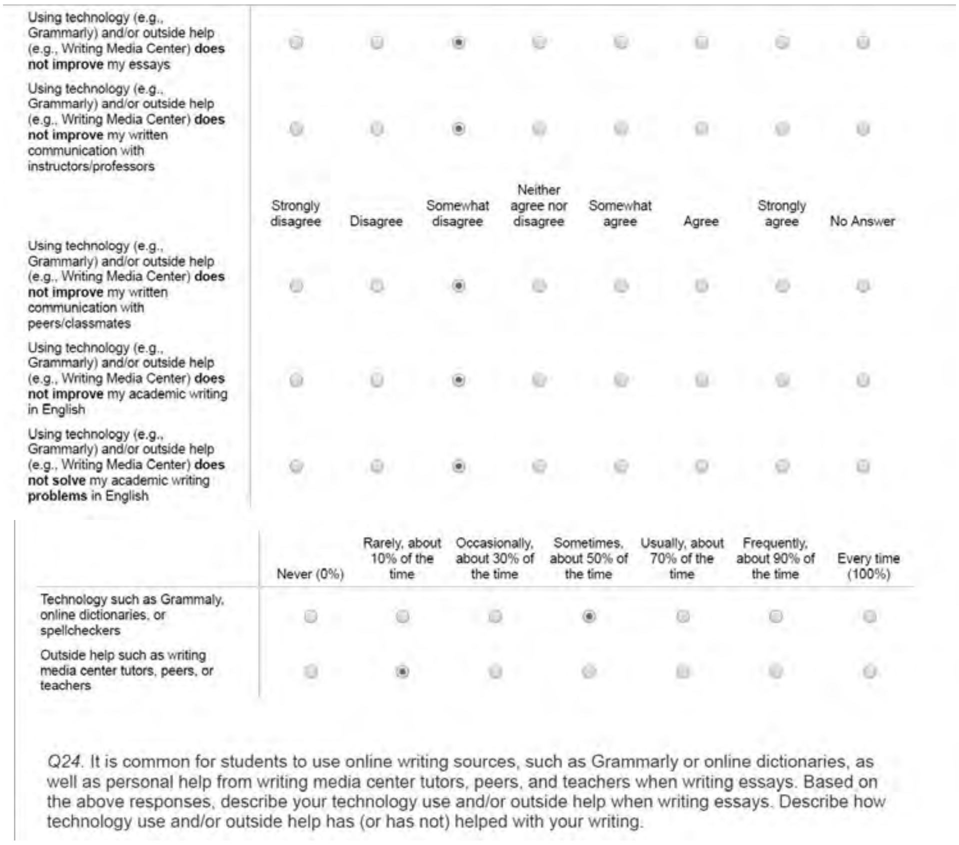
Tigchelaar, M., & Polio, C. (2017). Language-focused peer corrective feedback in second language writing. In H. Nassaji & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (pp. 97-113). New York: Routledge.

Weigle, S. C., & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing, 21*(2), 118-133.

Williams, J., & Severino, C. (2004). Tutoring and revision: Second language writers in the writing center. *Journal of Second Language Writing, 13*(3), 173-201.

APPENDIX A

Questionnaire Prompting Students’ Uses and Attitudes toward CBWT and FFHAWT



Using technology (e.g., Grammarly) and/or outside help (e.g., Writing Media Center) **does not improve** my essays	○	○	●	○	○	○	○	○
Using technology (e.g., Grammarly) and/or outside help (e.g., Writing Media Center) **does not improve** my written communication with instructors/professors	○	○	●	○	○	○	○	○
	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree	No Answer
Using technology (e.g., Grammarly) and/or outside help (e.g., Writing Media Center) **does not improve** my written communication with peers/classmates	○	○	●	○	○	○	○	○
Using technology (e.g., Grammarly) and/or outside help (e.g., Writing Media Center) **does not improve** my academic writing in English	○	○	●	○	○	○	○	○
Using technology (e.g., Grammarly) and/or outside help (e.g., Writing Media Center) **does not solve** my academic writing **problems** in English	○	○	●	○	○	○	○	○

	Never (0%)	Rarely, about 10% of the time	Occasionally, about 30% of the time	Sometimes, about 50% of the time	Usually, about 70% of the time	Frequently, about 90% of the time	Every time (100%)
Technology such as Grammaly, online dictionaries, or spellcheckers	○	○	○	●	○	○	○
Outside help such as writing media center tutors, peers, or teachers	○	●	○	○	○	○	○

Q24. It is common for students to use online writing sources, such as Grammarly or online dictionaries, as well as personal help from writing media center tutors, peers, and teachers when writing essays. Based on the above responses, describe your technology use and/or outside help when writing essays. Describe how technology use and/or outside help has (or has not) helped with your writing.

APPENDIX B

Interview Questions Prompting Students’ Uses and Attitudes toward CBWT and FFHAWT

1. What technology do you use to write your essays for the ESL course (and for other courses)?
2. What do you think about this particular technology? (good/bad, easy/difficult, interesting/boring, useful/not useful)?
3. What outside help (e.g., peers, teachers, tutors) do you use to write your essays for the ESL course (and for other courses)?
4. What do you think about this kind of outside help? (good/bad, easy/difficult, useful/not useful, accessible/inaccessible)?
5. Why do you say so?
6. How often do you use technology? Help from others while writing?