

Critical Thinking in the Context of Speaking Assessment

Ali Ebrahimpourlighvani
Teachers College, Columbia University

INTRODUCTION

Most university oral English placement tests in the United States focus on assessing language ability, yet they often overlook the analytical and strategic skills that are central to effective academic communication (Yin et al., 2024). While language is the primary vehicle for communicating values, abstract ideas, and complex problems, the majority of scholarship has emphasized critical thinking (CT) in the context of reading and writing, leaving a significant gap in speaking assessment (Wilson, 2016; Yin et al., 2024). Because academic discourse is purposeful and requires “focused thinking,” the intentional direction of an argument with conscious intent, oral communicative competence, and CT are indispensable and tightly connected skills.

To address this gap, this paper explores CT in the context of speaking assessment in order to propose the “Critical-Thinking-to-Interact” construct, which integrates linguistic proficiency with the capacity to analyze, evaluate, and respond to information. This integrated approach is grounded in an interactionist perspective of defining second language ability (Bachman, 2007; Chapelle, 1999), suggesting that communicative competence involves both knowledge of language and the capacity to implement that knowledge through metacognitive strategies within specific contexts. By prioritizing the quality and soundness of arguments alongside linguistic accuracy, this framework seeks to better reflect the real-world demands of university-level communication. Consequently, this literature review synthesizes scholarship in performance assessment, oral communicative competence, the role of CT in evaluation, and computer-based testing to provide a theoretical rationale for moving beyond traditional linguistic assessments toward a more holistic model of academic oral proficiency.

LITERATURE REVIEW

This literature review synthesizes and reviews six areas of scholarship that serve as the foundation of the current study: (a) performance assessment in ESL contexts, (b) oral communicative competence in speaking assessment, (c) CT and its role in performance-based evaluation, (d) assessing speaking and critical thinking skills using computer-based testing, (e) defining and assessing speaking and critical thinking as an integrated skill, and (f) current assessments of academic speaking ability and critical thinking.

Performance Assessment in the ESL Context

Performance assessments are defined as test activities that require test-takers to produce language that closely resembles what occurs in real-world target language use situations (Ockey, 2025). Messick (1994) states that performance is the vehicle of assessment, meaning that the task itself is of less interest than what the performance reveals. The underlying knowledge and abilities demonstrated through performance are the actual targets of assessment. Performance assessments are commonly used to evaluate productive language skills, such as speaking and writing. These assessments can also integrate multiple language skills, such as speaking and listening. In the context of an assessment of speaking and CT, integrated skills refer to speaking and critical thinking. The results of such tests should serve as indicators of how test-takers would perform in similar real-world situations.

The term *performance* can be understood in two ways: as real-world tasks and activities or, in a testing context, as assessments that involve direct simulations of real-world roles and tasks (McNamara, 1996). Related to this type of assessment, Carroll (1961) focuses on two aspects of an ideal English language proficiency test. First, these tests should differentiate test-takers based on their performance level, and also, the dimensions of the performance should be relevant to a similar situation in which the examinees could find themselves. Second, performance tests can predict the success of the test-takers in learning tasks and social situations in which the test-takers will be exposed. Moreover, performance assessments should provide a description in which an authentic situation is being reflected. In other words, as Ockey (2025) says, performance assessments aim to provide target language use opportunities for test-takers in realistic contexts.

In the context of speaking and CT testing, there is ongoing debate regarding whether critical thinking is a non-linguistic factor in performance assessment or an integral part of performance itself. Jones (1985) explores non-linguistic factors and their implications for performance-based language tests. He argues that, in performance assessments, language is only one of several factors being evaluated. The overall criterion for assessment is the successful completion of a task in which language use is essential. Moreover, a performance test is more than a basic proficiency test of communicative competence and it is related to some kind of performance task. For instance, it is possible for some test-takers to compensate for lower language proficiency by excelling in other areas. For example, a test-taker with limited language proficiency may still perform well in a task due to strong analytical skills or familiarity with the subject matter. Conversely, some test-takers with higher language proficiency may struggle due to difficulties in understanding or analyzing test tasks and prompts. Similarly, test-takers with prior experience in a given context may perform better due to their background knowledge.

Defining Oral Communicative Competence in Speaking Assessment

Language use relies on the meaningful production of linguistic forms in the target language context. However, the soundness and effectiveness of language use call for a test that is appropriate for measuring the abilities related to its purposes. In other words, language needs a tool to measure its elicitation in a real-world context; as Douglas (1998) said, “[a] language test is a second language acquisition (SLA) elicitation device” (p. 141). Language ability is not directly observable, and we have to infer it based on observed performance (Bachman, 1990). It

means we need to define a trait in a context. This reminds us of the interactionist approach regarding the construct definition. To define the interactionist approach, we shall refer to Chapelle (1999). Chapelle (1999) distinguishes among three perspectives on construct definition: a trait, a behavior, and a combination of trait and behavior. In a trait definition, a person's consistent performance depends on the knowledge and speech production process. A trait is defined as "a relatively stable characteristic of a person - an attribute, enduring process, or disposition - which is consistently manifested to some degree when relevant, despite considerable variation in the range of settings and circumstances" (Messick, 1989, p. 15). On the other hand, defining a construct as a behavior depends on the context in which the behavior is going to be observed. This is viewed from SLA researchers' perspective, i.e., conditions might include the linguistic environments or the sociolinguistic contexts in which particular interlanguage forms are most likely to appear (Ellis, 1989).

Both trait and behavior definitions did not support the need for theories of language in use, such as communicative language ability, because, as Bachman (1990, p. 84) has emphasized, communicative language ability contains both knowledge and "the capacity for implementing or executing that competence" in specific contexts of use. So, this could be a considerable desire to move to the interactionist definition by Messick (1989) and Chapelle (1999). In SLA, the interactionist perspective has been fueled primarily by the theory of communicative language use, which suggests that communicative competence refers to both knowledge of language and the ability to put language to use in context (Hymes, 1972; Canale & Swain, 1980; Widdowson, 1983). More explicitly, as Chapelle (1999) said, "performance is viewed as a sign of underlying traits and is influenced by the context in which it occurs and is therefore a sample of performance in similar contexts" (p. 43). Thus, for a speaking and CT test, the test-takers' underlying language abilities and capacity for critical thinking are going to be the trait; this trait will be shaped by the context in which the assessment takes place, which is the test tasks. Furthermore, to incorporate a dimension of interaction between trait and context, an interactionist definition must include metacognitive strategies responsible for mediating between the two (Chapelle, 1999, p. 43). As a result, in a speaking and CT test context, the language user's metacognitive strategies (e.g., analyzing the problem) will mediate between language use and the test-taker's knowledge during the performance.

Critical Thinking and its Role in Performance Assessment

In the area of critical thinking, there are several approaches to defining critical thinking. These approaches are rooted in philosophy and psychology (Lewis & Smith, 1993) and education (Sternberg, 1986). In philosophy, as Paul (1992) wrote, critical thinking is: "disciplined, self-directed thinking that exemplifies the perfections of thinking appropriate to a particular mode or domain of thought" (p. 9). In the view of psychology, Sternberg (1986) states that critical thinking is: "the mental processes, strategies, and representations people use to solve problems, make decisions, and learn new concepts" (p. 3), and last but not least, in the education discipline, we shall refer to Bloom's taxonomy (Anderson & Krathwohl, 2001; Bloom, 1956). Bloom's taxonomy is hierarchical, with "comprehension" at the bottom and "evaluation" at the top. The three highest levels (analysis, synthesis, and evaluation) are frequently said to represent critical thinking (Kennedy et al., 1991). As Ennis (1962) mentioned, "critical thinking is the correct assessing of statements" (p. 81). In other words, critical thinking includes evaluating or

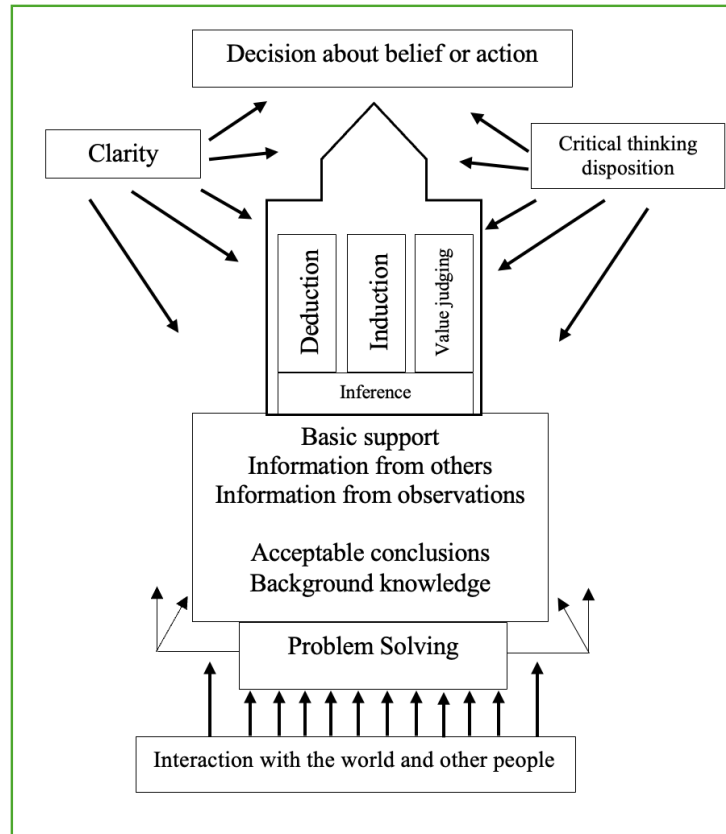
judging statements, claims, or arguments to determine their truthfulness, validity, or soundness. CT assessment requires individuals to analyze the information critically rather than accept it at face value. Moss and Koziol (1991) advocate for evaluating students on the basis of the quality of the arguments underlying their position. Although there are different perspectives on each approach, there exist areas of alignment among these three approaches:

- analyzing arguments, claims, or evidence (Ennis, 1985; Facione, 1990; Halpern, 1999; Paul, 1992);
- making inferences using inductive or deductive reasoning (Ennis, 1985; Facione, 1990; Paul, 1992; Willingham, 2007);
- judging or evaluating (Case, 2005; Ennis, 1985; Facione, 1990; Lipman, 1988; Tindal & Nolet, 1995); and
- making decisions or solving problems (Ennis, 1985; Halpern, 1999; Willingham, 2007).

Critical thinking in a performance assessment context encompasses multiple dimensions. The scenarios and tasks presented should reflect real-life situations. The test should align with these authentic contexts, ensuring that tasks are engaging and invite test-takers to actively contribute. It should also encourage the application of critical thinking skills, such as evaluation, analysis, and synthesis, in line with Bloom's taxonomy.

As seen in Figure 1, Norris and Ennis (1989) explain the process of critical thinking. Although the direction is from bottom to top, this does not imply that critical thinking follows an order. This figure depicts the relationships between the components of critical thinking in practice. In other words, critical thinking in practice can proceed and be elicited in various directions and without specific patterns. According to Norris and Ennis (1989), critical thinking is fundamentally a decision-oriented process, centered on determining what one should believe or do. To make a decision, we rely upon some information, background knowledge, and previously accepted conclusions. These form the basis of support for the decision, and the decision itself is the process of inference. The process of critical thinking begins with interactions and then leads to the problem-solving box to go through the process to accomplish the final decision about belief or action. The basic support box is the engine of critical thinking, which connects the problem with the background knowledge, and on top of that, there are logical inferences. The substantive boxes are clarity and critical thinking dispositions; in the former, the arrows emanating from it reflect the pervasion of clarity in of critical thinking process, and the latter informs us of the importance of critical thinking dispositions.

FIGURE 1
Pictorial Representation of Critical Thinking



Note. Adapted from Norris & Ennis (1989).

Assessing Speaking and Critical Thinking Skills Using Computer-Based Testing

There are various approaches to capturing test-takers' critical thinking skills, but the context in which performance takes place should provide circumstances in which the critical thinking skills of test-takers are elicited while they are speaking. Individual interviewing is one of the techniques to gather data on test-takers' critical thinking, leaving heavy demands and resources to build background knowledge connected with contemporary scenarios, but this can serve as a practical approach. Norris and Ennis (1989) later mention this technique, if used properly, can be better than the other approaches to acquiring critical thinking skills of test-takers because the inferences that are being built based on test-takers' performance are trustworthy. There are some concerns regarding the invalidity of this approach because of the presence of the interviewer and the student, but there are two points in this study that address this issue: 1) initiating tasks with open-ended instructions, that is, test-takers build their responses based on the initial instructions and instructions have no intended impact on the elicitation of critical thinking skills, and, 2) the interview approach implemented in this study is based on using a computer and collecting test-takers' performance with the help of an online testing interface.

As mentioned by Segalowitz and Trofimovich (2012), language and speaking consist of volitional and social dimensions and a large amount of contextual knowledge. Speakers need to identify volition, which is “when speakers use a language, they behave as active agents” (p. 182); in other words, they need a reason to initiate interaction. Segalowitz and Trofimovich wrote that, in this situation, a speaker tends to be persuasive to the listener, and “this means that the processing underlying speech output includes the processing underlying the formation of communicative intentions” (p. 182). In a computer-based testing situation, this means test-takers are supposed to imagine the interlocutor and create a context for their expectation, taking the correct agent position to construct appropriate speech.

In a computerized test, the social dimension does exist artificially, and test-takers consider it as the context. As noted by Segalowitz and Trofimovich, “the social dimension of communication has processing implications for speakers, especially in the L2” (p. 184). The interlocutors try to mitigate the processing load to proceed with the interaction. Wray (2002) mentioned that normally, speakers try to help each other by minimizing the processing loads they place on each other. They can do this, for example, by using formulaic expressions and partially fixed strings. In a computerized test, because of the absence of interlocutors, this difference is met by providing context in the test tasks. So, the test prompt provides the volition (the reason to speak) and the social dimension (the target audience), and by test inquiries like asking for summarizing a speech, explaining a picture, or responding to a scenario, the processing load will be reduced. Thus, understanding the directions (and the testing context) is crucial in a computerized speaking test; clear and transparent instructions play a critical role in reducing the processing load and creating a social dimension.

Segalowitz and Trofimovich (2012) also stressed that “speakers must understand the contexts in which the L2 is used” (p. 184). Due to their explanation, there are two general environmental contexts that shape speech processing: closed-skill contexts or open-skill contexts. In a closed-skill context, “variability in the conditions under which performance takes place has negligible impact on performance, and ... the goal of performance is to repeat some action (physical or cognitive) as precisely as possible to meet some standard” (p. 185). In a computer-based speaking test, examples of closed contexts are listing favorite activities or describing doing something on a daily basis. Segalowitz and Trofimovich (2012) later explain that open-skill contexts offer variability, so there are unforeseeable events, while this doesn’t exist in closed-skill contexts. Open skill depends heavily on quick responses and noticing the changes rapidly. Thus, some of the speaking tests are open skills, and some are closed. As the focus of the study is the closed-skill context, another study by Lee and Winke (2017) explains cognitive operations involved in a closed computer-based test that need to be satisfied to accomplish a successful testing process, outlined in Table 1 below.

TABLE 1
Cognitive Operations Involved in Closed Computer-Based Speaking Tests

Cognitive operation	Test-taker’s process	Rater’s assumption
1. Understanding of the volition of the task and the imagined context/scenario	Test-takers must understand the test directions and prompt to better address the test target, that is, understanding the volition of speaking and CT test tasks	Raters may assume test-takers understand the test directions and prompt, and they follow test directions to provide relevant responses

2. Understanding of the speaking context (closed-skill task)	Test-takers need to understand that the speaking tasks are closed; that they will not receive instant feedback and they need to follow the time frames and prompts to successfully complete the tasks.	Raters understand the task is closed, that there is no variability in the conditions under which performance takes place. Raters believe test-takers understand the task parameters, including time frame and transition between test tasks. Thus, each performance can be directly compared against the assigned criteria.
3. Performance of the speech act	Test-takers must demonstrate linguistic knowledge along with the critical thinking skills and the ability to access it to <i>actively speak</i> after (or while) performing the first two cognitive operations.	Raters may assume that if a test-taker does not speak or does not speak correctly (or about the correct things), they do not have the linguistic knowledge or ability to do so, resulting in a low score.

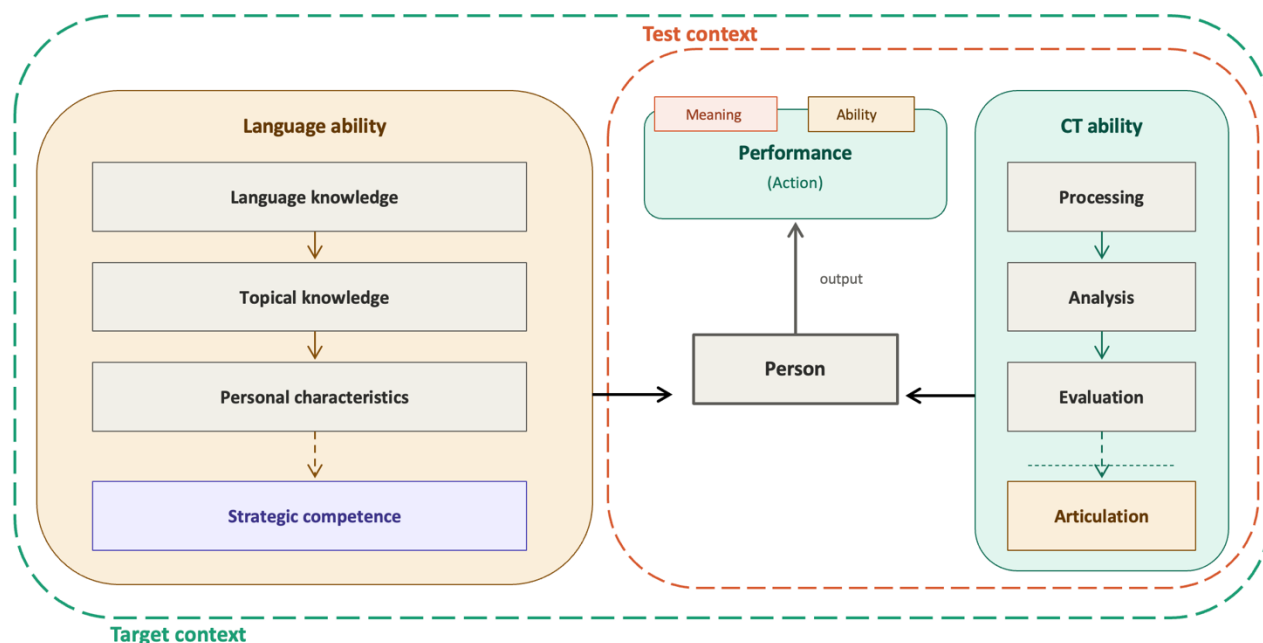
Note. Adapted from Lee and Winke (2017).

Defining the Critical Thinking-to-Interact Construct in Speaking Assessment

Understanding speaking ability and critical thinking is essential in crafting a construct definition. Hence, the intention of this study is to define a new construct based on the integration of speaking and CT skills and grasp them as one unique construct rather than two distinct constructs. More importantly, some assessment views would help me develop a definition in the assessment environment. Weissberg (2013) highlights that CT is not just about gathering facts. Instead, it involves using reasoning skills to delve deeper and uncover more profound meanings behind those facts. This means going beyond simply knowing information to understanding its significance, implications, and the broader context in which it exists. Moreover, Liu et al. (2014) provide a comprehensive review of critical thinking assessments in higher education, highlighting their definitions, psychometric qualities, and implementation challenges.

The core of the construct definition of speaking and CT assessment is rooted in Bachman's (1990) model of language ability, Chapelle's (1999) interactionist approach to defining second language ability, Paul and Elder's (2001) CT Model, and Norris and Ennis's (1989) evaluation of critical thinking. Bachman's model of language competence is a framework that describes the ability to use a language effectively for communication. The model hypothesizes language ability to emerge from the use of a speaker's language knowledge, topical knowledge, and personal characteristics mediated by their strategic competence. Paul and Elder's (2001) CT model focuses on three sections: elements of thought, intellectual standards, and intellectual traits. Purpose, interpretation, inference concepts, implications and consequences, and point of view will represent the elements of thought and clarity, relevance, depth, logic, and significance as intellectual standards. Figure 2 depicts the relationship between language ability and critical thinking in the context of speaking assessment.

FIGURE 2
Construct Definition Model for a Speaking and Critical Thinking Assessment



As seen in Figure 2, the proposed model for the test construct involves five building blocks. First, the *target context*, i.e., the context in which the test context happens; for a speaking and CT test, consider the university or academia as the target context, and the test context would be the oral communication test that is required by the university or academia. The other four components are inside the target context. The second refers to *language ability*, which includes *topical knowledge*, *language knowledge*, *personal characteristics*, and *strategic competence*. The third area is *CT ability*, comprising *processing*, *analysis*, *evaluation*, and *articulation*. The next component is the *test context*, in which the test and elicitation of language and CT abilities happen. The fifth building block refers to the *person* who will implement both language and CT abilities in the test context.

At the center of the framework, *performance* is represented as action, the moment in which language ability and critical thinking ability converge in real communicative behavior. Two terms are positioned alongside performance to reflect this dual origin: *meaning* derives from the language ability dimension, capturing the idea that spoken performance is inherently an act of meaning-making through linguistic knowledge, topical knowledge, and strategic competence; while *ability* derives from the CT ability dimension, signaling that performance is simultaneously an exercise of cognitive capacity. Together, meaning and ability illustrate that performance cannot be reduced to either linguistic form or cognitive skill alone; it is the intersection of both.

So, to make this model meaningful, let us consider language ability, CT ability, and performance as vertices of a triangle. Each vertex's presence in the model is necessary. We cannot have a performance that reflects the Critical-Thinking-to-Interact construct with only one of the abilities, and these abilities need context. Furthermore, language ability and CT ability exist in the person who produces language. The arrows pointing toward the Person from both the

Language ability and CT ability zones represent the inputs a speaker draws upon before and during speech: language ability feeds into the person as the communicative resources available to them, while CT ability feeds into the person as the cognitive operations that shape how those resources are deployed. *Output*, in this model, refers to both the spoken product and the observable communicative behavior that results from this convergence, not merely what is said, but how meaning and ability are made manifest through speech. Performance relies on implementing both critical thinking and linguistic skills. Thus, the Critical-Thinking-to-Interact construct of a speaking and CT assessment is defined as the ability to effectively use English in interactive communicative situations, demonstrating both linguistic proficiency and critical thinking skills. This construct emphasizes the quality and soundness of responses. The components of the Critical-Thinking-to-Interact construct are:

- *Linguistic proficiency*: The ability to articulate thoughts clearly and accurately in English.
- *Critical thinking*: The capacity to analyze, evaluate, and respond thoughtfully to information and arguments.

This construct evaluates how well individuals can combine their language skills with critical thinking to participate meaningfully in interactive settings. The linguistic component measures the test-takers' ability to communicate clearly, fluently, and appropriately in various contexts, while the CT component assesses the ability to analyze and evaluate information, articulate well-reasoned arguments and make inferences using inductive or deductive reasoning.

This review has synthesized foundational literature supporting the development of a performance-based test of speaking and CT. It reviewed the theoretical and pedagogical rationale for performance assessment in ESL contexts, highlighted essential components of oral communicative competence, and explored the centrality of critical thinking in academic communication. It further examined the potential of computer-based testing to assess these skills and considered emerging frameworks that conceptualize speaking and critical thinking as integrated constructs.

Assessments of Academic Speaking and 'Critical-Thinking-to-Interact'

Building on the proposed 'Critical-Thinking-to-Interact' construct, it is worth examining how existing L2 speaking assessments operationalize CT as an explicit construct component. Despite the growing theoretical recognition that CT is central to academic speaking, very few operational L2 speaking assessments have explicitly incorporated CT as a defined construct component (Yin et al., 2024). For instance, Yin et al. (2024), in their analysis of three standardized English for Academic Purposes (EAP) speaking tests, including IELTS, TOEFL iBT, and the College English Test Spoken English Test (CET-SET6), found that CT tends to appear only implicitly in rating scale descriptors. The absence of an explicit, systematic operationalization of CT as a measurable construct results in construct underrepresentation. Against this backdrop, this section reviews the small number of assessments that have explicitly attempted to operationalize the CT construct in speaking, as a way of grounding and illustrating the proposed 'Critical-Thinking-to-Interact' construct in concrete examples.

To review these assessments systematically, this section draws on Evidence-Centered Design (ECD), a framework developed by Mislevy et al. (2003). According to Mislevy et al. (2003), ECD is grounded in three core premises: (1) assessment must be built around important

knowledge in the domain; (2) reasoning from what test-takers do must follow evidentiary principles; and (3) purpose must drive all design decisions. For the purposes of this review, two components of ECD are particularly relevant: The Task Model and the Student Model. The Task Model is a generic description of a family of tasks that includes features such as the task, task prompt, and test-taker output. The Student Model, on the other hand, outlines a set of variables that represent a test-taker's knowledge, skills, and abilities (Yin & Mislevy, 2022; Mislevy et al., 2003). This section examines how each assessment is structured around the Task Model and the Student Model.

Among the few assessments that have explicitly attempted to operationalize CT in speaking contexts, two are particularly relevant to the proposed 'Critical-Thinking-to-Interact' construct. The first is the Speaking Assessment and Critical Thinking test (SA & CT test), which was developed to assess speaking and CT skills of university-level international students (Ebrahimpourlighvani, 2025). The second is the CET-SET6, a computer-based English speaking test administered in China at the Band 6 level, in which the paired discussion task requires test-takers to exchange ideas, debate, explain, and compare perspectives on a given topic. These assessments are reviewed below using three ECD lenses: task features, proficiency claims for speaking, and proficiency claims for CT, summarized in Table 2.

TABLE 2
Assessments of Academic Speaking and Critical-Thinking-to-Interact Using ECD Framework

Test Example	Task Features	Proficiency Claims for Speaking	Proficiency Claims for Critical Thinking
SA & CT Test (Ebrahimpourlighvani, 2025)	<ul style="list-style-type: none"> • Three-section computer-based speaking test: <ul style="list-style-type: none"> ○ Section 1: Descriptive prompt task ○ Section 2: Picture-based task ○ Section 3: Scenario-based persuasion task • Tasks reflect authentic university contexts • Closed-skill, time-limited format 	<ul style="list-style-type: none"> • Fluency • Interactional competence • Grammar and vocabulary • Pronunciation and comprehensibility 	<ul style="list-style-type: none"> • Problem identification and comprehension • Solution generation • Logical reasoning • Soundness of responses
CET-SET6 (Yin et al., 2024)	<ul style="list-style-type: none"> • Four-task computer-based EAP speaking test: <ul style="list-style-type: none"> ○ Task 1: Warm-up questions, ○ Task 2: Individual presentation on a given topic, ○ Task 3: Paired discussion on a given topic ○ Task 4: Further-check questions • Dialogic task requires exchanging ideas, debating, explaining, and comparing 	<ul style="list-style-type: none"> • Contribution and discourse management • Sustained and coherent speech • Flexibility and appropriacy • Active engagement in discussions 	<ul style="list-style-type: none"> • Six CT skills: interpreting, analyzing, evaluating, explaining, inferencing, regulating • Six CT criteria: clarity, relevance, logicity, breadth, depth, fairness • CT dispositions: cognitive, personality, and motivational components

Note. Task Features correspond to the Task Model in ECD; Proficiency Claims correspond to the Student Model in ECD (Mislevy et al., 2003; Yin & Mislevy, 2022).

CONCLUSION

In conclusion, this literature review demonstrates that oral communication in academia is not merely a matter of linguistic form but is fundamentally a reflection of deep-seated critical thinking processes. By synthesizing foundational theories from philosophy, psychology, and education, such as Bloom's taxonomy and the interactionist perspective, it becomes evident that assessing language in isolation from the ability to articulate reasoned arguments fails to capture the full scope of academic communicative competence. The proposed "Critical-Thinking-to-Interact" construct offers a more holistic framework for evaluating how university students use language as a vehicle for analysis, inference, and problem-solving. Furthermore, the move toward computer-based performance assessments that simulate real-world scenarios provides a scalable and efficient means of capturing these complex interactions. As higher education becomes increasingly globalized, adopting assessment models that prioritize the quality and soundness of arguments alongside linguistic accuracy will be essential for supporting the academic success of international students and standardizing global oral communication criteria.

REFERENCES

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 41–71). University of Ottawa Press.
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay Company.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
<https://doi.org/10.1093/applin/i.1.1>
- Carroll, J. B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. In *Testing the English proficiency of foreign students* (pp. 30–40). Center for Applied Linguistics.
- Case, R. (2005). Moving critical thinking to the main stage. *Education Canada*, 45(2), 45–49.
- Chapelle, C. A. (1999). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge University Press.
<https://doi.org/10.1017/CBO9781139524711.004>
- Douglas, D. (1998). Testing methods in context-based second language research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 141–155). Cambridge University Press.
<https://doi.org/10.1017/CBO9781139524711.008>

- Ebrahimpourlighvani, A. (2025). *Speaking assessment and critical thinking test: An innovation in assessing oral communication ability of international university-level students* (Publication No. 3224180533) [Master's thesis, Iowa State University]. ProQuest Dissertations & Theses Global.
- Ellis, R. (1989). Sources of intra-learner variability in language use and their relationship to second language acquisition. In S. Gass, C. Madden, D. Preston, & L. Selinker (Eds.), *Variation in second language acquisition: Psycholinguistic issues* (pp. 22–45). Multilingual Matters.
- Ennis, R. H. (1962). A concept of critical thinking. *Harvard Educational Review*, 32(1), 81–111.
- Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44–48.
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. American Philosophical Association.
- Halpern, D. F. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning*, 1999(80), 69–74. <https://doi.org/10.1002/tl.8005>
- Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–285). Penguin.
- Jones, R. L. (1985). Some basic considerations in testing oral proficiency. In Y. Lee, A. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing* (pp. 77–84). Pergamon.
- Kennedy, M., Fisher, M. B., Ennis, R. H., Jones, B. F., & Idol, L. (1991). Critical thinking: Literature review and needed research. In L. Idol & B. F. Jones (Eds.), *Educational values and cognitive instruction: Implications for reform* (pp. 11–40). Routledge. <https://doi.org/10.4324/9781315044392>
- Lee, S., & Winke, P. (2017). Young learners' response processes when taking computerized tasks for speaking assessment. *Language Testing*, 35(2), 239–269. <https://doi.org/10.1177/0265532217704009>
- Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice*, 32(3), 131–137. <https://doi.org/10.1080/00405849309543588>
- Lipman, M. (1988). Critical thinking: What can it be? *Educational Leadership*, 46(1), 38–43.
- Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series*, 2014(1), 1–23. <https://doi.org/10.1002/ets2.12009>
- McNamara, T. F. (1996). *Measuring second language performance*. Longman.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11. <https://doi.org/10.3102/0013189X018002005>
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23. <https://doi.org/10.3102/0013189X023002013>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02
- Moss, P. A., & Koziol, S. M. (1991). Investigating the validity of a locally developed critical thinking test. *Educational Measurement: Issues and Practice*, 10(3), 17–22. <https://doi.org/10.1111/j.1745-3992.1991.tb00199.x>

- National Academies of Sciences, Engineering, and Medicine. (2020). *A principled approach to language assessment: Considerations for the U.S. Foreign Service Institute*. The National Academies Press. <https://doi.org/10.17226/25748>
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Critical Thinking Press and Software.
- Ockey, G. J. (2025). *Introducing second language assessment*. Cambridge University Press.
- Paul, R. (1992). Critical thinking: What, why, and how. *New Directions for Community Colleges*, 1992(77), 3–24. <https://doi.org/10.1002/cc.36819927703>
- Paul, R., & Elder, L. (2001). *Critical thinking: Tools for taking charge of your professional and personal life*. Pearson Education.
- Segalowitz, N., & Trofimovich, P. (2012). Second language processing. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 179–192). Routledge.
- Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement, and improvement*. National Institute of Education.
- Tindal, G., & Nolet, V. (1995). Curriculum-based measurement in middle and high schools: Critical thinking skills in content areas. *Focus on Exceptional Children*, 27(7), 1–22. <https://doi.org/10.17161/foec.v27i7.6847>
- Weissberg, R. (2013). Critically thinking about critical thinking. *Academic Questions*, 26(3), 317–328. <https://doi.org/10.1007/s12129-013-9375-2>
- Widdowson, H. G. (1983). *Learning purpose and language use*. Oxford University Press.
- Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator*, 31, 8–19.
- Wilson, K. (2016). Critical reading, critical thinking: Delicate scaffolding in English for Academic Purposes (EAP). *Thinking Skills and Creativity*, 22, 256–265. <https://doi.org/10.1016/j.tsc.2016.10.002>
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Yin, C., & Mislevy, R. J. (2022). Evidence-centered design in language testing. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed., pp. 289–305). Routledge. <https://doi.org/10.4324/9781003220756-23>
- Yin, S., Fan, J., Jin, Y., & Stapleton, P. (2024). Towards a framework of critical thinking for assessing EAP speaking. *Journal of English for Academic Purposes*, 71, 101426. <https://doi.org/10.1016/j.jeap.2024.101426>

Ali Ebrahimpourlighvani holds an M.A. in Applied Linguistics (Language Assessment) from Iowa State University. His research interests include AI in Applied Linguistics, NLP, agentic AI, critical thinking, and language assessment. He is currently a doctoral student in Applied Linguistics at Teachers College, Columbia University, specializing in Language and Technology. Correspondence should be sent to Ali Ebrahimpourlighvani, E-mail: ae2936@tc.columbia.edu.