

Japanese Language and Literature

Journal of the American Association of Teachers of Japanese

jll.pitt.edu | Vol. 60 | Number 1 | April 2026 | DOI: 10.5195/jll.2026.309

ISSN 1536-7827 (print) 2326-4586 (online)

Exploring the Validity of the Advanced Placement (AP) Japanese Computer-Simulated Interpersonal Speaking Task: Domain Definition Inference

Nana Suzumura-Smith

Abstract

This study examined the domain definition inference for the AP Japanese exam, which aims to measure language proficiency in situations reflecting real life. For the domain definition inference, this exam attempts to claim that (a) features of examinee performance that provide evidence of language proficiency have been identified, and (b) the assessment tasks are relevant to and representative of the real-life domain. Focusing on the computer-simulated conversation task of the AP Japanese exam, four assumptions underlying these claims were investigated: (1) the target domain is clearly identified; (2) critical language skills and knowledge needed in the target domain are identified; (3) assessment tasks that are representative of the target domain are identified; and (4) assessment tasks that require critical skills and are representative of the target domain are simulated. Backing for these assumptions was gathered through analyses of exam-related documents and test task characteristics, including context, input language, and scoring criteria. This study found reasonable backing for the first two assumptions and partial backing for the others. Therefore, this paper argues that the two claims that the exam attempts to make for the domain definition inference are not fully supported. It also discusses possible ways to improve the validity argument.

1. Introduction

Computer-based language assessments have become common, especially for large-scale testing. Examples include the ACTFL Assessment of Performance toward Proficiency in Languages (AAPPL) and the Standards-Based Measurement of Proficiency—4 Skills (STAMP4S). Many of these tests include tasks that assess oral communication skills. However, they often lack an interpersonal component—a shortcoming that



New articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by [Pitt Open Library Publishing](https://pittopenlibrarypublishing.com).

Roever and Ikeda (2022) identify as a case of construct underrepresentation. Most tests target only one-way/presentational oral communication skills using monologic tasks. Against this backdrop, more recent computer-based tests have started to include tasks assessing interpersonal skills to achieve broader construct representation (Ockey and Neiriz 2021). Nevertheless, the validity of such assessment tasks has not been sufficiently investigated. In this context, validity refers to “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment” (Messick 1989, 13; italics in original). This paper explores the validity of computer-based assessment of interpersonal speaking through an analysis of the computer-simulated conversation task on the AP Japanese Language and Culture Exam. Specifically, the study focuses on the domain definition inference as a component of the broader validation process.

1.1. AP Japanese Language and Culture Exam

The AP Japanese exam was first administered in 2007. More than 3,000 students take the exam every year (College Board 2024). The intended test-takers are high school students enrolled in the AP Japanese course whose first language is not Japanese and who have received the equivalent of 300 hours of college-level introductory Japanese language instruction. The exam is designed to assess communicative language skills and learning objectives outlined for the AP Japanese Language and Culture course (College Board 2020, 19–21). The AP Japanese course “emphasizes communication (understanding and being understood by others) by applying the interpersonal, interpretive, and presentational modes of communication in real-life situations” (College Board 2020, 7). The interpersonal mode involves spontaneous, two-way communication. The interpretive mode refers to one-way communication, where listeners or readers interpret meaning. In the presentational mode, the speaker or writer prepares and shares information with an audience in one-way communication. Key aspects of communication skills emphasized in the course include vocabulary usage, language control, communication strategies, and cultural awareness.

The exam is approximately two hours long and consists of six sections: Listening, Reading, Interpersonal Writing, Presentational Writing, Interpersonal Speaking, and Presentational Speaking. Each section targets specific modes of communication in the *World-Readiness*

Standards (National Standard Collaborative Board 2015), as shown in Table 1.

Table 1: The AP Japanese exam format (2009–2024)

Skill	Task type	Target mode of communication	Number of items	Weight	Time
Listening	Multiple choice	Interpretive	30–35	25%	20 min. (Response time: 12 sec. per question)
Reading	Multiple choice	Interpretive	35–40	25%	60 min.
Writing	Text chat	Interpersonal	6	12.5%	10 min. (Response time: 90 sec. per question)
Writing	Compare and contrast article	Presentational	1	12.5%	20 min.
Speaking	Conversation	Interpersonal	4	12.5%	3 min. (Response time: 20 sec. per question)
Speaking	Cultural perspective presentation	Presentational	1	12.5%	Preparation time: 4 min. Response time: 2 min.

The AP Japanese Conversation is an asynchronous oral communication task designed to assess interpersonal speaking. Students first see the name and photo of their simulated interlocutor, along with a brief explanation of the situation in English. No other non-verbal cues, such as gestures or eye gaze, are available during this task. Students then participate in a simulated conversation by responding to four pre-recorded prompts (utterances and/or questions) that they hear from the computer (see Appendix). Students have twenty seconds to respond to each prompt without any preparation time.

After the exam, the series of student responses is separated by prompts and scored individually by human raters—teachers of Japanese at either the high school or college level—who have received rater training from the College Board. The College Board’s standard practice is to assign one rater per response. The scoring rubric (College Board 2023) uses a 7-point holistic scale, ranging from Score 0 to Score 6. A rater assigns a single score based on three aspects: task completion, delivery, and language use. Table 2 shows sample criteria for the conversation prompt score of 6, which indicates the highest level of spoken interpersonal communication skills.

Table 2: Scoring criteria for the AP Japanese conversation prompt score of 6 (College Board 2023)

Task completion	<ul style="list-style-type: none"> • directly addresses [the] prompt and provides a very thorough and appropriate response; includes elaboration and detail
Delivery	<ul style="list-style-type: none"> • natural, easily flowing expression • natural pace with minimal hesitation or repetition • pronunciation virtually error free • consistent use of register and style appropriate to [the] situation
Language use	<ul style="list-style-type: none"> • rich vocabulary and idioms • excellent use of grammar and syntax with minimal or no errors

Note: Each conversation prompt score ranges from 0 to 6.

The College Board claims to use a variety of validity measures for all AP exams (College Board 2020; n.d.-a; n.d.-b), including college curriculum surveys, test development by subject matter experts, and college comparability studies of test scores. For the AP Japanese exam, the College Board (Baum 2007b; Tohsaku 2009) states that it periodically conducts a college curriculum survey to ensure that the targeted topics, skills, and knowledge covered on the AP Japanese exam are equivalent to those covered in a college-level intermediate Japanese language course. However, the results of the college curriculum survey have never been made public.

Regarding test development, Tohsaku (2009) documented the brief history of the creation of the AP Japanese exam. In 2004, the College Board formed the AP Japanese Task Force with target-language experts to define the target settings and draft the objectives, format, and specifications for the AP Japanese exam based on the college curriculum survey, the *Standards for Foreign Language Learning in the 21st Century* (National Standards in Foreign Language Education Project 1999), and the ACTFL's proficiency guidelines (Breiner-Sanders et al. 2000). The College Board then formed the AP Japanese Development Committee in 2005 with target-language experts. This committee, together with ETS assessment specialists, finalized the AP Japanese exam specifications and

scoring criteria. This committee is currently responsible for reviewing periodic curriculum surveys and writing new exam questions based on the test specifications.

Regarding college comparability, Baum (2007b) reported that the College Board conducted a study by administering a subset of the 2007 AP Japanese exam to college students and confirmed that the cut scores of the exam grades were comparable to the grading standards of corresponding college courses (e.g., A, B, C, D). However, Baum (2007b) is the only document that reports the results of the AP Japanese exam comparability study. As a result, it is unknown whether other comparability studies have been conducted for the AP Japanese exam in subsequent years.

Despite the College Board's claim, the validity of the AP Japanese exam has rarely been publicly discussed by the test developers or scholars. In fact, even for other AP world language exams, publicly available validity studies and documents explaining specific validity measures and results are extremely limited. Several previous studies have examined various aspects (e.g., college curriculum, psychometric structure, score comparability) related to the validity of AP world language exams, including Spanish, French, German, and Chinese (Arendt and Morgan 1995; Baum 2007a; Bischof 2005; Bischof et al. 2004; Ginther and Stevens 1995; Liskin-Gasparro et al. 1979; Modu et al. 1975; Modu and Weeks 1982; Morgan and Mazzeo 1988; Wimmers and Morgan 1990). However, all of these studies, with the exception of Baum (2007a) for Chinese, were conducted using older test formats. Furthermore, no previous study has comprehensively examined the free-response sections such as speaking and writing. To address this gap, the author initiated a validation research project for the AP Japanese exam, focusing on the speaking tasks (Suzumura 2020; 2022). This paper reports findings from the initial phase of this project.

1.2. Theoretical Framework

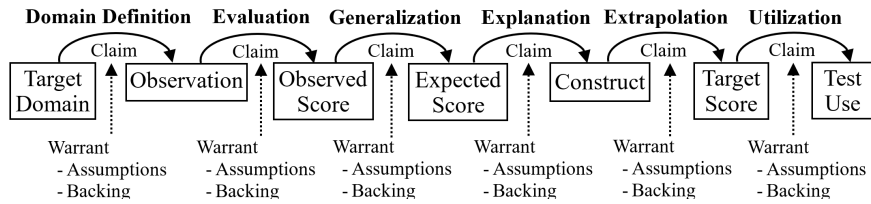
The present study adopted an argument-based approach to validity (Chapelle et al. 2008b) as an overarching framework. In the field of language testing, validity is considered a multifaceted yet unitary concept (Messick 1989). Validity concerns not only the traditional sense of construct validity (i.e., whether the test measures what it intends to measure) (Lado 1961) and the use of test scores, but also the value implications and social consequences of the test. Several validation

models have been proposed based on Messick's unitary validity framework. One of the prominent models developed using this framework is the argument-based approach, first proposed by Kane (1992) and extended by Chapelle et al. (2008b).

An advantage of the argument-based approach to validity over checklist-style evidence-gathering approaches—such as the test usefulness model (Bachman and Palmer 1996) and socio-cognitive model (Weir 2005)—is that it provides a clearer roadmap for logical and systematic validation (Chapelle and Lee 2021; Shepard 1993). Although both approaches view validity as a multifaceted concept (e.g., content, construct, social consequences, etc.) and aim to accumulate evidence, the checklist approach does not provide a framework for determining where to begin, how to proceed, and how to integrate findings related to different facets. It is also unclear how much supporting evidence needs to be gathered to make a validity claim. The argument-based approach, on the other hand, breaks down the validation process into stages; these will be further explained in the next section. This approach suggests that all claims and underlying assumptions should be laid out and examined by gathering supporting evidence at each stage. This roadmap allows researchers to organize the validation process and integrate findings more systematically. It also helps researchers organize findings from other studies and clarify the interpretive arguments and assumptions targeted in those studies.

The argument-based approach conceptualizes validity as a multi-stage interpretive argument linked by different types of inferences, including domain definition, evaluation, generalization, explanation, extrapolation, and utilization, as shown in Figure 1. Each stage of an argument consists of a claim, warrant, assumption, and backing. A warrant for a claim is based on some assumptions. Therefore, each underlying assumption must be examined for its backing in a validation study. All the evidence gathered in the assumption-checking process is integrated to make an evaluative judgment of a claim. The evaluative judgment at one stage serves as the ground for the next stage.

Figure 1: Interpretive Argument Structure for the AP Japanese Exam (Chapelle et al. 2008b)



Domain definition is the very first stage of inference for an interpretive argument. It aims to establish a link between performance in the target language use (TLU) domain and performance observed on the test. A TLU domain is “a specific setting outside of the test itself that requires the test taker to perform language use tasks” (Bachman and Palmer 2010, 60). Language ability in the TLU domain is what we want to generalize based on inferences drawn from the test score.

In the context of the AP Japanese exam, for example, domain definition inference generally concerns with how (clearly) the TLU domain is identified and how well the assessment tasks developed represent tasks in the target domain. In other words, it pertains to the situational and interactional authenticity of the assessment task (Bachman 1990). Since the domain definition inference serves as the basis for subsequent stages of inference, strong evidence for the domain definition is critical to building a solid foundation for the validity argument. Evaluation inference aims to ensure the intended relationship between observed performances and scores on the exam. In other words, it generally concerns whether the established rating criteria and scale are appropriate and have been applied as intended. Generalization inference addresses how reliable the observed scores on one exam or task are in estimating the expected scores on parallel test forms or tasks. Explanation inference addresses how the expected scores are attributed to a construct of language proficiency. Extrapolation inference is concerned with how the expected scores account for the quality of Japanese language performance in real-life situations (target scores). Utilization inference links the target scores to the decisions based on the scores, such as awarding a certain number of college credits or placing in an advanced course.

1.3. Domain Definition Inference

The focus of the present study is on domain definition inference. For large-

scale tests, studies and documents related to this type of inference may exist but are often not publicly available due to concerns such as test security and intellectual property. As a result, it is difficult to find out what kind of supporting evidence exists for many of these tests. One exception, however, is Chapelle et al. (2008a), which publicly presented evidence for domain definition inference for the Test of English as a Foreign Language (TOEFL) iBT using an argument-based approach.

In their project, an interpretive argument for domain definition is presented as follows:

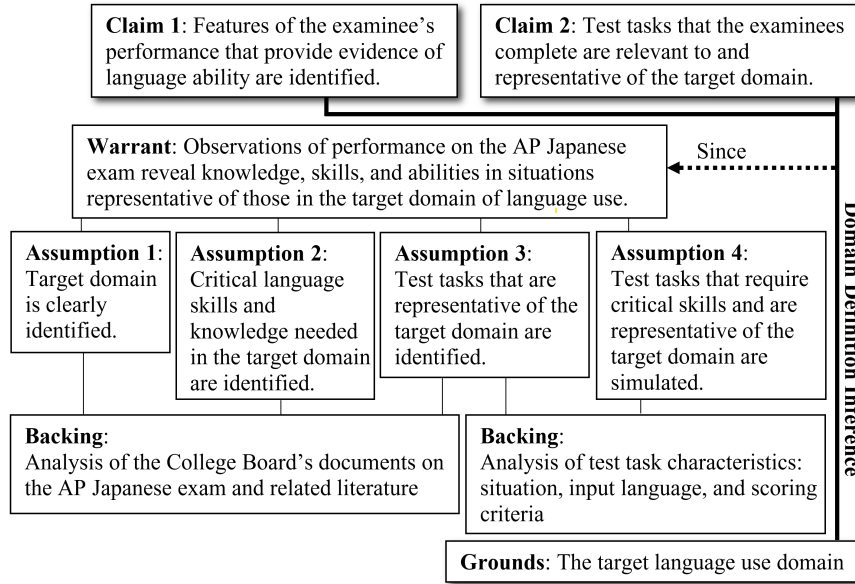
The domain definition inference is based on the warrant that observations of performance on the TOEFL reveal relevant knowledge, skills, and abilities in situations representative of those in the target domain of language use in English-medium institutions of higher education. This warrant, in turn, is based on the assumptions (a) that assessment tasks representing the academic domain can be identified, (b) that critical English language skills, knowledge, and processes needed for study in English-medium colleges and universities can be identified, and (c) that assessment tasks requiring important skills and representing the academic domain can be simulated as test tasks. (Chapelle 2008, 322)

To support these assumptions, the test developers, including content experts, conducted several studies. First, they analyzed the domain of interest (academic life) and identified the test domain; the TOEFL iBT targets “the communicative language ability of people whose first language is not English” and measures their “English-language proficiency in situations and tasks reflective of university life in North America” (Jamieson et al. 2000, 10). They then reviewed relevant theories of communicative language ability and language tasks, identified task characteristics (e.g., situation, text material, and test rubric), and operationalized the variables involved in each characteristic (Butler et al. 2000; Douglas 1997). Using this information, the test developers designed test task prototypes that simulated target tasks in non-test situations (Jamieson et al. 2008). These prototypes were piloted, and examinees’ performance was analyzed to adjust the test tasks (Enright et al. 2008). Based on the results of these studies, Chapelle (2008, 325) concluded that “test users can be confident that the performance test takers display as they take the test reflects their academic language ability and use in the academic domain.”

Adapting the argument structure used in Chapelle et al. (2008a), the

present study proposes the interpretive argument shown in Figure 2 for the AP Japanese exam. Note that the College Board did not specify claims, warrants, and assumptions for the AP Japanese exam using an argument-based approach, so the author instead proposed those based on Chapelle et al. (2008a) and available information from the College Board.

Figure 2: Overview of the domain definition inference for the AP Japanese exam (adapted from Chapelle (2008, 323))



There are two claims that the author assumes the College Board wants to make with the target language use domain as grounds: (a) features of the examinees' performance that provide evidence of language ability are identified, and (b) the test tasks that the examinees complete are relevant to and representative of the target domain. The former relates to the construct, and the latter relates to task and content authenticity. These claims are based on a warrant that examinees' performance on the AP Japanese exam reveals their knowledge, skills, and abilities in situations representative of those in the target domain of language use. There are four assumptions underlying this warrant: (1) the target domain is clearly identified; (2) critical language skills and knowledge needed in the target domain are identified; (3) test tasks that are representative of the target domain are identified; and (4) test tasks that require critical skills and are

representative of the target domain are simulated. The validity of the domain definition inference for the AP Japanese exam depends on whether there is sufficient support for all of these assumptions.

1.4. Context and Purpose of the Present Study

This study is part of a larger research project examining the validity of the AP Japanese speaking tasks (Suzumura 2020; 2022). The author is an independent researcher not affiliated with the test developers. To investigate the four assumptions in the context of the simulated conversation task, the present study addresses the following research questions:

1. What is the target domain of the AP Japanese exam? Is it clearly identified? (Assumption 1)
2. What are the critical skills and knowledge in the target domain? Are they clearly identified? (Assumption 2)
3. How is the conversation task operationalized? Is it representative of the target domain? (Assumption 3)
4. How well do the characteristics of the conversation task correspond with the characteristics of the target domain task? (Assumption 4)

2. Method

2.1. Document Collection and Analysis

For this study, documents related to the AP Japanese conversation task were collected from the College Board website (<https://apcentral.collegeboard.org/courses/ap-japanese-language-and-culture/exam>). Table 3 lists the documents collected and the research questions for which they were primarily used.

Documents related to the 2020 conversation task were not included in this study since they are not published on the College Board's website. Information about scoring procedures was also gathered through personal communication with the Director of the AP World Languages and Cultures at the College Board and with AP Japanese Readers. The scoring criteria were used to identify the critical language skills and knowledge specified by the College Board for the conversation task. The criteria were also examined to assess their relevance and representativeness in the target domain.

Table 3: Documents Collected and Research Questions Addressed

Document	RQ1	RQ2	RQ3	RQ4
<i>AP Japanese Language and Culture: Course and Exam Description</i> (College Board 2020)	✓	✓	✓	
<i>AP Japanese Language and Culture: Teacher's Guide</i> (Tohsaku 2009)	✓	✓	✓	
<i>AP Japanese Language and Culture Scoring Guidelines</i> (College Board 2023)		✓	✓	✓
15 sets of the AP Japanese conversation task (College Board 2007a; 2008a; 2009a; 2010a; 2011a; 2012a; 2013a; 2014a; 2015a; 2016a; 2017a; 2018a; 2019a; 2021a; 2022a) (see Appendix)			✓	✓
Sample student responses and scoring commentary (College Board 2007b; 2008b; 2009b; 2010b; 2011b; 2012b; 2013b; 2014b; 2015b; 2016b; 2017b; 2018b; 2019b; 2021b; 2022b)			✓	✓
Chief Reader reports on student performance (College Board 2007c; 2008c; 2009c; 2010c; 2011c; 2012c; 2013c; 2014c; 2015c; 2016c; 2017c; 2018c; 2019c; 2021c; 2022c)			✓	✓

For Research Questions 1 and 2, the author primarily examined *AP Japanese Language and Culture: Course and Exam Description* and *AP Japanese Language and Culture: Teacher's Guide* to find out how the College Board defines the target domain and the skills and knowledge required in the target domain. For Research Questions 3 and 4, the author

primarily examined the *AP Japanese Language and Culture Scoring Guidelines*, fifteen sets of the AP Japanese conversation task, sample student responses and scoring commentaries, and Chief Reader's reports on student performance to find out how the College Board operationalized the conversation task, whether the task is representative of the target domain, and how well the test task corresponds to the target domain task. More specific procedures for the task analysis are described in the next section.

2.2. Analysis of Task Characteristics

Characteristics of the AP Japanese conversation task were examined to address Research Questions 3 and 4. A variety of models for describing test characteristics have been proposed by various researchers (Bachman and Palmer 1996; Fulcher 2003; Weir 1993); however, they all state that no model is exhaustive or perfect for all test contexts, and thus recommend selecting aspects and categories that are salient to a test context of interest and useful for subsequent analysis. Following this recommendation, context and situation, input language, and scoring criteria were selected as aspects to focus on in this study, because the preliminary analysis of the test task suggested that these aspects would change each year for the conversation task. Other aspects, such as setting, structure/format, instructions, and time allotment, were not focused on primarily because the AP Japanese exam is already in operation and many of these are already established as common features across the AP world languages and culture exams and are consistent from year to year.

2.2.1. Context and Situation

Before starting the simulated conversation, the AP Japanese conversation task provides examinees with contextual information in English. The types of contexts represented in the task were analyzed based on this information. Specifically, the study looked for information about the interlocutor, setting, topic, familiarity, occasion or purpose, and register. The author categorized the occasion or purpose.

2.2.2. Input Language

The Japanese used in the conversation prompts was examined for its pragmatic characteristics. These were analyzed in terms of speech acts (Searle 1979) and discourse features. All fifteen sets of prompts (see Appendix) were coded by the author and the second coder. When there was a disagreement, the coders discussed the rationale for the codes and

resolved the disagreement. When the agreement between the author and the second coder was not high, the third coder was involved. All coders were native Japanese speakers and were trained in the relevant concepts, categories, and coding procedures.

2.2.2.1. *Speech Acts*

There are three major types of speech acts that are relevant to this study: directive, commissive, and expressive. Directive speech acts generally prompt the listener to react in some way, such as asking information questions, making requests for information or action, and extending invitations. An information question is defined in this study as a speech act that expects examinees to verbally provide information related to facts, descriptions, opinions, preferences, reasons, and so on. A request for information also expects examinees to provide information, but in the form of a request, such as *~ oshiete kudasai* “Could you tell me ~?” In commissive speech acts, speakers commit themselves to some future action, such as stating a plan, *~tai to omoimasu* “(I) would like to do ~.” Expressive speech acts include greeting, thanking, expressing feelings, and encouraging.

On the AP Japanese exam, each conversation prompt contains a few utterances; therefore, one prompt may consist of multiple, differing speech acts. In real life, speakers more likely react to each speech act as it occurs (Sacks et al. 1974; Tanaka 1999). However, this is not possible with the current design of the conversation task. In a context where multiple speech acts are presented at once and examinees have a time limit for responding, it seems natural for examinees to think that some speech acts are more worthy or urgent to respond to than others. Thus, this study coded the most salient speech act when multiple speech acts were observed in one prompt. For example, Form 2016 Prompt 2 (see Appendix), *Arigatoogozaimasu dooshite kono nihongo kyanpu ni hairitai to omotta n desu ka* “Thank you. Why do you want to come to this Japanese camp?,” was coded as a directive information question. Although this prompt consists of a thank you and a question, the thank you was used here to close the interactional sequence initiated in Prompt 1, *Hajimemashite nihongo kyanpu no Suzuki desu mazu hajimeni kantanni jikoshookai o shite kudasai* “Nice to meet you. (I am) Suzuki from the Japanese camp. First, please briefly introduce yourself.” Therefore, directive information question was judged to be the most salient speech act in this prompt. Coding of speech acts was done based on the pragmatic function and meaning of the utterance. In other

words, this study does not consider speech acts to be inherent to grammatical forms.

The present study analyzed the utterances by the speaker of the prompts without considering how examinees reacted to them. This is a different approach from the “next-turn proof procedure” (Hutchby and Wooffitt 1998, 15) used in Conversation Analysis, which aims to understand the meaning or function of an utterance by examining how the next speaker responds to it. However, it is debatable whether hearers or analysts can identify a speaker’s intentions with certainty (Streeck 1980) since “one and the same utterance may constitute the performance of several different illocutionary acts” (Searle 1969, 70). Therefore, in the present study, when the coders identified two equally salient speech acts, they were combined to form a single category rather than two separate categories. Form 2008 Prompt 1 is an example of such a prompt: *Konnichiwa Misaki desu senshuu kita bakaride gakko no kurasu no koto toka yoku wakaranai n desu iroiro oshiete kudasai* “Hello. (I) am Misaki. (I) just came (here) last week and (I) do not know much about classes at school or other things. Please advise me on many things.” The prompt ends with the *-te kudasai* “please do ~” structure, which is typically associated with a directive request, but the content of the request is vague. Considering that this utterance was the very first prompt that examinees heard, it may also function as a greeting. As a result, “greeting/request” was added as a new category. Two coders applied one speech act code to each prompt. The inter-coder agreement, defined as the agreement rate across all sixty prompts, was 96.9%. The numbers of different types of speech acts were tallied and compared across test forms.

2.2.2.2. Discourse Features

The conversation prompts were also analyzed in terms of the sequential organization of talk, such as the opening and (pre-) closing of talk, topic shifts and speech acts, and acknowledgment the other speaker’s response (Schegloff 1982). When these moves occur in real-life interaction, they are often accompanied by certain types of linguistic resources, including formulaic expressions and discourse markers, to achieve coherence (Tanaka 1999; Tateyama 2000). Therefore, this study examined what types of discourse moves related to sequential organization occurred in the conversation prompts and whether they were accompanied by linguistic resources that are commonly observed in spoken interaction. Two coders examined the prompts, and a total of 291 discourse moves and

accompanying linguistic resources were identified by at least one rater. The inter-coder agreement was 89.7%.

2.2.3. Scoring Criteria

Scoring criteria are often examined at the evaluation inference stage within the argument-based approach to validity. At that stage, test developers typically aim to confirm that the established rating criteria and scale function as intended. However, the present study examines the scoring criteria as a resource to identify the critical language skills and knowledge specified by the College Board for the conversation task, as well as to assess their relevance and representativeness in the target domain. Thus, the scoring criteria were examined as part of the domain definition inference in this study.

To evaluate how well the scoring criteria reflect the characteristics of language use in the target domain, the scoring guidelines, sample student responses and commentaries, and Chief Reader reports on student performance listed in Table 3 were examined. Preliminary analysis of these documents revealed that thoroughness, elaboration, and detail were emphasized as key features of task completion, regardless of prompt type. In other words, the test developers seem to expect substantive utterances that are typically produced in extended turns at talk (Schegloff 2007). Therefore, the present study specifically explored whether elaboration and detail were appropriate or expected for each prompt in a given context.

Specifically, the prompts were coded into one of the following three categories: (a) a detailed and elaborate response is hardly expected; (b) a detailed and elaborate response may be given but is not commonly expected; and (c) a detailed and elaborate response is highly expected. The author initially planned to report the number of prompts for each category based on agreement between two coders. However, the exact agreement between the two coders was 66.7%. A third trained coder was then asked to code the same data independently. Inter-coder reliability was assessed using Krippendorff's alpha (Marzi et al. 2024) for ordinal data with more than two coders. The resulting value was 0.554 (95% CI: [0.393, 0.676]), indicating moderate but not strong agreement among the coders. In fact, only 41.6% of items had full agreement across all three coders, and 56.7% showed partial agreement between any pair of coders. These results suggest that the expectation of thoroughness, elaboration, and detail varies even among native speakers. Therefore, the present study instead reports when the coders agreed or disagreed and discusses the implications for the

relevance and representativeness of the critical skills assessed in the test task in relation to the target domain.

3. Results

3.1. Target Domain

Two documents (i.e., the *AP Japanese Language and Culture: Course and Exam Description* and the *AP Japanese Language and Culture: Teacher's Guide*) published by the College Board were the primary sources of information for determining how the College Board defines the target domain of the AP Japanese exam. The College Board states that the AP Japanese exam is intended to be a criterion-referenced achievement test (2020, 2), which evaluates examinees' performance against a specific set of criteria, rather than comparing it to the performance of other students. The exam is designed for high school students who have completed the AP Japanese Language and Culture course, which is designed to be the equivalent of 300 hours of introductory college-level Japanese language instruction (Tohsaku 2009, 4). Based on the unpublished survey of college curricula, the College Board (Baum 2007b; College Board 2020; Tohsaku 2009) broadly defines the target domain as real-life situations in which Japanese language learners with proficiency levels ranging from the Intermediate Low to Intermediate Mid on the ACTFL Proficiency Guidelines (2012) come into contact with native speakers of Japanese.

The College Board organizes the target domain by modes of communication—interpersonal, interpretive, and presentational modes based on the *World-Readiness Standards for Learning Languages* (National Standard Collaborative Board 2015). These standards explain that interpersonal communication is two-way direct communication “characterized by active negotiation of meaning among individuals” (32) and that it “focuses on exchanging information, reactions, feelings, and opinions by negotiating and clarifying meaning” (47). Drawing on this framework, the College Board (2020) operationalizes spoken interpersonal communication with the conversation task, which requires the integration of listening and speaking skills. However, no further description of the target domain of the conversation task is provided.

3.2. Critical Skills and Knowledge in the Target Domain

Three documents (i.e., the *AP Japanese Language and Culture: Course and Exam Description*, and the *AP Japanese Language and Culture: Teacher's Guide*, and the scoring guidelines) were the primary sources of

information for finding out how the College Board defines the language skills and knowledge needed in the target domain. Following Bachman's (2002) recommendation, information related to the critical skills and knowledge was examined in terms of construct and task.

In terms of the construct, the College Board's documents make no direct reference to any specific theoretical model of language abilities. However, documents such as College Board (2020) and Tohsaku (2009) emphasize the importance of cultural knowledge, communication strategies, and language competence related to vocabulary, grammar, syntax, pronunciation, intonation, organization, cohesion, and register. These documents also state that the AP Japanese exam was developed based on the *Standards for Foreign Language Learning in the 21st Century*, which were later updated as the *World-Readiness Standards for Learning Languages*. The World-Readiness Standards state that knowing a language means being able to “carry out a large variety of tasks in the language” by “controlling the linguistic system (the syntax, morphology, phonology, semantics, lexis) of a language” and by “access[ing] the pragmatic, textual, and sociolinguistic aspects of language, including how to use the language to achieve communicative goals in ways that are appropriate to a particular cultural context” (National Standard Collaborative Board 2015, 32), referencing Canale and Swain (1980), Bialystok (1981), Savignon (1983), Hymes (1985), and Bachman (1990). These documents suggest that the AP Japanese exam construct is broadly tied to theories of communicative competence, although it is not clear which model of communicative competence the AP Japanese exam draws upon.

In terms of task, the author found the skills and objectives outlined for the AP Japanese course (College Board 2020, 19–21). They are generally organized according to the modes of communication, and those identified for spoken interpersonal communication are presented in Table 4.

Some of these skills are related to grammatical and textual competence (e.g., use of syntax, transitional expressions, and cohesive devices). Others are related to pragmatic and sociocultural competence (e.g., initiate, maintain, and close spoken exchanges, comprehend an interlocutor's message, provide and obtain information, use appropriate register and greeting). Still others are related to strategic competence (e.g., circumlocution, clarification, and self-correction).

Table 4: Critical Skills Identified for Spoken Interpersonal Communication in the AP Japanese Course (College Board 2020, 20)

-
1. Understand and apply appropriate communication strategies in interpersonal speaking.
 - a) Initiate, maintain, and close spoken exchanges.
 - b) Comprehend an interlocutor’s message in spoken exchanges.
 - c) Provide and obtain relevant information in spoken exchanges.
 - d) Use appropriate register and greeting for the intended target culture audience in spoken exchanges.
 - e) Use pronunciation and pitch that are comprehensible when interacting with speakers of the target language in spoken exchanges.
 - f) Use effective intonation patterns, pacing, and delivery when interacting with speakers of the target language in spoken exchanges.
 - g) Use communication strategies such as circumlocution, requesting clarification, asking for repetition, and paraphrasing to maintain spoken exchanges.

 2. Understand and apply appropriate and varied syntactical expressions in interpersonal speaking.
 - a) Use a variety of grammar and syntax in spoken exchanges.
 - b) Use transitional expressions and cohesive devices in spoken exchanges.
 - c) Comprehend and produce simple, compound, and complex sentences in a variety of time frames in spoken exchanges.
 - d) Monitor language production; recognize errors and attempt self-correction in spoken exchanges.
-

Based on these identified skills, the College Board operationalizes the skills to be assessed in the conversation task in the form of the scoring guidelines (College Board 2023). Examinee’s performance is assessed on the ability to “thoroughly address all aspects of the prompt with attention to detail,” “connect sentences smoothly,” “demonstrate natural pace and intonation, accurate pronunciation, and consistent use of register appropriate to the situation,” “use varied and appropriate vocabulary, including idiomatic language,” and “use a wide range of grammatical structures” (College Board 2020, 168). In particular, the task completion criteria emphasize thoroughness, elaboration, and detail as key features for distinguishing different levels of language proficiency. This suggests that the College Board views the ability to produce substantive utterances as critical regardless of the language function of the prompt.

Communicative language use involves not only language ability, but

also topical knowledge (Bachman and Palmer 1996). The College Board (2020, 14) suggests six major themes to be covered in the AP Japanese course: families and communities, personal and public identities, beauty and aesthetics, science and technology, contemporary life, and global challenges. Each of these themes has several subthemes as shown in Table 5.

Table 5: Themes to be covered in the AP Japanese Course (College Board 2020, 14)

-
1. Families and communities
 - a) Family structure/roles within families
 - b) Social customs, traditions, values, rites of passage
 - c) Urban and rural communities and their development
 - d) Relationships, friendship and social networking
 - e) Perspectives of age, ethnicity, class, religion
 - f) Citizenship, social welfare, and social justice

 2. Personal and public identities
 - a) National and ethnic identities
 - b) Heroes and national figures
 - c) Gender and gender roles
 - d) Personal interests
 - e) Self-image and the role of the individual in society
 - f) National and global responsibilities

 3. Beauty and aesthetics
 - a) Architecture
 - b) Defining beauty: perspectives of beauty in Japanese culture
 - c) Traditional and contemporary arts and music
 - d) Literature
 - e) Fashion and design
 - f) Visual and performing arts, film

 4. Science and technology
 - a) Technology and access to technology
 - b) Health and wellness
 - c) Climate and the physical world
 - d) Innovations and inventions
 - e) Effects of science and technology on self and society
 - f) Science and ethics

 5. Contemporary life
 - a) Education and careers

- b) Lifestyles and pop culture
 - c) Entertainment, sports, and leisure
 - d) Travel, transportation, and tourism
 - e) Holidays and celebrations
 - f) Food, nutrition, and meal-taking culture
6. Global challenges
- a) Population and demographics
 - b) Obtaining, managing, and protecting resources
 - c) Endangered species
 - d) Environmental issues
 - e) Economic trends
 - f) Japan's role in international trade
-

3.3. Context and Situation

Table 6 shows a summary of the contexts and situations used for the conversation task. Information about the interlocutor, setting, and topic was provided in the pre-task instructions in English, and information about familiarity, occasion, and register was provided through the Japanese prompts. The task included interlocutors with various statuses and occupations. In some test forms, the descriptions of the interlocutor (e.g., the director of an after-school program at a Japanese elementary school), together with the register, provided more clues about the social distance and power difference between the interlocutor and the examinee than those in other test forms (e.g., a Japanese student). Information about the setting was not provided, except for the airport (Form 2014) and the Japan Fair (Form 2022).

The task covered various topics in daily life. However, some of the topic descriptions were contextually ambiguous and insufficient, leading to different interpretations of the situation among examinees. This may raise a validity issue because examinees are expected to use their imagination to compensate for the ambiguous and insufficient context, suggesting that the task inadvertently requires imagination of the context, although it is construct-irrelevant. In addition, the resulting different interpretations of the situation may also affect the difficulty of the task and quantity and quality of responses, and thus their scores, although this requires empirical confirmation.

The Chief Reader reports documented many examples that illustrate this point. For instance, in 2010, some examinees treated an upcoming party as an event that they had already previously agreed to attend, while

others treated it as an event that they might or might not attend. However, Form 2010 Prompt 4 (*Aa naruhodo itsumo arigatoo ja mata paatii no tesudai yoroshiku* “Oh, I see. Thank you as always. Okay, then I will ask you for help later. Thanks!”) clearly assumes the former interpretation. Then, for those examinees who would say that they might not attend the party, this prompt may be more difficult because they have to suddenly change their stance. To make the interpretation of the context similar across examinees, the instructions should state that the examinees are having a conversation about an upcoming party that they have agreed to co-host.

In 2016, some examinees responded as campers, while others responded as camp counselors. In this case, the former interpretation may make the task easier, as the role of camper is likely to be more easily associated with high school students, especially those with less work experience, than a camp counselor. If the test developers intended examinees to play the role of a camper, the instructions should state that they are having a conversation about a Japanese language camp that they are attending as a camper.

In 2018, some examinees interpreted a situation as them being tutored by the interlocutor, while others interpreted it as them tutoring someone else along with the interlocutor. Similar to 2016, the former interpretation may make the task easier, as it is easier for high school students to associate with the role of being tutored. If the test developers intended examinees to play the role of a student being tutored, the instructions should state that they are having a conversation about the Japanese tutoring they will receive.

These reports suggest that the task needs to provide clearer contextual descriptions regarding identity of the interlocutors and their relationship to the examinees, the purpose of the interaction, and the presuppositions about the examinees’ actions and knowledge prior to the interaction in order to make the interpretation of the context more consistent across examinees and to minimize the unintended effects of differing levels of imagination of the context.

Table 6: An overview of the conversation task

Form	Interlocutor	Setting	Topic	Relationship	Occasion/Purpose	Register
2007	Youth center representative		Student activities	First-time	Survey interview	Formal
2008	Study-abroad program coordinator		Staying with a family in Japan	Not clear	Survey interview	Formal
2009	New foreign exchange student		School life	First-time	Everyday interaction	Formal
2010	Friend		Upcoming party	Acquainted	Everyday interaction	Informal
2011	New neighbor		Recycling garbage	First-time	Everyday interaction	Formal
2012	Friend		Club activities	Acquainted	Everyday interaction	Informal
2013	Japanese student		Working	Not clear	Survey interview	Formal
2014	Study aboard program assistant	Airport	Plans for the afternoon	First-time	Everyday interaction	Formal
2015	Japanese student		Upcoming event	Acquainted	Everyday interaction	Informal
2016	Camp director		Japanese language camp	First-time	Selection interview	Formal
2017	President of an environmental club		Global warming	Not clear	Survey interview	Formal
2018	Japanese-speaking student		Tutoring in Japanese	First-time	Survey interview	Formal
2019	Homestay coordinator		Upcoming homestay	First-time	Survey interview	Formal
2021	Director of an after-school program at a Japanese elementary school	(Phone)	Volunteering in the program	Acquainted	Selection interview	Formal
2022	Booth staff	Japan Fair	Japanese technology	First-time	Survey interview	Formal

Note: Form = test form labeled by the year it was administered; Relationship = the relationship between the interlocutor and examinee; First time = a first-time encounter; Acquainted = known relationship. A survey interview is a meeting at which information and/or opinion is obtained from examinees by the interlocutor. A selection interview is a formal consultation to evaluate qualifications of the prospective candidate. Everyday interaction is an informal interaction on common or ordinary topics in daily life such as daily routine, school, the weekend, etc. Register = speech style; Informal = a plain speech style.

As for occasions, nine out of the fifteen forms (60.0%) were interviews, while six (40.0%) were everyday interactions. Interviews tended to be presented as first-time encounters with a formal speech style, although the relationship was not clear for Forms 2008, 2013, and 2017. Everyday interactions were equally divided between the first-time encounters with a formal speech style and familiar encounters with an informal speech style.

3.4. Input Language

3.4.1. *Speech Acts*

The present study identified three major types of speech acts (Searle 1979) in the conversation prompts, as shown in Table 7: directive, commissive, and expressive. Information-seeking speech acts, consisting of information questions (60.0%) and requests for information (6.7%), were by far the most common speech acts. Information-seeking speech acts were observed in all positions in the discourse (i.e., the first prompt, second prompt, etc.), although they occurred most frequently in the second and third prompt. Requests for action (13.3%) occasionally occurred in the third and/or fourth prompt. An invitation occurred once (1.7%) in the fourth prompt. Expressive speech acts were observed occasionally. All greetings (15.0%) appeared in the first prompt; expressing feelings (1.7%) and encouraging (1.7%) appeared in the last prompt. Overall, information-seeking speech acts dominated the conversation prompts. In particular, Form 2021 consisted entirely of information questions. Most test forms consisted of two to three information-seeking speech acts and one or two other speech acts.

Table 7: Frequency of speech acts in the conversation task

Form	Interlocutor	Occasion/Purpose	Directive			Expressive + Directive		Commissive + Expressive		Expressive			
			Information question	Request for information	Request for an action	Invitation	Greeting + Request	For an action	Statement of a plan + Greeting	Greeting	Greeting + Thanking	Expressing feelings	Encouragement
2007	Youth center representative	Survey interview	3							1			
2008	Study-abroad program coordinator	Survey interview	3										1
2009	New foreign exchange student	Everyday interaction	1	1		1							
2010	Friend	Everyday interaction	2	1	1								
2011	New neighbor	Everyday interaction	2	1	1								
2012	Friend	Everyday interaction	3	1	1								
2013	Japanese student	Survey interview	2	1	1								
2014	Study aboard program assistant	Everyday interaction	3									1	
2015	Japanese student	Everyday interaction	2		1								1
2016	Camp director	Selection interview	2	2									

Form	Interlocutor	Occasion/Purpose	Directive		Expressive + Directive		Commissive + Expressive		Expressive			
			Information question	Request for information	Request for an action	Invitation	Greeting + Request for an action	Statement of a plan + Greeting	Greeting	Greeting + Thanking	Expressing feelings	Encouragement
2017	President of an environmental club	Survey interview	3				1					
2018	Japanese-speaking student	Survey interview	2	1					1			
2019	Homestay coordinator	Survey interview	2	1					1			
2021	After-school program director	Selection interview	4									
2022	Booth staff	Survey interview	2	1					1			
Total			36	4	8	1	1	2	5	1	1	1
%			60.0	6.7	13.3	1.7	1.7	3.3	8.3	1.7	1.7	1.7

Note: Form refers to the test form, which is labeled by the year it was administered.

3.4.2. Discourse Features

The analysis of the conversation prompts identified five major interactional moves in the conversation task: (a) opening of the interaction, (b) shifting topics and speech acts, (c) developing a topic in hand with a follow-up question or comment, (d) showing acknowledgment and assessment of an examinee's utterance, and (e) possible pre-closing of the interaction.

Out of the fifteen forms, eleven included an interaction opener such as a greeting (e.g., *konnichiwa* "Hello," *hajimemashite* "How do you do?," and *yoroshiku onegaishimasu* "Nice to meet you.") in the first prompt. A self-introduction (e.g., *~ desu* "I am ~") and the introduction of a general topic or reasons for the interaction (e.g., *~ nitsuite okikishitai to omoimasu* "I would like to ask you about ~") were often observed with greetings. The remaining four test forms did not include a greeting or interjection to get examinees' attention. They instead started with the introduction of a general topic using conjunctive particles such as *~ kedo* "but" and *~ n da kedo* "but," then they asked the first question, immediately after the preface within the same prompt (see Form 2008, Prompt 1, in Appendix as an example). Among these four forms without greetings, three forms (2010, 2012, and 2015) were presented as interactions between two friends or acquaintances. The context for one form (2008) was not clear. The lack of a greeting may not necessarily be an issue for interactions between friends, especially if they are already in the same space prior to the interaction. Nevertheless, the issue is that none of these four forms started with any method of verbally getting attention (e.g., *nee* "hey," *anoo* "umm") from the other speaker. In a computer-based test environment where non-verbal contextualization cues are not available, it would be more natural for the beginnings of interactions to be verbally marked.

Numerous shifts in topics and speech acts were observed among the conversation prompts. Here a shift in speech act means a shift in discourse orientation and organization, such as from greeting to asking questions, from asking questions to a request, and from asking questions to a pre-closing sequence. The majority of shifts in topics and speech acts were marked by discourse markers such as conjunctions (e.g., *ja* "then" and *sorekara* "and then"), adverbs (e.g., *mazu* "first"), fillers (e.g., *anoo* "umm"), and conjunctive particles for prefaces (e.g., *~ kedo* "but" and *~ n desu ga* "but"). Meanwhile, seven shifts occurred without any discourse markers (see Form 2013, Prompts 3 and 4, in Appendix as examples).

Acknowledgments (e.g., *soo desu ka* "I see," and *wakarimashita* "I

understand”) and assessments (e.g., *ii ne* “(That sounds) good”) of an examinee’s utterance were observed but not consistently. In real-life interactions, acknowledgments and assessments are commonly used to close question-answer and request sequences (Schegloff 2007). In fact, acknowledgment and assessment in the third-turn position is a minimum conversational obligation since it shows that a questioner or requester has heard what was being said to them (Clark 1996). Therefore, the absence of the third-turn utterance is a violation of this tacit rule of interaction. Among thirty-seven question-answer or request sequences, twenty sequences (54.0%) were missing the third-turn part (see Form 2008, Prompts 1–2, in Appendix as examples). Five test forms (2012, 2013, 2014, 2019, 2022) were completely missing an acknowledgment and/or assessment. Furthermore, there were four prompts that included neither the third-turn part nor a discourse marker to indicate a shift in the discourse; this does not reflect the features of real-life spoken interaction. The lack of third-turn parts and discourse markers needs to be addressed to increase the authenticity of the test discourse.

Topic development was observed five times in the fifteen test forms. Four of the instances were follow-up *dooshite* “why” questions asking for elaboration (see Form 2012, Prompt 3, in the Appendix as an example); one instance (Form 2015, Prompt 4) was a statement of feelings towards the general topic of the conversation. Among the four elaboration-seeking *why* questions, two appeared in an interview context, and the other two appeared in a casual conversation with friends (see Form 2010, Prompt 3, in Appendix as an example). Asking explicitly for reasons or an explanation may be common or somewhat expected in an interview context; however, *why* questions are not often used in everyday interaction in Japanese (Mori 2005). Usually reasoning and explanations are instead voluntarily supplied by speakers or induced by a listener’s response token such as *e?* “huh?” or *honto?* “really?” (Ford and Mori 1994; Mori 1999). Moreover, native speakers of Japanese often avoid using *why* questions because they tend to carry an accusatory or disagreeing tone (Mori 2005). However, the two *why* questions in the casual conversations on the AP Japanese exam did not carry such a tone; they were instead neutral. These elaboration-seeking *why* questions in the casual conversations on the exam seem very similar to *why* questions that are often used by classroom teachers as a pedagogical tool to encourage more elaborated thinking and production (McCormick and Donato 2000). In light of previous studies, it seems that *why* questions are overused when the task is presented as a

casual everyday interaction; the way these questions are used corresponds more to the characteristics of institutional interactions (e.g., classroom interactions) rather than to non-institutional interactions. The authenticity of the test discourse in the everyday interactional setting could be improved if the topic development were done without the use of explicit elaboration-seeking *why* questions.

The last major discourse move observed in the conversation prompts was the pre-closing of the interaction. A pre-closing sequence usually refers to a stage where speakers confirm that there is no more to talk about prior to a terminal exchange (Iwasaki 1997; Schegloff and Sacks 1973). In the present study, however, pre-closing has a broader scope and refers to steering the direction of the interaction towards closing or signaling that the closing is approaching using a variety of strategies, including wrapping up the conversation and summarizing what has been discussed. In the pre-closing stage, speakers do not usually introduce a new topic into a conversation (Schegloff and Sacks 1973); therefore, this notion was also used to identify pre-closings. Out of fifteen test forms, two test forms (2008 and 2010) included pre-closing utterances. They were marked with the conjunction *ja* “then,” signaling a shift in discourse orientation, and they summarized or wrapped up the conversation. A greater use of signs of possible pre-closings would be helpful to make the test discourse more authentic.

3.5. Expectation of Response Elaboration

Table 8 summarizes the three coders’ judgments about the expectation of response elaboration. The mean (*M*) indicates the average ordinal score across all three coders. Higher means indicate a stronger expectation of response elaboration. As explained in the methodology section, out of sixty prompts, only twenty-five prompts (41.7%) had full agreement among all three coders. Meanwhile, thirty-four prompts (56.7%) had partial agreement from two coders. One prompt (Form 2022, Prompt 2) received different ratings from all three coders. These results suggest that the expectation of a detailed and elaborate response varies even among native speakers. Therefore, it is inadvisable to assume that this criterion uniformly fits all types of prompts.

Table 8: Expectation of response elaboration

Form	Prompt	Speech Act	M	
Full inter-coder agreement 3/3				
2007	3	それはどうしてですか。	Directive Info. Q	3.00
2017	3	どうしてそう思うのですか。	Directive Info. Q	3.00
2021	2	小学校でボランティアをしたらしいんですけど、それはどうしてですか。	Directive Info. Q	3.00
2010	3	あ、そう。どうしてそう思うのか、もうちょっと詳しく話してくれる？	Directive Req. for Info.	3.00
2018	2	まず、なぜ日本語を勉強しているか教えて下さい。	Directive Req. for Info.	3.00
2019	2	どうしてホームステイをしたのか教えてください。	Directive Req. for Info.	3.00
2013	3	高校生がアルバイトをするのとどちらがいいと思いますか。	Directive Info. Q	2.00
2017	2	まず、車ですが、ガソリンを使う車と電気を使う車とどちらがいいと思いますか。	Directive Info. Q	2.00
2012	1	クラブのボランティア活動のことなんだけど、公園のそうじをするのと小学生の勉強を手伝ってあげるのと、どっちにする？	Directive Info. Q	2.00
2021	3	なるほど、このプログラムには小学校1年生から6年生までの子供たちがいますが、何年生のグループがいいですか。	Directive Info. Q	2.00
2014	3	今日はお昼ご飯のあと、少し時間があるんですが、何がしたいですか。	Directive Info. Q	2.00
2015	2	あ、そう。それで、コンサートの後、どこに食べに行こうか。	Directive Info. Q	2.00
2014	1	日本へようこそ。プログラムアシスタントの山下アキです。初めまして。	Expressive Greeting	1.00
2022	1	車のブースによろこそ！私がこのブースをご案内します。	Expressive Greeting	1.00
2011	1	あ、初めまして。日本から隣に引越してきた中村です。どうぞよろしくお願ひします。	Expressive Greeting	1.00
2018	1	初めまして。今度日本語の勉強のお手伝いをすることになった高木ゆみと申します。よろしくお願ひします。	Expressive Greeting	1.00
2019	1	はじめまして、ホームステイ・コーディネーターの中村です。よろしくお願ひします。	Expressive Greeting	1.00
2007	1	初めまして。日本ユースセンターの森山です。先日はアンケートにお答えくださいましてありがとうございます。	Expressive Greeting + Thanking	1.00
2008	4	ありがとうございました。じゃ、私もなるべくいいホストファミリーを見つけてますから、あなたも日本語の勉強頑張ってってくださいね。	Expressive Encouraging	1.00
2013	1	今日は高校生のアルバイトについてお聞きしたいと思ひます。よろしくお願ひします。	Commissive Plan + Expressive Greeting	1.00
2017	1	地球温暖化についてお聞きしたいんですが、よろしくお願ひします。	Commissive Plan + Expressive Greeting	1.00

2009	1	こんにちは。美咲です。先週来たばかりで、学校のクラスの事とかよく分からないんです。色々教えてください。	Expressive Greeting + Directive Req. for Action	1.00
2022	4	後で、もっと意見を聞きたいんですが、オンラインのアンケートか、紙のアンケートに答えてください。	Directive Req. for Action	1.00
2010	4	ああ、なるほど。いつもありがとう。じゃまたパーティーの手伝いよろしく。	Directive Req. for Action	1.00
2021	1	ああ、もしもし、お電話待っていました。こちらはひどい雨なんですけど、そちらのお天気はどうですか。	Directive Info. Q	1.00
Partial inter-coder agreement 2/3				
2016	2	ありがとうございます。どうしてこの日本語キャンプに入りたいと思いますか。	Directive Info. Q	2.67
2012	3	その曜日と時間がいいのはどうして？	Directive Info. Q	2.67
2011	2	あのう、明日はゴミの日ですよね。この町ではゴミとカリサイクルとかどうしているんですか。	Directive Info. Q	2.67
2018	3	そうですね。日本語の勉強の中でもっと上手になりたいことは何ですか。	Directive Info. Q	2.67
2007	4	あ、そうですね。それから、今度うちのセンターで高校生の活動について話をしてもらえるというんですが、どんなお話でしょうか。	Directive Info. Q	2.33
2008	1	日本のホストファミリーのことでですけど、どんな家族がいいですか。	Directive Info. Q	2.33
2008	3	そうですね。ホストファミリーとどんなことがしたいですか。	Directive Info. Q	2.33
2010	2	あ、いいね。クラスのみんなが楽しめるパーティーにしたいんだけど、何をしたらいいかな。	Directive Info. Q	2.33
2013	2	あなたの高校の生徒はどんなアルバイトをしていますか。	Directive Info. Q	2.33
2014	4	ところで、日本語が上手ですね。どのように日本語を勉強しているんですか。	Directive Info. Q	2.33
2016	3	ああそうですね。では、このキャンプではどんなことをしたいですか。	Directive Info. Q	2.33
2009	2	あのう今学期、数学が科学。を取らなくちゃいけないんですけどどっちの方が難しいと思いますか。	Directive Info. Q	2.33
2013	4	アルバイトをしたい人に何かアドバイスをしてください。	Directive Req. for Action	2.33
2016	1	はじめまして。日本語キャンプの鈴木です。まずはじめに簡単に自己紹介をしてください。	Directive Req. for Action	2.33
2009	3	そうですね。それから、私友達をいっぱい作りたいと思っていますんですけど、どんなクラブ活動があるか教えてください。	Directive Req. for Info.	1.67

Of the sixty prompts, six were rated by all coders as “detail and elaboration are highly expected” ($M = 3.00$); these prompts all asked for reasons directly. Another set of six prompts were rated by all coders as “detail and elaboration may be given, but are not common” ($M = 2.00$). Of these, four prompts asked the test-taker to make a choice (i.e., which); two were information questions about description and location (i.e., what and where). There were thirteen prompts that were rated by all coders as “detail and elaboration are hardly expected” ($M = 1.00$); they all appeared as either the first or the last prompt of the task. Of these, ten prompts carried an expressive speech act function, such as greeting or encouragement; two were requests for action.

Similar tendencies were observed when two out of three coders agreed on their ratings. The prompts directly asking for reasons had a higher expectation of elaboration ($M = 2.67$). The prompts asking for descriptions generally had a moderate expectation of elaboration, ranging from 2.67 to 2.33. The prompts asking the test-taker to make a choice (e.g., which, when, how many times, and which day of the week) had a relatively low expectation, ranging from 2.33 to 1.33. Prompts that expected a simple yes/no response (e.g., yes/no questions, requests for action, and invitations) generally had low expectations, ranging from 1.67 to 1.33. One expressive prompt also had a low expectation ($M = 1.33$).

These results suggest that detailed and elaborate responses were clearly expected for only a handful of the information-seeking prompts. Many of the information-seeking prompts were rated as having a moderate expectation of elaboration. Furthermore, all the expressive prompts were rated as having a low expectation for elaboration. These results illustrate that there is a mismatch between the critical skills identified in the target domain and the critical skills required in the test task.

4. Discussion

4.1. RQ 1: What is the target domain of the AP Japanese exam? Is it clearly identified? (Assumption 1)

Support for Assumption 1 comes from the results of the document analysis. The primary sources of information used to examine RQ1 (Assumption 1) were the *AP Japanese Language and Culture: Course and Exam Description* and the *AP Japanese Language and Culture: Teacher’s Guide* published by the College Board. According to these documents, the target domain of the AP Japanese exam is real-life situations in which Japanese language learners with proficiency levels ranging from Intermediate Low to Intermediate Mid on the ACTFL Proficiency Guidelines (2012) interact

with native speakers of Japanese. For the conversation task in particular, the target domain is spoken interpersonal communication in real-life situations. Although this definition is relatively broad, it seems reasonable as a general domain definition, given the variety of knowledge and skills required in introductory college-level courses. As Kane et al. (1999, 8) note, “[t]he target domain should be defined broadly enough to represent the kinds of performances that are of interest to parents, teachers, students, and state legislators (if appropriate), even if this leads to some fuzziness in the boundary of the domain.” Moreover, the level of specificity in the domain definition of the AP Japanese exam seems similar to that of a well-researched test such as the TOEFL cited in the literature review. Thus, the author argues that the target domain for the AP Japanese exam is reasonably well defined. In other words, there is reasonable support for Assumption 1.

4.2. RQ 2: What are the critical skills and knowledge in the target domain? Are they clearly identified? (Assumption 2)

Support for Assumption 2 comes from the review of the *AP Japanese Language and Culture: Course and Exam Description*, the *AP Japanese Language and Culture: Teacher’s Guide*, and the scoring guidelines. The document analysis revealed that the AP Japanese course and exam aim for a construct- and task-based design. The critical elements for the construct were identified somewhat in relation to theories of communicative competence. Critical language skills in the TLU tasks were identified in terms of modes of communication. Critical knowledge in the target domain was identified in terms of major themes and subthemes, as shown in Table 5. Specifically for the domain of spoken interpersonal communication, several specific skills were listed as learning objectives for the AP Japanese course, as shown in Table 4. This list suggests that not only skills related to grammatical and textual competence, but also those related to pragmatic, sociocultural, and strategic competence were identified as critical skills for the target domain. Based on this supporting evidence, the author argues that the critical skills and knowledge in the target domain of the AP Japanese exam is reasonably well defined. In other words, there is reasonable support for Assumption 2.

4.3. RQ 3: How is the conversation task operationalized? Is it representative of the target domain? (Assumption 3)

Assumption 3 primarily concerns the task design and characteristics that serve as a common framework across the different test forms. Support for

Assumption 3 comes from the author's analysis of the fifteen sets of the conversation task and the review of the scoring guidelines. The College Board operationalized the conversation task to represent authentic real-life tasks in the interpersonal speaking domain (Tohsaku 2009). Several pieces of evidence were found to support Assumption 3. In the conversation task, examinees first receive contextual information to frame the subsequent interaction. They then listen to a series of four related prompts and respond to them without any planning time. The interlocutors for this task were people of different social statuses and occupations whom examinees would possibly meet in real life. The overarching topics of the task varied within the six themes listed on the *AP Japanese Language and Culture: Course and Exam Description*. Regarding the input language, both formal and informal speech styles were observed in the fifteen forms. Most test forms included information-seeking speech acts and at least one other type of speech act, such as making a request for an action, inviting, greeting, encouraging, or expressing feelings. The results suggest that the conversation task reflects features of the real-life interpersonal speaking domain.

Despite the positive evidence, there were a few areas of concern. First, the College Board seems to use "conversation" in a very broad sense of spoken interaction and does not distinguish between different types of spoken interaction. This is evidenced by the fact that the occasion or purpose for 60.0% of the conversation task was an interview, while for 40.0% it was an everyday interaction. These results suggest that the test developers are treating interviews and everyday interaction as equivalent speech events on the AP Japanese exam. This is a concern, since previous studies of oral proficiency interviews (Johnson 2000; Johnson and Tyler 1998; Kasper and Ross 2007) have repeatedly pointed out that interviews are a distinctly different type of interaction from ordinary conversation in terms of turn-taking, topic nomination, and topic development, among other features. For example, interactional roles are fixed in interviews, while the roles can change spontaneously in ordinary conversation. Therefore, it would be uncommon for an interviewee to initiate an interaction or ask questions, although the initiation of interaction is one of the critical skills in the target domain. In addition, these two types of interaction do not share the same response expectations in real life. Thus, when used for assessment, they may affect examinees' performance and scores, and the interpretation of the scores, although this needs to be confirmed empirically. Therefore, the task definition should be clarified,

and the type of speech event should be set accordingly to increase the validity of the conversation task.

The second concern is how critical language skills in spoken interpersonal communication were operationalized in the test task. Grammatical and textual competence are clearly referenced in the scoring guidelines (College Board 2023). In contrast, it is not clear to what extent pragmatic competence (Bachman 1990) is considered in the test task, aside from pragmatic aspects such as register and style, which are briefly addressed in the scoring guidelines. The learning objectives for the AP Japanese course (see Table 4) list knowledge and skills related to the structure of talk—such as the opening and closing of an interaction and adjacency pairs (e.g., how to use or respond to formulaic expressions)—but criteria for these elements are not included in the scoring guidelines. This suggests that grammatical and textual competence (Bachman 1990) are emphasized much more than pragmatic competence in the test task, relative to the list of critical skills identified in the target domain shown in Table 4. In addition, asking for repetition (e.g., “Please repeat.”) receives a score of 0 if it is the only utterance in the response, despite being recognized as a communication strategy. In this test task, asking for repetition in the target language is not considered evidence of competence because the pre-recorded prompts in the computer-based testing environment can only accommodate one turn per prompt.

Based on the above evidence, the author argues that the conversation task design reflects many relevant features of the target domain; however, it does not fully represent the target domain in terms of critical skills. In other words, Assumption 3 is partially supported.

4.4. RQ 4: How well do the characteristics of the conversation task correspond with the characteristics of the target domain task? (Assumption 4)

Assumption 4 concerns how specific task characteristics were realized across different test forms. Support for Assumption 4 comes from the author’s analyses of the fifteen sets of the conversation task and the review of the scoring guidelines, the Chief Reader’s reports on student performance, and sample student responses and scoring commentary. The results suggest that the characteristics of the conversation task partially correspond to those of the target domain task. Similar to real-life interaction, most of the conversation test discourse had an opening such as a greeting. A variety of speech acts were included in the test discourse, such as asking for information, requesting information, requesting action,

inviting, greeting, thanking, expressing feelings, and encouraging. Furthermore, utterances acknowledging and assessing examinees' responses were also included, though only occasionally. Shifts in topics and speech acts were mostly marked with discourse markers.

Nevertheless, the conversation test task fell short in several respects. The contextual information provided in the test task was generally limited. The insufficient interactional context in some test forms was reported to lead to different task interpretations among examinees. The lack of contextual information is inevitable to some extent for a test task for the sake of practicality; however, it is essential for contextual information to be sufficiently clear to avoid different task interpretations that could contribute to response irregularity. Furthermore, elaboration-seeking *dooshite* 'why' questions were overused in the everyday interactional setting, but are infrequent in real-life everyday interaction (Mori 2005). Another issue is related to the general lack of consistency in spoken language features across test forms. For instance, 26.7% of the test forms did not include an opening utterance. One test form consisted entirely of information questions. Many of the question-response and request sequences lacked an acknowledgment and/or assessment as the third-turn part. Few test discourse interactions ended with a possible pre-closing utterance. The lack of consistency across multiple test forms should be addressed, as it may pose serious questions not only for the domain definition inference, but also for the generalizability inference. Prompts that elicit different types of responses should be included in a more principled and balanced manner across test forms.

The current scoring guidelines implicitly assume that substantive utterances are always supportive evidence for communicative competence. However, this expectation is overgeneralized in the context of the AP Japanese conversation task. The present study found that the expectation of substantive utterances may not be appropriate for many of the conversation prompts that were expressive, expected a simple yes/no response, or required examinees to make a choice. In other words, the current descriptors have a low correspondence with real-life response expectations.

For instance, the opening and closing of the test discourse often comprised phatic communion. According to Malinowski (Malinowski 1953 [originally published in 1923]), phatic communion is "a type of speech in which ties of union are created by a mere exchange of words. ... [L]anguage does not function here as a means of thought" (315). The main

goal of phatic communion is to create or confirm the ties and relationship between speakers and hearers (Suzuki 1981). Therefore, as long as a hearer provides some response to confirm the tie, the response meets the social expectation. In other words, thoroughness and elaboration of responses may be observed in phatic communion, but they are not the central concern. Moreover, phatic talk often occurs at the beginning and end of interactions (Laver 1975), so speakers tend to avoid lengthy utterances and resort to formulaic or other concise expressions (Schegloff 2007; Schegloff and Sacks 1973). Thus, expecting thorough and elaborate utterances for phatic communion does not reflect real-life expectations.

Furthermore, for utterances that expect a yes/no, agreement/disagreement, or acceptance/refusal response, elaboration and detail are mere options, especially when speakers accept the request, invitation, or assumption. Explanations or justifications of a response are typically expected when the speaker has a dispreferred response, such as rejecting a request or invitation (Schegloff 2007). Even then, a long and detailed excuse may not necessarily be desirable. Similarly, for questions about examinees' preferred time (and date) for an activity, examinees are expected to respond with their preference; however, they are not usually expected to state the reasoning behind the preference unless there is no time available.

These findings suggest that the appropriateness and completeness of responses should not be judged simply based on thoroughness, detail, and elaboration. Such characteristics are not typical of interactional turns as opposed to transactional turns (Brown and Yule 1983) that primarily focus on exchanging information and accomplishing tasks. To increase the correspondence between the test task and real-life tasks, the task definition should be clarified, as discussed earlier. If the College Board decides to assess communicative competence in everyday interaction, the scoring criteria should reflect its characteristics. Based on the above evidence, the author argues that the characteristics of the conversation task partially correspond to those of the target domain task, and thus Assumption 4 is partially supported.

One way to improve the correspondence between the test task and the real-life tasks is to refer to empirical studies on real-life spoken language, discourse, and interaction (e.g., Ford and Mori 1994; Mori 1999; Mori 2005; Suzuki 1981) during task development (e.g., Biber et al. 2004; Kyle et al. 2021; Kyle et al. 2022; Nishizawa 2024). Another way may be to prepare different sets of scoring criteria to reflect the nature of each prompt

in a given context; however, this solution is less practical since it would be labor intensive. In addition, it would be difficult to standardize response expectations for some types of prompts. In fact, inter-coder agreement for the response expectation was low in this study. A more practical solution may instead be to revise the scoring procedure to treat each series of responses to a set of four prompts as one language sample instead of four individual language samples. This may alleviate the impact of different prompt compositions across test forms and the mismatch between the nature of individual prompts and the response expectation set in the current scoring criteria. Further research is needed to investigate how this alternative scoring procedure would contribute to the validity argument of the simulated conversation task.

5. Conclusion

The AP Japanese exam is a criterion-referenced performance test designed to measure Japanese language proficiency in tasks and situations reflective of real life. The present study explored the validity argument for the AP Japanese exam at the domain definition inference stage. Focusing on the computer-simulated conversation task, this study specifically examined four underlying assumptions: (1) the target domain is clearly identified; (2) critical language skills and knowledge needed in the target domain are identified; (3) test tasks that are representative of the target domain are identified; and (4) test tasks that require critical skills and are representative of the target domain are simulated. The backing for these assumptions was collected through document review and the analysis of test task characteristics, including context and situation, input language, and scoring criteria. The present study found reasonable backing for Assumptions 1 and 2 and partial backing for Assumptions 3 and 4. These results provide reasonable support for the claim that features of the examinee performance that provide evidence of language ability are identified. However, the test task was found to have somewhat limited relevance to and representation of real-life tasks due to the limited correspondence.

To build stronger validity arguments, this study recommended that the College Board provide more contextual descriptions regarding the identity of the interlocutors and their relationship to the examinees, the purpose of the interaction, and the presuppositions about the examinees' actions and knowledge prior to the interaction to make the interpretation of the context more consistent across examinees and to minimize the unintended effect

of differing level of imagination of the context. However, this must be balanced with practicality. This study also recommended that the College Board review the role of pragmatic competence in the conversation task and clarify the task definition to situate the speech event accordingly. If the College Board decides to assess communicative competence in everyday interaction, the test developers should draw on empirical studies of real-life spoken interaction and discourse when writing prompts to increase the authenticity of the test task. In particular, elaboration-seeking *dooshite* ‘why’ questions should be avoided except for the purpose of generating arguments. In addition, this study recommended that prompts that elicit different types of responses should be included in a more principled and balanced manner across test forms taking into account different response expectations in real life. This study also suggested that responses to a set of four prompts be treated as one sample rather than four individual samples in order to mitigate the effects of different prompt compositions across test forms and the mismatch between the nature of individual prompts and the response expectation set in the current scoring criteria.

One limitation of this study was its scope in analyzing the input language. The speed, amount, and syntactic structures of the input were not analyzed in the present study. Another limitation was the low inter-coder agreement for response expectation. Nevertheless, the results of this study provide a starting point for further validation research for the AP Japanese exam. In general, validation studies for tests of languages other than English have been scarce. In addition, this study could serve as a useful reference when researching the validity of other computer-simulated conversation tasks in tests such as the AP Chinese exam and the ACTFL Assessment of Performance toward Proficiency in Languages (AAPPL).

APPENDIX

Conversation Prompts from the 2007–2022 AP Japanese Exam

Form 2007

You will have a conversation with Naoto Moriyama, a representative from a youth center, about student activities.

1. 初めまして、日本ユースセンターの森山です。先日はアンケートにお答えくださりまして、ありがとうございました。
“How do you do? I am Moriyama from the Japan Youth Center. Thank you for answering the questionnaire the other day.”
2. アンケートについてもう少し詳しくお伺いしたいんですが、あなたはボランティア活動をするのと、アルバイトをするのではどちらが好きですか？
“I would like to ask more about the questionnaire. Between doing volunteer activities and having a part-time job, which do you like better?”
3. それはどうしてですか？
“Why is that?”
4. あ、そうですか。それから、今度うちのセンターで高校生の活動について話をしてもらえるということですが、どんなお話でしょうか？
“Oh, I see. And then it seems that at our center next time we will have you talk about high school students’ activities. What kind of talk will it be?”

Form 2008

You will have a conversation with Yumiko Tanaka, a study-abroad program coordinator, about staying with a family in Japan.

1. 日本のホストファミリーのことですけど、どんな家族がいいですか。
“Regarding the Japanese host family, what kind of family would you like to have?”
2. いつからいつまでホームステイしたいですか。
“From when to when would you like to do your homestay?”
3. あ、そうですか。ホストファミリーとどんなことがしたいですか。
“Oh, I see. What would you like to do with your host family?”
4. 分かりました。じゃ、私もなるべくいいホストファミリーを見つけますから、あなたも日本語の勉強頑張ってきてくださいね。
“Okay. Then I will do my best to find a good host family, so you should also work hard with your Japanese studies.”

Form 2009

You will have a conversation with Misaki Saito, a new foreign exchange student, about school life.

1. こんにちは、美咲です。先週着たばかりで、学校のクラスの事とか、よく分からないんです。色々教えてください。
“Hello. (I) am Misaki. (I) just came (here) last week, so (I) do not know much about classes at school or other things. Please advise me on many things.”
2. あのう、今学期、数学か科学を取らなくちゃいけないんですけど、どちらの方が難しいと思いますか。
“Umm this semester, (I) have to take math or science, but which do (you) think is more difficult?”
3. そうですか。それから、私、友達をいっぱい作りたいと思っているんですけど、どんなクラブ活動があるか教えてください。
“I see. And then I think I want to make a lot of friends, but please tell me what kind of club activities there are.”
4. ああ、そうですか。ありがとうございます。じゃ明日、よかったら、また、カフェテリアで昼ご飯、一緒に食べませんか。
“Oh, I see. Thank you. Then, if (you) would like, why don't we have lunch together again?”

Form 2010

You will have a conversation with Masao Hirano, a friend, about an upcoming party.

1. 今度のパーティーのことなんだけど、夜がいい？それとも、昼がいい？
“Speaking of the next party, do you prefer night time? Or do you prefer day time?”
2. あ、いいね。クラスのみんなが楽しめるパーティーにしたいんだけど、何をしたらいいかな？
“Oh, that sounds good. (I) would like to have a party that everyone can enjoy, but I wonder what we should do?”
3. あ、そう。どうしてそう思うのか、もうちょっと詳しく話してくれる？
“Oh, I see. Can you tell me more about why you think that?”
4. ああ、なるほど。いつも、ありがとう。じゃ、またパーティーの手伝い、よろしく。
“Oh, I see. Thank you as always. Okay, then I will ask you for help later. Thanks!”

Form 2011

You will have a conversation with Keizo Nakamura, your new neighbor, about recycling garbage.

1. あ、はじめまして。日本から隣に引越してきた中村です。どうぞよろしくお願いします。
“Oh, how do you do? I'm Nakamura, from Japan, who moved in next door. Nice to meet you.”

2. あのう、明日はゴミの日ですよ。この町ではゴミとリサイクルとか、どうしているんですか。
“Um, tomorrow is the day to take the trash out, isn't it? In this town, what do you do with garbage and recycling items?”
3. あのう、うちにコーラの空き缶なんかがたくさんあるんですけど、やっぱり、リサイクルした方がいいでしょうか。
“Um, I have a lot of empty coke cans at home, should I recycle them?”
4. そうですね。あの、もしできれば、明日、ゴミを出す時、ちょっと手伝ってもらいたいんですけど……。
“I see. Um, if it's possible, would you be able to help me when I take the trash out tomorrow?”

Form 2012

You will have a conversation with Yumiko Tanaka, your friend, about club activities.

1. クラブのボランティア活動のことなんだけど、公園のそうじをするのと小学生の勉強を手伝ってあげるのとどっちにする？
“About the volunteer activity for the club, which should we do: cleaning a park or helping elementary school students with their studies?”
2. 毎週一回することになってるんだけど、何曜日の何時にしようか。
“We are supposed to do that once a week, but what time on what day of the week should we do that?”
3. その曜日と時間がいいのはどうして？
“Why do you think that time and day of the week are good?”
4. じゃ、今度のクラブのミーティングで、今話したことをみんなに伝えてくれる？
“Well then, at the next club meeting, can you tell everyone what we talked about?”

Form 2013

You will have a conversation with Yumiko Tanaka, a Japanese student, about working.

1. 今日は、高校生のアルバイトについてお聞きしたいと思います。よろしくお願いします。
“Today, (I) would like to ask you about part-time jobs for high school students. Thank you for your cooperation.”
2. あなたの高校の生徒は、どんなアルバイトをしていますか。
“What kind of part-time job do students at your school do?”
3. 高校生がアルバイトをするのとしないのと、どちらがいいと思いますか。
“Which do (you) think is better? For high school students to have a part-time job or not to have a part-time job?”
4. アルバイトをしたい人に、何かアドバイスをしてください。

“Please give some advice to students who want to have a part-time job.”

Form 2014

You will have a conversation with Aki Yamashita, a study abroad program assistant who is meeting you at the airport, about plans for the afternoon.

1. 日本へようこそ。プログラムアシスタントの山下アキです。初めまして。
“Welcome to Japan. (I) am Aki Yamashita, Program Assistant. How do you do?”
2. お昼御飯ですが、ハンバーガーがいいですか。それとも、日本のお弁当がいいですか。
“About your lunch, would you like a hamburger? Or would you like a Japanese bento box?”
3. 今日は、お昼ご飯のあと、少し時間があるんですが、何がしたいですか。
“Today after lunch, we have some time, what would you like to do?”
4. ところで、日本語が上手ですね。どのように日本語を勉強しているんですか。
“By the way, (you) are good at Japanese. How are you studying Japanese?”

Form 2015

You will have a conversation with Hiro Yamamoto, a Japanese student, about an upcoming event.

1. 金曜日のコンサートなんだけど、Tシャツとジーンズでいいかな？それともフォーマルな服の方がいいかな？
“Regarding the concert on Friday, would a T-shirt and jeans be fine? Or I wonder if a more formal outfit would be better?”
2. あ、そう。それで、コンサートの後、どこに食べに行こうか？
“Oh, Okay. Then, where should we go eat after the concert?”
3. ところで、コンサートの場所がよく分からないから一緒に行ってくれない？
“By the way, I don’t know the concert location very well. Can you come with me?”
4. 金曜日のコンサート楽しみだね。わくわくするよ。
“I am looking forward to the concert on Friday. I am very excited about it.”

Form 2016

You will have a conversation with Daisuke Suzuki, a camp director, about a Japanese language camp.

1. はじめまして、日本語キャンプの鈴木です。まずはじめに、簡単に自己紹介をしてください。

- “How do you do? (I am) Suzuki from the Japanese camp. First, please briefly introduce yourself.”
2. ありがとうございます。どうしてこの日本語キャンプに入りたいと思ったんですか。
“Thank you. Why do you want to come to this Japanese camp?”
 3. ああ、そうですか。では、このキャンプではどんなことをしたいですか。
“Oh, I see. So what do you want to do at this camp?”
 4. 分かりました。ではこのキャンプについて何か質問をしてください。
“Okay. So please ask some questions.”

Form 2017

You will have a conversation with Hiroko Kikuchi, the president of an environmental club, about global warming.

1. 地球温暖化についてお聞きしたいんですが、よろしくお願ひします。
“(I) would like to ask you about global warming.”
2. まず、車ですが、ガソリンを使う車と電気を使う車と、どちらがいいと思いますか。
“First, about cars, between cars that use gasoline and cars that use electricity, which do you think is better?”
3. どうしてそう思うのですか。
“Why do you think so?”
4. そうですか。面白いですね。もっと話したいんですが、来週はいつがいいですか。
“I see. (That) is interesting. I would like to talk with you more; when will be good next week?”

Form 2018

You will have a conversation with Yumi Takagi, a Japanese-speaking student, about tutoring in Japanese.

1. 初めまして。今度日本語の勉強のお手伝いをする事になった高木ゆみといいます。よろしくお願ひします。
“How do you do? I am Yumi Takagi who will be assisting you with your Japanese studies. Nice to meet you.”
2. まず、なぜ日本語を勉強しているか教えてください。
“First, please tell me why you are studying Japanese.”
3. そうですか。日本語の勉強の中で、もっと上手になりたいことは何ですか。
“I see. What do you want to get better at in your Japanese studies?”
4. 分かりました。じゃ、これからの予定ですが・・・週に何回会いましょうか。

“All right. So, about our schedule...how many times a week shall we meet?”

Form 2019

You will have a conversation with Takeshi Nakaura, a homestay coordinator, about your upcoming homestay.

1. はじめまして、ホームステイ・コーディネーターの中村です。よろしく
願います。
“How do you do? I am Nakamura, your homestay coordinator. Nice to
meet you.”
2. どうしてホームステイをしたいのか教えてください。
“Please tell me why you would like to do a homestay.”
3. ホームステイは、一週間と一か月のプログラムがありますが、どちらが
いいですか？
“We have both one-week and one-month homestay programs...which do
you prefer?”
4. ホストファミリーについて、来週もう一度話したいんですが・・・いつ
がいいですか？
“I would like to talk to you again next week about your host
family...when is a good time?”

Form 2021

You will have a conversation with Daisuke Yamamoto, director of an after-school program at a Japanese elementary school, about volunteering in the program.

1. ああ、もしもし、お電話待ちしていました。こちらはひどい雨なんです
が、そちらのお天気はどうですか。
“Oh, hello, I've been waiting for your call. It's raining terribly here. How
is the weather there?”
2. 小学校でボランティアをしたいそうですが、それはどうしてですか。
“I heard that you want to volunteer at an elementary school, why is that?”
3. なるほど、このプログラムには小学校1年生から6年生までの子供たち
がいますが、何年生のグループがいいですか。
“Okay, this program has children in Grades 1 through 6. What grade
group would you like to work with?”
4. 分かりました。では、来週一度見に来てもらいたいんですが、いつ来ら
れますか。
“Okay. So, I would like to have you come once next week to take a look
(at our program). When can you come?”

Form 2022

You will have a conversation with Naomi Nakanishi, a booth staffer at a Japan fair, about Japanese technology.

1. 車のブースによろこそ！私がこのブースをご案内します。
“Welcome to the car booth! I will show you around this booth.”
2. ここには、新しいテクノロジーを使った車が色々ありますが、どんな車に興味がありますか。
“We have a variety of cars here that use new technology. What kind of cars are you interested in?”
3. これから、テクノロジーで、私達の生活はどう変わるとおもいますか。
“How do you think technology will change our lives in the future?”
4. 後で、もっと意見を聞きたいんですが、オンラインのアンケートか、紙のアンケートに答えてくださいませんか。
“We would like to hear more about your opinions later. Could you fill out the online survey or the paper survey?”

REFERENCES

- American Council on the Teaching of Foreign Languages. 2012. “ACTFL Proficiency Guidelines.” http://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf.
- Arendt, Ulli, and Rick Morgan. 1995. Comparing the Performance of High School and College Students on the Listening, Speaking, and Reading Sections of the Advanced Placement German Language Examination. *Die Unterrichtspraxis/Teaching German* 28 (1): 40–45. <https://doi.org/10.2307/3531332>.
- Bachman, Lyle F. 1990. *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, Lyle F., and Adrian S. Palmer. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford Applied Linguistics. Oxford University Press.
- Bachman, Lyle F., and Adrian S. Palmer. 2010. *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford Applied Linguistics. Oxford University Press.
- Baum, David I. 2007a. “Establishing AP Chinese Language and Culture Exam Validity Using College Professors’ Grading Standards.” *Educational Testing*

- Service.
http://apcentral.collegeboard.com/apc/public/repository/collegeCompStudy_chineseLang_07.pdf.
- Baum, David I. 2007b. "Establishing AP Japanese Language and Culture Exam Validity Using College Professors' Grading Standards." Educational Testing Service.
http://apcentral.collegeboard.com/apc/public/repository/collegeCompStudy_japaneseLang_07.pdf.
- Bialystok, Ellen. 1981. The Role of Linguistic Knowledge in Second Language Use. *Studies in Second Language Acquisition* 4 (1): 31–45. <https://doi.org/10.1017/S0272263100004265>.
- Biber, Douglas, Susan M. Conrad, Randi Reppen, et al. 2004. *Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic*. MS-25. TOEFL Monograph. Educational Testing Service. https://www.ets.org/research/policy_research_reports/publications/report/2004/ibyq.
- Bischof, Deborah Lokai. 2005. Validating the AP® German Language Exam through a Curricular Survey of Third-Year College Language Courses. *Die Unterrichtspraxis / Teaching German* 38 (1): 74–81. <http://www.jstor.org/stable/3531296>.
- Bischof, Deborah Lokai, David I. Baum, Jodi M. Casabianca, Rick Morgan, Kathleen A. Rabitea, and Krishna Tateneni. 2004. Validating AP Modern Foreign Language Examinations through College Comparability Studies. *Foreign Language Annals* 37 (4): 616–22. <https://doi.org/10.1111/j.1944-9720.2004.tb02427.x>.
- Breiner-Sanders, Karen E., Pardee Lowe, John Miles, and Elvira Swender. 2000. ACTFL Proficiency Guidelines—Speaking: Revised 1999. *Foreign Language Annals* 33 (1): 13–18. <https://doi.org/10.1111/j.1944-9720.2000.tb00885.x>.
- Brown, Gillian, and George Yule. 1983. *Teaching the Spoken Language: An Approach Based on the Analysis of Conversational English*. Cambridge Language Teaching Library. Cambridge University Press.
- Butler, Frances A., Daniel Eignor, Stan Jones, Tim McNamara, and Barbara K. Suomi. 2000. *TOEFL 2000 Speaking Framework: A Working Paper*. TOEFL Monograph No. 20. Educational Testing Service. <http://www.ets.org/Media/Research/pdf/RM-00-06.pdf>.
- Canale, Michael, and Merrill Swain. 1980. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics* 1 (1): 1–47. <https://doi.org/10.1093/applin/1.1.1>.

- Chapelle, Carol A. 2008. The TOEFL Validity Argument. In *Building a Validity Argument for the Test of English as a Foreign Language*, edited by Carol A. Chapelle, Mary K. Enright, and Joan M. Jamieson, 319–352. Routledge.
- Chapelle, Carol A., Mary K. Enright, and Joan M. Jamieson, eds. 2008a. *Building a Validity Argument for the Test of English as a Foreign Language*. ESL & Applied Linguistics Professional Series. Routledge.
- Chapelle, Carol A., Mary K. Enright, and Joan M. Jamieson. 2008b. Test Score Interpretation and Use. In *Building a Validity Argument for the Test of English as a Foreign Language*, edited by Carol A. Chapelle, Mary K. Enright, and Joan M. Jamieson. ESL & Applied Linguistics Professional Series. Routledge.
- Chapelle, Carol A., and Hye-won Lee. 2021. Understanding Argument-Based Validity in Language Testing. In *Validity Argument in Language Testing: Case Studies of Validation Research*, edited by Carol A. Chapelle and Erik Voss, 1–25. Cambridge University Press. <http://doi.org/10.1017/9781108669849.004>.
- Clark, Herbert H. 1996. *Using Language*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511620539>.
- College Board. 2007a. “AP Japanese Language and Culture 2007 Free-Response Questions.” <https://apcentral.collegeboard.org/media/pdf/ap-2007-japanese-language-free-response-questions.pdf>.
- College Board. 2007b. “AP Japanese Language and Culture 2007 Scoring Commentary.” https://secure-media.collegeboard.org/apc/ap07_japanese_is_conversation.pdf.
- College Board. 2007c. “Student Performance Q&A: 2007 AP Japanese Language and Culture Free-Response Questions.” https://secure-media.collegeboard.org/apc/ap07_japanese_q_a.pdf.
- College Board. 2008a. “AP Japanese Language and Culture 2008 Free-Response Questions.” <https://secure-media.collegeboard.org/apc/ap-2008-japanese-language-free-response-questions.pdf>.
- College Board. 2008b. “AP Japanese Language and Culture 2008 Scoring Commentary.” https://secure-media.collegeboard.org/apc/ap08_japanese_interpersonal_speaking_conversation.pdf.
- College Board. 2008c. “Student Performance Q&A: 2008 AP Japanese Language and Culture Free-Response Questions.” https://secure-media.collegeboard.org/apc/ap08_japanese_qa.pdf.
- College Board. 2009a. “AP Japanese Language and Culture 2009 Free-Response Questions.” <https://secure-media.collegeboard.org/apc/ap-2009-japanese-language-free-response-questions.pdf>.

- College Board. 2009b. "AP Japanese Language and Culture 2009 Scoring Commentary." https://secure-media.collegeboard.org/apc/ap09_japanese_interpersonal_speaking.pdf.
- College Board. 2009c. "Student Performance Q&A: 2009 AP Japanese Language and Culture Free-Response Questions." https://secure-media.collegeboard.org/apc/ap09_japanese_qa.pdf.
- College Board. 2010a. "AP Japanese Language and Culture 2010 Free-Response Questions." <https://secure-media.collegeboard.org/apc/ap-2010-japanese-language-free-response-questions.pdf>.
- College Board. 2010b. "AP Japanese Language and Culture 2010 Scoring Commentary." Author. https://secure-media.collegeboard.org/apc/ap10_japanese_interpersonal_speaking.pdf.
- College Board. 2010c. "Student Performance Q&A: 2010 AP Japanese Language and Culture Free-Response Questions." https://secure-media.collegeboard.org/apc/ap10_japanese_qa.pdf.
- College Board. 2011a. "AP Japanese Language and Culture 2011 Free-Response Questions." <https://secure-media.collegeboard.org/apc/ap-2011-japanese-language-free-response-questions.pdf>.
- College Board. 2011b. "AP Japanese Language and Culture 2011 Scoring Commentary." https://secure-media.collegeboard.org/apc/ap11_japanese_interpersonal_speaking.pdf.
- College Board. 2011c. "Student Performance Q&A: 2011 AP Japanese Language and Culture Free-Response Questions." https://secure-media.collegeboard.org/apc/ap11_japanese_language_scoring_statistics.pdf.
- College Board. 2012a. "AP Japanese Language and Culture 2012 Free-Response Questions." https://secure-media.collegeboard.org/apc/ap_2012_frq_japanese.pdf.
- College Board. 2012b. "AP Japanese Language and Culture 2012 Scoring Commentary." https://secure-media.collegeboard.org/apc/ap12_japanese_interpersonal_speaking.pdf.
- College Board. 2012c. "Student Performance Q&A: 2012 AP Japanese Language and Culture Free-Response Questions." https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap12_japanese_qa.pdf.
- College Board. 2013a. "AP Japanese Language and Culture 2013 Free-Response Questions." https://secure-media.collegeboard.org/digitalServices/pdf/ap/apcentral/ap13_frq_japanese.pdf.
- College Board. 2013b. "AP Japanese Language and Culture 2013 Scoring

- Commentary.” https://secure-media.collegeboard.org/digitalServices/pdf/ap/apcentral/ap13_japanese_language_interpersonal_speaking.pdf.
- College Board. 2013c. “Student Performance Q&A: 2013 AP Japanese Language and Culture Free-Response Questions.” https://secure-media.collegeboard.org/digitalServices/pdf/ap/apcentral/ap13_japanese_language_qa.pdf.
- College Board. 2014a. “AP Japanese Language and Culture 2014 Free-Response Questions.” https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap14_frq_japanese_language.pdf.
- College Board. 2014b. “AP Japanese Language and Culture 2014 Scoring Commentary.” https://secure-media.collegeboard.org/digitalServices/pdf/ap/apcentral/ap14_japanese_language_interpersonal_speaking.pdf.
- College Board. 2014c. “Student Performance Q&A: 2014 AP Japanese Language and Culture Free-Response Questions.” <https://secure-media.collegeboard.org/digitalServices/pdf/ap/apcentral/ap14-japanese-qa.pdf>.
- College Board. 2015a. “AP Japanese Language and Culture 2015 Free-Response Questions.” https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap15_frq_japanese_language.pdf.
- College Board. 2015b. “AP Japanese Language and Culture 2015 Scoring Commentary.” https://secure-media.collegeboard.org/digitalServices/pdf/ap/apcentral/ap15_Japanese_interpersonal_speaking_convo.pdf.
- College Board. 2015c. “Student Performance Q&A: 2015 AP Japanese Language and Culture Free-Response Questions.” <https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap15-japanese-student-performance-qa.pdf>.
- College Board. 2016a. “AP Japanese Language and Culture 2016 Free-Response Questions.” https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap16_frq_japanese_language.pdf.
- College Board. 2016b. “AP Japanese Language and Culture 2016 Scoring Commentary.” https://secure-media.collegeboard.org/digitalServices/pdf/ap/apcentral/ap16_japanese_interpersonal_speaking_convo.pdf.

- College Board. 2016c. "Student Performance Q&A: 2016 AP Japanese Language and Culture Free-Response Questions." https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap16_japanese_student_performance_qa.pdf.
- College Board. 2017a. "AP Japanese Language and Culture: Free-Response Questions." <https://apcentral.collegeboard.org/media/pdf/ap-japanese-language-frq-2017.pdf>.
- College Board. 2017b. "AP Japanese Language and Culture: Sample Student Responses and Scoring Commentary." <https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap17-japanese-interpersonal-speaking-conversation.pdf>.
- College Board. 2017c. "Chief Reader Report on Student Responses: 2017 AP Japanese Language and Culture Free-Response Questions." <https://secure-media.collegeboard.org/ap/pdf/ap17-chief-reader-report-japanese-language.pdf>.
- College Board. 2018a. "AP Japanese Language and Culture: Free-Response Questions." <https://apcentral.collegeboard.org/media/pdf/ap18-frq-japanese-language.pdf>.
- College Board. 2018b. "AP Japanese Language and Culture: Sample Student Responses and Scoring Commentary." <https://secure-media.collegeboard.org/ap/pdf/ap18-japanese-interpersonal-speaking-conversation.pdf>.
- College Board. 2018c. "Chief Reader Report on Student Responses: 2018 AP Japanese Language and Culture Free-Response Questions." <https://apcentral.collegeboard.org/media/pdf/ap18-japanese-language-chief-reader-report.pdf>.
- College Board. 2019a. "AP Japanese Language and Culture: Free-Response Questions." https://apcentral.collegeboard.org/media/pdf/ap19-frq-japanese-language_0.pdf.
- College Board. 2019b. "AP Japanese Language and Culture: Sample Student Responses and Scoring Commentary." https://apcentral.collegeboard.org/media/pdf/ap19-apc-japanese-language-t3-interpersonal-speaking_1.pdf.
- College Board. 2019c. "Chief Reader Report on Student Responses: 2019 AP Japanese Language and Culture Free-Response Questions." <https://apcentral.collegeboard.org/media/pdf/ap19-chief-reader-report-japanese-language.pdf>.
- College Board. 2020. "AP Japanese Language and Culture: Course and Exam Description." <https://apcentral.collegeboard.org/pdf/ap-japanese-language->

- and-culture-course-and-exam-description.pdf?course=ap-japanese-language-and-culture.
- College Board. 2021a. “AP Japanese Language and Culture: Free-Response Questions.” <https://apcentral.collegeboard.org/media/pdf/ap21-frq-japanese-language.pdf>.
- College Board. 2021b. “AP Japanese Language and Culture: Sample Student Responses and Scoring Commentary.” <https://apcentral.collegeboard.org/media/pdf/ap21-apc-japanese-language-task-3.pdf>.
- College Board. 2021c. “Chief Reader Report on Student Responses: 2021 AP Japanese Language and Culture Free-Response Questions.” <https://apcentral.collegeboard.org/media/pdf/ap21-chief-reader-report-japanese-language.pdf>.
- College Board. 2022a. “AP Japanese Language and Culture: Free-Response Questions.” <https://apcentral.collegeboard.org/media/pdf/ap22-frq-japanese-language.pdf>.
- College Board. 2022b. “AP Japanese Language and Culture: Sample Student Responses and Scoring Commentary.” <https://apcentral.collegeboard.org/media/pdf/ap22-apc-japanese-language-task-3.pdf>.
- College Board. 2022c. “Chief Reader Report on Student Responses: 2022 AP Japanese Language and Culture Free-Response Questions.” <https://apcentral.collegeboard.org/media/pdf/ap22-cr-report-japanese-language.pdf>.
- College Board. 2023. “AP Japanese Language and Culture Scoring Guidelines.” <https://apcentral.collegeboard.org/media/pdf/ap23-sg-japanese-language.pdf>.
- College Board. 2024. “AP Japanese Language and Culture Student Score Distributions – Global AP Exams – May 2024.” <https://apcentral.collegeboard.org/media/pdf/ap24-japanese-language-score-distributions.pdf>.
- College Board. n.d.-a. “Exam Development – AP Central.” Accessed June 6, 2024. <https://apcentral.collegeboard.org/courses/how-ap-develops-courses-and-exams/exam-development>.
- College Board. n.d.-b. “Score Setting and Scoring – AP Central.” Accessed June 6, 2024. <https://apcentral.collegeboard.org/courses/how-ap-develops-courses-and-exams/score-setting-and-scoring>.
- Douglas, Dan. 1997. *Testing Speaking Ability in Academic Contexts: Theoretical Considerations*. ETS Research Memorandum RM-97-01. Educational Testing Service. <http://www.ets.org/Media/Research/pdf/RM-97-01.pdf>.

- Enright, Mary K., Brent Bridgeman, Daniel Eignor, et al. 2008. Prototyping New Assessment Tasks. In *Building a Validity Argument for the Test of English as a Foreign Language*, edited by Carol A. Chapelle, Mary K. Enright, and Joan M. Jamieson, 97–143. Routledge.
- Ford, Cecilia E., and Junko Mori. 1994. Causal Markers in Japanese and English Conversations: A Cross-Linguistic Study of Interactional Grammar. *Pragmatics* 4 (1): 31–61. <https://doi.org/10.1075/prag.4.1.03for>.
- Fulcher, Glenn. 2003. *Testing Second Language Speaking*. Applied Linguistics and Language Study. Routledge.
- Ginther, April, and Joseph Stevens. 1995. Language Background, Ethnicity, and the Internal Construct Validity of the Advanced Placement Spanish Language Examination. In *Validation in Language Assessment: Selected Papers from the 17th Language Testing Research Colloquium, Long Beach*, edited by Antony John Kunnan, 169–194. Lawrence Erlbaum Associates.
- Hutchby, Ian., and Robin Wooffitt. 1998. *Conversation Analysis: Principles, Practices and Applications*. Polity Press.
- Hymes, Dell. 1985. Toward Linguistic Competence. *Revue d'AILA: AILA Review* 2: 9–23.
- Iwasaki, Shoichi. 1997. The Northridge Earthquake Conversations: The Floor Structure and the 'Loop' Sequence in Japanese Conversation. *Journal of Pragmatics* 28 (6): 661–93. [https://doi.org/10.1016/S0378-2166\(97\)00070-2](https://doi.org/10.1016/S0378-2166(97)00070-2).
- Jamieson, Joan M., Daniel Eignor, William Grabe, and Antony John Kunnan. 2008. Frameworks for a New TOEFL. In *Building a Validity Argument for the Test of English as a Foreign Language*, edited by Carol A. Chapelle, Mary K. Enright, and Joan M. Jamieson, 55–95. Routledge.
- Jamieson, Joan M., Stan Jones, Irwin Kirsch, Peter Mosenthal, and Carol Taylor. 2000. *TOEFL 2000 Framework: A Working Paper*. TOEFL Monograph No. 16. Educational Testing Service. <http://www.ets.org/Media/Research/pdf/RM-00-03.pdf>.
- Kane, Michael T. 1992. An Argument-Based Approach to Validity. *Psychological Bulletin* 112 (3): 527. <https://doi.org/10.1037/0033-2909.112.3.527>.
- Kyle, Kristopher, Ann Tai Choe, Masaki Eguchi, Geoff LaFlair, and Nicole Ziegler. 2021. A Comparison of Spoken and Written Language Use in Traditional and Technology-Mediated Learning Environments. *ETS Research Report Series* 2021 (1): 1–29. <https://doi.org/10.1002/ets2.12329>.
- Kyle, Kristopher, Masaki Eguchi, Ann Tai Choe, and Geoff LaFlair. 2022. Register Variation in Spoken and Written Language Use across Technology-Mediated and Non-Technology-Mediated Learning Environments. *Language Testing* 39 (4): 618–48. <https://doi.org/10.1177/02655322211057868>.

- Lado, Robert. 1961. Linguistics and Foreign Language Teaching. *Language Learning* 11 (s2): 29–52. <https://doi.org/10.1111/j.1467-1770.1961.tb00200.x>.
- Laver, John. 1975. Communicative Functions of Phatic Communion. In *Organization of Behavior in Face-to-Face Interaction*, edited by Richard M. Harris, Adam Kendon, and Mary Ritchie Key, 215–238. Mouton. <https://doi.org/10.1515/9783110907643.215>.
- Liskin-Gasparro, Judith E., Christopher C. Modu, and Joseph Schraibman. 1979. The Validity of the Multiple-Choice Component of the Advanced Placement Spanish Language Examination. *Hispania* 62 (1): 98–105. <https://doi.org/10.2307/340084>.
- Malinowski, Bronislaw. 1953. The Problem of Meaning in Primitive Languages. In *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*, 8th ed., edited by C. K. Ogden and I. A. Richards, 296–336. Harcourt, Brace, & Co.
- Marzi, Giacomo, Marco Balzano, and Davide Marchiori. 2024. K-Alpha Calculator–Krippendorff’s Alpha Calculator: A User-Friendly Tool for Computing Krippendorff’s Alpha Inter-Rater Reliability Coefficient. *MethodsX* 12 (June): 102545. <https://doi.org/10.1016/j.mex.2023.102545>.
- McCormick, Dawn E., and Richard Donato. 2000. Teacher Questions as Scaffolded Assistance in an ESL Classroom. In *Second and Foreign Language Learning through Classroom Interaction*, edited by Joan Kelly Hall and Lorrie Stoops Verplatse, 183–201. Erlbaum.
- Messick, Samuel. 1989. Validity. In *Educational Measurement*, 3rd ed., edited by R. L. Linn, 13–103. Macmillan.
- Modu, Christopher C., Mariette Reed, and Robert L. Coon. 1975. How Valid Is the Advanced Placement French Language Examination? *The French Review* 48 (6): 1013–22. <http://www.jstor.org/stable/389176>.
- Modu, Christopher C., and C. Andrew Weeks. 1982. The Validity of the Multiple-Choice Section of the Advanced Placement German Language Examination. *Die Unterrichtspraxis / Teaching German* 15 (1): 116–26. <https://doi.org/10.2307/3530349>.
- Morgan, Rick, and John Mazzeo. 1988. A Comparison of the Structural Relationships among Reading, Listening, Writing, and Speaking Components of the AP French Language Examination for AP Candidates and College Students. *ETS Research Report Series* 1988 (2): i–46. <https://doi.org/10.1002/j.2330-8516.1988.tb00315.x>.
- Mori, Junko. 1999. *Negotiating Agreement and Disagreement in Japanese: Connective Expressions and Turn Construction*. Vol. 8. Studies in Discourse

- and Grammar. John Benjamins Publishing Company. <https://doi.org/10.1075/sidag.8>.
- Mori, Junko. 2005. Why Not Why? The Teaching of Grammar, Discourse, Sociolinguistic and Cross-Cultural Perspectives. *Japanese Language and Literature* 39 (2): 255–89. <https://doi.org/10.2307/30038902>.
- National Standard Collaborative Board. 2015. *World-Readiness Standards for Learning Languages*. 4th ed. American Council on the Teaching of Foreign Languages.
- National Standards in Foreign Language Education Project. 1999. *Standards for Foreign Language Learning in the 21st Century*. 2nd ed. Author.
- Nishizawa, Hitoshi. 2024. Authenticity of Academic Lecture Passages in High-Stakes Tests: A Temporal Fluency Perspective. *Language Testing* 41 (4): 792–816. <https://doi.org/10.1177/02655322241262453>.
- Ockey, Gary J., and Reza Neiriz. 2021. Evaluating Technology-Mediated Second Language Oral Communication Assessment Delivery Models. *Assessment in Education: Principles, Policy & Practice* 28 (4): 350–68. <https://doi.org/10.1080/0969594X.2021.1976106>.
- Roever, Carsten, and Naoki Ikeda. 2022. What Scores from Monologic Speaking Tests Can(Not) Tell Us about Interactional Competence. *Language Testing* 39 (1): 7–29. <https://doi.org/10.1177/02655322211003332>.
- Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50 (4): 696–735. <https://doi.org/10.2307/412243>.
- Savignon, Sandra J. 1983. *Communicative Competence: Theory and Classroom Practice*. Addison-Wesley Publishing Company.
- Schegloff, Emanuel A. 1982. Discourse as an Interactional Achievement: Some Uses of ‘Uh Huh’ and Other Things That Come between Sentences. In *Analyzing Discourse: Text and Talk*, edited by Deborah Tannen, 71–93. Georgetown University Press.
- Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Vol. 1. Cambridge University Press. <https://doi.org/10.1017/CBO9780511791208>.
- Schegloff, Emanuel A., and Harvey Sacks. 1973. Opening up Closing. *Semiotica* 8: 289–327. <https://doi.org/10.1515/semi.1973.8.4.289>.
- Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139173438>.
- Searle, John R. 1979. *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge University Press.

- <https://doi.org/10.1017/CBO9780511609213>.
- Shepard, Lorrie A. 1993. Evaluating Test Validity. *Review of Research in Education* 19: 405–50. <https://doi.org/10.2307/1167347>.
- Streeck, Jürgen. 1980. Speech Acts in Interaction: A Critique of Searle. *Discourse Processes* 3 (2): 133–53. <https://doi.org/10.1080/01638538009544483>.
- Suzuki, Takao. 1981. Aisatsu Towa Nanika [What Is Aisatsu?]. In *Aisatsu to Kotoba [Greetings and Words]*, edited by Agency for Cultural Affairs, vol. 14, 34–46. Printing Bureau of the Ministry of Finance.
- Suzumura, Nana. 2020. The Usefulness of the Computer-Based Speaking Tasks of the AP Japanese Exam. Doctoral dissertation, University of Hawai‘i, Mānoa. ProQuest Dissertations & Theses Global (Publication No. 22624584).
- Suzumura, Nana. 2022. Content Analysis of Test Taker Responses on an AP Japanese Computer-Simulated Conversation Test: A Mixed Methods Approach for a Validity Argument. *Language Assessment Quarterly* 19 (5): 480–502. <https://doi.org/10.1080/15434303.2022.2130326>.
- Tanaka, Hiroko. 1999. *Turn-Taking in Japanese Conversation: A Study in Grammar and Interaction*. Pragmatics & Beyond New Series 56. John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.56>.
- Tateyama, Yumiko. 2000. Use of Japanese Connectives, Sorede, Dakara, and Ja in JFL Classroom Discourse [Paper Presentation]. In *Proceedings 2000: Selected Papers from the Fourth College-Wide Conference for Students in Languages, Linguistics and Literature*, edited by G. Sibley, N. Ochner, and K. Russell, 58–67. National Foreign Language Resource Center, University of Hawai‘i, Mānoa.
- Tohsaku, Yasu-hiko. 2009. “AP Japanese Language and Culture: Teacher’s Guide.” College Board. http://apcentral.collegeboard.com/apc/public/repository/japanese_teachers_guide.pdf.
- Weir, Cyril J. 1993. *Understanding and Developing Language Tests*. Language Teaching Methodology Series. Phoenix.
- Weir, Cyril J. 2005. *Language Testing and Validation: An Evidence-Based Approach*. Research and Practice in Applied Linguistics. Palgrave Macmillan.
- Wimmers, Eric, and Rick Morgan. 1990. Comparing the Performance of High School and College Students on the Advanced Placement French Language Examination. *The French Review* 63 (3): 423–32. <http://www.jstor.org/stable/394487>.