Enhancing Student Feedback for Eiken Interviews through a Criterion-Referenced Framework

Tim Saito
Toyama College of Welfare Science

Reference Data:

Saito, T. (2024). Enhancing student feedback for Eiken interviews through a criterion-referenced framework. In B. Lacy, R. P. Lege, & P. Ferguson (Eds.), *Growth Mindset in Language Education*. JALT. https://doi.org/10.37546/JALTPCP2023-44

This paper reports on the development of a criterion-referenced framework (CRF) designed to provide feedback to Japanese students preparing for the Eiken second-stage speaking test. Drawing from the works of Griffin (2018) and Hattie and Timperley (2007), it documents the first stage of an action research project aimed at prototyping a CRF to provide clear feedback to students. The CRF is proposed to provide advice to students constructing narratives from image sequences. The tool integrates the Eiken Foundation of Japan's (2008) can-do list and the Japanese Ministry of Education, Sports, Science, and Technology's (2018) High School Foreign Language Goals for the English Communication Curriculum, outlining essential skills for each proficiency level. This approach aims to develop an effective alternative to extemporaneous speaking test preparation. The CRF illustrates a method for providing formative, structured feedback for Eiken speaking test preparation. This research has implications for ESL/EFL educators advising students preparing for speaking tests.

本論は、英検二次試験のスピーキングテストを受験する日本人学生にフィードバックを提供するために設計された基準参照フレームワーク（CRF）の開発について報告するものである。Griffin (2018) ならびに Hattie と Timperley (2007)の先行研究を基に、アクションリサーチ手法により、学生に明確なフィードバックを提供することを目的としたCRFのプロトタイピングの第一段階を記録している。また、CRFは一連のイラストからナラティブを構築する学生に対して助言を提供するための提案である。このツールは、日本英語検定協会のcan-doリスト(日本英語検定協会, 2008)と高等学校学習指導要領「外国語科の目標」(文部科学省, 2018)を統合し、各習熟度レベルに必要なスキルを概説している。このアプローチは、即興の英検二次試験準備に代わる効果的な方法を開発することを目的としている。CRFは、英検二次試験の準備のための形成的で構造化されたフィードバックを提供する方法を示している。本研究は、ESL/EFLの教育者がスピーキングテストの準備を行う学生にアドバイスする際に示唆を与えるものである。

The importance of the Eiken test has grown significantly since 2018, when a panel from the Ministry of Education officially recognized Eiken as one of the approved private English tests for university admissions. However, a scarcity of detailed information regarding the interview scoring system, the pass rates, and questions with high failure rates—along with an emphasis on traditional approaches to teaching and learning—make preparation for the Eiken a continuing challenge. This point is pertinent in Japanese high schools, where "…clear learning outcomes and assessment criteria are new aspects of pedagogy for many Japanese teachers and students" (Yamamoto, 2016, p. 106).

In recent years, Japan's Ministry of Education, Culture, Sports, Science and Technology (MEXT) has set ambitious student Eiken proficiency targets, aiming for 50% of all 3rd-year middle school and high school students to achieve Grade 3 and Grade Pre-2, respectively. Although the Eiken is not officially part of the national curriculum, it is backed by MEXT and supported by local education boards and schools (Tatsukawa, 2018, p. 90). Despite incremental improvements since 2011, data from MEXT indicate that these targets have remained unmet for over a decade (Nippon Communications Foundation, 2023).

Thus, one challenge facing English education in Japanese high schools is to improve students' Eiken scores, particularly their ability to pass the speaking interview section. While criticisms of standardized speaking tests such as the Eiken abound (e.g., Borg, 2023; Wood, 2019), research supports the view that the capacity to speak English remains a significant barrier for Japanese students (Humphries et al., 2015; Masutani, 2021). Consequently, the interview component of the Eiken test emerges as a critical hurdle to overall test success. Approximately 92% of applicants who applied for external entrance examinations in 2021 used Eiken results to supplement the Common Test for University Admissions (Nichibei Eigo Gakuin, 2023). However, pass rate data for Eiken have been withheld from the public since 2015, and the Eiken interview's scoring system, aside from the minimum passing score, remains largely opaque. This lack of transparency hinders effective student preparation.

In response to the significant anxiety and difficulty that Japanese students experience with speaking English, various strategies have emerged. For instance, Takeno and Matsuura (2018) suggest analyzing model responses to compile lists of useful verbs and vocabulary, while others have collected interview questions from past tests to produce a reference guide to coach students on answering them (Whitley, 2017). Similarly, educational resources frequently adopt a scattergun approach, aiming to expose test-takers to a large volume of sample questions and model answers. However, more emphasis must be placed on aligning test preparation with fostering the capabilities that the test ostensibly aims to measure, rather than focusing exclusively on scores.

Adjacent research in this field predominantly addresses the complexities of assessing speaking ability, ensuring consistency in assessor judgments, establishing the validity of test construction, and developing pedagogical approaches to enhancing generalized communication skills. However, instructors remain largely ill equipped to provide students with effective feedback (Kriewaldt & Turnidge, 2013). One possible way forward is through the work of Patrick Griffin (2018), whose comprehensive guide *Assessment for Teaching* is instrumental to understanding the theory and practice of establishing a criterion-referenced framework (CRF). Hattie and Timperley (2007) then provide authoritative commentary regarding how such tools enhance the feedback process.

It was with the principles of those particular works in mind that the feedback mechanism described here for Question 2 of the Eiken interview was developed. This question requires interview participants at Grade 3 and above to produce an extended response to an image or sequence of images. Performing well on this question is crucial, as it carries significant weight in the interview. However, in the absence of specific, publicly available criteria, EFL teachers in Japan are usually left with no other option but to adopt extemporaneous approaches to interview preparation.

This paper is part of an ongoing action research project aimed at developing a comprehensive CRF for English-as-a-second-or-foreign-language (ESL/EFL) teachers to advise students and deliver feedback about their preparedness for the Eiken second-stage speaking test (Eiken interview). It details the steps taken to create a prototype CRF (currently in piloting). The prototype is a step toward addressing two research questions: 1) whether a more structured method exists to offer students feedback and advice to enhance their performance in the Eiken interview and, more broadly, 2) how teachers can ensure that their test preparation feedback is effective.

Literature Review

Why Not Traditional Rubrics?

Traditional rubrics—whether holistic, analytical, descriptive, single-point, or checklist—generally take one of four common approaches, which Brown (2018) refers to as "all-or-nothing," "target-level," "matter-of-degrees," and "multiple-features" (p. 10). Although each has its merit, a general reliance on counts, scores, and scales often proves to be punitive rather than developmental. In short, rubrics tend to focus on *quantifying* performance rather than *qualitatively* reflecting on a student's progress, and this approach can lead to assessments that are more about penalizing errors than fostering growth.

Furthermore, the use of negative and ambiguous expressions in such rubrics can demotivate learners by providing feedback that is neither constructive nor actionable. This tendency is especially problematic in language learning, where continuous engagement and positive reinforcement are crucial. It is for this reason that the current research digresses from such established methods to advocate for the progressive rules for developmental rubric writing established by Griffin and Francis (2018b).

In short, while traditional rubrics may be useful for *scoring* purposes (i.e., helping evaluators determine a test-taker's score), they fall short when the goal is to provide students with advice on how to improve their performance. As a result, they are not well suited for instructors aiming to help students prepare effectively for tests.

Why Criterion-Referenced Frameworks?

In contrast to traditional rubrics, Patrick Griffin's criterion-referenced framework (CRF) measures student performance against specific, well-defined capabilities, so as to reduce the propensity for rater biases such as norm-referenced comparisons, leniency/severity, central tendency, and halo effects, which occur when criteria are ambiguous. The CRF is particularly suitable for performance-based, formative assessment in ESL/EFL classrooms as it focuses on individual achievement and skill mastery on a developmental continuum. It stands out for its emphasis on assessing proficiency and its position as a student-centered educational approach.

Griffin (2018) argues that the prevailing "deficit model" (p. xv) in traditional rubrics often frames assessment as a tool for identifying what students lack, tasking teachers with fixing these gaps without providing adequate scaffolding. This approach can be particularly counterproductive in ESL/EFL contexts, where it might penalize rather than support learners. Griffin challenges this model by advocating for a developmental

approach, which shifts the focus from deficiencies to readiness for learning. By contrast, a traditional rubric, with its binary proposition, is unsuitable for providing feedback on test performance as it fails to offer students the detailed information needed to formulate better responses.

Teachers can use the descriptive, indicative behaviors in a CRF to identify specific points for improvement, set goals based on students' current abilities and potential for growth, and recommend strategies for improvement. In this way, the process encourages teachers to make clinical judgements and "...[consider] and [interpret] the learner's response to gauge the effect of the intervention on learning" (Clarke et al., 2020, p. 388).

This approach is crucial in language learning, where clear, actionable feedback is essential for progress. Griffin's framework aligns capabilities with learning objectives, ensuring that assessments are transparent. This developmental methodology thus transforms assessment instruments into tools that can facilitate dialog between teacher and student.

Methodology

Narrowing the focus on the problem through a process outlined by Wallace (1998), the action research project reported on here focuses on the author's previous experience working in a Japanese high school, where a perceived lack of interview preparation support was noted. These observations led to a consideration of potential tools that could be provided to assist teachers and students to bridge the "...gaps between what is actually happening in our teaching situation and what we would ideally like to see happening..." (Burns, 2009, p. 2).

In short, this paper is a report on the prototype development process. Specifically, it describes a preliminary iteration of the cyclical plan-act-observe-reflect action research model based on Kemmis and McTaggart (1988, as cited in Burns, 2009). The research design has been modeled after the instructive paper published by Kashio et al. (2019), who developed a similar rubric using an approach adapted from Brown (2012).

Stage 1: Plan

Griffin et al. (2018) recommend several critical steps for those preparing to develop a CRF. The first step is to identify the construct that the framework will measure. Once the construct is defined, it should be described in terms of observable evidence, known as manifest criteria. This step involves outlining the specific behaviors or responses that indicate the various target proficiencies, after which teachers can "collect evidence of proficiency, interpreting it as representative of a stage of a stage or level... and repeating these actions after a teaching intervention or learning period" (Griffin et al., 2018, p. 46).

Identifying the Construct

To identify the construct for the context outlined above, Dreyfus and Dreyfus's (1980) model of skill acquisition was used as a suitable developmental framework for the prototype CRF. Then curriculum progressions were identified (Griffin et al., 2018, p. 60) to derive domain, strand, capability, indicative behaviors, and quality criteria. In this model, the progression of learning skills is presented through five stages: novice, advanced beginner, competent, proficient, and expert. Each stage is characterized by an increasing level of speaking proficiency and a shift from reliance on fixed rules to intuitive responses. The design of the rubric is also informed by the proficiencies identified in the construct map, with the aim of assisting the identification of learners' current skill levels and the provision of nuanced feedback.

Construct Map

The levels of proficiency in the construct map in Table 1 are a cross-reference of the Eiken Foundation of Japan's (2008) can-do list of examination standard descriptions and the Japanese Ministry of Education, Culture, Sports, Science and Technology's *High School Foreign Language Goals for the English Communication Curriculum* (MEXT, 2018; see Appendix A). In accordance with established precedence (see Fujita et al., 2016), the resultant levels of proficiency have been worded to capture the core ideas central to assessing a student's progress, consistent with the standards established by both interdependent stakeholders.

Table 1
Construct Map

	Levels of proficiency	MEXT High School English Communication Curriculum	Jitsuyo Eigo Gino Kentei (Eiken Test)
Outstanding	Students discuss complex social issues reflecting various perspectives employing figurative language. Students can tell spontaneous or unrehearsed stories.	Higher education and beyond	Grade 1
High	Students speak about a variety of topics expressing empathy and self-reflection. They demonstrate sophisticated vocabulary. Students can tell stories with little preparation.		Grade Pre-1
Medium	Students are able to speak about common social issues and reflect on how those issues affect others. They demonstrate variety in vocabulary and sentences. Students tell stories which describe both a scene and the characters when given time to prepare.	English Communication III (H.S. Year 3) English Communication II (H.S. Year 2)	Grade 2
Low	Students are able to speak about everyday topics. They can reproduce basic vocabulary and simple sentences, tell stories in a logical manner and can describe a scene when given time to prepare.	English Communication I (H.S. Year 1)	Grade Pre-2

The construct map informs the development of a criterion-referenced framework for assessing student performance in the Eiken interview and allows the developer to target specific skills—in this case those pertinent to Question 2. Although high school students typically take Grade Pre-2 and Grade 2 tests, the range of individual English proficiency varies broadly based on a variety of external factors. To accommodate students operating at advanced levels, this construct map also outlines skills necessary to demonstrate proficiency at Grades Pre-1 and 1. Another reason for this inclusion is that universities in Japan may award points to applicants based on a higher Eiken grade as part of the admission process or may consider externally conducted test results as an alternative to the English component of the entrance examination. As such, high school students in their final year are highly motivated to pass higher levels of the Eiken test. In Table 1, each level of proficiency requires students to demonstrate a) familiarity with a variety of increasingly complex topics, b) increasing command over a variety of language features, c) an ability to construct stories spontaneously, and d) an ability to present ideas logically to enhance communication.

Rules for Writing Quality Criteria

In constructing the criterion-referenced framework for this project, the methodology was guided by established rules for writing *quality criteria*—a term used to describe criteria that comply with each of the 10 rules presented in Table 2. These rules are crucial for ensuring clarity and consistency and for demonstrating a developmental progression.

The quality criteria were written to distinguish between varying levels of performance. Care was taken to ensure that each criterion contained a single, central idea that was directly observable. The criteria were also written to address just one question in the Eiken interview to ensure that the framework was robust and that it provided targeted feedback to students about specific demonstrated capabilities.

Table 2

Rules for Writing Quality Criteria

Rule 1:	Enable an inference to be made about developmental learning – there should be no counts of things right and wrong or pseudo-counts (e.g., some, many, etc.)
Rule 2:	Avoid language that is ambiguous or contains comparative terms (e.g., appropriate, suitable, adequate) to define quality of performance
Rule 3:	Discriminate between performances of increasing quality without procedural steps in a sequence of operations
Rule 4:	Describe performances such that each successive description implies a progressively higher level of performance quality
Rule 5:	Contain one central idea that can be recognized
Rule 6:	Be directly observable (do, say, make, write) and avoid negatives
Rule 7:	Reflect typical behaviours that cover a diverse range of quality, including a stretch for the most proficient
Rule 8:	Self-weight based on capacity to separate by performance quality, i.e. no weightings are to be used
Rule 9:	Have four or fewer criteria for any indicator (to support consistency of judgements)
Rule 10:	Be transparent so persons assessed can verify their assessment – no jargon

(Assessment Research Center, 2015, February 9; Griffin & Francis, 2018b, pp. 124–131)

Stage 2: Act (Prototype)

In this stage of the action research, a pairwise comparison process was employed, as described by Griffin and Francis (2018a, pp. 170–173) and best illustrated by Lawless (2022) in a recorded demonstration. The process begins by evaluating the quality criteria (developed in planning) for each capability regarding their relative complexity. Teachers typically work in pairs within professional learning teams (PLTs) to consider two items at a time and ascertain which is more advanced.

This decision-making method can help in simplifying complex distinctions by breaking them down into a series of simpler choices. However, it becomes increasingly complex and less practical as the number of indicative behaviors grows. For this reason, Griffin (2018b) suggests that CRFs should be limited to six indicative behaviors, with no more than four quality criteria each.

For this specific project, the process was undertaken during the author's postgraduate studies at the University of Melbourne, and peer input was gathered through an online PLT established as part of a clinical teaching course. Over the course of a semester, an iterative approach to development with modifications based on peer feedback was undertaken. The details of these revisions were documented and later presented at a conference by Saito (2023), highlighting the method's application in an educational setting.

The result of application of the process delineated in the plan and action stages above is the Eiken Interview (Picture Description Task) Criterion-Referenced Framework (Appendix B).

Stage 3: Observe (Test)

To evaluate the effectiveness of the CRF prototype, a trial was conducted with a small group of 2nd-year students ($N = 5$) at a private high school in Tokyo, all of whom gave their informed consent after the school granted authorization to gather a small sample of anonymized responses. For the purpose of illustrating the analysis, a single response representating common errors and typical elements of the stimulus that students commonly chose to describe appears in Appendix C.

As the students articulated their responses to Eiken interview Question 2, the interviewer annotated the CRF by circling the corresponding criteria. Following the interview, the students were provided with a copy of the annotated CRF, accompanied by a verbal, bilingual (Japanese/English) explanation of each demonstrated capability in relation to the relevant question. Detailed feedback was later sent to the students via email. A sample of this feedback for Student A is provided (Appendix D) and has been color coded to show adherence to the task level, process level, and self-regulation level aspects of the *feed-up*, *feedback*, and *feedforward* as delineated in Hattie and Timperley (2007).

In Appendix D, feed-up focuses on clarifying learning objectives of the Eiken interview and emphasizing criteria for success, thereby enhancing student engagement with the task. Feedback is mainly achieved through the marked-up CRF. However, specific

examples are recorded to illustrate which aspects of a response correspond with the criteria in each rubric (denoted numerically). Finally, feedforward, which offers guidance on future actions, recommends an appropriate interview level based on these insights and addresses the query "Where to next?"

Stage 4: Reflect

In the reflection stage of the action research cycle, the focus is on evaluating and learning from the experiences and challenges encountered during the initial cycle of implementation. In this instance, the reflection cycle unearthed the following limitations of the prototype CRF.

One significant challenge was the students' limited comprehension of the CRF, primarily due to the language barrier, as it was written in English. During the assessment of students at Eiken Pre-2 and 2 levels, it was anecdotally observed that all students struggled with higher order verbs such as "speculates," "distinguishes," and "makes inferences," as well as metalanguage concepts such as reported speech, personal pronouns, and present participles. To address this issue, the CRF was revised to include language samples. However, no student in the small group could confidently comprehend all the quality criteria. In recognition of this shortcoming, a translated version of the CRF was developed.

Another reflection centered on the need for providing bilingual feedback. When providing bilingual, verbal feedback, a structured approach would enhance the feedback process. A translated feedback matrix, inspired by Brooks et al. (2019), is planned for use with the next iteration of this project to ensure that the feedback is accessible and actionable.

Additionally, it was observed that students were unfamiliar with criteria and rubrics, which required foundational instruction on how to read and interpret them. When queried, all students responded that they had not ever encountered a rubric in their English classes during the entirety of their formative schooling. Each participant needed to be taught how to interpret a rubric, including understanding levels of progressive complexity in tasks and descriptions of competencies across two axes in the table. This instruction was provided on an individual basis, which proved to be time-consuming. In future iterations of the study, it would be more efficient to explain rubrics as a group activity.

Discussion

Implications for Educators and Policy Makers

For regular classroom teachers who have not received training from the Eiken Foundation of Japan to interview test candidates, the standards by which a student's response is judged can be difficult to discern.

Coupled with the knowledge that "…about one-quarter (24%) of Japanese teachers reported that they do not feel prepared to teach the content, pedagogy and practical components of the subjects they teach" (Organisation for Economic Co-operation and Development [OECD], 2018, p. 55) and that "…numerous studies have also discussed the difficulties for [Japanese Teachers of English] of implementing methodologies that have a communicative focus while fulfilling examination preparation needs" (Thompson & Woodman, 2019, p. 49), it becomes clear that a significant disconnect exists between the objectives of the Eiken interview and the practical abilities of teachers to effectively prepare students for it. Therefore, tools such as the prototype CRF presented in this paper for formative assessment may foster a positive backwash effect, ensuring greater cohesion between the test and the curriculum. In addition, educators should be trained in the development of their own CRFs, thereby enabling them to provide feedback to meet their students' specific needs.

Educational Impact

The strength of this research lies in its demonstration of how a CRF, as opposed to a traditional rubric, can be used to provide feedback to students preparing for various standardized speaking tests and tasks. For example, the speaking components of the TOEIC and GTEC tests include similar image description tasks. This versatility extends to all productive skills, providing formative feedback tailored to specific learning objectives. Grounding that feedback in criterion-referenced frameworks can also promote more consistent evaluations compared to traditional rubrics. Arguably a CRF, centered around specific learning objectives, provides an objective method to assess performance-based tasks.

Challenges and Limitations

The CRF developed in this research, while robust in its current form, must be viewed as a living document, subject to ongoing review and refinement. Continued research, building on the insights from this paper, will extend the current project by replacing

the existing action research methodology with the more efficacious collaborative action research approach. Embracing such a collaborative focus in action research is significantly beneficial since, as Burns (1999) explains, "[c]ollaborative action is potentially more empowering than action research conducted individually as it offers a strong framework for whole-school change" (p. 13).

Another challenge is the need for continuous adaptation and revision of tools in response to the evolving standards set by both test providers and MEXT. A recent announcement by the Eiken Foundation of Japan (2023), introducing a new level or standard between the current Grade Pre-2 and Grade 2, exemplifies this need for flexibiility. The change affects the structure of the Eiken test and has significant implications for the CRF developed in this research. Namely, the can-do list, initially developed in 2008, serves as a foundational element of the CRF and so must be revisited to reflect the subsequent changes to the test structure.

These changes will necessitate additional quality criteria and the addition of another level in the progression. The process of continual adaptation is crucial to maintaining the relevance and effectiveness of the CRF, ensuring that it aligns with the latest standards and accurately guides students in their preparation for the Eiken interview.

Conclusion

In conclusion, this paper presented an illustrative model for creating a comprehensive criterion-referenced framework aimed at supporting students' efforts to improve specific speaking skills. The research process has led to an innovative approach that exemplifies the clinical model of teaching (Clarke et al., 2020; Kriewaldt & Turnidge, 2013). Furthermore, this model can be adapted as a tool for use in ESL/EFL classrooms. Its significance is underscored by its practicality as an instrument for enhancing feedback in performance-based tasks. By advocating for the use of criterion-referenced frameworks, this paper seeks to inspire educators to adopt a more structured, objective, and meaningful approach to student feedback.

Finally, and perhaps more modestly, the future development of this project will aim to refine the current CRF prototype on the basis of input from teachers and students, with the intention of producing a valuable tool for educators charged with preparing students for the Eiken interview.

Bio Data

Tim Saito is an associate professor at Toyama College of Welfare Science with a master's degree in international education from the University of Melbourne, Australia. He is also an experienced high school and university educator. His research interests include assessment and evaluation, curriculum design, digital and information literacy, and international and comparative education. He can often be found trying and failing to hit the target at his local *kyūdō dōjō*.

References

Assessment Research Center. (2015, February 9). Rules for writing quality criteria. *Reliable Rubrics*. Retrieved May 6, 2023, from https://reliablerubrics.com/2015/02/09/rules-for-writing-quality-criteria/

Borg, P. (2023). Has school English language education in Japan been a "failure?" *The Journal of Gifu Kyoritsu University*, *56*(2), 127–142. http://gku-repository.gku.ac.jp/bitstream/11207/2211/1/ronshu_56%282%29_127_borg.pdf

Brooks, C., Carroll, A., Gillies, R. M., & Hattie, J. (2019). A matrix of feedback for learning. *Australian Journal of Teacher Education (Online)*, *44*(4), 14–32. https://doi.org/10.14221/ajte.2018v44n4.2

Brown, J. D. (2012). *Developing, using, and analyzing rubrics in language assessment with case studies in Asian and Pacific languages*. National Foreign Language Resource Center, University of Hawaii.

Brown, J. D. (2018). Developing rubrics: What steps are needed? *Shiken: Testing and Evaluation SIG Newsletter*, *22*(1), 7–13. https://teval.jalt.org/sites/default/files/22_01_07_Brown_Statistics_Corner.pdf

Burns, A. (1999). *Collaborative action research for English language teachers*. Cambridge University Press.

Burns, A. (2009). *Doing action research in English language teaching: A guide for practitioners*. Routledge. https://doi.org/10.4324/9780203863466

Clarke, K., Truckenbrodt, A., Kriewaldt, J., Angelico, T., & Windsor, S. (2020). Beginning teachers' developing clinical judgement: Knowledge, skills and attributes for clinical teaching. *Teaching Education*, *31*(4), 381–403. https://doi.org/10.1080/10476210.2019.1594755

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. Operations Research Center, University of California, Berkeley Berkeley, CA. Retrieved from https://apps.dtic.mil/sti/pdfs/ADA084551.pdf

Eiken Foundation of Japan. (2008). The Eiken can-do list: English translation. Retrieved January 10, 2024, from https://www.Eiken.or.jp/Eiken/exam/cando/pdf/Eiken_CandoList_translation.pdf

Eiken Foundation of Japan. (2023, September 29). *Eiken-ni shinsetsu-kyū (jun-2-kyū-to 2-kyū-no aida) dōnyū-no oshirase* (英検に新設級（準 2 級と 2 級の間）導入のお知らせ) [Announcement of the introduction of a new Step Test level beteween grades Pre-2 and 2] https://www.Eiken.or.jp/Eiken/info/2023/pdf/20230929_info_newgrade.pdf

Fujita, R., Yokouchi, Y., Matsuoka, D., Nakamura, K., & Hirai, A. (2016). Eiken 2-kyū-no tesuto mondai-no bunseki: CEFR reberu, gakushū tōtatsu mokuhyō, hakyū kōka-no kanten-kara (英検2級のテスト問題の分析—CEFR レベル，学習到達目標，波及効果の観点から) [Examination of the Eiken Grade 2 test items: Focusing on CEFR levels, achievement goals, and washback effects]. 関東甲信越英語教育学会誌 *[KATE Journal]*, *30*, 85–97. https://doi.org/10.20806/katejournal.30.0_85

GRAS Group Inc. (2014). *Eiken 2-kyū paato-betsu 1–2 e-no byôsha*（英検2級パート別1–2 絵の描写）*[Eiken Grade 2 Part 1–2 Drawing Description]*. Retrieved March 30, 2021 from https://eikaiwa.weblio.jp/information/test-based-material/Eiken/Eiken2-batch1-2

Griffin, P. (Ed.). (2018). *Assessment for teaching* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108116053

Griffin, P., & Francis, M. (2018a). Building a development progression. In P. Griffin (Ed.), *Assessment for teaching* (pp. 161–197). Cambridge University Press. https://doi.org/10.1017/9781108116053

Griffin, P., & Francis, M. (2018b). Writing rubrics. In P. Griffin (Ed.), *Assessment for teaching* (pp. 113–140). Cambridge University Press. https://doi.org/10.1017/9781108116053

Griffin, P., Rickards, F., & Francis, M. (2018). Competence assessment: A clinical approach. In P. Griffin (Ed.), *Assessment for teaching* (pp. 43-76). Cambridge University Press. https://doi.org/10.1017/9781108116053

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Humphries, S. C., Burns, A., & Tanaka, T. (2015). "My head became blank and I couldn't speak": Classroom factors that influence English speaking. *The Asian Journal of Applied Linguistics*, *2*(3), 164–175. https://caes.hku.hk/ajal/index.php/ajal/article/view/230

Kashio, F., Ueda, T., Minami, C., & Takemoto, S. (2019). Jiki gakushū shidō yōryō-ni okeru mittsu-no shishitsu, nōryoku-o hakaru supiikingu hyōka ruuburikku-no yūyōsei (次期学習指導要領における 3 つの資質・能力を測るスピーキング評価ルーブリックの有用性) [The usefulness of the speaking evaluation rubric to measure three qualities and abilities in the next curriculum guidelines] *Eiken Bulletin 31*, 32–57. https://www.eiken.or.jp/center_for_research/pdf/bulletin/vol31/vol_31_p32-p57.pdf

Kriewaldt, J., & Turnidge, D. (2013). Conceptualising an approach to clinical reasoning in the education profession. *Australian Journal of Teacher Education, 38*(6), 103–115. http://doi.org/10.14221/ajte.2013v38n6.9

Lawless, B. (2022, November 5). *Pairwise comparison* [Video]. YouTube. https://www.youtube.com/watch?v=IqXPPTeXdZQ&t=421s

Masutani, Y. (2021). The foreign language anxiety of Japanese EFL learners: Focusing on anxiety when speaking English. *LET Kansai Chapter Collected Papers*, *19*, 1–14. https://doi.org/10.50924/letkansai.19.0_1

Ministry of Education, Culture, Sports, Science and Technology–Japan. (2018). Furoku 11: Kōtō gakkō "gaikokugo-no mokuhyō"-no kamoku kaidan-bestu ichiranhyō (付録11: 高等学校「外国語の目標」の科目段階別一覧表) [Appendix 11: High school "foreign language goals" list by subject stage]. Retrieved March 30, 2021, from https://www.mext.go.jp/content/1407196_29_1.pdf

Nichibei Eigo Gakuin. (2023, April 12). Eiken®-wa daigaku juken-de tsukaenai? Eiken® shutoku-ni yoru yūgū naiyō-ya nan-kyū-ga hitsuyō-ka dō-ka kaisetsu! (英検®は大学受験で使えない?英検®取得による優遇内容や何級が必要かどうか解説!) [Can Eiken® be used for university entrance exams? Explanation of the benefits of obtaining Eiken® and what grade is required!] *KotsuKotsu Web Magazine*. https://www.nichibeieigo.jp/kotsukotsu/exam/5856/

Nippon Communications Foundation. (2023, May 31). English proficiency targets still not being met in Japanese schools. https://www.nippon.com/en/japan-data/h01685/

Organisation for Economic Co-operation and Development. (2018). Education policy in Japan: Building bridges towards 2030. *Reviews of National Policies for Education*. https://www.oecd-ilibrary.org/content/publication/9789264302402-en

Saito, T. (2023, November 24–27). *Better than a score: Reimagining English rubrics in Japan* [Conference presentation]. JALT 2023 International Conference, Tsukuba, Ibaraki Prefecture, Japan.

Takeno, J., & Matsuura, K. (2018). Eiken nikyū mensetsu shiken "jun-ni-kyū"-no e-no byōsha-de mochiirareru dōshi-no tokuchō (英検二次面接試験「準 2 級」の絵の描写で用いられる動詞の特徴) [An analysis of verbs used in the picture description tasks for the Grade Pre-2 Eiken interview test]. *JAILA Journal*, *4*, pp. 85-96. https://doi.org/10.57359/jailajournal.4.0_85

Tatsukawa, K. (2018). The expected oral proficiency level for Japan's secondary school English teachers: Analysis of the Eiken pre-1st grade interview exam. *Journal of Pan-Pacific Association of Applied Linguistics*, *22*(1), 89–104. https://doi.org/https://doi.org/10.25256/PAAL.22.1.5

Thompson, G., & Woodman, K. (2019). Exploring Japanese high school English teachers' foreign language teacher efficacy beliefs. *Asia-Pacific Journal of Teacher Education*, *47*(1), 48–65. https://doi.org/10.1080/1359866X.2018.1498062

Wallace, M. J. (1998). *Action research for language teachers*. Cambridge University Press.

Whitley, H. (2017). Eiken interview: Tips for ALTs & students. Retrieved March 30, 2021, from https://hiroshimajet.files.wordpress.com/2017/06/eiken-prep.pdf

Wood, D. J. (2019). Testing English speaking in Japan. *Journal of Chikushi Jogakuen University*(14), 13–28. https://chikushi-u.repo.nii.ac.jp/record/977/files/14-02.pdf

Yamamoto, B. A. (Ed.). (2016). *Implementation and impact of the dual language IB DP programme in Japanese secondary schools*. https://www.ibo.org/globalassets/new-structure/research/pdfs/research-dp-in-japan-full-eng.pdf

Appendix A

High School English Communication Syllabus Objectives (Unoffical Translation by Author)

Ministry of Education, Culture, Sports, Science and Technology (MEXT)
High school learning guidelines handbook (2018 edition)
Foreign language edition and English edition
Appendix 11: High school "foreign language goals" list by subject stage

Objectives			
	English Communication I	English Communication II	English Communication III
Speaking (Expression)	A. Able to speak about everyday topics. Demonstrates ability to use basic vocabulary and simple sentences in a logical manner to convey information, thoughts and feelings when given time to prepare. Able to speak at a communicative level. B. Able to speak about socially relevant topics. Demonstrates ability to reproduce basic terms and sentences based on what the student has heard and read. Able to synthesize information with personal opinion in a logical manner when supported.	A. Able to speak about everyday topics. Demonstrates ability to use common vocabulary and sentences in a logical manner to convey detailed information about thoughts and feelings when given time to prepare. Able to speak and communicate. B. Able to speak about socially relevant topics. Demonstrates ability to reproduce a variety of terms and sentences structures based on what the student has heard and read. Able to synthesize information with personal opinion in a logical manner with some support.	A. Able to speak about everyday topics. Demonstrates ability to use a variety of vocabulary and sentences in a logical manner to convey detailed information about thoughts and feelings with limited preparation. B. Able to speak about socially relevant topics. Demonstrates ability to reproduce terms and sentences taking into account the purpose and context of the conversation and integrating what the student has heard and read. Able to independently synthesize information with personal opinion in a logical manner with limited support.

(MEXT, 2018, pp. 316–317)

Appendix B

Eiken Interview (Picture Description Task) Criterion-Referenced Framework

Eiken Interview (Picture Description Tasks)
Criterion-Referenced Framework

Japanese Eiken Interview Level	Quality Criteria Informing Theory: Patrick Griffin's (2018) Guidelines for writing quality developmental rubrics Dreyfus & Dreyfus' (1980) Model of Skill Acquisition						
Level 1		**1.2.4** Draws on visual cues to make inferences about the character's emotional state or thoughts *(e.g. "Mr. Ito was angry, his hands were shaking and his face was red.")*	**1.3.3** Invents names or other characteristics and attributes them to new characters *(e.g. Naming minor characters who fulfill small roles such as hotel staff, cashiers etc.)*	**2.1.4** Composes additional dialog consistent with the original style and context *(e.g. a greengrocer might say: "Would you like to try these juicy tomatoes from Kumamoto prefecture?")*	**2.2.4** Speculates about undepicted events which likely occurred during the course of the narrative *(e.g. If the character is wearing a different color shirt: "When Mayumi got home she had a bath and changed her clothes.")*	**3.1.3** Incorporates evaluative vocabulary *(e.g. Implies judgement through positive or negative adjective choice: "The kind man...", "The naughty child..", "The mean boss.")*	At **Level 1** a student can consider the entire scenario (beginning to end) and intuitively selects the most important aspects. The student can also deviate from typical storytelling patterns. The student makes inferences about the stimuli.
Level Pre-1	**1.1.3** Includes information about the setting of the story *(e.g. "at the supermarket...", "inside the cave...")*	**1.2.3** Distinguishes and references facial expressions, gestures, and other visual cues *(e.g. "Mrs. Sato smiled.")*		**2.1.3** Converts existing dialog to reported speech *(e.g. Adjusting verb tense, point of view etc.)*	**2.2.3** Integrates key information like time and place in recounting events *(e.g. "when they arrived...", "before they left the house...")*	**3.1.2** Applies vocabulary from the stimulus test in Task 1 to the new context *(e.g. If Task 1 includes a paragraph about sustainable energy, uses words like 'sustainable' in the story.)*	At **Level Pre-1** a student can extract nuanced information from the stimuli. They can demonstrate a planned or considered response. The student shows reasoning but may not be confident.
Level 2	**1.1.2** Describes people and objects *(e.g. "Mr and Mrs Yamamoto are a young energetic couple.")*	**1.2.2** Describes the characters motivations for their interactions with the environment *(e.g. "Kota thought it looked delicious so he put the fruit in his basket.")*	**1.3.2** Substitutes names and personal pronouns to avoid repetition *(e.g. "Kenji put his hat on the shelf. He decided not to go outside.")*	**2.1.2** Reads dialog bubbles or scene transition arrows as presented	**2.2.2** Creates a correctly sequenced narrative which expresses the passage of time *(e.g. "first", "then", "later", "after that", "finally")*		At **Level 2** a student makes minor adjustments to familiar stories about routine situations (eg. telling a story about a trip to the supermarket). They can link the stimuli to cause and effect.
Level Pre-2	**1.1.1** Recognizes simple objects, people, and places *(e.g. "cup", "plate", "man", "floor")*	**1.2.1** Selects correct present participle verb to convey the actions of each character *(e.g. "The woman is listening to music.")*	**1.3.1** Identifies character traits: gender, age, number with appropriate article *(e.g. "A young girl is...")*	**2.1.1** Describes the situation using 'so' and 'because' to communicate cause and effect *(e.g. "The man has no money **so** he cannot buy a drink.")*	**2.2.1** Identifies the situation, problem or main issue as depicted *(e.g. Notices that the cross over the man's wallet indicates he has no money and his intent.)*	**3.1.1** Uses everyday expressions and set phrases *(e.g. A: "Hi, how are you?" B: "I'm fine thank you.")*	At **Level Pre-2** a student can tell a pre-rehearsed or conventional story which draws on a set of plain, concrete, and observable characteristics of the stimuli.
Insufficient evidence	**1.1.0** Insufficient evidence	**1.2.0** Insufficient evidence	**1.3.0** Insufficient evidence	**2.1.0** Insufficient evidence	**2.2.0** Insufficient evidence	**3.1.0** Insufficient Evidence	**Insufficient Evidence**
Indicative behaviours:	**1.1** Identifies explicit key information	**1.2** Identifies implicit key information	**1.3** Assigns traits to characters	**2.1** Selects appropriate storytelling devices	**2.2** Develops a narrative related to stimulus images	**3.1** Chooses appropriate language	**Level Statements**
Capabilities:	**1.** Interprets a series of stimulus images			**2.** Create a short, logical, verbal narrative response		**3.** Communicates verbally	**Golden thread:** Considers all stimulus images, the contents, and the context to develop their narrative

英検二次試験の絵関係タスク向け

Criterion-Referenced Framework

英検面接レベル	判断基準 情報理論：ドレフュス(1980)のスキル習得モデルとグリフィン(2018)の評価基準ルール						
1級		**1.2.4** 視覚的な手がかりを利用して、キャラクターの感情状態や思考について推測します *(e.g. "Mr. Ito was angry, his hands were shaking and his face was red.")*	**1.3.3** 名前やその他の特性を発明し、それらを新しいキャラクターに帰属させる *(e.g. Naming minor characters who fulfill small roles such as hotel staff, cashiers etc.)*	**2.1.4** 元のスタイルとコンテキストと一致する追加のダイアログを作成する *(e.g. a greengrocer might say: "Would you like to try these juicy tomatoes from Kumamoto prefecture?")*	**2.2.4** 物語の過程で起こった可能性のある描写されていない出来事について推測する *(e.g. If the character is wearing a different color shirt: "When Mayumi got home she had a bath and changed her clothes.")*	**3.1.3** 評価語彙を組み込む *(e.g. Implies judgement through positive or negative adjective choice: "The kind man...", "The naughty child..", "The mean boss.")*	レベル1**では**、学生はシナリオ全体(最初から最後まで)を検討し、最も重要な側面を直感的に選択できます。生徒は、典型的なストーリーテリングパターンから逸脱することもできます。生徒は刺激について推論します。
準1級	**1.1.3** ストーリーの設定に関する情報が含まれています *(e.g. "at the supermarket...", "inside the cave...")*	**1.2.3** 顔の表情、ジェスチャー、その他の視覚的な手がかりを区別して参照する *(e.g. "Mrs. Sato smiled.")*		**2.1.3** 既存のダイアログを報告音声に変換する *(例:動詞の時制、視点の調整)*	**2.2.3** 時間や場所などの重要な情報をイベントのカウントに統合 *(e.g. "when they arrived", "before they left the house")*	**3.1.2** タスク1の刺激テストの語彙を新しいコンテキストに適用します *(例:タスク1に持続可能なエネルギーに関する記事が含まれている場合、ストーリーで「持続可能」などの単語を使用します)*	レベル Pre-1 **では**、学生は刺激から微妙な情報を抽出できます。彼らは計画的または考慮された応答を示すことができます。生徒は推論を示しますが、自信がないかもしれません。
2級	**1.1.2** 人物と物を記述する *(e.g. "Mr and Mrs Yamamoto are a young energetic couple.")*	**1.2.2** 人物が環境と相互作用する動機を描写する *(e.g. "Kota thought it looked delicious so he put the fruit in his basket.")*	**1.3.2** 繰り返しを避けるために名前と人称代名詞を置き換える *(e.g. "Kenji put his hat on the shelf. He decided not to go outside.")*	**2.1.2** ダイアログバブルまたはシーン遷移矢印を表示されたとおりに読み取ります	**2.2.2** 時間の経過を表現する正しく順序付けられた物語を作成します *(e.g. "first", "then", "later", "after that", "finally")*		レベル2**では**、学生は日常的な状況についてのおなじみの話を微調整しますたとえば、スーパーマーケットへの旅行についての話をします)。それらは刺激を原因と結果に結びつけることができます。
準2級	**1.1.1** 単純な物体、人、場所を認識する *(e.g. "cup", "plate", "man", "floor")*	**1.2.1** 各人物の行動を伝えるために正しい動詞を選ぶ *(e.g. "The woman is listening to music.")*	**1.3.1** 性別、年齢、数字を適切な冠詞で表す *(e.g. "A young girl is...")*	**2.1.1** 'so'や'because'を活用して状況を説明し、原因と結果を伝える *(e.g. "The man has no money **so** he cannot buy a drink.")*	**2.2.1** 描かれている状況、問題、主要な問題を特定する *(e.g. Notices that the cross over the man's wallet indicates he has no money and his intent.)*	**3.1.1** 日常表現や一般的なセットフレーズを使う *(e.g. A: "Hi, how are you?" B: "I'm fine thank you.")*	レベルPre-2**では**、学生は刺激の平易で具体的で観察可能な一連の特徴を利用した、事前にリハーサルされた、または従来のストーリーを語ることができます。
不十分な証拠	**1.1.0** 不十分な証拠	**1.2.0** 不十分な証拠	**1.3.0** 不十分な証拠	**2.1.0** 不十分な証拠	**2.2.0** 不十分な証拠	**3.1.0** 不十分な証拠	**不十分な証拠**
指標となる動作:	**1.1** 明示的なキー情報を識別する	**1.2** 暗黙的なキー情報を識別する	**1.3** キャラクターに特性を割り当てる	**2.1** 適切なストーリーテリングデバイスを選択する	**2.2** 刺激イメージに関する物語を展開する	**3.1** 適切な言語を選択する	**レベルステートメント**
資格：	**1.** 一連の刺激画像を解釈します			**2.** 短く、論理的で、口頭での物語の応答を作成します		**3.** 口頭でコミュニケーションをとる	**共通テーマ:** すべての刺激画像、内容、およびコンテキストを考慮して、物語を展開します

Appendix C

Simulated Student Sample

(GRAS Group Inc, 2014)

Student A

One day, Mr. Sato was talking to his wife about an advertisement. Mr Sato said to his wife "The Farmer's market will be open next weekend". He asked his wife if she was free the next weekend. The next weekend at the farmer's market, Mr Sato found saw some ~~a~~ tomatoes that looked delicious and he thought about ~~to~~ buying it. That evening, his wife ~~cooked~~ made tomato salad. So Mr Sato decided to wash the dishes after dinner.

Appendix D

Model Feedback

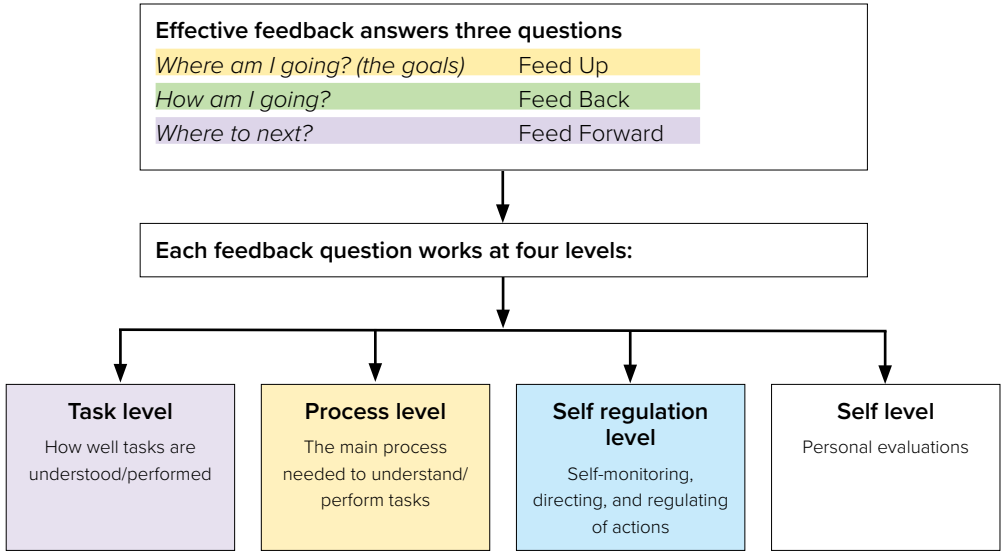
Student A, the goal of the spoken interview portion of the Eiken test is to demonstrate your ability to tell a story using the stimulus images. The feedback below details the skills which you have demonstrated and gives advice about how to advance to the next level.

Student A, you were able to recognize simple objects (nouns) like tomatoes, dishes, Mr. Sato and his wife (1.1.1). You have described the reason that Mr. Sato bought the tomatoes and why he decided to wash the dishes (1.2.1). You could use Mr. Sato's name and you also used words like 'he' and 'his' to refer to him (1.3.2). You were able to read each of the transitional arrows and the dialog bubbles (2.1.1). The story was told in the correct order and shows the passing of time using phrases like 'after dinner' (2.2.2). Your story also uses phrases from everyday life such as 'wash the dishes' (3.1.1).

*Based on this feedback you could take a **Grade Pre-2 interview** with a very high chance of success.*

The next step for you is to give more information about the things and people. You can check you're headed in the right direction by trying to describe their appearance. To do this, you can add adjectives which describe appearance (e.g., handsome), character or personality (e.g., polite), or feelings and emotions (e.g., nervous). *You should also try to convert the dialog to reported speech. For example: She said, 'I saw him.' (direct speech)*

▸ *She said that she had seen him (reported speech). If you can master these two skills you will be ready to challenge the **Grade 2 interview**.*



(Hattie & Timperley, 2007, p. 87)