

Stillman, Toni
FLT 808 - Dr. Winke
August 17, 2021

Assessment Design Project

Informing with Assessments

Background Information

A recent experience afforded me the opportunity to work with Boys and Girls Clubs of the Gulf Coast (Bay St. Louis, Pass Christian, D'Iberville, and East Biloxi Centers). The organization hosted a summer literacy workshop, featuring Introduction to Japanese language. (Sato) The program was short-term, serving a mixed population of students from the four centers. I worked from a lesson plan that would allow the learners sufficient time for introduction to the Kana syllabaries (Hiragana and Katakana) two days per week, two sessions, one hour per session. We also engaged in form-focused vocabulary and grammar. This brief teaching experience encouraged me to assess the effectiveness of my lesson plan, tasks and activities. However, there are other considerations for developing the assessment. I want to know if what is the washback and overall impact on improving the novice-level Japanese students' abilities in reading, and writing. Martel (2019) describes washback as "the effect of test on teaching and learning." Therefore, the intended purpose of classroom assessment is to improve learning outcomes.

Current trends in the field of language assessment are shifting away from narrowly defined perceptions of assessment as standard operating procedure in the field of education. The focus is on a broader interpretation.. Therefore, tests should be designed to improve the teaching and learning process. The purpose of this report is to evaluate assessment designs and instruments for use in testing novice-low to -mid-level Japanese language learners. The assessment instruments are formative and diagnostic. These tests are designed to assess students' reading comprehension writing ability. Another reason for the assessment is to inform lesson planning in the classroom context.

Stillman, Toni
FLT 808 - Dr. Winke
August 17, 2021

Context

This tests will be given to novice-low to novice-mid-learners of Japanese language. The assessment is based on the assumption that students have met requirements for reading Japanese script, Hiragana and Katakana. It also assumes that students will attend a daily class schedule for a minimum of forty-five minutes. With a frequent schedule, most students at the novice level will be able to create basic sentences and recognize the vocabulary that will be used on the tests. The tests are designed for formative assessment of the students' ability to apply their knowledge of reading and writing and demonstrate understanding in listening and speaking. Some test items are intent on expanding the learners' zone of proximal development or moving the student beyond the their current linguistic and content knowledge, Edwards (2013).

Overall Design

Assessments are a major role or the language teacher, whose responsibilities extend to designing, writing, administering, and scoring of assessments. Kunnan (2017) further confirms that the assessments may be administered as part of a unit or weekly test. It will generally consist of several task types: selected response, limited production, extended production, gap-fill. One test format, discussed in this report is limited production. To prepare students for the limited production assessment, which assesses grammar and vocabulary words, students should be scaffolded in giving "selected responses" Brown and Abeywickrama, 2018 p. 268. Furthermore, a limited production test will assess the use of single words (vocabulary), as well as the part of speech called Japanese particles, as used in completing written and spoken sentences. It is important to precede this test by activating the students background knowledge, with reading practice that uses the targeted vocabulary and particles, and scaffold sentences with fill-in-the blank quizzes. For the tests associated with this assessment, it is suggested that the test be administered by the teacher (may also be assisted by a teacher assistant) during and after a lesson unit on introducing self and other and Time or Days of the Month lessons.

Stillman, Toni
FLT 808 - Dr. Winke
August 17, 2021

Production Tests

In the Grammar Production instrument, students are asked to complete each conversation by selecting the correct form of a word and reproduce it in writing. The form of the word is shown, enclosed in parenthesis: examples: (____1____), (____2____); whereas 1 may require a particle, and 2 may require a verb. There are five questions in Part I of the test, with twelve fill-in-the-blank items. Brown and Abeywickrama (2018) describes this type of task as limited production. The students' response will provide a limited amount of language production. However, some items require that students produce a single word or several words, making up a phrase or idiomatic expression. The test will reveal whether the student can read the question in Japanese. For this test, the test-takers will answer questions related to the theme of introducing oneself and others, writing their response in the Japanese Hiragana script. Yamashita and Izuka (2017) suggests that tasks providing structured input will raise the students' ability to notice grammatical form. This test is designed to measure the students' knowledge of grammar and writing.

Interpretive Reading Picture Matching

Additionally, Part II of the Production test is designed for reading and vocabulary assessment. There are a total of three questions, with content language to recall days of the week. The students are instructed to read each sentence and choose one out of four multiple choice items that best describes the situation. Three answers have incorrect spellings, and the student must know the correct word, pronunciation and kana, in order to select the correct spelling of the day/month. The test items are meant to assess the learners' ability to read as well as grapheme recognition.

Furthermore, the Interpretive Reading Assessment is designed to assess the learners' ability to make grapheme and morpheme connections and sentence formation, without audio. The second part of the assessment is in T/F format, with audio and picture matching, it can classify as interpretive listening, as well. It is expected that the absolute novice learner may

Stillman, Toni
FLT 808 - Dr. Winke
August 17, 2021

experience some difficulty with the (audio) picture matching task. However, an additional task may be more suitable in that it will assess writing and speaking without audio. The final result of this reading and listening assessment will provide the teacher with insight into the learners' ability to produce vocabulary through a listening and reading test. It is presented in Hiragana and Romaji (Roman alphabet).

Scoring and Rating

In selecting what to assess, standards-driven principals of task design should also be considered. Test design ideas should improve positive washback for teachers and students. According to Shrum and Gilsan (2010), this indicates a paradigm shift in the educational practice of assessment (Adair ref). The authors suggests that previously assessment was limited in its purpose. It was once considered only to evaluate student achievement in the classroom, resulting in issuing grades. In an explanation of assessment and testing, Brown and Abeywickrama (2018) propose that assessment is an ongoing process that encompasses wide-range of methodological techniques. In addition, Cox Winke (2018) raise awareness of the importance of proficiency guidelines and its world-wide influence on language learning, which is also important to assessment design.

On the topic of assessment design, test instruments must also meet standards, which play an important role in education and instruction. Standardized testing is mainly used to assure the capability of the test-taker, from students, teachers, administrators and employees (public and private sectors) on all levels. Standards-based assessment (SBA), though laden with negative and unintentional impacts, evaluates achievement or failure in many areas. SBA certifies competence in secondary, trade schools, and higher education. Countries all around the world have standards that endorse everything from accreditation to validation of classroom assessment.

On the topic of classroom assessment, performance standards exist and relied upon to evaluate all four language skills: listening, speaking, reading, and writing. The American Council on the Teaching of Foreign Language is the professional organization that sets standards for performance in language education ([Home | ACTFL](#)). The organization provides guidelines on achievement, performance, and language proficiency. The World Readiness Standards lays out

Stillman, Toni
FLT 808 - Dr. Winke
August 17, 2021

specifics goal areas to be achieved within the ACTFL standards. These goals are known as the four Cs: cultures, connections, comparisons, communities. Foremost, included in the range of techniques for creating effective assessment design are the topics of scoring, grading, and feedback, which are discussed below.

Reporting and Feedback

Feedback and evaluation of curriculum are linked to the instructional function of assessment, according Bygate (2013), who suggest that feedback should be planned as a post-task feedback procedure. Tests are associated with scoring and grading, but feedback can be aligned with the tests. For the novice-level learners, feedback for assessments will be given in the form of a checklists. Students will be informed of areas that reveal their strengths and weaknesses. Instead of reporting scores, assessments of production and written tests will be received through peer assessment-evaluations, followed by whole-class discussion, which will provide input for future tests, as Edwards (2013) recommends for improving classroom or cooperative learning. Peer assessment is not limited to the pair format but can also be conducted in groups. It is possible to use peer feedback as part of the grading strategy. The greatest advantage to learners is that it creates student interaction. For teachers, there is the added advantage of saving administrative time.

Validity and Reliability

The final analysis of this report reflects on how validity is established for the tests, which will be used to assess learner performance and measured against goals and objectives designed around standards. To understand the value of test validation, a reference to Tsagari and Cheng (2017) is helpful:

It is, then, essential to establish the link between test validity and test consequences theoretically and empirically. It is imperative that washback and impact researchers work together with other language testing researchers, as well as educational policy makers and test agencies, to address the issue of validity, in particular, fairness and ethics of language tests, Tsagari and Cheng, 2017 p. 370.

Stillman, Toni
FLT 808 - Dr. Winke
August 17, 2021

The quote refers to large-scale testing, but there are similar requirements to establish validity in the classroom context. For this reason, the class will establish a post-test feedback discussion, and notes can be compiled for each students' comments, then shared. Additional validation will be confirmed by student checklists and self-evaluation. Another step in the validation process may require conducting some type of item analysis on specific questions to determine which questions were easy vs difficult, based on the item responses and student self-evaluation. The goal is to make every effort to design reliable and valid tests that achieve measurable outcomes. In sum,

Assessment is invaluable to improving student abilities in language reading and writing. Carefully designed tests on the large-scale can provide information that informs the broader community, i.e., leaders, policy and decisionmakers; and can be used to inform teaching and learning in the classroom context.

Stillman, Toni
FLT 808 - Dr. Winke
August 17, 2021

References

- Adair-Hauck, B., Donato, R., Cumo-Johanssen, P., Shrum, J. L., & Glisan, E. W. (2010). *Teacher's handbook: Contextualized foreign language instruction*.
- Cox, T. L., Malone, M. E., & Winke, P. (2018). Future directions in assessment: Influences of standards and implications for language learning. *Foreign Language Annals*, 51(1), 104-115.
- De Gruyter Mouton. Cohen, A. D. (1994a). *Assessing language ability in the classroom*. 2nd Edition. Boston: Newbury House
- Douglas Brown and Priyanvada Abeywickrama, Pearson Education, ISBN 0134860225
- Edwards, J. G. H. (2013). Peer assessment in the classroom. *The companion to language assessment*, 2, 730-750.
- Kunnan, A. J. (2017). *Evaluating language assessments*. Routledge.
- Martel, J. (2019). Washback of ACTFL's Integrated Performance Assessment in an Intensive Summer Language Program at the Tertiary Level. *Language Education & Assessment*, 2(2), 57-69.,
- Phakiti et al. (eds.), *The Palgrave Handbook of Applied Linguistics Research Methodology*, https://doi.org/10.1057/978-1-137-59900-1_3
- Yamashita, T., & Iizuka, T. (2017). The effectiveness of structured input and structured output on the acquisition of Japanese comparative sentences. *Foreign Language Annals*, 50(2), 387-397.
- Brunfaut, T. Assessing Listening (2016). In D. Tsagari & J. Banerjee (Eds.) *Handbook of Second Language Assessment*.
- Youn, S. J. (2018). Second language pragmatics. In *The Palgrave handbook of applied linguistics research methodology* (pp. 829-844). Palgrave Macmillan, London.

Stillman, Toni
FLT 808 - Dr. Winke
August 17, 2021

[Assessing Language Ability in the Classroom, 2nd ed. \(tesl-ej.org\)](#)

www.carla.umn.edu/assessment/vac/CreateUnit/p_1.html

[The Handbook of Language Teaching - Google Books](#)

[World-Readiness Standards for Learning Languages | ACTFL](#)

Stillman, Toni
 FLT808 - Dr. Winke
 August 17, 2021

Self -Assessment Rubric Design

Stillman, Toni FLT808 - Dr. Winke August 17, 2021	Self Assessment Rubric Grammar Production and Interpretive Reading		Japanese Language Novice low and Novice Mid-level Learners		
STRONG		GOOD	DEVELOPING	EMERGING	Low
My Strengths			My Goals		
	Comprehensibility Was I understood?				
	Vocabulary: I am knowledgeable of the test vocab				
	Language Control: I am mostly accurate when I use structures, time frames and word order that we practiced.				
	Pronunciation: My pronunciation is mostly accurate. I try to use native sounds Introduction of self and others				
	Fluency: I try to speak at a consistent rate.				
	Interaction How well did I communicate?				
	Content: My listening skills on the picture matching test are sufficient for understanding.				
	NOVICE Interaction: I use sentences, phrases and words. I				
	I am confident about the picture matching tasks				
	I am comfortable with grammar and vocabulary used on the Writing and grammar test.				
(January 2017)					
Adapted from Ohio Department of Education Rubric Instrument					