I used to develop tests only for the purpose of assigning students grades. When I was ready to assess my students’ knowledge of the content, I would think about what my co-teacher and I had taught in class and what students were supposed to know, and I would then make a test. The test would have matching, fill-in-the-blank, and sentences to translate. I built the tests so that they would be easy to grade. I did not consider that the students had little practice with written language or that they were not being assessed on their oral ability, even though oral communication had been the focus of class and their main way of practicing.

Learning about authentic assessments changed my assessment development process. Authentic assessment measures are formulated to measure classroom goals, curricula and instruction (O’Malley & Pierce, 1996). The assessments aim to discover what students have learned, their motivation and their attitudes toward class content. Authentic assessments can include portfolios, oral picture exams in which students use provided pictures to tell a story, or performing a skit, to name a few. The assessments should be measured with a rubric, checklist, or other type of scale. In addition, students need to be aware of the expectations of the assessment and the grade scale to be used before assessment begins.

I wanted to ensure that I was creating assessments that would inform my teaching practice and provide feedback to students. In the PACE Model unit developed for the story “qateryuk” or “grouse girl,” I developed an oral exam, a comprehension exam, self-assessments, and class evaluations. In order to provide grades and feedback to the students, I employed rubrics and checklists. The assessments developed for this unit are different than any I have developed before. These are authentic assessments.

It is important to ensure the assessments are both valid and reliable. Assessment validity can be established by testing what you have intended to measure. For example, if you want to measure your students’ ability to talk about the weather, then have them talk about the weather. A written exam or multiple-choice test on the weather will tell you more about a student’s reading and writing ability than their oral ability. In addition, students hope that tests will cover what is taught in class. Content validity means students are tested on what they were taught. Consequential validity is where the assessment outcomes are used to inform teaching (O’Malley & Pierce, 1996).Many types of assessments can have consequential validity as long as the teacher analyzes the assessment results, looking for ways to improve the class.

One of the assessments created for this unit is a paired oral exam where the students were asked to tell parts of the story. They are allowed to ask for help from their peer and the teacher, but it is best if they did as much as they can on their own first. This will show student’s individual ability to perform the task. It is common for the students to practice in pairs or groups in class, and students are typically encouraged to ask for help when needed. My co-teacher and I chose to incorporate paired oral assessments because the oral exam mirrors the classroom practice. This assessment measure has content validity because the students are able to communicate with their partner and the teacher in a simulated conversation about content we covered in class.

The oral exam also contains features of dynamic assessment. Dynamic assessment is a method of assessment that assesses the students’ language ability with assistance from the assessor. Poehner (2004) describes the process, “The challenge faced by mediators is to provide support that is neither too implicit and therefore ineffective, nor too explicit such that it threatens learner agency and self-regulation” (p.35). In a dynamics assessment like the paired oral exam, the goal is to both assess the students’ progress toward the unit objectives and continue to develop their language during the assessment (Poehner, 2004). In the exam described above, the assessor, my co-teacher Michael Bach, assisted the students through the exam by asking the students questions about the picture to inspire the student to continue talking. Where the student made errors, my co-teacher would correct the error and ask another question about the picture or narration. Michael used many mediating skills with the students during the exam, including asking clarifying questions, answering questions from the student, cueing them to add endings by only providing the root word, providing recasts, and asking the partner if she can help her peer. He also provided gestural cues to put the ending in past tense and to help students remember the words. This description of the oral exams sounds more like group work with an advanced student than an oral exam, but the speaking sample from the students provided enough information to determine where the student fell in the rubric.

Incorporating features of dynamic assessment added to the exam’s cultural validity. Cultural validity takes into account the different funds of knowledge that students bring to the classroom, and addressing the issues with the use of “one size fits all” testing and expectations. Alaska Native students are taught to be modest and not boastful, while many forms of assessment require students to talk about themselves or highlight their own work in a way that is incongruent with their upbringing. Anecdotal records, comprehension exams, portfolios and dynamic assessment can be culturally valid as they each accommodate students who are not boastful. In the dynamic assessment described above, there were students who needed to be asked many questions to get a quality sample of language for grading purposes. The ability to stray from a script by asking students more questions persuaded students to speak more. By helping students where they struggled, Michael gave them permission to try to use language they are not comfortable with, make mistakes and keep talking. This and the other assessment types above consider the cultural differences of the students in the assessment process.

The rubric used to grade the students for the oral exam measures the students’ progress toward the unit objectives. My co-teacher facilitated and video recorded the assessments so that we could both grade the students. Reliable assessments are those that when taken within a couple days of each other, the test taker would receive a very similar score (Pierce, 1996, p.19-25). Clearly defined guides such as rubrics and checklists decrease the amount of personal judgment that is included in the grading of assessments. When Michael and I graded the students, we assigned the students within half-point of each other. The rubric clearly described the abilities of a student earning each score.

The rubric used for the oral exam rates the students’ ability on Marzano’s four-point scale on three of the unit objectives (Marzano, 2009). In Marzano’s four-point scale, a score of three signifies that the student is proficient on the learning target. A score of four is assigned when the student goes above and beyond what was expected. A score of two is representative of a student who needs help completing the task but can almost complete it independently. A score of one indicates that a student cannot complete the task with help, and a score of zero is reserved for when a student makes no effort. Marzano’s scale will help me, as the instructor, to discover the areas where students have met the learning targets and the topics on which the class needs more practice. The objectives were created from the content of the unit to include vocabulary, past tense endings and intelligibility. These are all unit objectives that fit into the Alutiiq 1 and Alutiiq 2 sections of the Alutiiq course objectives and are parallel to the world language standards for the district.

I developed a comprehension exam to measure how much of the vocabulary and the story they understood in Alutiiq and to see if they could apply the cultural material we covered in the pre-storytelling activities to the story. This assessment was facilitated at the end of the co-construction phase. Students were asked to locate the vocabulary in the pictures, determine which scene from the story the assessor described in Alutiiq, and retell the story to the assessor in English. After students retold the story in English, the assessor asked them to talk about the lessons the story teaches. Students were graded using a checklist at the end of which included a space where I recorded the values or morals of the story identified by the students. This test had cultural validity because part of the assessment was to discuss the cultural values that the student learned in the story. As described before, cultural validity not only takes into account the students’ funds of knowledge but also gives their funds of knowledge a place in the curriculum and assessment (Nelson-Barber &Trumbull, 2007). I witnessed two students leave the test room with smiles on their faces due to the comprehension exam’s cultural validity. One student enjoyed the exam so much that she told me after class that she loved the exam. This particular student connected culturally and emotionally with the story and was able to articulate many cultural values that the story taught. The other student hears traditional stories from his great grandmother and related the experiences from her storytelling to the comprehension assessment.

In order to encourage students to reflect on their learning, I included a self-assessment in the form of a pre and post-test of the vocabulary from the unit, titled vocabulary self-assessment. This assessment was administered at the beginning of the unit to evaluate what vocabulary the students knew. The vocabulary self-assessment may have helped bring their attention to the vocabulary. Students heard and saw these words for the first time on the pre-test, encouraging students to watch for that specific vocabulary later in the unit. The vocabulary self-assessments include words from the story that mark important events, short communicative phrases, and other key vocabulary.

Self-assessment encourages students to take an active role in the learning environment. Pre- and post-tests inform students of their progress toward the objectives and provide great feedback to the instructor. Butler and Lee state, “through self-assessment, students can become aware of goals and expectations, monitor their learning processes and progress, and evaluate their own state of understanding against the goals and standards that are defined by the curriculum” (2010, p.8). Self-assessments are a low-stakes inventory of perceived achievement that encourages reflection on the content. I had been looking for something to inspire students to become more engaged in the learning process, and I knew that self-assessments would work for some of my students. After reading about the benefits of self-assessment for language learners, I developed reflective assessment measures. The self-assessment results informed me of changes I needed to make in the delivery of content, the grading structure, and the curriculum. I used information from the vocabulary self-assessment pre-test to determine how much time we spent on vocabulary and which students might need more help. I analyzed the post-test to inform my teaching of the vocabulary for the next time I teach this unit. Student success at learning the vocabulary reflects on my teaching skills and the lessons in the unit. If there are deficiencies that the whole group experienced, I have to look back at my lessons and make adjustments for the next time I teach the unit. Test results can inform me of changes I need to make in the delivery of content, the grading structure, or the curriculum.

While I incorporated assessments which evaluate students’ skills and retention of the material covered, I also include assessment measures that will evaluate the development of the PACE Model unit lessons and implementation. The phase assessments are adaptable to future lessons, as they were made with the overall model in mind, not specific to this particular unit of lessons. The assessment measures include a Likert scale, asking students to assess the teacher on a scale of 1-5. A Likert scales is a way of allowing respondents to rate statements based on how much they agree with the statement. The middle number represents a neutral response. The phase assessments contain a checklist of tasks for the instructor to cover in each phase. There is plenty of room on the form for students to offer written feedback. When we completed the presentation phase in class, I had my students fill out the phase assessment. Looking back at the presentation phase assessments, they included suggestions such as “name some words that we already know that are in the story,” “cut the story into segments,” “She did pretty well at engaging the students, but more props could provide a better visual.” They provided a lot of feedback and were specific enough that I know what they want me to spend more time on next time I teach this unit.

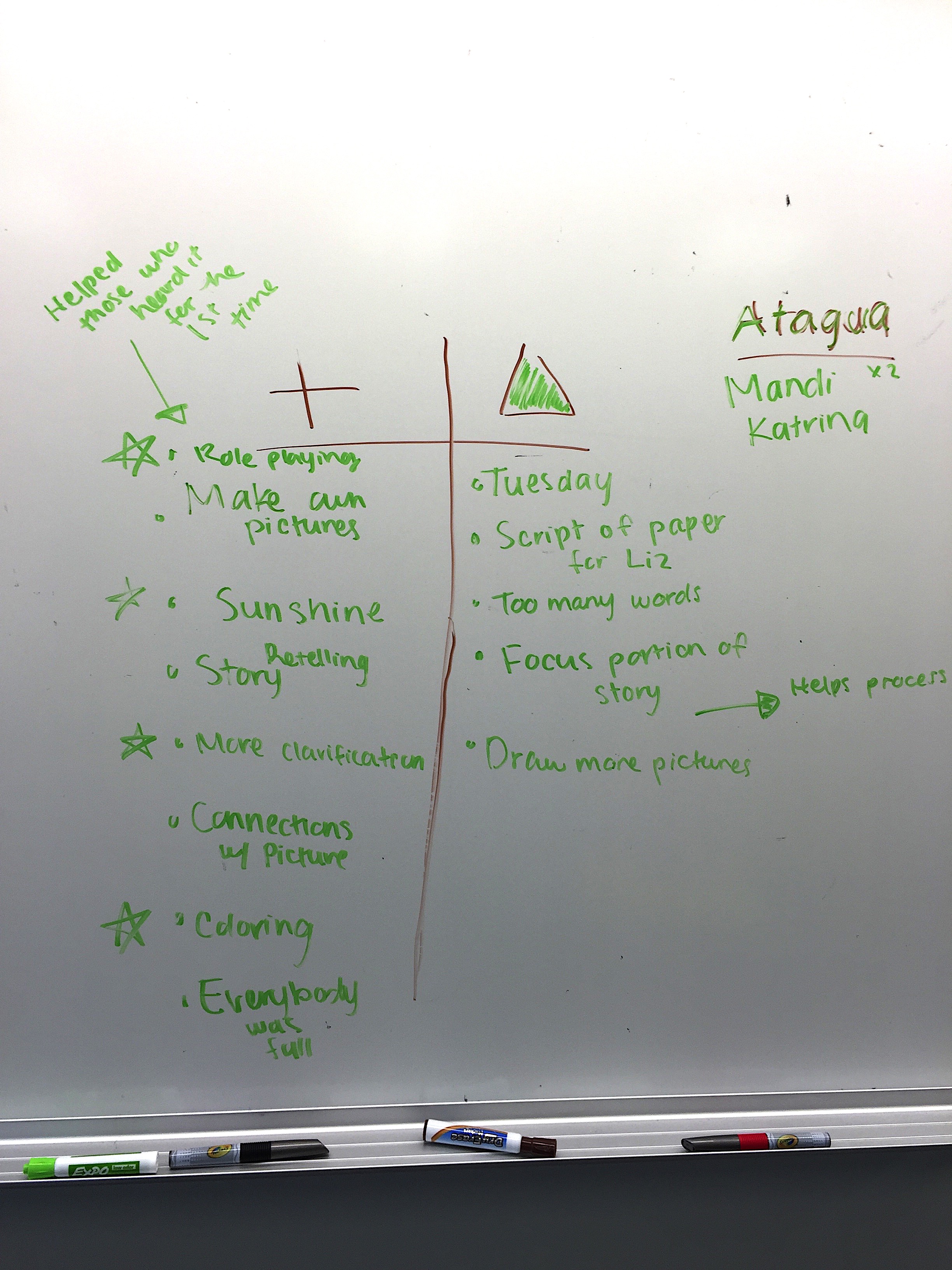
I also use what we call Plus Deltas almost daily to assess how students perceive the classroom environment, the teaching style, lessons, and their own progress. My students are accustomed to providing feedback during the Plus Deltas at the end of most class periods. Plus Deltas is a short reflective assessment procedure to evaluate the day, week, or lesson to inform the teachers of student preferences and needs. Students lead this activity as facilitators to encourage their peers to be open and honest. They are evaluating the classroom environment, the teachers’ delivery of the content, the lessons and games, and their perception and enjoyment of the class. A full step-by-step description is provided in the Teacher Guide document. Figure 1 is a picture of the whiteboard after completing a Plus Delta during the Presentation Phase of the Qateryuk Unit. This process has been very important for the growth of my teaching skills, the curriculum and activities that my co-teacher and I develop for class, as well as the success of the students. At first, students are uncomfortable participating in this process ~~at first~~, but after a few weeks of consistent practice they see its value. I try to make time for this activity daily at the end of class, by setting aside 10 minutes for the students to put stuff away and facilitate this activity. The results from these assessments inform my teaching on a daily basis and inspire many conversations between my co-teacher and I as we try to address student needs or concerns.

Figure 2. Plus Delta on whiteboard

The assessments designed for this unit of instruction are effective at accurately grading students’ ability, are culturally appropriate and add to a positive learning environment. I have included a table (Figure 2) of the assessments used in the unit and the elements that have made them valuable in the classroom. They provide me with adequate examples of student work and student feedback to adapt the curriculum to meet students where they are and improve my teaching practice.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Assessment** | **Assessment type** | **Measures** | **Modality** | **Rating Scale** |
| **Vocabulary Self Assessment** | Self-assessment | Written word recognition | Written, can be oral if instructor reads the words to students | Compare as pre and post- test |
| **Plus Delta** | Class evaluation | Classroom activities, environment and progress. | Oral, whole group | None |
| **Comprehension Checks** | Self-assessment | Student confidence in story comprehension | Oral, whole group | None |
| **Phase Assessment** | Teacher and self-assessment | Lessons and student progress. | Written | None |
| **Comprehension Exam** | Oral and comprehension |  | Oral, visual, aural, tactile | Checklist |
| **Podcast** | Oral |  | Oral, visual, aural, tactile, | Rubric |
| **Oral Exam** | Oral, paired, dynamic |  | Oral, visual, aural | Rubric |
| **Student Work Checklist** |  | Work completion | Combination of different assignments | Checklist |

Figure 2. Table of assessments used in the Qateryuk Unit.

**Reference List**

Butler,Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing, 27*(1), 5-31.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, UK: Cambridge University Press.

Marzano, R. (2009). Formative assessment & standards-based grading. Bloomington, IN: Marzano Research.

Nelson-Barber, S. & Trumbull, E. (2007). Making assessment practices valid for indigenous American students. *Journal of American Indian Education, 46*(3), 132-147.

O’Malley, J. M. & Pierce, L. V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Boston, MA: Adison-Wesley.

Poehner, M. E. (2004). Both sides of the conversation: The interplay between mediation and learner reciprocity in dynamic assessment. In Lantolf, J. P. and Peohner, M. E. (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 33-56). London: Equinox Publishing Ltd.