Webinar Series

Content or Language: Do you Know Which One You're Assessing?

Ildi Porter-Szucs, 13 May 2021

As time did not allow for all the participants' questions to be answered during the webinar, Ildi has kindly provided comments and answers below.

In order to help our students practice English, we might give them assignments or projects that we don't assess yet we check to make sure that they practice the language skills. It kills me when weak students avoid doing them knowing that, according to the educational system here, they will get no marks on them. Should I just lie and fake assessing their assignments to make the students do them?
I would like to begin by distinguishing assessing/assessment from grading/marking/evaluating. When you give the students assignments or projects, do you, the instructor, check how the students did on them? Or do you encourage the students to look up the answer or to reflect on what did and did not work for them? If the answer is yes, then there is (self-)assessment, the formative kind. I am probably not alone in arguing that this is the most important kind of classroom assessment. If the learners also receive a grade/mark or there are consequences attached to the assessment results, then we can speak of evaluation. Thus, if I understand your situation correctly, because some assignments are not evaluated, some learners choose not to do the work. We as teachers – and even some of our learners – know that they learn for themselves and not for us or for the system. However, many learners are externally motivated, and they will only do the work if it is interesting to them or if there are consequences. The consequences or incentives may come in the form of the educational system imposing penalties or rewards, but they may come from other sources as well. Young learners can be motivated with small gifts like stickers or watching a favourite movie during class-time or having extra recess time. Older learners also enjoy movies, free food, field trips, or a cancelled test. I have found that peer pressure can be very effective. If the prize is appealing to the vast majority of the class but can only be obtained if all students do the work, then the students can pressure each other to perform. Your class might also come up with projects whose results can be published, such as a book or a video or an exhibition in the classroom or the school's hallway. Whatever solution you and your students come up with would need to be culturally appropriate. I would prioritize incentivizing the students over outright lying to them though. :-)

What is your guidance about creative and collaborative writing tasks assessment in the most objective way in university online classes?
In both cases (creative and collaborative assessment tasks), rating scales are a must. The evaluation criteria need to be carefully chosen and should be fully understood by the students. The criteria

might originate from an accrediting body or other professional organization (e.g., one that governs the field the students are entering). The criteria may be jointly arrived at by experts at your institution. Alternatively, they may be co-constructed by you the instructor and your students. If this is a content-and-language-integrated task, the criteria will need to represent both language and content (including creativity).

When creating the criteria and later the rubric/rating scale, I would recommend looking at numerous writing samples at all the different levels. While examining the writing samples, ask yourselves which pieces of writing meet the minimum standard and why, and which ones do not and why. The creation of a rating scale is a somewhat lengthy process and is best undertaken as a team. At the very least, if you alone create the rating scale, your colleagues should weigh in before it is finalized. I would also recommend creating a separate tool to assess collaboration. A web search for peer-evaluation tools could yield examples. I find it beneficial when group members have to score each other's and their own contribution to the group effort and also narratively describe what part of the total project they each were responsible for and how satisfied they are with the effort and result.

In the case of a content-based collaborative task, I would recommend one analytic content rubric, one analytic language rubric, and one anonymous peer-evaluation instrument (Google Forms is well suited to this). The weighting between the content and language grades/marks will need to be calculated carefully to reflect the course's learning outcomes. The individual group members' scores can be totalled or averaged for each student and the final peer score can be added to the content-language assignment score. It can either be extra credit (potentially resulting in a final assignment score of more than 100%) or be the difference between a mark of 'good' vs. 'excellent.' The specifics will need to be worked out for your particular setting.

I'd like to know where/how you think context-/subject-specific skills fit into your scale (e.g., academic reading skills or seminar communication skills).

I am assuming that you are referring to the proposed language-content continuum I showed in slide 6. The guiding principle behind the continuum is the amount of language vs. content focus as well as the extent to which allowances or accommodations are made for language learners. Thus, the instruction of academic reading skills or seminar communication skills alone would not sway a programme's placement on the continuum, but the extent to which the programme targets language instruction to non-native English speakers (NNESs). For instance, a university EAP class may teach these skills. If the class is open to all ESL/EFL students, if the organizing principle in the class is around academic success skills, and if content is presented to learners as a vehicle to practice reading skills, then I would place such a class toward the language end of the continuum. On the other hand, if the class is only open to NNESs who will pursue a STEM field (such as the ESLN 562 course in slide 5), then study skills and technical skills are central but so is content. I would place this class toward the middle. Finally, if the course on academic success skills is open to all university students and the instructor makes no effort to accommodate for the needs of NNESs, then study skills become the content of the class and thus the class is content-focused, hence its placement toward the content-end of my continuum.

If the content and language are so closely intertwined, e.g. in university assignments, how possible is it to separate them in analytic marking criteria, e.g. a separate mark for language and content?

… the attendees continued the conversation on the question of how neatly it is possible to separate content and language if the two are intertwined. I would agree that it is not entirely possible. But as one of the respondents stated, the students are not concerned with precisely which rubric is being used or under which evaluation criterion the language of content (e.g., content-specific vocabulary) might fall. I would argue that what is much more important is intra- and inter-rater reliability. All raters/evaluators must be able to determine clearly, consistently and reliably which evaluation criteria should be applied and what score point a particular test-taker performance is worth.

What could be the most appropriate way for eliciting relevant vocabulary and getting students to complete the task within the language area?

I am not sure if I understand the context and the question precisely, but let me attempt a response. I am imagining a B1-level language-focused situation, where the learners have been exposed to content-specific vocabulary (such as by reading or listening/viewing about a topic like climate change or fashion marketing). The assessment is aimed at eliciting the new vocabulary. Depending on how authentic the task, the assessment can be more or less authentic as well. At the authentic, performance-based end of the spectrum, students can be asked to create a public service announcement/infomercial/YouTube or TikTok video/debate/survey/interview/skit/PowerPoint presentation/blog post/letter to the editor/article in the school newspaper/article for the school bulletin board, etc. The instructions and the rating scale need to clearly state the vocabulary requirement. If there is a list of new words (I personally rarely teach words without the formulaic sequences they are part of, thus a list of new formulaic sequences), the instructions may state how many must be used accurately for each score point/rating.

Do you know of a usable rating rubric or similar for evaluating "critical thinking"?

Much depends on the definition of critical thinking. Here is a good resource:
https://www.aacu.org/value/rubrics/critical-thinking
From youngest learner to oldest, here are four rubrics that I've used. Some of them require free registration:

1. https://cpb-us-w2.wpmucdn.com/blogs.acpsk12.org/dist/0/1149/files/2019/02/PBLworks-Critical-Thinking-Rubric-Grades-K-2-1k3ep8m.pdf
2. 2. https://my.pblworks.org/system/files/documents/PBLWorks-3-5-Critical-Thinking-Rubric-CCSS.pdf
3. https://my.pblworks.org/system/files/documents/Critical_Thinking_Rubric_for_PBL_%286-12%29%20CCSS.pdf
4. http://www.deanza.edu/slo/icctaskforce/sample_rubric_gittens.pdf

Have you detected a common denominator in evaluation/assessment criteria (in terms of changes, adjustments) from your professional experience?

Analytic rating scales are displacing holistic ones. Single-point and primary-trait rubrics are gaining in popularity. Checklist assessment and training learners to assess themselves and their peers and other forms of formative assessment are also replacing more traditional forms of assessment. As for particular evaluation criteria, in general I have observed that larger elements are finding their way into rating scales, such as suprasegmentals, discourse features, and task accomplishment.