

Assumptions

- Proficiency levels vary within each class.
- Proficiency will not change significantly over one semester—only 22½ hours.
- They probably won't ever need to take another language test.

Therefore:

- Assessment should be formative rather than summative—we should assess the learning process rather than the product of learning.
- Quantity of input is more important than quality of output.

My goals

- Engage students with the English language.
- Maximise input, in and out of the classroom.
- Give students tools to learn languages.

Student goal

Engage with a lot of English...

... how much is a lot?

Per semester:

- Read at least 100,000-200,000 words?
- Watch 2-3 movies 4-5 times each?
- Write 10,000 words?
- Speak for 10 hours?

This is more than one teacher can possibly keep track of!

Blended Learning

Computers as Teaching Assistants.

How technology can help:

- Instant scores possible (quizzes).
- Learning is collaborative and social.
- Scores and grades visible to students.

Moodle is an open-source Learner Management System (LMS). Moodle allows many activities and

keeps scores and records. The most useful are *forums*, *quizzes* and *wikis*. Forums can be used for different tasks and activities, for example: introductions, generating topics or ideas, choosing topics, projects, peer assessment, as well as traditional forum roles such as exchanging ideas.

Grading forum posts

Moodle allows forum posts to be graded, and scores from each post are aggregated, either using the highest score, lowest score, average score, count of scores or (most usefully) sum of scores.

The following criteria can be used for scoring each post:

- Appropriate
- Original
- Substantial
- Suitably titled
- Extended
- Intelligible

The big picture

Teachers can put grading items into different categories, and weight items within categories and categories within the total grade. Students can see their own scores to see how they are progressing.

Improvements

There are many areas in which the system could be made more effective or easier to use:

- Navigation
- Grading flexibility
- Global settings
- Anonymous forums

Tracking

As well as formally scored items, Moodle records

every mouse click, and every keystroke from students. These copious records are available. For example, around 8,000 records were kept for a class of 10 students over a semester. From these records, students' activity could be assessed. As well as total activity or frequency of activity, regularity of activity could be measured, encouraging regular study that we know leads to improvements in proficiency.

Challenges

- Computers have a steep learning curve.
- Teacher's role changes from communicating with students to managing communication between students.
- We can't see how big the pile is—it's easy to get an idea of how much grading there is with a pile of essays to mark, it can be difficult to see how much there is to do online.
- Are we testing computer literacy?
- More transparency means more complaints from students.

Strategies

- What to make transparent
- When to make it transparent

Assessment concepts

Summative or Formative?

For example, in extensive reading, the goal is for students to read many books. Summative assessment may measure improvements in reading speed, vocabulary size or comprehension.

Formative assessment, on the other hand, would measure the reading diet (what and how much students are reading), their attitudes towards reading and awareness of strategies.

Backwash

The effect of the assessment on students' activities. For example, university entrance exams are replete with tortuous passages where obtuse grammar is saturated with arcane vocabulary. As a result, high school students study obtuse grammar and arcane vocabulary. They do not practice simple grammar and high-frequency vocabulary, which make up the vast majority of language use, and many high school graduates cannot engage in simple conversations.

Reliability or Construct Validity?

No assessments are ideal. There is often a trade-off between reliability (getting an accurate measurement) and construct validity (measuring desirable outcomes). For example multiple-choice items can have very high reliability, but no relevance to the real world. Spoken interviews can closely resemble real-world tasks, but are expensive in time and subject to more variation.

References

- Brierley, M. & Orlandini, O. (2010). CALL Implementation Strategies—A Case Study. <https://soar-ir.shinshu-u.ac.jp/dspace/handle/10091/10220>
- Brierley, M., Ruzicka, D., Sato, H., & Wakasugi, T. (2010). The measurement problem in extensive reading: Students' attitudes. <http://jalt-publications.org/archive/proceedings/2009/E052.pdf>
- Brierley, M. (2010). Assessing extensive reading through written responses and comprehension tests. <http://jalt.org/pansig/2009/PDF/Brierley.pdf>