Test Design, Administration, and Grading

[from LATN 4770 courseguide; R.A. LaFleur; rev. 8-27-08]

There are countless ways to design tests, and to administer and grade them for that matter. The guidelines we use for our college-level beginning Latin tests are posted on the course website (www.classics.uga.edu/courses/latn4770/tests.htm), where you should not only examine those guidelines (which have at least a general applicability to tests for beginning Latin classes even at the K–12 level) but also to look at the variety of sample tests posted on the site (if you are already a teacher, send an e-mail to wheelocks@harpercollins.com and ask for access to the online "Teacher's Guide and Answer Key for*Wheelock's Latin*," where you can look at several actual tests and answer keys for classes using Wheelock. The site also contains sample lesson plans, quizzes, worksheets, handouts, etc.).

One thing is certain about testing: all your important teaching objectives can be undermined if you design poor "assessments," i.e., quizzes and tests: your assessments can be effective motivators for your students as well as measures of what they've learned. Every veteran teacher will tell you that if you want to motivate students to learn the sorts of skills that you think are important, to meet your learner objectives, then you need to test them on those skills, on what they need to know and be able to do. Who hasn't heard students ask, "Will this be on the test?" The answer to that question is simple: if it's important, then—*ita vero*!—it will!

So, learning objectives are the first, most important factor in test design, and you'll remember most of those objectives if you remember your SANDALS. The LEGITE objective of course involves reading about Roman culture, first and foremost, so don't forget that. Not every test will involve all objectives, but try to include as many as possible, and of course be sure to include questions that connect to each new topic in the particular chapters being tested. The LEGITE and SCRĪBITE objectives are among the easiest to assess. Some simple composition will test their ability to write in Latin. Or, combine writing with listening by asking them a question or two orally and having them write out their answers in Latin, ideally in complete Latin sentences.

To test their reading skills give them continuous narrative passages to read (not disconnected sentences) that have content reasonably familiar to them from their classwork (but not word-for-word the same as passages gone over in class), that are interesting, and that provide authentic insights into Roman culture, history, literature, etc. (in large part, of course, this will be a product of your textbook choices). Be sure to include comprehension and discussion questions and not merely translation; in fact, a good option is to include one passage for translation and another passage that students must read and answer questions on.

The AUDĪTE objective may be addressed with something as simple as a sentence dictation: I include one on each and every test, worth 5 to 10 points. SPECTĀTE, DĪCITE, and AGITE can be a bit more challenging to test in a classroom setting, but use your imagination: you could, for example, bring in a couple of props and perform an action or two with them, all involving familiar vocabulary, and then ask students to write out a sentence or two in Latin describing what you are doing (e.g.,*Magister ad mēnsam ambulat et librum in mēnsā pōnit*). And you can always assess students' oral reading and pronunciation individually (e.g., in five-minute sessions every

four to six weeks) or in small groups.

To prepare students for listening, speaking, reading, and writing, you must of course train, drill, and test them frequently on vocabulary, form production and recognition, and grammatical analysis (conjugation, declension, case usage, etc.). And don't forget another important "real-world" objective of English word-building: I recommend including etymology items on every test.

Doubtless you can't include all these sorts of items or assess all your learner objectives on every test, but you can certainly address them all over the course of a series of tests and quizzes. Think of your assessments in combination, as a total package, and not in isolation.

What are some other important considerations in test design?

> Cover all the objectives, yes, but keep in mind what you (and your syllabus) have told students to expect; tests shouldn't contain unpleasant "surprises." Let students know exactly the sorts of things that will be on the test, and they'll strive to learn them.

> How long is the testing period? Don't give students more work than they can reasonably complete in that amount of time. Be sure your questions are clearly worded and unambiguous.

> Use a 100-point scale (at least that's my recommendation—most students seem to prefer it, and it makes grading easier); indicate on the test the points or percentage that each item counts, so students know where to concentrate their efforts if time runs out (this means, by the way, that you will have to think in advance about how you want to weigh each item, and this can be most useful when it comes time for you to do the grading).

> Strive for a clean, elegant layout, and give students plenty of room for their answers; for some item types (declensions, conjugations, identification of case uses, etc.) lines or even charts to fill in can help students organize their answers and make your grading easier.

> One last thing: don't forget to proofread. In the process of designing, revising, and redesigning a test, or any classroom document, errors will crop up, and teacher errors on a test can be more than a little disconcerting.

Obviously, it is important to design a test that addresses objectives, is clear, and fair. How you administer tests is important too: arrive early, be cheery, help students relax, set up an environment that does not tempt students to copy from others, stay alert when proctoring and never leave the classroom while students are still working. I advise walking around the room right after distributing a test, just to be sure no one has inadvertently left a textbook or notebook open on the floor or has any notes in open view: when it comes to academic dishonesty, I think almost anyone at some point or other can be tempted: an important part of your role is to set up a temptation-free environment, because—*crede mihi*—in such matters, an ounce of prevention is worth a pound of cure!

Finally, your very best efforts in teaching and in designing and administering tests can be undermined by failures in grading. First, grade and return quizzes and tests promptly (quizzes the

next day or two, tests within a week at most). When I have graded tests and am ready to return them, I usually just mention them at the beginning of the period, with comments on how well students did (remember: use lots of positive reinforcement*COTĪDIĒ*!), and then I reserve the last five to ten minutes of the class period first to remark briefly on problem areas and what students need to review, and then to return their papers. Returning major tests at the beginning of class typically presents a major distraction, as students will want to keep looking them over while you're trying to teach, and those who did poorly will of course be in a funk for the entire remainder of the period!.

When I return graded tests, I always provide a thorough answer key that includes answers, with options, and occasional notes on problem areas: for samples, look at some of my LATN 1001 tests and keys on [the course website](www.classics.uga.edu/courses/latn4770/tests.htm) (www.classics.uga.edu/courses/latn4770/tests.htm). The purposes of testing are undermined if you don't provide feedback: you could, of course, write in all the correct answers for every error on every single student paper as you grade them, but—*heu*!—now that would take an awful lot of your precious time (you'll find out just how precious, once you begin teaching!), or you could go over every single test item during a class period, but that takes quite a bit of equally precious time, and in that case any students who are absent don't get the feedback.

I think a detailed key is the best option: I return tests at the end of class, along with the key and a few remarks on any common errors, and then tell students to look over their tests along with the key at home that evening and bring test and key back the next day. This procedure puts the onus on the students to analyze their own errors, and typically there are few if any questions on the test the next day in class. So, a little time preparing a key is a great learning device for the kids and saves you a lot of time grading and going over tests in class; you might choose to focus five to ten minutes of class time on a particular section of the test that next day, after your students have read through their own tests and the key at home, the translation passage for example, but you won't need to spend time on every item or even on most items.

You'll see from my tests and keys that I use the same precise sort of grading system for tests as I do for quizzes. You actually plan your grading system while designing the test. What are your most important objectives? Listening skills are pretty basic, so maybe put the dictation first: a five-word sentence is worth 5 points (and, *ita vērō*, spelling counts in dictations!). Certainly reading is paramount: make that your second test item, and weight it in accordance with its importance: I might count the reading/translation passage 30–35 points (out of 100) on the first test, but after that the passages get longer and count more, usually 40–50 percent of the entire test. Most other items can be 1 point each: a conjugation item requiring students to produce twelve forms (two tenses, singular and plural) might simply be 12 points, or a declension with ten forms might be 10 points: clear and easy to grade.

Sometimes I'll go with ½ point per item, let's say a noun-adjective declension requiring them to produce twenty forms, where I want to weight this item only 10 percent of the test. (By the way, you'll notice from my sample tests that when asking for conjugations/synopses, I always provide all principal parts, and for declensions I provide genitive as well as nominative: this keeps the items "clean," by which I mean that the items are not testing vocabulary but rather whether or not a student can conjugate a verb or decline a noun/adjective, given the information they need to

determine the verb stems and noun/adjective bases.) I tend to avoid other fractions other than ½ , except maybe in assessing their translations (where, e.g., I might have a 105-word passage worth 35 percent = 1/3; point per word).

Which raises the question, how*does* one grade translations? (And remember, please, that you can assess the reading objective by other means than translation, e.g., by reading comprehension questions.) Not holistically, I'd say, and not "impressionistically," please! Few things can agitate or demoralize students more than a grading system that seems ambiguous and which therefore can appear unfair. As an academic department head for over twenty years, I found that 90 percent of all student complaints were related to grades they felt were unfair, and nearly all of those involved instructors who used "holistic" letter-grading for tests and other assessments that could easily have been graded using a clear, relatively exact numerical system—easily done with Latin test items like declensions, conjugations, short-answer questions, and the like, and not difficult with translations either.

On the first tests in Latin I, I suggest a reading passage of 30–40 words and scored 30–40 points: the math is simple! Oh, and by the way, remember that the passage should be a continuously, logically flowing narrative, and though the topic should be somewhat familar, e.g., based on readings assigned for class, it should be a sight passage and not something you've actually assigned and gone over verbatim with students in class. If you do that very often, you'll be training students to scribble down translations feverishly during class and then memorize these English "transcripts" for test time. And be careful to use only words from the chapter vocabularies or that have otherwise been more or less formally introduced, though it's okay occasionally to use words whose meanings and function are easily deduced from context and from English derivatives (e.g., if they've had *patria* but not *antiqua*, they can nevertheless deduce the meaning of *patriam antiquam* in context), or to use a few other words they haven't seen but for which you provide glosses at the end of the passage.

Oh, but back to grading: make your passage approximately 40 words, let's say, and weight it 40 points/40 percent: one point per word. As the term progresses, the passages should become longer: 50 words at 50 points, for example. But here's where you can get into fractions: you may give them a 60-word passage and still want it to count only 40 percent—that's 2/3 point per word—or a 100-word passage that is 50 percent of the test = ½ point per word.

What about partial credit? I've developed a system that over the years has proven effective and that students understand and regard as quite fair. It's a pretty simple system: put an "X" (use red ink or some color other than the student has used) through the translation of a word that misses **two or more** "aspects" of that word's meaning, and draw a squiggly line under the translation of a word that misses only a **single** aspect of that word's meaning. The "X" loses full credit, the "squiggle" loses only half. Example:

Puer canem videt. (a 3-point test item)

The boy sees the dog.	Earns full credit
The boy sees the dogs.	Loses half credit, - ½ (missed one aspect: number)
The boy sees the cat.	Loses half credit, - ½ (missed only one aspect: root meaning)

The boy sees the ca**X**ts.	Loses full credit, -1 (missed two aspects: root meaning and number)
The boy <u>saw</u> the dog.	Loses half credit, -½ (missed one aspect: tense)
The <u>girl</u> sees the ca**X**ts.	-1 ½ (½ for girl, root meaning; 1 for cats, root meaning + number)

With just a little practice, it becomes second-nature to read through a passage and **X** the translation of a word that misses two or more aspects and squiggle-line a translation that misses just one; doing so allows very systematically for awarding partial credit, and students perceive it as rational and fair.