



# Marking Latin Unseen Translations

by John Tuckfield

The unseen translation - translation of a passage of Latin that the student has not seen before, under constraints of time and with limited access to resources - is a persistent element of Latin courses, especially at school level. It is present in A Level courses in England (for example, OCR 2017), in the Scottish Highers (SQA, 2017), in the New Zealand curriculum (NZQA, 2017), and in Australia (VCAA, 2004; Board of Studies, 2009), to name but a few examples. In Victoria, courses have undergone various changes in the last 30 years, but the unseen has remained a constant: there seems to be a consensus among teachers and examiners that the ability to translate a passage of Latin on the spot is a rigorous and enduring test of at least one aspect of a student's skills in Latin.

Similarly, going by the overseas examples and the Victorian experience, there has been general consensus as to what is looked for in a successful unseen translation. Oxford Cambridge and RSA (OCR) state that 'this component is designed to enable learners to demonstrate their linguistic competence in Latin' (OCR 2017, p. 7); in Victoria, this has been broken down into a series of skills (VCAA, 2004, p. 23):

To achieve this outcome the student should demonstrate the knowledge and skills to

- use a dictionary to determine meaning, including nuances of meaning;

- provide fluent English equivalents for Latin idioms and expressions;
- convey the author's meaning in English;
- identify and translate Latin grammatical constructions accurately;
- reflect the style and purpose of the author.

A similar approach has been taken in Scotland (SQA, 2014, p. 6):

This question paper will give learners an opportunity to demonstrate the following skills, knowledge and understanding:

- translate a detailed and complex unseen Latin prose text into English
- apply knowledge and understanding of vocabulary, accidence and syntax
- convey the meaning of the text in English using appropriate language, style and structure.

In all of the example jurisdictions, the scope of the task is clearly stated, usually with reference to whether the translation will be prose or poetry, the length of the passage, access to a dictionary or word list, and sometimes even the author of the piece (for example, OCR specify Livy for the A level examination).

What has received much less attention is how to mark the unseen, and yet this would seem critical to helping students achieve success. Student teachers as well as experienced teachers have asked just how to mark unseens so that students can learn from them - and the marking and returning process is crucial to that. This article will review the author's own journey in approaching different ways of marking translations.

## Assessment and feedback

It may be useful to review some of the theory behind assessment. There are many reasons why we assess our students - to 'keep them on their toes'; to give an end point to spur them to revise; to sort students; to generate a grade - but the reason which should be dominant is to improve student learning. By assessing the student, we can advise them on where they are, and how they can improve. It is sometimes easy to lose sight of this in schools' relentless drives for numbers and marks, but teachers must never forget that the primary purpose of assessment is to improve student learning - and this should be the litmus test for any assessment method.

Assessment theory is an area that can seem daunting, but at its heart are some

simple concepts that should underpin all assessments:

- validity;
- reliability; and
- fairness.

To these, I would add practicality.

**Validity** can be interpreted as the degree to which an assessment task actually measures the skills that the teacher wants to assess. This might seem a bit obvious, but it can evade some of the most skilled teachers. Consider a physics examination with a worded problem: did the student get the wrong result because they did not understand the law of physics being tested, or because their English vocabulary was weak? In the Victorian HSC, students were not allowed a dictionary; if a student failed to translate words, was it because they could not work out the grammar, or did not know the vocabulary? Teachers may reply that they want the students to learn both, and this is true; but the problem with testing more than one variable at a time is that it can be difficult to determine which one has caused the error, and therefore how the student can be helped to move ahead.

**Reliability** is the measure of how consistently an assessment task can be marked. This breaks into intra-rater reliability, and inter-rater reliability. Intra-rater reliability measures how consistent the same marker is. Do I mark harder at the start of a batch, or at the end? If I know the student, will that affect the mark I give? Might my interpretation of the answer change over time, so if I marked the same test a week later, I might give a different answer? Inter-rater reliability occurs where more than one person marks an assessment task. Do they mark in exactly the same way? How much personal judgement comes into play? Is someone a 'hard' marker, while someone else is an 'easy' marker?

**Fairness** includes the sometimes unconscious bias that can occur in a test, which for example assumes a particular cultural background, and disenfranchises a group of students.

**Practicality** is usually omitted from the assessment textbooks, but my experience leads me to think that it is one of the most important factors. An assessment scheme that is valid, reliable

and fair but is laborious is at risk of being ignored by teachers in place of something more expedient but with perhaps reduced validity and reliability. British education expert Dylan Wiliam admonishes teachers that they should never work harder than the students when it comes to marking (Wiliam, 2011). This is not otiose. If assessment is onerous, it will be done less often, with negative consequences: there is less feedback for students, while in students' minds, fewer assessment tasks means that those that do occur might take on a larger importance than they merit. If, again following Wiliam, we strive to have feedback that is frequent, then any system must be very user-friendly for the teacher.

**Feedback** is the partner to assessment - the means by which information is returned to the student. In education, feedback is the provision of information about students' performance that will direct their future actions in positive ways (Wiliam, 2011). Wiliam describes feedback:

Feedback should focus on the specific features of the task, and provide suggestions on how to improve, rather than focus on the learner; it should focus on the 'what, how and why' of a problem rather than simply indicating to students whether they were correct or not ... feedback should not be so detailed and specific that is 'scaffolds' the learning to such an extent that the students do not need to think for themselves. (OECD, 2010, p. 141)

However, not all feedback is equal. Nyquist (2003) grouped feedback into five categories:

1. weaker feedback only: students are given their own score/grade;
2. feedback only: students are given their own score/grade, and the right answer;
3. weak formative assessment: students are given information about the right answer, together with some explanation;
4. moderate formative assessment: students are given information about the right answer, some explanation and some specific suggestions for improvement;

5. strong formative assessment: students are given information about the right answer, some explanation and some specific activities to improve.

Not surprisingly, Nyquist found that the effect size of the feedback increased, culminating in the stronger formative assessment.

The aim, therefore, should be for a method of assessing students' translations that is highly practical, valid, fair and has high reliability - and one that gives meaningful feedback in the form of strong formative assessment to students to help them to improve.

With this in mind, I will review some of the different methods used for marking unseen and giving students feedback.

## Method #1: Gut feeling

When I first started marking unseen translations, back in the 1980s, I am afraid I had no real system. Instead, I marked by what I could generously call the 'gut feeling' method: I would annotate the students' translation, covering their efforts in red ink, with crossings out, arrows and wild circling. Once that was done, I would take a step back and say, 'That's worth 16 out of 20'.

The problems with this method are obvious. The consistency of my own marking was doubtful, let alone any consistency with another marker - low intra- and inter-rater reliability. If I were to be given the same answer, but perhaps in neater handwriting, would I correct the same mistakes, and would I give it the same mark? The meaningfulness of any mark was also highly doubtful - just what does 16/20 mean? Does it indicate that the student has achieved what I would expect, or are they more advanced than the average student at that level of instruction? Given the fairly shaky grounds on which I gave such judgements, it is doubtful if the overall judgement had much real meaning anyway. And yet that is what students would focus on - my complicated squiggles and corrections in red ink would be ignored, as students went straight to the overall mark, which they then compared with their friends. These corrections did not make it clear where



the students' deficiencies lay; instead, they would need me to go over the test with them to help them identify what areas they needed to work on. But the corrections were vital - even though they took considerable time and effort - for it was only by looking at the overall picture of these that I could arrive at a mark.

## Method #2: Highlighter

My first breakthrough was in the discovery of the highlighter pen. Now, instead of correcting and annotating the students' answers, I would mark the Latin text. To begin with, I highlighted every word in the Latin passage the students got right (my thinking was that the highlighting would be a positive stimulus); I soon found this tedious and misleading for myself, and switched to highlighting any Latin word the students got wrong. I would then count the number of words wrong, subtract that from the overall number of words, and give that to the students as a percentage.

There were some advantages to this method over my previous one. First, reliability was greatly improved. Latin is a language where often it is easy to tell if a word is right or wrong: has the student understood the *-tur* ending on that verb? Have they worked out that this subjunctive transforms the *qui* into the marker of a purpose clause? Certainly, there are judgement calls to be made, but in general at a school level of translation these matters are clear. My intra-rater reliability was much higher, and with agreed principles, inter-rater reliability should be high as well. This method also fulfilled the precept of Dylan Wiliam: the teacher should not work harder than the students. I can correct over 20 unseens in one spare period, with reasonable accuracy. I do not write in any corrections, and usually the students' translations were left completely untouched by me; when we go over the translation in class, students work out for themselves or with peers why their translation was wrong, and what it should have been. Drawbacks are that this method is sometimes blunt, as it does not distinguish between orders of magnitude of errors; words are either right or wrong, so it treats a total disaster of a translation ('the forts have monkey faces' for *simus fortes*) the same as ones

closer to the mark ('let them be happy' for *simus fortes*).

## Method #3: Rubric

For this method, the teacher reviews the passage chosen for translation, and identifies the key areas of grammar. These are written down the left hand column of a table, with descriptors for the achievement students might typically attain. The more specific the areas, the more useful they are to students. For example,

(There may be a whole page of such grammatical structures.)

When marking, the teacher then checks the box that best corresponds to the student's overall achievement. These

can be given a numerical value if there is a need to generate a mark.

There are advantages to this method. It clearly indicates what a student can do, and what they need to work on next; an extra column could be added leading students to further exercises to help them. If the same rubric is used over a period of time, it can be used to show a student's growth over that period of time (and teachers should always remember that just because a student can successfully translate an ablative absolute one week does not mean they will always do so from then on). It is tedious to create but quick to mark, simply checking the appropriate box, and it can be converted to a numerical mark if that is desired. However, there are also drawbacks. Unless some form of weighting is used, a point that is minor

Area				
<b>Translate singular and plural ablatives</b>	Not shown	The student can translate ablative propositional phrases	The student can translate ablative with and without propositions	The student can translate ablative with and without propositions, including ablative absolutes
<b>Translate deponent verbs</b>	Not shown	The student translates deponent verbs as passive	The student translates deponent verbs as active	The student translates deponent verbs as active, correctly translating tense, mood and number

Figure 1. | Marking Rubric Descriptors.

Area				
<b>Dictionary Skills</b> Finding the right headword, and choosing the most appropriate meaning	This is something I really need to focus on	OK, but I could improve on these	Generally OK, but a few small details need attention	I've got these under control
<b>Nouns #1</b> Singular and plural, subjects and objects, accusative of time and with propositions	This is something I really need to focus on	OK, but I could improve on these	Generally OK, but a few small details need attention	I've got these under control

Figure 2. | Marking Rubric Directions.

may seem as important as a major feature of the language; there may not be sufficient examples of a grammar point to enable the teacher to make a meaningful judgement; features of grammar will be scattered throughout a passage, making the teacher's job of juggling the on-balance assessment of a student's skill in a particular area difficult.

An alternative to the type of descriptors given above is to give direction to the student - helping them prioritise what they should work on next:

## Method #4: Block translations

In this method, a passage is divided into a series of chunks or blocks (for example, a clause). Each clause is allocated a mark range, perhaps 1-3 marks, depending on the complexity of the block. This is the method used for the VCE examination, and also in use in Scotland. They describe their method:

Marks will be awarded for accuracy in translation of each block of text and for conveying the essential ideas of the blocks.

Credit will be given for high quality of translation and use of appropriate style and structure including use of synonyms and alternative translation of phrases provided the translation of essential ideas/full blocks is appropriate.

Two marks are available for each block, including the essential idea being correctly or almost correctly translated. For the award of 2 marks for correct translation of the block learners will be expected to translate all the words in the block and show recognition of the overall structure and meaning of the block. However, 2 marks may also be awarded if a minor error occurs, such as an error of tense or syntax which does not detract from an accurate understanding of the full meaning of the block.

1 mark is awarded for translating the essential idea of the block correctly.

No marks are awarded for the block if the essential idea is not translated correctly. (SQA, 2014, pp. 6-7)

	Reliability	Practicality	Feedback
Gut Feeling		★	★
Highlighter	★★★	★★★	★★ (dependent on students making their own notes while reviewing the passage)
Rubric	★★★	★★	★★★
Block Translations	★★	★★	★★ (dependent on students making their own notes while reviewing the passage)

Figure 3. | Ratings for different marking schemes.

Note that they allocate 2 marks for every block: there must be an effort to ensure that all blocks are of comparable difficulty.

Advantages are that this approach eschews an overly picky approach, and gives students credit for getting the essential meaning, even if there are some errors in the exact translation; it also allows for more leeway in what to subtract for an error than the highlighter method mentioned above, which treats all errors the same, whether a student translates *ducem* as 'leaders' or 'duck'. However, this flexibility might also lead to lower reliability (this can be reduced with consensus training beforehand), and it can be a slower method of marking, especially if used for very frequent unseens, rather than just the end of year effort.

## Which method to use?

With the exception of the gut feeling method, all of these methods have something in their favour. Validity and fairness are constant for all methods, as they are inherent in the setting of the task. Reliability, practicality and feedback differ for each method. Based on my own experience, I have tried to rate each one:

I moved from the Gut Feeling, to Highlighter, to Rubric, and now use a combination of Rubric and Highlighter; ultimately, I found the gains in practicality of the Highlighter method edged out the other considerations for older students (who do a lot of unseens), although I like the strength of the rubric and use that for younger years, especially when a specific point of grammar is under focus. For whatever method is chosen, the review of the passage is crucial for the provision of meaningful feedback. I spend as much

time reviewing a passage as the students do in doing the actual unseen; they get back their marked answers, and have to write in their own corrections; we use this time to discuss different approaches, and how we could go about solving the problems a word presents; finally, students fill in a log of the grammar they need to brush up on, and list any vocabulary that caused problems.

## Future directions

It is worth contemplating the future of the unseen - and it is a future that many teachers will, I suspect, see in the next few years. Computer programs - Automated Essay Scoring - currently exist that will mark an essay: these programs examine an essay not just for spelling and grammar - like an elaborate version of the spell-checker that is built in to most word-processing programs - but also for its style, coherence, even sophistication. It is not a wild leap to imagine this software being applied to a Latin translation, especially when translations come from a known pool of passages. An ideal translation could be plugged in by the teacher, but the program would be astute enough to allow for individual differences that still translate the passage accurately.

The next step is where things get interesting. The program can detect what types of mistake the student has made: for example, the student might consistency struggle with noun number, or the voice of verbs; they might not be getting ablative absolutes or indirect statements. Instead of simply giving a grade, the program should be able to give more detailed feedback on what areas of grammar the student has struggled with, and therefore needs to work on.

Great work James! You did well with the indirect statements, but I see you're still having trouble with active and passive, and ablative absolutes

This could then be easily linked to interventions. Let us imagine the student has submitted a translation, and the program has identified that the student is mixing up the imperfect and perfect tenses. The feedback the program gives might be a link to a Khan Academy-style video tutorial explaining the difference, and then another link to some targeted exercises - done and marked online - that give the student practice in distinguishing between the tenses and translating correctly. The student might then be guided to submit another translation of a similar passage, and see how well they have understood.

I suggest you watch this quick [video](#) and then try these [exercises](#), and we'll [check](#) to see if you've got these concepts

The implications of this technology should not be underestimated:

- They will make it much easier for students to progress at their own pace - they signal the end of the era when

all students in a class had to be working on the same thing at the same time

- They will make it much easier for students to learn Latin without a specialist teacher present
- They will profoundly change the role of the teacher.

**John Tuckfield is a Latin teacher at Camberwell Grammar School, Australia. He has been teaching Latin in Australian schools for 30 years. He is an executive member of the Classical Association of Victoria and teacher in Latin Method in the University of Melbourne's Master of Teaching programme.**  
**[jwt@cgs.vic.edu.au](mailto:jwt@cgs.vic.edu.au)**

## References

- Board of Studies (New South Wales) (2009). *Latin Syllabus*. [http://www.boardofstudies.nsw.edu.au/syllabus\\_hsc/pdf\\_doc/latin-continuers-st6-syl-from2010.pdf](http://www.boardofstudies.nsw.edu.au/syllabus_hsc/pdf_doc/latin-continuers-st6-syl-from2010.pdf).
- Greenwood, E., Irwin, E., Lovatt, H., Low, P., Rogerson, A. and Weeks, A. (2003). *Rethinking*

*'Unseen' Translation: a pilot scheme for developing students' reading skills in Greek and Latin*. Cambridge: University of Cambridge.

Nyquist, J. (2003). *The Benefits of Reconstructing Feedback as a Larger System of Formative Assessment: A Meta-Analysis*. Nashville, Tennessee: Vanderbilt University.

NZQA (2017). *Latin Subject Resources*. <http://www.nzqa.govt.nz/ncea/subjects/latin/levels/>

OCR (2017). *AS/A Level GCE Latin - H043, H443* <http://www.ocr.org.uk/qualifications/as-a-level-gce-latin-h043-h443-from-2016/>

OECD (2010). *The Nature of Learning: Using Research to Inspire Practice*. [www.sourceoecd.org/education/9789264086470](http://www.sourceoecd.org/education/9789264086470)

SQA (2014). *Higher Latin Course Assessment Specification*. [https://www.sqa.org.uk/files\\_ccc/CfE\\_CourseAssessSpec\\_Higher\\_Language\\_Latin.pdf](https://www.sqa.org.uk/files_ccc/CfE_CourseAssessSpec_Higher_Language_Latin.pdf)

SQA (2017). *Higher Latin*. <https://www.sqa.org.uk/sqa/47907.html>

VCAA (2004). *Latin Study Design*. <http://www.vcaa.vic.edu.au/Documents/vce/latin/LatinSD.pdf>

Wiliam, D. (2011). *Embedded Formative Assessment*. Bloomington Ill: Solution Tree Press.