VAC Henmon
The measurement of ability in Latin. *Journal of Experimental Psychology* VIII,9(1917), 515-538.

This note is the first in series that will deal with very early research on vocabulary. Most readers will be familiar with the main trends in the research post 2000, say, but there is a lot of early research which people seem not to be aware of, research which is no longer cited, and no longer part of our collective memory. This is something of a shame, since some of this work is not just good, but really quite astonishing. Often, modern research is just a reprise of studies that were carried out in the first part of the twentieth century.

This note summarises some very early work by VAC Henmon (Henmon 1917). Henmon was an influential psychologist, who went on to become the Founding Director of the School of Education at the University of Wisconsin. This was the culmination of a long and distinguished career which involved research on a huge range of topics. Henmon was particularly interested in the measurement of intelligence (he had worked in Wundt's laboratory in Berlin where Galton had also been active for a while). He was also involved in officer selection during the 1914 War, and carried out some of the very earliest work on Pilot selection. (See Figure 1.) He also published a number of papers on "prognosis" - what would nowadays be referred to as language aptitude.

Figure 1: An early French flight simulator.

Summary
In his 1917 paper, Henmon starts out by asking how can we assess learner's progress in learning a foreign language – in this case Latin. He argues that a good way to do this would be to assess how many words the learner has learned, on the grounds that we could reasonably expect the best learner in group to be the one who scores best in a Latin vocabulary test. This seems obvious (to Henmon at least), but he notes that it's less obvious how you might construct a reasonable vocabulary test for Latin. His first attempt at this is a short test consisting of 50 English words to be translated in Latin,

and 50 Latin words to be translated into English. These words were selected because they were all found in twelve commonly used Latin text books.

Henmon reports that the scores on this test correlated fairly highly with scores on an end of year assessment test. (252 students were tested in 10 different schools.) However, a more detailed analysis of the data suggests that some words are clearly easier than others. 99% of the students know that **pulcher** means **beauty**, while only 55% know that **pedes** means **foot**, so Henmon argues that it would be a mistake to give these words equal value. Henmon therefore proposes a weighting system which gives higher marks for words that only a few students know.

Henmon then decides that more work on the selection of words for the test needs to be implemented, so he analyses the entire vocabulary of a set of 13 beginners' text book. This relatively modest corpus contains a total of 3350 word types, but only 319 of these types occurred in all of the text books. Henmon reduced this total further by eliminating any word that was not found in the three main literary texts that learners at the time would be expected to read (Julius Caesar, Cicero and Virgil). 80 words were eliminated in this way, leaving a total of 239 words. Henmon argues that these words "seem to be in a real sense a standard vocabulary for High School Latin...all pupils, no matter what books are used, will have come in contact with them in their first year and will continue in contact with them throughout the four years of high school Latin." A further study involving 847 learners of Latin in 19 schools did indeed show that student's knowledge of these words increased steadily by grade, with a particularly large increase between Year I and Year II.

Further analysis of these words showed that some words were not reliably recognised even by the higher level students. Henmon argues that these words could be removed from the data set, leaving a set of 200 reliable words. Finally, he suggests that this 200 word list can be subdivided into smaller lists, so that each of the lists contains words with similar discrimination values. This would mean that the smaller lists are all equivalent tests, making it possible to evaluate progress as a result of study.

Commentary
Three things impressed me most about this paper when I first read it.

The first was the idea that vocabulary (size?) might be used as a surrogate for overall language competence. As far as I know, this is the earliest paper in which this idea appears, and the earliest one for which any hard empirical data to support the idea was published. It came as a real shock to realise that what I had thought of as a revolutionary 1980s idea was expressed so clearly some sixty years earlier. And it's worth noting that Henmon's empirical data is not just a quick one-off study. Unlike much modern research, his instruments are carefully developed and theoretically justified, and the number of learners he tests are substantial.

The second thing that impressed me was the amount of work that Henmon put into developing his word list. Nowadays, this is the sort of work that we can do relatively easily thanks to machine-readable texts and word-counting software (although it is not uncommon to find recent PhDs which have been awarded for work that just counts the number of words in common across a mere three or four text books.) Henmon, of course, writing in 1917, did not have access to machine-readable textbooks. He would have had to count the words used in each text by hand, and tabulate each text against every other text by hand as well. This must have been a massive task. Then in the follow up study, Henmon checks that each of his 250 words also occur in the classical Latin texts. Again, with modern computers this would be an easy task, but in the absence of concordances for Caesar, Cicero and Virgil, we would be looking at a painstaking comparison of the 319 target words against each of these texts - and they are not short texts: Caesar's *Gallic War* alone consists of eight

volumes.

Finally, the idea of using a weighted value in the scoring system is astonishingly modern, presaging some of the more recent work that is standard in IRT analysis, but predating it by fifty or sixty years. Surprisingly, though, no-one seems to have picked on this very clever idea. Equally impressive is the idea of developing shorter equivalent tests which can be used to measure vocabulary growth and progress over time.

Conclusion
Overall, this is very impressive piece of work, even by modern standards, and this makes you wonder why it hasn't been copied by later researchers. Sadly, this study was cited only a few times in the years after publication, four times in the 1920s, and once each in 1931 and 1940. After that it seems to disappear from the collective consciousness, along with the innovative methodological approaches that it pioneered. I suppose that, in part at least, this loss can be attributed to the decline of Latin teaching in schools in the 1950s and 1960s. Henmon comments tantalisingly that "Studies of vocabularies ... in French and German will be published in the near future..." , but so far I have not been able to find this further work. If any reader knows where it appeared, I would be glad to know.