

**Q: How do test developers deal with the issue of timing in reading tests?
Does the L1 have an effect on timing considerations?**

A: Clifford and Cox define proficient reading as the active, automatic, far-transfer process of using one's internalized language and culture expectancy system to comprehend, efficiently, an authentic text for the purpose for which it was written.

A key part of this definition is the notion of automatic processing. On Tuesday, Branka mentioned the development of automaticity in language proficiency. Yesterday, Troy talked about communicative competence, distinguishing between foundational and higher-order skills.

Some view developing automaticity as the bottom-up approach to improving, for example L2 Reading, proficiency. This is consistent with structural linguistics in that language is viewed as a self-contained, self-regulating system. Simplistically, letters to words to sentences to paragraphs to extended texts. Of course, as Ray and others have pointed out, it is not just about learning to read – it is, more importantly, reading to learn. So what are we missing?

Verbal Efficiency Theory includes higher-order language skills, with Pragmatics at the highest level. In addition, foreign-language readers employ other knowledge, skills, and abilities to realize successful reading comprehension. For example, readers use background knowledge and tools of strategic competence (cognitive and meta-cognitive techniques) to construct meaning from what they are reading. Additionally, personal characteristics such as interest, inquisitiveness, and tolerance for ambiguity can influence reading proficiency. Many of these knowledge, skills, and abilities are established in the L1 and are transferred to the L2. This means that your reading ability in a foreign-language, generally, cannot exceed your native-language reading ability.

All of these factors vary greatly from person to person, even in the L1, so it is difficult to establish what normal reading speed is. Consequently, determining L2 reading rates at the various STANAG 6001 Reading levels requires making some educated guesses and practical decisions. After all, practicality is one of the cornerstones of language testing.

Because we focus on testing the given construct, we have to ask what the STANAG 6001 Reading Level descriptors say about reading rate. Level 1 does not address reading rate. Level 2 states “Can readily understand prose that is predominantly constructed in high frequency sentence patterns.” There are, however, numerous caveats – for example, “may be slow.” Level 3 states, “Reading speed may be somewhat slower than that of a native reader.” One question is, “What is normal reading speed for natives.” Another question is, “Does reading speed vary from text to text, depending, not just on content and task, but, reader idiosyncrasies and individual backgrounds (including L1), as well?” The easy solution would be to refrain from placing timing constraints on readers, but that is not very practical within a military testing context. So what are our options?

Traditionally, foreign-language reading tests have been linear in design, and test developers have set timing limits for the overall test. For example, test takers might have 120 minutes to complete a 50-item reading test. If the test is computer-delivered, another option might be to set the timing for individual items. This begs the question, “Should the timing be the same for all items?” We would have to know, at least, the text lengths and demands of the test tasks.

Nowadays, more and more test developers are interested in Computer Adaptive Testing. Since overall test length depends on individual test-taker performance, setting an overall time limit is complicated. Some testing organizations, nevertheless, set a time limit for practical reasons. Anyone who has taken the English Comprehension Level (ECL) test at DLIELC is familiar with the 45-minute time limit. If the CAT algorithm cannot produce a reliable result in this time, there is no resulting score.

CAT can be item-level or multistage adaptive. The ECL is an item-level adaptive test, so the algorithm makes an item-selection decision after every item. One option for such a test would be setting timing constraints on individual items, which might be determined through pre-testing. Multinational cooperation might also help with this; otherwise, you could end up with different reading rates for population A and population B, especially considering linguistic distance between the native languages and English.

Multistage adaptive testing offers an additional option. An overall testing time limit can be set, individual items can be timed, or timing can be set for the various testlets. A reading testlet might, for example, have five items. Test takers could have a set time to answer the five items (before the algorithm makes a testlet-selection decision). Those of you familiar with BAT2 for Reading know that the items were individually timed. Post-BAT2 surveys reflected a variety of opinions on the use of individually-timed items. Some test takers did not feel that timing affected their performance, while others contended that timing prevented them from demonstrating their actual reading proficiency.

Whatever test developers decide, it must be construct based. If shareholders (including test takers) do not understand why there are timing constraints, test score validity and reliability could be questioned.