

Title: Measuring Second Language (L2) Reading Comprehension Ability Using SVT Tests

Authors: François Pichette (Université du Québec à Montréal, Télugu; pichette.francois@teluq.uqam.ca), Linda de Serres (Université du Québec à Trois-Rivières), and Marc Lafontaine (Université Laval)

Statement of the Problem

As teachers and researchers, we are sometimes confronted with immigrant students who show limited reading ability that is not always easy to assess. Among the different methods that may be used for such assessment, one that deserves attention is the Sentence Verification Technique (SVT) (Royer, Hastings, & Hook, 1979). This technique has been used for three decades to assess the ability of specific groups to comprehend certain texts. The targeted audience usually consists of students who are tested by their teacher.

The method requires that the teacher select one or more 12-sentence passages from a book or text. From each passage the teacher builds 16 items of the following types:

- four sentences are paraphrased, conveying the same meaning as in the original text but with as many different words as possible;
- four sentences undergo meaning change, while keeping as many words intact as possible;
- four sentences are left intact, and
- four more sentences are added as distractors.

These 16 items are displayed on a separate page, shown to the student after the text is read. Without referring to the text, the student reads each item and checks or writes “yes” if it says the same thing as in the text, and “no” if the sentence has a different meaning or mentions something that was not stated in the text.

The typical SVT test is based on four texts of 12 sentences each, for a total of 64 items. SVT scores help the teacher determine the suitability of the book or text for his class in terms of readability. If students score between 70% and 80% that means the text is at the right level for them.

These tests have been shown over the years to be reliable and valid: reliabilities for 64-item tests like ours are in the .7 to .8 range, and correlations with standardized reading tests range from .52 to .73 (Royer, in press), which means that people who score high on standardized reading tests also score high on SVT tests, and vice versa. As further demonstrated by Royer (in press), SVT tests seem to demonstrate sensitivity to differences in reading ability, and to variation in text difficulty. They measure global text comprehension, and they have been used successfully in first and second languages.

Summary of Research

This paper introduces the very first SVT test developed with the goal of measuring general reading comprehension ability. The creation process follows recommendations formulated by Royer (2007, personal comm.), in order to come up with an instrument expected to be simpler, faster to administer, and more affordable for educators than current standardized reading tests.

In our quest for the best instrument possible, we carried out two important modifications to the standard procedure for creating SVT tests:

1. We submitted our texts to two readability scales before creating our instrument:
 - a. The Flesch-Kincaid Formula (Flesch, 1950), which is a scale based on vocabulary frequency and sentence length. Included in the Word software, it is recognized as the most reliable readily available formula (DuBay, 2004; Rogers, Hazelwood, Sewell, Harrison, & Shuman, 2008).
 - b. The Sentence Complexity Index (Wampler & Williams, 1991), which is a scale based on text structure and the density of clauses. It is included in the Grammatik component of the WordPerfect software.
2. We modified the traditional items ratio. Instead of creating four items of each category, our 16 items consist of two originals, four distractors, five meaning changes and five paraphrases. This modification is based on the fact that paraphrases and meaning changes have proved to be better discriminators for external validity than originals and distractors (Marchant, Royer, & Greene, 1988).

To reach our goal, four short English texts of general interest were selected, to be subsequently modified so that they reach 12 sentences each and with readability levels that were more or less equal in terms of difficulty on both readability continua. The following table shows our text characteristics.

Table 1
Texts and their readability

TEXT	Flesch-Kincaid (100 = very easy)	Interpretation	SCI (100 = very complex)
<i>Volunteer</i>	78	Fairly easy	27
<i>Titanic</i>	66	Standard	40
<i>Parrot</i>	49	Difficult	42
<i>Frog</i>	28	Very difficult	75

To enhance motivation, we inserted the most difficult texts neither at the beginning nor at the end of the test. The sequence chosen for administering the test was the following: standard, difficult, very difficult, and fairly easy. This new instrument was administered to 161 adult Canadian university students. Levels of English proficiency ranged from high beginners to near-natives. All were French-speaking ESL learners except for six participants who were native speakers of English. Including the latter did not affect the profile of results. Table 2 shows a comparison of SVT scores obtained for the four texts.

Table 2
SVT scores for all texts

TEXT	Flesch-Kincaid (100 = very easy)	Level	SVT Mean (%)	SD	SE	Conf. Interval (95%)	
						Lower	Upper
<i>Volunteer</i>	78	Fairly easy	86.1	11.25	.89	84.35	87.85
<i>Titanic</i>	66	Standard	90.6	8.50	.67	89.32	91.97
<i>Parrot</i>	49	Difficult	84.5	9.51	.75	83.03	85.99
<i>Frog</i>	28	Very difficult	74.1	11.67	.92	72.29	75.92

As evidenced in Table 2, means for our texts increased as their established levels of difficulty decreased, except for the text *A special volunteer*. This could be due to the fact that it was given as the fourth and final text. Therefore, it is possible that participants paid less attention or made less effort toward the end, due to some fatigue effect and/or the fact that some were anxious to finish.

A multivariate analysis indicates significant differences between texts ($F(3,158) = 89,689, p < .001$), confirming that the texts chosen are all of different levels of difficulty, the only exception being for *Volunteer* and for *Parrot* which are not significantly different ($p > .05$). This suggests that scores on this instrument vary with text readability as was shown in previous research.

The overall mean for our test is 83.7% which is slightly above the 70%-80% range. However, our participants were all university students who, being highly educated, are assumed to have above-average reading skills. We could thus expect the scores to fall within the ideal range when it comes to testing non-university participants.

Internal consistency for the test was evaluated using Cronbach's alpha, which yielded a value of .70, suggesting good reliability of the instrument. This means that the level of difficulty of our items is consistent throughout the whole test. The only exception is one of the 64 items, which was automatically rejected by the SPSS software because 159 of the 161 participants had the right answer. This question will be replaced with a more valid item.

Practical Suggestions

A reliable test of reading comprehension can help teachers identify immigrant students whose L2 reading ability is lower than that of other immigrant students, since this ability is not always visible through general language proficiency scores. As was mentioned, our ultimate goal is to create an SVT-based reading test, which would present three possible advantages for teachers who may be interested in using it:

1. The test is faster to administer than usual standardized tests, with a testing time of about 22 minutes. This factor is of high interest given the teachers' frequently tight schedules, and the fact that students tend not to like long tests.
2. The test would be free of charge, which is undeniably a great advantage for teachers and researchers with little budget, if any, given that some standard tests cost several dollars for each participant.
3. The test is motivating for students, who have to read short, interesting texts while having nothing to write.

References

- DuBay, W. W. (2004). *The principles of readability*. Costa Mesa, CA: Impact Information. Retrieved June 17, 2009, from <http://www.impact-information.com/impactinfo/readability02.pdf>
- Flesch, R. (1950). Measuring the level of abstraction. *Journal of Applied Psychology, 34*, 384–390.
- Marchant, H. G., Royer, J. M., & Greene, B. A. (1988). Superior reliability and validity for a new form of the Sentence Verification Technique for measuring comprehension. *Educational and Psychological Measurement, 48*, 827-834.
- Rogers, R., Hazelwood, L. L., Sewell, K. W., Harrison, K. S., & Shuman, D. W. (2008). The language of Miranda warnings in American jurisdictions: A replication and vocabulary analysis. *Law and Human Behavior, 32*, 124-136.
- Royer, J. M., Hastings, C. N., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. *Journal of Reading Behavior, 11*, 355-363.
- Royer, J. M. (in press). Uses for the sentence verification technique for measuring language comprehension. *Progress in Education*.
- Wampler, B. E., & Williams, M. P. (1991). *Grammatik Windows (Computer program)*. San Francisco, CA: Reference Software.