

Assessing Progress in an Extensive Reading Program Using a Cloze Test

Stan PEDERSON

(Received October 3, 2011)

Abstract

This paper investigated the suitability of a pre-test post-test cloze procedure as a measure of reading progress in an extensive reading class at a high-level Japanese high school. Results showed the test to be reliable, consistent with a construct of reading fluency and sensitive to improvement in reading proficiency.

Introduction

In 2004, I took over a class titled "Rapid Reading" at a high-level public high school in Japan. The elective, open to all years, was attended by 21 students—with one entering part-way through and another leaving. Most students had excelled in the regular Japanese public school system and were expected to enter top universities in Japan, or to study overseas after graduation. There were also some returnees in the class. The previous teacher had used book reports and reading journals as the basis for evaluation since he also had most of the same students in a writing class, allowing him to disambiguate writing skill from progress in reading. In my case, I saw these students only in this reading class so I wanted to test progress directly through reading rather than indirectly through writing.

At the time, I believed that the benefits of extensive reading depended mostly on the student interacting with the material, not with the teacher or with classmates. The teacher may guide students' expectations and strategies, give them information on selecting materials, offer encouragement, monitor progress and model the practice of extensive reading itself, but should otherwise stay out of the way and let students read. In addition to the limited opportunities for interaction, and the window for evaluation they might afford, the fact each student was reading different books, had widely varying abilities, and that there was no visible product, placed unusual demands on designing assessment. After considering my options, I decided to test the target skill as directly as possible using a cloze test. In order to calibrate progress to the ability level of each student, a test was administered at the start of the course and a comparison was drawn between this and the final test using Rasch analysis.

This paper investigated the suitability of the cloze test as a measure of reading proficiency in the classroom context. Results showed it to be reliable, consistent with a construct of reading fluency and sensitive to improvement in reading proficiency.

Literature review

In order to judge the validity of a test, one must analyze the target skill and its projected use in the real world and the extent to which the test resembles or reproduces the same skill or set of sub-skills. In addition, it is desirable for testing to reflect the skill as it was trained. I will now summarize some major perspectives on the reading process as background for assessing the appropriateness of the cloze test for evaluating reading.

LaBerge and Samuels (1974) conceived of reading as being composed of two processes, decoding and comprehension.

A fluent reader, through automatic decoding, is able to devote almost constant attention to comprehension processes while a non-fluent reader must alternate attention between decoding and comprehension. Comprehension, therefore, depends on fluent decoding.

But automaticity in decoding is only one piece of the puzzle. Anderson (1984) showed the interaction between the text, as decoded, and the reader's background knowledge in what he called schema theory. He posited that the reader searches for a match between the message as decoded and an interpretation consistent with his existing knowledge. He lists six functions of schema in the reading process: 1) schema may provide a slot for certain information such as a slot for the murder weapon in a mystery novel, 2) it directs attention to important parts of a text 3) it allows a reader to go beyond the explicit and literal, to make inferences 4) it guides searches of memory, permitting quick recall of related elements like the foods likely to appear when a fine restaurant schema is activated, and 6) it can help generate hypotheses about gaps.

Goodman (1976) portrayed reading as a "psycholinguistic guessing game" with three elements in play: graphic, syntactic and semantic. Similarly, Just & Carpenter (1980) posited that reading involves contemporaneous processes of: 1) lexical access 2) assigning case, and 3) integrating clauses. Each may be processed serially or automatically.

Cloze tests involve a highly similar set of skills and processes as identified in the reading process above. Readers must decode fluently enough to allow attention to shift to comprehension, which forms the basis for making the inferences needed to provide a suitable word for each slot. They must access words they know, matching the correct meaning and assigning the proper case. They must take into account the context of the local clause and sentence, and beyond this, must fit it with the broader demands of the discourse pattern and the schema of the topic. It seems reading theory would support the cloze test as a good approximation of the target skill.

Instrument

The instrument selected had to meet several criteria. First of all, it had to be able to assess students who had been reading different materials, which meant it had to test reading ability in general. Second, it had to be sensitive enough to detect the progress of each student over the period of the course, not just show ranking of one student against another. Third, it had to have face validity. In particular, I wanted a test that was seen as a credible measure of reading ability so that my unusual approach of mostly reading rather than "teaching" could be justified to administrators, colleagues, and to students themselves. For this reason, and for the sake of construct validity, I wanted a test that was patterned as closely as possible on the target skill. And lastly, it had to be easy to create, to administer, and to mark.

The random cloze test was chosen by a combination of its own merits and by other possibilities being easily eliminated. It should be noted that long and careful consideration of each possibility was not possible. I had taken this job on recently, and it was one among many courses so preparation time was limited. Furthermore, assessment had to be set before the course began if I was going to do a pre-test which I had determined early on must be done in order to gauge individual progress. Reading prosody tests were not a good candidate due the subjectivity of the marking. Multiple choice tests have a variety of flaws, chiefly that they do not reproduce the target skill well and are notoriously difficult to construct. And reading speed tests were eliminated for face validity—it can be argued that one can read speedily without the requisite comprehension.

On the positive side, the random cloze quickly made itself obvious as a strong candidate. First of all, it was close to the target skill as outlined in the literature above, but more than that, it resembled extensive reading itself. Students read an extended text and infer meaning and function to unknown words just as they do in a cloze test. Moreover, with the recommendation that students choose books where they can understand 95% of the words on a page, the cloze could be adjusted to reproduce this ratio by blanking 1 of 20 words rather than the conventional 1 of 7 or 1 of 14. In addition, a large number of questions can be put, improving reliability. Research shows that at least 50 deletions are needed to assure reliability (Alderson, 2000).

It is easily administered, with the whole class taking the same test at the same time; and is easily marked with answers being either right or wrong. Furthermore, the data would easily be analyzable using Rasch analysis which

could further enhance reliability by assigning more difficult questions more weight, by making equivalent tests comparable, and by eliminating badly performing items which confuse the results.

Ultimately, the test was constructed of four sections, each with 25 blanks. Each section was excerpted from a graded reader with each moving to a higher level. Blanks were assigned every 20 words with exceptions for the first word in a sentence and proper nouns. When a blank fell on one of these, the blank was moved 20 words forward.

An item was marked right if the answer was acceptable both in meaning and grammatical case and was marked wrong if it was incorrect in either sense, a marking approach which is fair to higher level students who are capable of coming up with multiple correct answers (Kobayashi, 2002). Correct answers were judged while marking, then recorded in a common answer bank for reference.

Results

As mentioned earlier, the Rasch model allows measures to be anchored in the results of equivalent tests as long as there is overlapping data. To make this comparison robust, one section of the test was retained in the post-test.

The person measures in Table 1 reveal a very high level of reliability (.92 in April and .89 in January with separation of 3.45 and 2.83 respectively) indicating that the ordering of ability levels is unlikely to change given the same students taking an equivalent test.

Table 1

Anchored person measures for April pre-test and January post-test

<u>Test</u>	<u>Measure</u>	<u>SD</u>	<u>Gain</u>	<u>Rel.</u>	<u>Sep.</u>
April pre-test	-2.73	10.75	--	.92	3.45
January final	17.28	8.95	20.01	.89	2.83

The post-test, anchored in the pre-test for comparison, shows a gain of 20 logits, or about 22% over the pre-test.

Construct validity is inferred in Rasch analysis by measuring the extent to which items lie along a single line, the putative construct. This measure is taken from fit statistics, A figure of less than 2 in the outfit mean square shows acceptable fit to a single line. April results yielded a figure of .96 and January produced an even 1.0 showing unidimensionality and, by inference, conformity to an underlying construct.

Discussion

First of all, the cloze test showed itself, by way of the outfit mean square, to be operating on a single construct. This is an inference, but one supported by the close relation of the cloze test to the target skill. It involves an extended stretch of reading (more than 2000 words) , which is comparable to the lengthy reading encountered in training. In terms of reading skills, the cloze forces students to produce words appropriate in both meaning and function to fit an overall topic schema, a discourse context, and more localized collocations, as in reading as conceptualized in theory.

Reliability was revealed to be very high in both tests and the separation estimates were small, meaning we can have confidence that the large gain, 22% over the first test, has a high probability of being reproduced. Given this accuracy and the ease of construction and administration, it would seem the cloze is highly applicable to the classroom situation.

Limitations

As mentioned at the outset, the school was high-level. But more than that, it was an elective class, open to all grades, thereby gathering only those who were the most motivated and probably the most skilled. I have given the easiest section of the text at a lower-level university, where students were mostly unable to answer more than a few questions, and at a high-level university, where second year specialists in English were scarcely able to match the performance of these high school students. I would say, therefore, that the sort of cloze test I conducted might be ill-suited to most Japanese school settings.

However, modifications might be possible. For example, students could choose from a limited set of perhaps 3 or 4 words. If these were embedded in sentences, and the student had only to circle the best answer, it could allow them to continue the flow of reading with the only minimal interruption, and to succeed at the much easier recognition level. In this way, one might preserve the many advantages of the cloze as a test of reading while adjusting to the more usual level we find among Japanese students.

References

- Alderson, J. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Anderson, R. (1984). Role of the reader's schema in comprehension, learning, and memory. In R. C. Anderson, R. Olson, & R. J. Tierny (Eds.), *Learning to read in American schools: Basal readers and content texts* (243-257). Hillsdale, NJ: Erlbaum.
- Goodman, K. (1976). Reading: A psycholinguistic guessing game. In H. Singer, & R. Ruddel (Eds.), *Theoretical models and processes of reading* (497-508). Newark, DE: International Reading Association.
- Just, M. A., Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension, *Psychological Review*, 87 (4), 329-354.
- Kobayashi, M. (2002). Cloze tests revisited: Exploring item characteristics with special attention to scoring methods, *The Modern Language Journal*, 86 (4), 571-586.
- LaBerge, D., Samuels, S. J. (1974). Toward a theory of automatic information processing in reading, *Cognitive Psychology*, 6, 293-323.