Inferencing Subskills Reading Test

L. D. Nicolas May

Northern Arizona University

Abstract

This project explored the possible performance differences between two types of inferences. Inferencing, an identified component skill of reading, is a cognitive skill that was assessed in achievement testing at the Program in Intensive English, an intensive English program within Northern Arizona University. The exploration of the possible performance differences is important because there is evidence of a discrepancy between expectations and how test takers actually perform. To investigate this apparent shortcoming, a limited-scope test was developed to gather data on high-intermediate students' performance on two subconstructs of inferencing, *bridging inferences* and *elaborative inferences*. These have been repeatedly identified in various taxonomies that categorize the types of inferences that are made by proficient readers. Ultimately, the test supported treating these two types of inferences differently in instruction and testing.

Keywords: L2 reading, inference, bridging inference, elaborative inference

Inferencing Subskills Reading Test

Background

One of the prioritized reading objectives of the Program in Intensive English (PIE) at Northern Arizona University is "Make inferences," an objective that aims to help language learners gain the ability to infer when reading. This ability is also included in theoretical inventories of reading component skills (Alderson, 2000, pp. 9-13; Grabe, 2009, pp. 68-70). While Alderson argues that not much empirical evidence exists for conclusively justifying tests that attempt to isolate and measure any one reading component skill (2000, p. 114), the fact that inferencing is included in the course objectives of the real scenario described above and that it receives some treatment throughout second-language reading literature (Chikalanga, 1992; Urquhart & Weir, 1998; Koda, 2004; Grabe, 2009) and cognitive psychology literature (Warren, Nicholas, & Trabasso, 1979; Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; Zwaan & Radvansky, 1998) provides some justification for deeper investigation of inferencing as a construct of reading, perhaps with its own component subskills that may be definable as subconstructs.

Unfortunately, the subconstructs of the inferencing construct are not well understood or defined, which leads to confusion about what inferencing is, how it can be learned, and how inferencing abilities can be tested. This was the premise that is the basis for this study's approach, which is to use a theoretical taxonomy for subdividing inferencing in reading into two subconstructs, *bridging inferencing* and *elaborative inferencing*.

Research Question

To investigate whether there is a relationship between English as a second language (L2) readers' ability to make bridging inferences and elaborative inferences, the following research

questions were addressed: How much correlation is there between English L2 readers' ability to bridging inferences and their ability to make elaborative inferences? How much of the variability of one ability is accounted for by the variability of the other?

Methods

This test was administered to forty-eight pre-academic intensive English students, but eight were removed from the participant pool on the basis of their expressed desire to not participate in research (N = 40).

The test was a narrow-scope reading test that aims to measure just two subconstructs, bridging inferences and elaborative inferences. The test document included a table of specifications, which is a test design plan document that helped balance the test along two dimensions: (1) course relevant topics and (2) subconstructs being tested (Miller et al., 2013).

The test itself included two testlets consisting of a reading passage of about three hundred words. Only one task type, a sentence-completion task, was used. There were two item types, multiple-choice items and short-answer items. For the test, each subconstruct was assessed with an equal relative importance.

Scoring for both multiple-choice and short-answer items is dichotomous with a strict key. Score reports were generated and given to test takers on November 26, 2013. Student identification information on the reports included test taker name, ID number, course and section. A table of results illuminates their performance by detailing both types of inference, a description of the inference types, their score on each inference type, and tailored feedback based on the percentage of correct answers for each inference type.

Results

The items were grouped into three sets and then analyzed using classical test theory methods. The three sets were bridging inference items ($K_B = 12$), elaborative inference items ($K_E = 12$), and the total item set ($K_T = 24$). The number of test takers for all three item sets was the same, N = 40.

The minimum score for the bridging inference items was 4, and the maximum was 12. The mean score for this item subset was 8.88 and the standard deviation was 1.99. The minimum score for the elaborative inference items was 3, and the maximum was 12. The mean score for this item subset was 8.18 and the standard deviation was 2.44. For the total item set, the minimum score was 7 and the maximum score was 23. The mean total score was 17.05, with a standard deviation of 4.04. (See Table 1.)

Table 1

Descriptive Statistics for Inferencing Subskills Reading Test

Item Set	N	K	Min	Max	$\bar{X}$	SD
Bridging Inference Items	40	12	4	12	8.88	1.99
Elaborative Inference Items	40	12	3	12	8.18	2.44
Total Items	40	24	7	23	17.05	4.04

Because the test results had a norm-referenced interpretation, Cronbach's alpha was calculated to measure the internal consistency of each item set, or the reliability of the test over each item set (Miller et al., 2013). This statistic was then used to calculate the standard error of measurement (SEM), which estimates the margin of error for an individual's test score (Doppelt, 1956). The internal consistency for the bridging inference item subset was .57 with an SEM of 1.31. For the elaborative inference item subset, the internal consistency was .64 with an SEM of

1.47. Finally, the total item set yielded an internal consistency of .76 and a SEM of 1.96. This is summarized in Table 2 below.

Table 2

Reliability Statistics for Inferencing Subskills Reading Test

Item Set	N	K	Cronbach's α	SEM
Bridging Inference Items	40	12	.57	1.31
Elaborative Inference Items	40	12	.64	1.47
Total Items	40	24	.76	1.96

It was necessary to find evidence that the test results support the hypothesis that bridging inference ability and elaborative inference ability are somewhat independent. This necessity is rooted in the definitions of the subconstructs, one which posits that bridging inferences are required for text comprehension, and the other that elaborative inferences enhance interpretation of a text. If this is the case, we can expect individual test takers to perform differently—although not entirely so—between the item sets that measure the two abilities. Because the abilities are not completely independent of each other, the results should indicate that some of the variability of one ability is accounted for by variability of the other (Pagano, 2010). To determine if there is shared variability between the results of the bridging and elaborative item sets, the Pearson product moment correlation was calculated. From Table 4, a correlation table for the three item sets (bridging, elaborative, and total), we see that the correlation between the item sets of interest is .66. To determine how much shared variance exists between the two abilities, we calculate the coefficient of determination (r^2) (Bachman, 2004). Here, the coefficient of determination was .44.

Table 4

Correlations between Item Sets

Item Set	Bridging Inf. Items	Elaborative Inf. Items	Total Items
Bridging Inf. Items	1.00	.66	.89
Elaborative Inf. Items		1.00	.93
Total Items			1.00

The original hypothesis stated that there would be some differences between how individual test takers will perform on the items that measure their abilities to make bridging inferences and elaborative inferences. The correlation coefficient ($r = .66$) and the coefficient of determination ($r^2 = .44$) between the two item subsets supports this hypothesis. Without a doubt, the ability to inference in general determines how successful a test taker is at making *subtypes* of inferences, but that the results support the idea that the subtypes are somewhat independent is evidence that the subconstructs are worth considering individually, especially when it comes to the skill and strategy instruction required to lead academic English learners toward successfully making appropriate inferences.

Relevance to the PIE

Ultimately, the test achieved some of it purpose. It was successful in that it (1) connected variability in the results to the proposed subconstructs, and (2) useed the results to help the test designer decide whether to further explore the subconstructs of inferencing. Where its success remains to be seen is in terms of impact. If other stakeholders determine it is appropriate to further explore the possibility of changing instructional approaches for teaching inferencing, the test has the potential to positively impact teachers, students, and the program.

References

Alderson, J. C. (2000). *Assessing reading.* New York, NY: Cambridge University Press.

Bachman, L. F. (2004). *Statistical analyses for language assessment*. New York, NY: Cambridge University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. New York, NY: Oxford University Press.

Buck, F. (2001). *Assessing listening.* New York, NY: Cambridge University Press.

Chikalanga, I. (1992). A suggested taxonomy of inferences for the reading teacher. *Reading in a Foreign Language*, 8(2), 697-709.

Doppelt, J. E. (1956). How accurate is a test score?. *Test Service Bulletin*, 50, 1-3.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice.* New York, NY: Cambridge University Press.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.

Graesser, A. C., Bertus, E. L., & Magliano, J. P. (1995). Inference generation during the comprehension of narrative text. *Sources of coherence in reading*, 295-320.

Jamieson, J. (2013). *ENG638: Using SPSS for dichotomously scored items*. [Class handout]. English Department, Northern Arizona University. Flagstaff, AZ.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Kispal, A. (2008). Effective teaching of inference skills for reading. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/222352/DCSF-RR031.pdf

Koda, K. (2004). *Insights into second language reading*. New York, NY: Cambridge University Press.

McKoon, G., & Ratcliff, R. (1992). Inferencing during reading. *Psychological Review*, 99(3), 440-466.

Miller, M. D., Linn, R. L., & Gronlund, N. (2013). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Pearson.

Pagano, R. R. (2010). *Understanding statistics in the behavioral sciences*. Belmont, CA: Wadsworth.

Stoynoff, S., & Chapelle, C. A. (2005). *ESOL tests and testing*. Alexandria, VA: TESOL.

Swender, E., Conrad, D. & Vicars, R. (2012). *ACTFL proficiency guidelines 2012*. Alexandria, VA: American Council for the Teaching Foreign Languages. Retrieved from http://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf

Urquhart, S., & Weir, C. (1998). *Reading in a second language: Process, product and practice*. New York, NY: Addison Wesley Longman.

Warren, W. H., Nicholas, D. W., & Trabasso, T. (1979). *New directions in discourse processing* (2nd ed.). Norwood, NJ: Ablex Publication Corporation.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162-185.