

## When One is Not Enough: Translation Rating and the Assessment of Partial Word Knowledge

### ABSTRACT

An important requirement in foreign language incidental vocabulary acquisition research is accurate assessment of partial word knowledge. Open-format L1 translation tests are increasingly used for this purpose. What level of precision is appropriate in the translation rating procedure? To answer this question, we analyze experimental data from a read-and-test study. We rate the pretest and posttest translations on an eleven-level scale. Through an approximation process, we derive equivalent binary, three-level, and six-level data. We apply the Mann-Whitney  $U$  Test to each of the four data sets (eleven-level, binary, three-level, and six-level) to identify the words for which subject knowledge improvement reached significance. Using the original, eleven-level data as a standard, we show that binary and three-level rating lead to false positives and false negatives. We draw two conclusions. 1. Not all partially correct translations deserve equal credit. 2. Multi-level rating is a more precise measure of translation accuracy than binary and three-level rating. We discuss practical rating issues and the advantages of using a pretest and posttest as opposed to a posttest only.

**Key words:** incidental vocabulary acquisition, English for specific purposes, partial word knowledge, translation rating, vocabulary assessment

---

## When One is Not Enough: Translation Rating and the Assessment of Partial Word Knowledge

### ABSTRACT

An important requirement in foreign language incidental vocabulary acquisition research is accurate assessment of partial word knowledge. Open-format L1 translation tests are increasingly used for this purpose. What level of precision is appropriate in the translation rating procedure? To answer this question, experimental data from a read-and-test study are analyzed. The pretest and posttest translations are rated on an eleven-level scale. Through an approximation process, equivalent binary, three-level, and six-level data are derived. The Mann-Whitney  $U$  Test is applied to each of the four data sets (eleven-level, binary, three-level, and six-level) to identify the words for which subject knowledge improvement reached significance. By using the original, eleven-level data as a standard, it is shown that binary and three-level rating lead to false positives and false negatives. Two conclusions are drawn: 1. Not all partially correct translations deserve equal credit; and 2. Multi-level rating is a more precise measure of translation accuracy than binary and three-level rating. Practical rating issues and the advantages of using a pretest and posttest as opposed to a posttest only are also discussed.

**Keywords:** incidental vocabulary acquisition, English for specific purposes, partial word knowledge, translation rating, vocabulary assessment

### Editor's Notes:

(1) "For clarity, restrict your use of **we** to refer only to yourself and your coauthors (Use **I** if you are the sole author of the paper). Broader

[Editor's Notes, continued]

uses of **we** leave your readers to determine to whom you are referring; instead substitute an appropriate noun or clarify your usage."

*(Publication Manual of the American Psychological Association, 5<sup>th</sup> Ed., p. 39)*

Since this abstract does not show the name(s) of the author(s), it is impossible for this reader to know how many there are. However, repeated usage of "we" by multiple authors soon becomes tediously redundant. The best way to avoid using "we" is to use the passive voice of the verbs. Although this practice is not universally recommended, it is supported in textbooks such as *Academic Writing for Graduate Students: A Course for Nonnative Speakers of English* (John M. Swales & Christine Feak, Michigan Series in English for Academic & Professional Purposes, The University of Michigan Press, 1994) and *Writing Up Research: Experimental Research Report Writing for Students of English* (Robert Weissberg & Suzanne Buker, Prentice Hall Regents, 1990).

(2) The keywords—actually *phrases*--highlighted in **red** do not occur in the text of this abstract. Since an abstract should be capable of standing alone without reference to the complete report, all keywords should occur in the abstract. Therefore, I recommend one of the following revisions: (a) Delete the keywords which do not occur in the abstract, or (b) revise the abstract to include the **highlighted** keywords in the text.