

## **Statistical procedures for research on L2 reading comprehension: An examination of ANOVA and Regression Models**

Cindy Brantmeier  
Washington University

### **Abstract**

Through a discussion of research that examines a plethora of variables involved in second language (L2) reading comprehension, the present study attempts to examine and analyze the statistical procedures utilized in studies of this nature. A review of recent research from the past five and a half years from four leading scientific journals of reading is offered. Research questions that motivate the selection of statistical procedures are examined for each study. Results show that Analysis of Variance (ANOVA) is utilized more than Regression Models (RM) primarily because researchers are asking questions about the variation between and within groups of variables and are not predicting performance on dependent variables via independent variables. The strong resemblances and differences between ANOVA and RMs are discussed in light of the review of research, and through a detailed critique of Brantmeier's (2003) study with different research questions and additional analysis of data, the relationship between statistical procedures is further exemplified. Explanation for the use of statistical procedures in light of recent theoretical models (Bernhardt, 2003) is included.

*Keywords:* second language reading comprehension, statistical procedures, analysis of variance, regression models, methods of reading research

### **Introduction**

In a discussion about the applied linguistics contribution to second language (L2) reading, Urquhart and Weir (1998) contended that there was no established body of experimental methods for applied linguists to rely on. Six years later we see that L2 reading research conducted by applied linguists continues to take many forms, but researchers who conduct experimental, quantitative investigations concerning second language reading comprehension engage in a number of similar activities.<sup>1</sup> The present study attempts to examine one component of the research process for investigations of this type: the stage where statistical procedures and techniques are selected and utilized. Selection of statistical procedures is an integral part of the research process, and this choice is motivated by research questions and validated through a discussion of results.

A mixture of prior investigations concerning research methods has been influential for researchers conducting studies concerning L2 reading comprehension. These experiments have

examined factors involved in the creation of data collections instruments such as passage type and structure (Bernhardt, 1984; Brown, 1987; Leow, 1993; Leow, 1997; Tsang, 1987; etc.), passage content (Brantmeier, 2002; Brantmeier, 2003; Bügel and Buunk, 1996; Carrell, 1984; Hudson, 1982; Johnson, 1981; Mohammed and Swales, 1984; Pritchard, 1990; Steffenson, Joagdev, and Anderson, 1979; Schueller, 1999; Young and Oxford, 1997, etc.), assessment tasks (Carrell, 1991; Lee, 1990; Shohamy, 1984; Wolf, 1993; etc.), language used for assessment (Lee and Ballman, 1987; Shohamy, 1982, 1984; Wolf, 1993; etc.), and procedures utilized for scoring instruments (Bernhardt, 1991).

More specifically, with regard to codifying and scoring data, Bernhardt (1991) argued that collected data must be scored consistently both within the study and across L2 reading studies in order to make appropriate generalizations. In the same vein, this present study attempts to show that statistical tests should also be utilized appropriately and consistently both within and across inquiries. Researchers agree that a solid research plan for L2 reading comprehension involves a description of intended data analyses including statistical procedures. To date, no investigation has reviewed and analyzed the statistical procedures most commonly utilized in research on L2 reading comprehension. Through a synthesis of prior L2 reading research that examines comprehension, the present study attempts to do the following: 1) demonstrate which statistical procedures are currently being utilized; 2) report the research questions that motivate the choice of statistical tests; 3) discuss the strong resemblance and difference between statistical tests utilized to analyze data; and 4) exemplify the relationship between statistical procedures through a critique of a recent study. L2 reading research that examines comprehension during the last five and a half years from the following leading scientific journals of reading is reviewed: *Journal of Literacy Research*, *Reading in a Foreign Language*, *The Reading Matrix*, and *Reading Research Quarterly*.

## **L2 reading models and studies on comprehension**

Before moving into an examination of recent research, a brief discussion of L2 reading models and comprehension is essential. Though interactive models of L2 reading emphasize different components involved in the process, all models include and underscore the importance of comprehension (Bernhardt, 1991; Coady, 1979). Throughout the years L2 reading researchers have defined and discussed comprehension while relying heavily on Bernhardt's (1991) model (Hammadou, 1991; Lee and VanPatten, 1995; Wolf, 1993; Young, 2000), and they all agree that comprehension is obviously a critical part of the multifarious interplay of mechanisms involved in L2 reading. It is not new news that different comprehension assessment tasks may be testing different abilities. Measures of comprehension consist of a variety of assessment tasks including free written and oral recalls, summaries, multiple choice, true/false, close-deletion items, open-ended questions, and sentence completions. Dating back to the 80's and up to the present day, L2 reading researchers have utilized a mixture of comprehension assessment tasks. For example, Block (1986) utilized verbal retellings and a written multiple choice test; Anderson (1991) echoed Block's comprehension measures, but he reversed the order; Sarig (1987) utilized verbal reports of main ideas and the overall messages of the passages; Barnett (1989) used a written recall and participants also chose the most appropriate continuation of the story; Carrell (1989) only used written multiple-choice questions to assess comprehension. More recently, researchers continue to utilize a variety of assessment tasks. Bügel and Buunk (1996) utilized multiple choice questions from a standardized exam; Young and Oxford (1997) utilized oral recalls;

Schueller (1999) used both multiple-choice and open-ended questions; and Brantmeier (2002; 2003) utilized both multiple-choice questions and written recall.

Discussions of varied comprehension assessment tasks across studies are not enough. The diversity in measurement tasks leads to the following questions: What statistical tests are utilized to analyze data in investigations that examine L2 reading comprehension? Are there variations across studies? Which procedures are appropriate for the research questions? How much confidence can we place in results and conclusions? The present study hopes to answer these questions and more. Bernhardt's (2000; 2003) model of L2 reading illustrates that 50% of L2 reading is accounted for by L1 literacy (20%) and L2 knowledge (30%), and she contends that more research is needed to examine the remaining 50% of variance that is unexplained. Current studies with appropriate research questions and corresponding statistical tests may contribute to the unexplained variance, and in addition, recent research may possibly examine which of the many interacting variables in L2 reading models best predicts successful comprehension.

### Review of research about L2 reading comprehension

Figure 1 lists the investigations in the aforementioned academic journals that examine L2 reading comprehension with adults and children, and it also reports research questions, statistical procedures utilized to analyze data, and findings.

Figure 1: Literature review on L2 reading comprehension

\*Articles are listed by year and then alphabetical order

| Author                                   | Research Questions  | Statistical Results                          | Procedures   |
|--|---|--|--|
| Droop and Verhoeven (1998)<br><i>JLR</i> | Does the cultural background of schoolbook texts influence first- and second-language reading comprehension? To what extent does the linguistic complexity of the text constrain the effects of different cultural schemata on first- and second-language reading comprehension?                              | MANOVA, Wilks Lambda                         | A facilitating effect of cultural familiarity was found for both reading comprehension and reading efficiency. For the minority children, this effect was restricted to linguistically simple texts, because of their limited knowledge of the target language, Dutch. |
| Tweissi (1998)<br><i>RFL</i>             | Does language simplification (LS) have a positive influence on reading comprehension? Does the difference or amount of LS and type of LS result in differences in the levels of reading comprehension? Which of the amounts and types of LS are superior in producing higher levels of reading comprehension? | One-Way ANOVA, Tukey Pairwise and Regression | The type of linguistic features involved in the process of simplification, not how many parts of the text receive simplification, will produce the needed modification to render a text more comprehensible to L2 learners.  |

|  |  |   |  |
|--|--|---|--|
| Wilkinson (1998)<br><i>RRQ</i>                     | What school and classroom factors moderate gender and home language gaps in reading achievement?   | Hierarchical Linear Model: random effects ANOVA model, random-coefficients regression model, and intercepts- and slopes-as-outcomes model; Chi-Square tests | The magnitudes of the gender gap for comprehension and of the home language gaps for comprehension and word recognition varied across schools. Factors that moderated the gaps were those that reflected teachers' capacities to handle diversity.   |
| Mori and Nagy (1999)<br><i>RRQ</i>                 | Does a student who appropriately uses one source of information (kanji clues) also use another source of information (context clues) successfully? Does the ability to use one source of information (either kanji or contextual clues) correlate with the ability to integrate information?           | 1-Way ANOVA, Tukey-Kramer HSD test, Correlations  | Students were most likely to obtain correct answers when both types of clues were available, demonstrating their ability to combine information from multiple sources to interpret unfamiliar words. Use of kanji clues and context use are not correlated, and proficiency correlates with context use, but not with kanji use. |
| Steffenson , Goetz, and Cheng (1999)<br><i>JLR</i> | Does decoding a foreign language make such heavy demands on attentional resources that it minimizes (or precludes) the formation of nonverbal (imagery, affect) representations, or is nonverbal representation an integral and obligatory part of reading, as proposed by dual coding theory?         | MANOVA, Correlations  | English readers produced fewer reports of imagery. English readers did not understand the passage as well as the Chinese readers did. Imagery and affect were formed even in the absence of total understanding. This shows that they are fundamental variables in foreign language reading.                                     |
| Hsueh-chao and Nation (2000)<br><i>RFL</i>         | Will different densities of unknown words result in differences in comprehension? In particular, as the number of unknown words increases, will comprehension decline? Is there a vocabulary coverage level which acts as a threshold between adequate and inadequate comprehension of a fiction text? | Regression and ANOVA  | This research does not support the idea of a 95% vocabulary knowledge threshold for comprehension of narrative text. On average, learners' comprehension scores increase to a predictable degree as the coverage of known words increases.   |

|                                      |  |  |  |
|--------------------------------------|--|--|--|
| Van den Branden (2000)<br><i>RRQ</i> | Does negotiation of meaning promote the comprehension of Dutch written input by primary school pupils, and under which conditions does negotiation of meaning optimally produce the comprehension of written input in the context of the real-life language classroom?   | Repeated Measures ANOVA, Post hoc analyses | Negotiating the meaning of unmodified written input led to higher comprehension than premodifying the same input. Meaning negotiation in which the teacher was involved was superior to peer negotiation. Comprehension scores were higher for students who had cooperated with a peer of a different level of language proficiency than for students who had cooperated with a peer of similar level of language proficiency. |
| Bell (2001)<br><i>RM</i>             | Will learners in the 'extensive' group achieve significantly faster reading speeds than those in the 'intensive' group as measured on relatively easy, non-problematic texts? Will learners in the 'extensive' group achieve significantly higher scores on a test of reading comprehension containing texts at an appropriate level, than those in the 'intensive' group? | t- test                                    | Subjects exposed to "extensive" reading achieved both significantly faster reading speeds and significantly higher scores on measures of reading comprehension.  |
| Carrell (2001)<br><i>RFL</i>         | Is there an interaction between purpose and task? In other words, will purpose for reading relate to the specific task which conforms to that purpose? And if so, what is the nature of that interaction?  | Two-Way ANOVA                              | Students perform better on a task which conforms to their purpose of reading. One purpose does not facilitate higher scores than another (reading-to-recall and reading-to-do).  |
| Liontas (2001)<br><i>RM</i>          | What reading strategies and pragmatic features govern and characterize the comprehension and interpretation process of Greek phrasal idioms during contextualized and acontextualized reading?   | Frequencies and Means calculated           | Idiom understanding involves more than recognizing a lexemic string as an idiom; it implies the syntactic and semantic processing and metaphorical extension of the lexemes forming the idiom which can be used with the surrounding context to generate further interpretations.  |
| Leung (2002)<br><i>RFL</i>           | Does extensive reading lead to vocabulary acquisition? Promote reading comprehension? Promote positive attitudes toward reading? What challenges does a beginning foreign language learner face in the extensive reading process and how did the learner deal with these challenges?   | Frequency Mean Scores                      | Results from vocabulary tests reveal that vocabulary knowledge increased 23.5% in one month. Data from the journal entries show that Wendy's reading comprehension gradually improved throughout the course of the study.  |

|  |   |  |  |
|--|---|--|--|
| Salataci and Akyel (2002)<br><i>RFL</i>  | Does strategy instruction in EFL reading affect EFL reading strategies and reading comprehension in English? Does strategy instruction in EFL reading affect reading strategies in Turkish?   | Wilcoxon<br>Matched-pairs<br>Signed-rank<br>tests,<br>Correlations | Strategy instruction had a positive effect on both Turkish and English reading strategies and reading comprehension in English.  |
| Sharp (2002)<br><i>RFL</i>               | Will the rhetorical organization of English (as represented by four different patterns) affect the reading comprehension of native Chinese school pupils? Will the gender or English language proficiency levels of the Chinese school pupils have a significant effect on reading comprehension? | 1-Way ANOVA  | Cloze testing indicated significant differences between the four rhetorically different texts. English proficiency levels appeared to have little effect on rhetorical preferences. There are no substantial differences between the texts for recall quantitative scores for either boys or girls, but the mean scores remain consistent across the four texts. |
| Stakhnevich (2002)<br><i>RM</i>          | What is the impact of the web instructional medium on L2 comprehension during independent reading versus the traditional print medium and a control?  | ANCOVA<br>ANOVA  | The medium of instruction does have an impact on the level of reading comprehension, with the web mode resulting in better performance when compared to the traditional print mode.  |
| Taguchi and Gorsuch (2002)<br><i>RFL</i> | Does the RR method significantly help foreign language readers improve their silent reading rate when reading a new passage? Does the RR method significantly help FL readers improve their reading comprehension when reading a new passage?   | T-test,<br>Mann Whitney<br>U test                                  | The silent reading rate of the experimental group improved significantly from the initial reading of the pretest passage to that of the posttest passage. The reading performances by the experimental group were not significantly different from those by the control group.   |
| Brantmeier (2003)<br><i>RFL</i>          | Are there gender differences in learners' topic familiarity? Are there gender differences in learners' second language reading comprehension? Does the gender-oriented passage content of the second language reading text affect learners' comprehension?  | 2 Way-ANOVA,<br>Kruskal-Wallis                                     | No significant difference between mean scores for males and females on overall comprehension of the passages. There was no difference in performance by gender across passages.  |
| Camiciottoli (2003)<br><i>RFL</i>        | Are L2 readers able to understand a text containing more metadiscourse better than one with less?   | t-test   | Some significant positive effect for metadiscourse on specific questions.  |

|  |  |   |   |
|--|--|---|---|
| Droop and Verhoeven (2003)<br><i>RRQ</i> | Do differences in the development of the oral language, word decoding, and reading comprehension skills of L1 versus L2 learners occur? Do differences between the two minority groups and the Dutch children from high versus low socio-economic backgrounds occur? And, if so, do the various differences remain the same, converge, or diverge over time? What interactions are found between the oral language capacities, word decoding capacities, and reading comprehension capacities of the L1 and L2 learners? | MANOVA,<br>Wilks Lambda,<br>Chi-Square,<br>Correlations | Minority children were faster decoders than Dutch-low-socio-economic children. Regarding reading comprehension and oral language proficiency, the minority children lagged behind the Dutch children in all respects. The development of reading comprehension was more influenced by top-down strategies than bottom-up processes for both L1 and L2 learners. The oral Dutch skills of the minority group played a more prominent role in the explanation of their reading-comprehension skills than the oral-language skills of the Dutch group. |
|--|--|---|---|

Note: *JLR = Journal of Literacy Research*  
*RFL = Reading in a Foreign Language*  
*RM = The Reading Matrix*  
*RRQ = Reading Research Quarterly*

As demonstrated in the review of studies about L2 reading comprehension, Analysis of Variance (ANOVA) is the most widely used statistical procedure in this type of research. This is because, as depicted by the research questions, L2 reading researchers often investigate the relationship of many different independent variables with dependent variables and are concerned about the variation between and within groups of variables. For example, the following research questions guided Carrell (2001): Is there an interaction between purpose and task? In other words, will purpose for reading relate to the specific task which conforms to that purpose? And if so, what is the nature of that interaction? Given these inquiries, Carrell selected the appropriate statistical test (ANOVA) to answer her questions. Only two of the 18 studies used a regression model (RM) to analyze data, and some studies utilized both ANOVA and multiple regression MR (e.g., Wilkinson, 1998; Tweissi, 1998). For instance, Wilkinson (1998) asked: What school and classroom factors moderate gender and home language gaps in reading achievement? Wilkinson selected a variety of statistical tests including RM to answer this question because he was interested in predicting which variable (e.g., gender and home language) best predicts reading achievement. To further exemplify choice of statistical tests, a detailed discussion of ANOVA and RM follows.

## Statistical procedures: ANOVA and RM

### *Assumptions underlying ANOVA and RM*

The general assumptions underlying the use of ANOVA are the following: 1) data are score or ordinal scale data that are continuous; 2) data are independent. The comparison is between

groups; 3) there is a normal distribution of scores in each group; 4) there are equal variances of scores in each group; 5) there is a minimum of five observations per cell; and 6) the F statistics allow the rejection or acceptance of the null hypothesis (Hatch and Lazarton, 1991).<sup>2</sup> If these assumptions cannot be met, then nonparametric tests can be utilized, such as the Kruskal-Wallis test which also can help determine whether there are significant differences between groups. If differences are found with this test, then the Ryan procedure is often used to understand the exact location of the differences.

The general assumptions underlying the use of RM are the following: 1) the variables are interval or truly continuous and the relationship is linear; 2) correlation values are accurate; 3) the variables entered in the regression formula should not be highly intercorrelated; 4) the more variables there are in the regression equation, the larger the N size for the study must be; and 5) if the procedure is used for inferential purposes then the sample must be drawn at random, normal distribution and equal variances must be found (Hatch and Lazarton, 1991).<sup>3</sup>

### *ANOVA*

The most common way for L2 reading researchers to find out if there are significant differences between the means of more than two groups is with the ANOVA procedure, which is actually a t-test that is appropriate to use with three or more groups. ANOVA examines the variation both within and between each of the groups. Technically, ANOVA compares two different estimates of the same variance under the null hypothesis. One variance estimate is based on the within-group variation of scores around group means (error variance). In the experimental designs in Figure 1 that utilized ANOVA tests, all of the people within a group are expected to be the same, except for random variations, because they have all been treated the same. This variance estimate is the denominator of the F test. The other variance estimate is based on the variation of group means around the grand mean and is the numerator of the F test. This variation can be due to two sources: a) random variation and b) systematic variation due to an experimental treatment. But, under the null hypothesis, the second source is assumed to be zero and so the F ratio will tend to be 1.00 if the null hypothesis is true. To the extent that the null hypothesis is false, the F ratio exceeds 1.00, and if it exceeds 1.00 enough, the null hypothesis is rejected (Hatch and Lazarton, 1991).<sup>4</sup>

There are several types of ANOVA tests. In the One-Way ANOVA there is exactly one dependent variable (always continuous) and exactly one independent variable (always categorical) (e.g., Tweisse, 1998; Mori and Nagy, 1999; Sharp, 1999). A Two-Way ANOVA procedure attempts to discover whether the interaction of two independent variables has an effect on the dependent variable (e.g., Carrell, 2001; Brantmeier, 2003). An Analysis of Covariance is a variation of ANOVA where the researcher adjusts mean scores on the dependent variable for each group to compensate for the initial differences between groups on another variable, which is the covariate (e.g., Stakhnevich, 2002). A Multivariate Analysis of Variance (MANOVA) uses two or more dependent variables in the same analysis, and this is used when the researchers believe that correlations exist among the dependent variables. (e.g., Fraenkel and Wallen, 1996). In the MANOVA, there may be multiple dependent variables and multiple independent variables (e.g., Droop and Verhoeven, 2003).



Finally, if significant differences are found among the means in the ANOVA procedures, then the researcher calculates post hoc comparisons (two-tailed tests) to identify more specifically where the difference lies.<sup>5</sup>

### *Regression Models*

Simply stated, regression is used to predict performance on the dependent variable via one or more independent variables (e.g., Tweisse, 1998; Wilkinson, 1998). In simple regression, researchers predict scores on one variable on the basis of scores on the second. In MR, the possible sources of prediction are expanded and tested to see which of many variables and which combination of variables allows the researcher to make the best prediction (Hatch and Lazarton, 1991). In other words, MR is a technique used to determine a correlation between a criterion variable and the best combination of two or more predictor variables. MR is the extension of a simple linear regression.

In second language reading research, simple regression has been used when researchers need to predict scores on a test on the basis of another test. According to Hatch and Lazarton (1991), MR is used when researchers want to know how much "weight" to give to a number of possible independent variables that relate to performance on the dependent variable. For example, prior research that examines a comprehension assessment test for L2 reading may have shown that success on the test is related to factors such as topic familiarity levels, gender, type of assessment task, etc. By using a MR model researchers can determine which of the variables best predicts achievement. A combination of these variables or which variables do not predict achievement can also be predicted. In MR analysis, the amount of explained variation is often contrasted with residual, which is unexplained variation. MR takes correlations among the predictors into account, and thus gives estimates of the unique variance accounted for in the outcome by the predictors.

### *Relation between ANOVA and Regression*

An ANOVA identifies whether the mean of one group differs significantly from the mean of another group or groups. Regressions identify whether two or more variables are significantly related to each other. Hatch and Lazarton (1991) offer a discussion about the resemblance between ANOVA and Regression. They contend that in ANOVA researchers account for the variance in a DV on the basis of two major components: the variance between groups (including the treatment effect and error) and the variance within groups (error only). In regression analysis, researchers can conceive of the sum of squares for the predicted value of Y as the sum of squares regression (the predicted variation) and the leftover variation as sum of squares residual (which is the variance left unaccounted for).

In MR researchers need as many non-redundant predictors as they have degrees of freedom for the main effect (or any effect for that matter). In ANOVA it is customary to have one source of variation for each main effect and one source of variation for each interaction, perhaps because ANOVA is really a special case of MR. Aiken and West (1996) carefully address the difference between ANOVA and MR in usual practice. They state:

In ANOVA with multiple levels of a factor and the use of usual approaches to variance partitioning, any curvilinear variation is automatically subsumed in the

variance partitions. In contrast, in MR the analyst specifically decides which terms need to be included: Terms to represent curvilinear relationships must be built systematically into the equation (Aiken and West, 1996: 71).

In other words, with ANOVA the multiple degrees of freedom for any multiple-degree-of-freedom effect are combined and tested together. In MR, each single degree of freedom is usually tested individually. The authors continue to discuss how ANOVA and MR do not differ mathematically. What they contend is that the conventional partitions of variance operationalized in common statistical packages for ANOVA are structured so that all components of an effect are subsumed in the omnibus term for that effect. In MR the structuring of the components of each effect is left to the analyst (Aiken and West, 1996: 71).

What all this means to the second language reading researcher is that both ANOVA and Regression are dealing with variance in the DV, and they account for as much variance as possible as an "effect of" (ANOVA) or "accounted for by" (regression) various independent variables (Hatch and Lazarson, 1991: 486).

### **A critique of a study**

Brantmeier (2003) employed ANOVA in analyzing data for a study on L2 reading comprehension. More specifically, Brantmeier's study examined the effects of readers' gender and passage content on L2 reading comprehension with participants from the intermediate level of language instruction. Seventy-eight participants read two different authentic passages, and two different measures were used to assess comprehension: written recall and multiple choice questions. The following research questions guided the study:

1. Are there gender differences in learners' topic familiarity?
2. Are there gender differences in learners' second language reading comprehension?
3. Does the passage content of the second language reading text affect learners' comprehension?

Findings revealed significant interactions between readers' gender and passage content with comprehension on both assessment tasks. The results of the study provided evidence that subject matter familiarity has a facilitating effect on L2 reading comprehension by gender at the intermediate level of Spanish language instruction.

In summary, Brantmeier's (2003) study was undertaken in order to examine the interaction effects of readers' gender and passage content on L2 readers' comprehension at the intermediate level of Spanish language instruction. In this research design, the independent variables were: 1) passage content (boxing and housewife) and 2) readers' gender. The two sets of dependent variables were: 1) comprehension (measured with the written recall protocol and multiple choice comprehension questions) and 2) topic familiarity.

In order to compare several means simultaneously and to assess interaction effects, for research questions two and three data were submitted to a two-way Analysis of Variance (ANOVA). The ANOVA procedure showed the between-subject main effect (e.g., gender) and the within-subject

main effect (e.g., passage content), as well as their interaction. The alpha level was set at .05 (Brantmeier, 2003: 8).

### *Evaluation of the appropriateness of statistical procedures*

An attempt to answer the following questions about the appropriateness of statistical procedures follows: What statistical tests are utilized to analyze data in Brantmeier's (2003) study about L2 reading comprehension? Were the procedures appropriate for the data? How much confidence can we place in Brantmeier's results and conclusions?

The goal of the ANOVA in Brantmeier's (2003) design was to explain the variance in the dependent variable (written recall or multiple choice) in terms of variance in the independent variables (reader's gender and passage content). The type of ANOVA (two-way) was used because more than one independent variable was involved in the separate designs. Brantmeier is careful to note that the two passages are not being compared. In other words, the effect of passage content was not considered as boxing content versus housewife content. Rather, the passages are treated and tested as separate entities in a single report.<sup>6</sup> Passage content (boxing passage and housewife passage) is included in the statistical design for each separate analysis, and therefore this study could actually be reported as two separate experiments. Because Brantmeier used the same participants and they followed the same procedures, the results of both experiments were reported in a single article.

The researcher expected that there would be variability in the performance of males and females on the comprehension tests. She wanted to know what effect the gender factor had on variability in the data, as well as what effect the passage content factor had on that variability. She also wanted to know the effect of the combination of passage content and gender on variability in comprehension test performance. In other words, Brantmeier (2003) examined the following:

1. Effect of gender: male versus female
2. Effect of passage content: boxing or housewife
3. Interaction effect (gender by passage content)

The advantage of using a two-way ANOVA in this study is that the researcher was able to look not only at the effect of each independent variable but also the interaction effect in the combination of independent variables. Results of the ANOVAs showed no significant difference between mean scores for males and females on overall comprehension of the passages. There was no difference in performance by gender across passages, however, results of the ANOVAs yielded significant interactions between independent variables readers' gender and the boxing passage content as they affect dependent variables recall ( $F(1,76) = 8.26, p = .01, \eta = .10$ ) and multiple choice questions ( $F(1,76) = 4.20, p = .04, \eta = .05$ ). Likewise, the results of the ANOVAs yielded significant interactions between readers' gender and housewife passage content as they affect recall ( $F(1,76) = 15.90, p = .00, \eta = .18$ ) and multiple choice ( $F(1,76) = 8.67, p = .00, \eta = .10$ ). Brantmeier includes the following footnote supporting her choice of statistical procedures:

A one-way ANOVA and a bivariate regression model with a dichotomous independent variable are precisely the same (King, 1986). The only substantive difference is that in the ANOVA case one only reports whether there exists a

significant difference or not, and therefore to answer the research questions in the present study the ANOVA was calculated. In a bivariate regression the magnitude of the difference is reported, but in the present study the reported sample means by group (e.g., gender) reveal the magnitude, and the ANOVA shows whether the difference is significant or not (Brantmeier, 2003: 12).

Findings revealed significant interactions between readers' gender and passage content with comprehension on both assessment tasks (written recall and multiple choice). Hatch and Lazarson (1991) and Kirk (1982) state that when interpreting results of a two-way ANOVA that the interpretation must emphasize the interaction effect when it is significant. If the interaction effects are not significant, then more powerful statements can be made about the effects of the independent variables on the dependent variables. Brantmeier's interpretation and discussion focused on the significant interaction effects. She states:

The results of the present study indicated that two important interacting factors in the L2 reading process of university students of intermediate Spanish are the readers' gender and passage content. Male and female readers were able to make connections to familiar passage content, and therefore were able to understand and comprehend better as they read (Brantmeier, 2003: 12).

How much confidence can we give the findings given the statistical procedures used? In Brantmeier's (2003) study, she was interested in accounting for as much variance in multiple choice and recall as possible as an "effect of" (ANOVA) the reader's gender and passage content. She was not interested in accounting for as much variance as possible in multiple choice and recall as "accounted for by" (regression) these independent variables.<sup>7</sup> The study was not an examination of how well the researcher could predict scores on the multiple choice and recall tests from the scores on two or more independent variables. The author was not interested in knowing what combination of variables best predicts scores on comprehension tests. If the researcher were interested in these inquiries, then both research questions as well as the overall research design would need to change accordingly.

An example of a question that would require regression analysis for Brantmeier's (2003) study would be: How much of the variance in multiple choice and recall did gender account for? In order to show this predictive relationship among the effect of Readers' Gender on the performance of readers, data are re-examined using regression analysis. Results show that overall (both passages combined) readers' gender accounts for 14% of variance in written recall and 7% of variance in multiple choice questions. To provide further analysis, both passages are analyzed separately. Results show that for the boxing passage, readers' gender accounts for 10% of variance in written recall and 5% of variance in multiple choice questions. For the housewife passage, readers' gender accounts for 17% of variance in written recall and 10% of variance in multiple choice. These results add intriguing dimension to Brantmeier's (2003) findings. One way to interpret these results is that readers' gender accounts for greater variance in the written recall assessment measure than in the multiple choice questions. Future investigations could examine this relationship even further.

Regarding Brantmeier's (2003) study, regression analysis can also show which of the IVs (readers' gender or topic familiarity) are superior (more influential) in producing higher scores on

reading comprehension. Through MR a test of the difference between two regression coefficients can be derived. Results are listed in Table 1:

Table 1: Regression Analysis

*Boxing Passage*

| Predictors (Constant) |        | R <sup>2</sup> | T-ratio | P    |
|-----------------------|--------|----------------|---------|------|
| Readers' Gender       | MC     | 0.05           | -2.10   | 0.00 |
|                       | Recall | 0.10           | -2.20   | 0.03 |
| Topic Familiarity     | MC     | 0.11           | -3.10   | 0.00 |
|                       | Recall | 0.05           | -1.90   | 0.05 |

*Housewife Passage*

| Predictors (Constant) |        | R <sup>2</sup> | T-ratio | P    |
|-----------------------|--------|----------------|---------|------|
| Readers' Gender       | MC     | 0.10           | 2.90    | 0.00 |
|                       | Recall | 0.17           | 4.00    | 0.00 |
| Topic Familiarity     | MC     | 0.14           | -3.50   | 0.00 |
|                       | Recall | 0.14           | -3.50   | 0.00 |

As depicted on Table 1, the number R<sup>2</sup> yields a value that depicts the proportion of variation in the dependent variable (either multiple choice or recall) that is explained by independent variables (readers' gender and topic familiarity). For example, findings indicate that with the boxing passage, readers' gender (RG) accounts for more variance than topic familiarity (TF) in recall (RG = 10%; TF = 5%), but the reverse is true for multiple choice (RG = 5%; TF = 11%). Likewise, with the housewife passage, results show that readers' gender accounts for more variance than topic familiarity in recall (RG = 17%; TF = 14%), and again, the reverse is true for multiple choice (RG = 10%; and TF = 14%). In summary, RG is more influential than TF in producing higher recall scores, but TF is more influential than RG in producing higher multiple choice scores. These results underline the need for more research on variables that influence performance on comprehension assessment tasks.

An excellent example of a study that emphasizes both the effect of independent variables as well as variance accounted for by independent variables is Tweissi (1998). This study formulated the following research questions: Does language simplification (LS) have a positive influence on reading comprehension? Does the difference or amount of LS and type of LS result in differences in the levels of reading comprehension? Which of the amounts and types of LS are superior in producing higher levels of reading comprehension? The researcher utilized a One-Way ANOVA, a Tukey Pairwise and a Regression procedure to analyze data. Tweissi (1998)

states that the study investigates the influence of one independent variable with five levels (language of the text) on one dependent variable (level of comprehension) and therefore the One-Way ANOVA was used. Because the null hypothesis about the effect of simplification on reading comprehension was rejected, the researcher applied two other statistical procedures to answer the other research questions: a Post-hoc analysis using the Tukey's pairwise comparisons, and a MR analysis. The MR analysis was used to answer the following question: Which of the amounts and types of simplification are superior (e.g., more influential) in producing higher levels of reading comprehension? The MR specifically showed predictive relationships among the effects of the five versions of text on performance of readers by predicting scores based on these versions (Tweissi, 1998: 197). Findings revealed the following: "The premise that the simpler the text the more comprehensible to L2 learners is unwarranted. LS in general has a positive influence, however, increasing the amount of LS alone does not lead to greater comprehension. The type of simplification, rather than the amount, may have a higher impact on reading comprehension" (Tweissi, 1998: 201). Given Tweissi's (1998) research questions and research design, the MR was necessary, as explained previously.

## Conclusion

As shown in the present study, the selection of appropriate statistical procedures driven by research questions is a critical part of the L2 reading research process. The summary of recent studies shows that ANOVA is the test most commonly used in experimental research of this type. The reviewed studies demonstrate that when ANOVA has been employed in analyzing data for inferential purposes, the appropriateness of the procedure for the study has been directly supported. In light of new issues about L2 reading (Bernhardt, 2003) perhaps more inquiries about L2 reading comprehension should be concerned with the amounts and types of variables that are superior, or more influential, in producing higher levels of reading comprehension. Studies that show predictive relationships among the effects of variables could contribute to the lacuna in the database concerning the 50% of unexplained variance in Bernhardt's (2001) model. Through a re-examination and further analysis of a published study, the present investigation attempts to exemplify the rationale behind ANOVA and MR. As a final point, although ANOVA and MR may be mathematically equivalent, analyses should be tailored to test specific research questions.

## Notes

1. Brantmeier (2004) offers a concise review of research methods commonly utilized in L2 reading research and includes a graphic presentation to show the typical sequence in which the mechanisms are usually executed and described in a study.
2. See Chapter 11, Hatch and Lazardon (1991) for more specific details and examples.
3. The present study does not attempt to examine whether researchers have violated the underlying assumptions for the use of ANOVA and RM, but rather attempts to identify and discuss statistical tests utilized.

4. The t test is simply a special case of the same technique, when there are only two groups. In that case,  $F = t^2$ . But, when there are more than two groups, there is no direct way to derive a t test.
5. See Kirk (1982) for a detailed description of post hoc comparisons.
6. When comparing the two passages, the Cortazar passage yielded lower comprehension scores on both multiple choice and recall, independent of gender. Text difficulty could be a limitation to the extent that it would be an intervening variable, or a variable that was not included in the present study. The author chose not to control for text difficulty because the study does not make comparisons of recall comprehension scores between the two passages, rather it examines the differences in recall comprehension scores by gender within each passage. Furthermore, to maintain authenticity, the researcher did not simplify the Cortazar text.
7. When using ANOVA, if the author was interested in determining the proportion of variability in multiple choice or recall that was accounted for by passage content or gender, a strength of association measure (omega squared) could have been calculated in a balanced design. For an unbalanced design, the eta squared formula could have been used.

### Acknowledgements

The author is thankful to Anthony Alvarez and Andy Levi for their assistance with research and to Mike Strube for his time spent discussing this article. The author would also like to thank the anonymous RFL reviewers for their constructive criticism and valuable feedback.

### References

- Aiken, L. & West, S. (1991). *Multiple regression: Testing and interpretations*. London: Sage Publications.
- Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *Modern Language Journal*, 75, 460-472.
- Barnett, M. A. (1989). *More than meets the eye: Foreign language reading theory and practice*. Englewood Cliffs, NJ: Prentice Hall.
- Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix*, 1(1), 1-9.
- Bernhardt, E. B. (1984). Towards an information processing perspective in foreign language reading. *Modern Language Journal*, 68, 322-31.
- Bernhardt, E. B. (1991). *Reading development in a second language*. Norwood, NJ: Ablex Publishing Corporation.

- Bernhardt, E. B. (2000). Second-language reading as a case study of reading scholarship in the 20<sup>th</sup> century. In M. Kamil, P. Mosenthal, D. Pearson, & R. Barr (Eds.). *Handbook of reading research Volume III* (pp. 793-811). Hillsdale, NJ: Earlbaum.
- Bernhardt, E. B. (2003). New directions in reading research: Second language perspectives. *Reading Research Quarterly*, 37(4), 112-117.
- Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20, 463-494.
- Brantmeier, C. (2002). The effects of passage content on second language reading comprehension by gender across instruction levels. In J. Hammadou Sullivan (Ed.). *Research in second language learning: Literacy and the second language learner* (pp. 149-176). Greenwich, CT: Information Age Publishing.
- Brantmeier, C. (2003). Does gender make a difference? Passage content and comprehension in second language reading. *Reading in a Foreign Language*, 15(1), 1-27.  
<http://nflrc.hawaii.edu/rfl/April2003/brantmeier/brantmeier.html>
- Brantmeier, C. (2004). Building a comprehensive theory of foreign language reading: A variety of variables and research methods. In C. Brantmeier, (Special Guest Editor). *Adult Foreign Language Reading: Theory, Research, and Implications. The Southern Journal of Linguistics*, 27(1).
- Brown, R. I. (1987). A comparison of the comprehensibility of modified and unmodified reading materials for ESL. *The University of Hawai'i Working Papers in ESL*, 6, 49-79.
- Bügel, K. & Buunk, B. P. (1996). Sex differences in foreign language text comprehension: The role of interests and prior knowledge. *Modern Language Journal*, 80, 15-31.
- Camiciottoli, B. C. (2003). Metadiscourse and ESP reading comprehension: An exploratory study. *Reading in a Foreign Language*, 15(1), 1-16.  
<http://nflrc.hawaii.edu/rfl/April2003/camiciottoli/camiciottoli.html>
- Carrell, P. L. (1984). Evidence of a formal schemata in second language comprehension. *Language Learning*, 34, 87-112.
- Carrell, P. L. (1989). Metacognitive awareness and second language reading. *Modern Language Journal*, 73, 121-133.
- Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency. *Applied Linguistics*, 12, 159-79.
- Carrell, P. L. (2001). Influence of purpose for reading on second language reading: Reading procedural texts in ESL. *Reading in a Foreign Language*, 13(2), 567-591.
- Droop, M. & Verhoeven, L. (1998) Background knowledge, linguistic complexity and second language reading comprehension. *Journal of Literacy Research*, 30(2), 253-271.



- Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first and second language learners. *Reading Research Quarterly*, 38(1), 78-103.
- Coady, J. (1979). A psycholinguistic model of the ESL reader. In R. Mackay, B. Barkman, & R. Jordan (Eds.), *Reading in a second language* (pp. 5-12). Rowley, MA: Newbury House Publishers.
- Fraenkel, J. & Wallen, N. (1996). *How to design and evaluate research in education*. New York: McGraw Hill.
- Hammadou, J. (1991). Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *Modern Language Journal*, 75, 27-38.
- Hatch, E. & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Boston, MA: Heinle & Heinle.
- Hsueh-chao, M. H. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- Hudson, T. (1982). The effects of induced schemata on the short-circuit in L2 reading: Non-decoding factors in L2 reading performance. *Language Learning*, 32, 1-32.
- Johnson, P. (1981). Effects on reading comprehension of language complexity and cultural background of a text. *TESOL Quarterly*, 15, 169-181.
- King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science*, 12, 666-687.
- Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2<sup>nd</sup> ed.). Belmont, CA: Brooks/Cole.
- Lee, J. F. (1990). A review of empirical comparisons of nonnative reading behaviors across stages of language development. In H. Burmeister & P. Rounds (Eds.), *Variability in second language acquisition* (pp. 453-72). Eugene, OR: University of Oregon Press.
- Lee, J. F. & Ballman, T. L. (1987) Learners' ability to recall and rate important ideas of an expository text. In B. Van Patten, T. R. Dvorkack, & J. F. Lee (Eds.), *Foreign language learning: A research perspective*. Rowley, MA: Newbury House, 108-17.
- Lee, J. F. & VanPatten, B. (1995). *Making communicative language teaching happen*. New York: McGraw-Hill.
- Leow, R. (1993). To simplify or not to simplify. *Studies in Second Language Acquisition*, 15, 333-355.

- Leow, R. (1997). The effects of input enhancement and text length on adult L2 readers' comprehension and intake in second language acquisition. *Applied Language Learning*, 8, 151-182.
- Leung, C. Y. (2002). Extensive reading and language learning: A diary study of a beginning learner of Japanese. *Reading in a Foreign Language*, 14(1), 1-16.  
<http://nflrc.hawaii.edu/rfl/April2002/leung/leung.html>
- Liontis, J. (2001). That's all Greek to me! The comprehension and interpretation of modern Greek phrasal idioms. *The Reading Matrix*, 1(1), 1-31.
- Mohammed, M. A. & Swales, J. M. (1984). Factors affecting the successful reading of technical instructions. *Reading in a Foreign Language*, 2, 206-217.
- Mori, Y. & Nagy, W. (1999). Integration of information from context and word elements in interpreting novel kanji compounds. *Reading Research Quarterly*, 34(1), 80-101.
- Pritchard, R. (1990). The effects on cultural schemata on reading processing strategies. *Reading Research Quarterly*, 25, 273-295.
- Salataci, R. & Akyel, A. (2002). Possible effects of strategy instruction on L1 and L2 reading. *Reading in a Foreign Language*, 14(1), 1-16.  
<http://nflrc.hawaii.edu/rfl/April2002/salataci/salataci.html>
- Sarig, G. (1987). High-level reading in the first and in the foreign language: Some comparative process data. In J. Devine, P. L. Carrell, & D. E. Eskey (Eds.), *Research in reading in English as a second language* (pp. 105-120). Washington: TESOL.
- Schueller, J. (1999). *The effect of two types of strategy training on foreign language reading comprehension: An analysis by gender and proficiency*. Unpublished doctoral dissertation. The University of Wisconsin, Madison.
- Sharp, A. (2002). Chinese L1 schoolchildren reading in English: The effects of rhetorical patterns. *Reading in a Foreign Language*, 14(2), 1-22.  
<http://nflrc.hawaii.edu/rfl/October2002/sharp/sharp.html>
- Shohamy, E. (1982). Affective considerations in language testing. *Modern Language Journal*, 66, 13-17.
- Shohamy, E. (1984). Does the testing method make the difference? The case of reading comprehension. *Language Testing*, 1, 147-70.
- Stakhnevich, J. (2002). Reading on the Web: Implications for ESL professionals. *The Reading Matrix*, 2(2), 1-8.
- Steffensen, M. S., Joag-dev, C., & Anderson, R. C. (1979). A cross-cultural perspective on reading comprehension. *Reading Research Quarterly*, 15, 10-29.

- Steffensen, M. S., Goetz, E. T., & Cheng, X. (1999). A cross-linguistic perspective on imagery and affect in reading: Dual coding in Chinese and English. *Journal of Literacy Research*, 31(3), 293-319.
- Taguchi, G. & Gorsuch, G. (2002). Transfer effects of repeated EFL reading on reading of new passages: A preliminary investigation. *Reading in a Foreign Language*, 14(1), 1-21. <http://nflrc.hawaii.edu/rfl/April2002/taguchi/taguchi.html>
- Tsang, W-K. (1987). *Text modifications in ESL reading comprehension*. Unpublished doctoral dissertation. The University of Hawai'i, Manoa.
- Tweissi, A. (1998). The effect of the amount and type of simplification on foreign language reading comprehension. *Reading in a Foreign Language*, 11(2), 191-207.
- Urquhart, S. & Weir, C. (1998). *Reading in a second language: Process, product and practice*. London: Longman.
- Van den Branden, K. (2000). Does negotiation of meaning promote reading comprehension? A study of multilingual primary school class. *Reading Research Quarterly*, 35(3), 426-443.
- Wilkinson, I. (1998). Dealing with diversity: Achievement gaps in reading literacy among New Zealand students. *Reading Research Quarterly*, 33(2), 144-167.
- Wolf, D. (1993). A comparison of assessment tasks used to measure FL reading comprehension. *Modern Language Journal*, 77, 473-89.
- Young, D. J. & Oxford, R. (1997). A gender-related analysis of strategies used to process input in the native language and a foreign language. *Applied Language Learning*, 8, 43-73.
- Young, D. (2000). An investigation into the relationship between L2 reading anxiety and L2 reading comprehension, and self-reported level of comprehension, topic familiarity, features of an L2 text and reading ability in the L1 and L2. R. Leow & C. Sanz (Eds.), *Current research on the acquisition of Spanish*. Somerville: Cascadilla Press, 15-33.

### **About the Author**

Cindy Brantmeier is an Assistant Professor of Applied Linguistics and Spanish at Washington University in St. Louis. She is Program Director of Advanced Spanish and Co-Director of the Graduate Certificate in Language Instruction. She also oversees language placement and departmental assessment. Her research interests include interacting variables in second language reading, reading research methodology, language learning and technology, and language and gender.