A Logistic Regression on the Qualities of Educational Research: Characteristics of Publishable Articles

Leping Liu, Towson University, Baltimore, MD
George C. J. Fernandez, University of Nevada, Reno, NV

ABSTRACT

In educational research, especially in the field of using computer technology in education, there has been a gap between researchers and readers that was caused by poor research design or writing. This study will present some major problems that influenced the quality of research and the characteristics of a publishable research article. A logistic regression was performed to analyze six research quality variables – theme and focus, research design, instrument, findings and conclusions, references, and writing – to determine which variable(s) would influence the probability for an article to be published. The data source was an editorial management database for a referred international journal in the field of educational technology. The findings of this study suggested some useful information for researchers and researchers to improve the quality of research.

INTRODUCTION

The term *research* is defined as " the systematic study or investigation in some filed of knowledge to discover facts, or establish principles" (Guralnik, 1972, p. 1208). In educational research, especially in the field of using computer technologies in education, these studies and investigations provide information, knowledge, or experience for educators to improve the quality of education (Johnson, 1997; Suter, 1998). Therefore, the value of educational research depends on not only what the researchers have discovered, but also to what extent they share their findings and knowledge with other educators. There are many ways to share findings, but the primary method is through writing reports, articles, or books (Kirt & Berger, 1993).

Sharing knowledge is a two-way communication process between authors and readers (Anderson, 1980; Becker, 1991; Beebe, 1993; & Blyler, 1995). That is, only when authors disseminate knowledge effectively, and readers understand and use what they learn effectively, can the communication be established. However, there has frequently been a communications gap between researchers and readers caused by poor research design or writing (Kirt & Berger, 1993; Smeltzer & Thomas, 1994; Sanger, 1995; Rundall, 1996; & Charney, 1998). This gap prevents researchers from communicating effectively with readers and practitioners. Furthermore, it also impedes communication among researchers.

It is likely that a similar communication gap exists between the producers and consumers of educational research. However, the factors that influence the dissemination of educational technology research have not been well documented. The primary purpose of this study is to identify research design and writing factors that limit the publication of articles within an educational technology journal, and to provide guidance to authors who wish to communicate through educational journals.

INDENTIFICATION OF FACTORS

The identification of limiting factors was facilitated from the data in an editorial management database (EMDB) for a refereed international journal in the field of educational computer technology. This database was created to manage the information from the authors, the journal reviewers, and the editorial staff. The EMDB contains the journal reviewers' comments for all the submissions that are sorted into six categories as six fields:

1. theme and focus
2. research design
3. measurements or instruments
4. research findings
5. references
6. writing

These categories follow the model suggested by traditional research textbooks (Gay, 1992). However, since comments from reviewers were occasionally difficult to categorize, the most commonly identified problems within each category were summarized. Each field was searched to establish the frequency of these common problems. If over 30% of the reviewers identified an issue as a problem, that issue was defined as a "frequently addressed" problem. The some major research quality problems summarized from the EMDB are discussed as the following.

Theme and Focus. An appropriate them should be based on the literature review and focus on an area that could produce meaningful outcomes. Some major problems in determining the theme and focus are: First, the themes had been studied repeatedly over many years, with similar findings, leading to similar suggestions. These themes may have been interesting in 1979, but not in 1999. If the article could not provide new theoretical or practical insights, it would not be publishable. Second, the authors' focus was narrow and shallow, and contributed no new insights to the literature. Third, the article titles were inconsistent with the content. The titles were ambiguous, disjointed with content, or using quantitative terms for a qualitative study.

Research Design. The first research design problem identified in the EMDB was sampling problem, such as lack of description of sampling processing, or seriously

skewed female-male ratio. The second problem was weak design. The research design was ambiguous because the author wrote a qualitative/quantitative design as a quantitative/qualitative piece; or used inappropriate statistical analysis method. The third problem was that some articles were cast in research terms and format, but they were not research. Authors wrote an opinion piece or discussed an internal project as though it was research.

Measurements and Instruments. Instrumentation was an important issue that influenced the validity and reliability of research findings. According to the information stored in the EMDB, the major instrumentation problem was that studies failed to provide evidence to substantiate the validity of their instruments. Some studies used two different instruments to measure the same variable for two groups without reporting the consistency of the instruments. Some measurements were based on surveys with very low return-rates (e.g., lower than 40%). Or, some instruments were inappropriate for the sample (e.g., the test was too difficult for subjects of that age group).

Findings and Conclusions. First, studies were often inconclusive. Some had important findings but failed to draw a conclusion that explained how these findings could make contributions, theoretically or practically, to the literature. Second, the conclusions were weak. The findings were either minimal, or unrelated to the introduction, or too ambiguous to support a strong argument.

References. First, the references were not up-to-date. In the field of educational technology, where change is a daily occurrence, current references are essential. Second, references were either insufficient or superfluous. Third, some literature reviews failed to create an integrated picture of current theory and knowledge, but rather consisted of a series of single sentence summaries of articles.

Writing. Writing deficiencies were most frequently addressed problem. Journal reviewers offered some suggestions that might be helpful to prospective authors: (a) write to the reader, (b) use terms consistently, (c) use verb tense consistently, and (d) avoid writing the article as if "someone was forced to write a course assignment."

RESEARCH QUESTIONS AND HYPOTHESES

We have identified the six major research quality variables. Next, according to the purpose of this study, we will determine whether or how these variables limit the publication of articles. The research questions examined in this study were:

1. Is there any relationship between each of the six research quality variables (theme and focus, research design, measurements or instruments, research findings, references, and writing) and the probability of an article being published?
2. Can one or more of the six research quality variables predict the probability of an article being published?

It was hypothesized that at least one of the six variables would have a relationship with and could be used to predict the probability of an article being published.

DATA ANALYSIS METHOD

To estimate the probability that an article will be published, a logistic regression was performed. Logistic regression is a form of statistical modeling that is often appropriate for categorical outcome variables. It describes the relationship between a categorical response variable and a set of explanatory variables (Stokes, Davis, & Koch, 1996; & Neter, Kutner, Nachtsheim, & Wasserman, 1996). The response variable is usually dichotomous; typically, the two outcomes are *Yes* and *No*. The explanatory variables in logistic regression can be categorical or continuous. The purpose of a logistic regression analysis is to detect the relationship between the explanatory variables and the probability of the response variable to have the outcome *Yes* (or *No*).

In this study, the explanatory variables were the six research quality items. They were all categorical variables with the outcomes 1 and 0. One (1) indicated that the variable met the quality criteria, and 0 indicated that the variable did not meet the quality criteria. The response variable was article status (publication decision), with the outcome 1 as published, and 0 as rejected (as shown in Table 1).

Table 1. Variables in the Logistic Regression

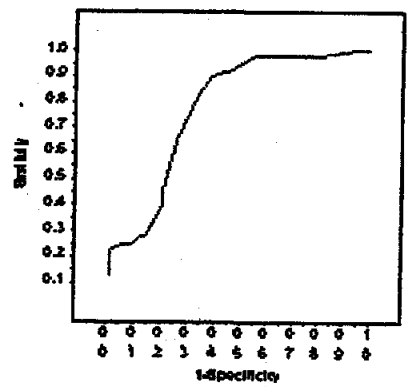
Variables E: explanatory variable R: response variable	1	0
Article Status (R)	Published	Rejected
Theme/Focus (E)	New and Interesting	Unupdated and Nonresearchable
Research Design (E)	Appropriate and Valid	Poor and Invalid
Instruments (E)	Valid and Reliable	Invalid and Irreliable
Findings/Conclusions (E)	New and Generative	Weak and Irrelevant
References (E)	Sufficient and updated	Insufficient and Unupdated
Writing (E)	Good	Not Good

LOGISTIC ANALYSIS AND RESULTS

One hundred and two articles submitted during the last three years were randomly selected from the EMDB. Each article was selected by matching its log number with the last two digits of a calculator-generated random number. Forty articles out of the 102 were actually published.

The analysis consisted of identifying a model that best fit the data and analyzing the data. The first logistic analysis produced a model that consisted of all six explanatory variables. In this initial model, only the Research Design variable was significantly ($P < 0.0117$) related to the response variable, Publication.

Figure 1. ROC Curve



In the second logistic analysis, Research Findings, Instrument, and Research Design variables were retained in the model. Research Findings and Instrument variables were included in the second model because they approached significance ($P < 0.1651$, and $P < 0.1104$, respectively) in the first model. It was hoped that including them in the second model would contribute significantly to the prediction of publication. However, this was not the case. In the second model Research Design was the only variable that contributed significantly ($P < 0.0006$) to the prediction of the response variable. A third logistic analysis was performed which included only the Research Design variable, and it was significant ($P < 0.0202$).

To determine the best model to use, we compared the AIC (Akaike Information Criterion) values, a goodness-of-fit measure. Since small AIC values indicate a more desirable model, the second model that contained the Research Findings, Instrument, and Research Design variables was selected (AIC value of 115.214 as compared to 117.082 and 135.061). Also, the receiver operating characteristic (ROC) curve plotted from this model, as in Figure 1, rises quickly and the area under the curve is considerably large, indicating that this logistic regression model has relatively high predictive accuracy. This model indicated that the combination of these three variables would provide an explanation of the relationship between the research quality variables and the probability of an article being published.

As shown in Table 2, the Research Design variable was a significant ($P < 0.0006$) predictor of article publication. Another useful value in this table that helped explain the probability of the response variable was the Odds Ratio. If the Odds Ratio was larger than 1, the probability of an article being published would increase and if smaller than 1, the probability would decrease. In this model, the Odds Ratio for the Research Design variable was 15.123. This meant that an article that met the Research Design quality criterion would be 15.123 times more likely to be published than an article that failed to meet the criterion.

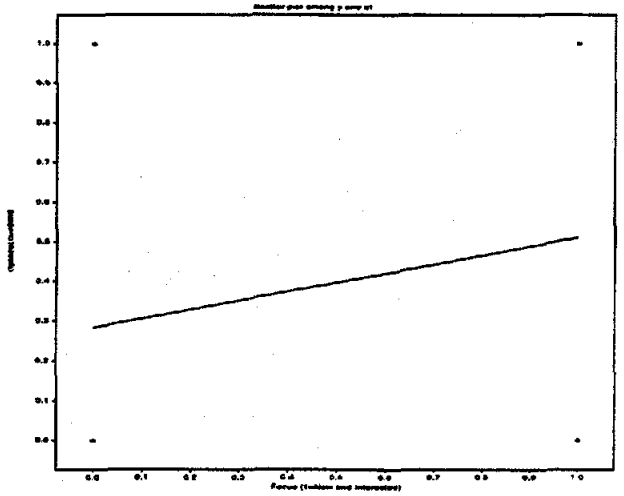
Table 2. Logistic Analysis Outputs

Variable	DF	Parameter Estimate	Wald Chi-square	P > Chi-square	Odds Ratio
Intercept	1	-2.7753	14.3821	0.0001	
Findings	1	0.2575	0.4722	0.4920	1.294
Instruments	1	1.4184	1.5658	0.2108	4.130
Design	1	2.7162	11.9086	0.0006	15.123

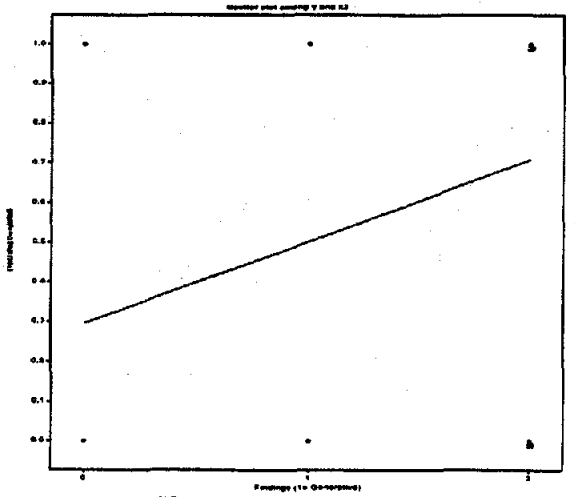
The analysis summarized in Table 2 indicated that the Research Design variable was the most important predictor of an article being published. Even though the Research Findings and Instrument variables did not make a significant contribution to the model, the Odds Ratios for these variables suggested that the probability of publication increased when the article met those quality criteria as well.

The scatter plots among the response variable (the probability of an article being published) and each of the six explanatory variables are shown in Figure 2. From these plots, we can see the trends of how the three variables selected in the model (Research Findings, Instrument, and Research Design) related to the response variable, Publication.

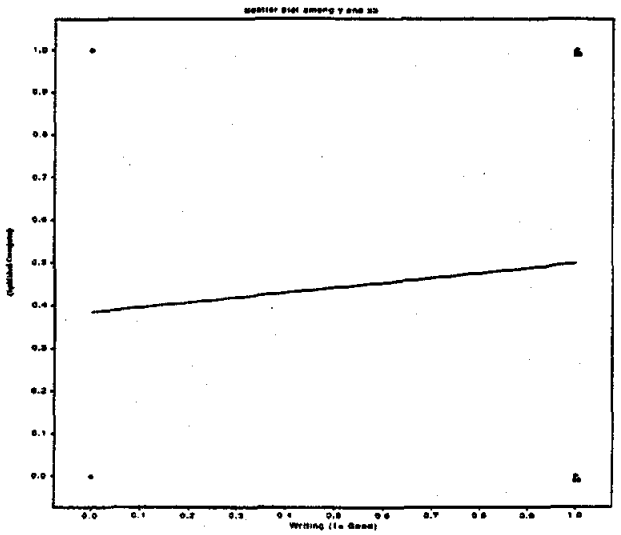
Figure 2. Scatter Plot Among Response Variable and Explanatory Variables



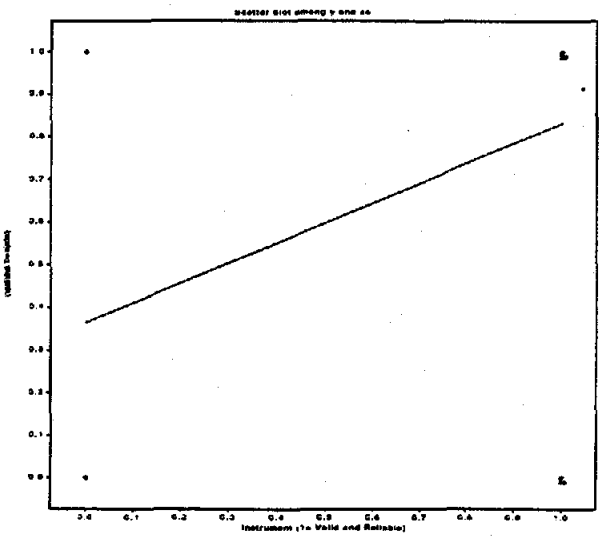
Y and X1—Theme and Focus



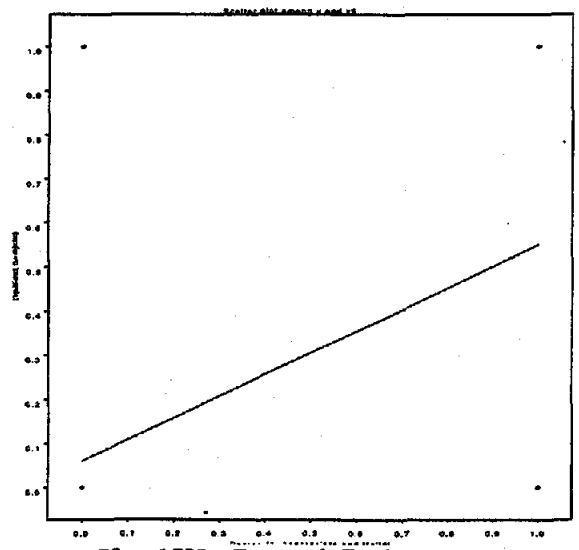
Y and X2 – Research Findings



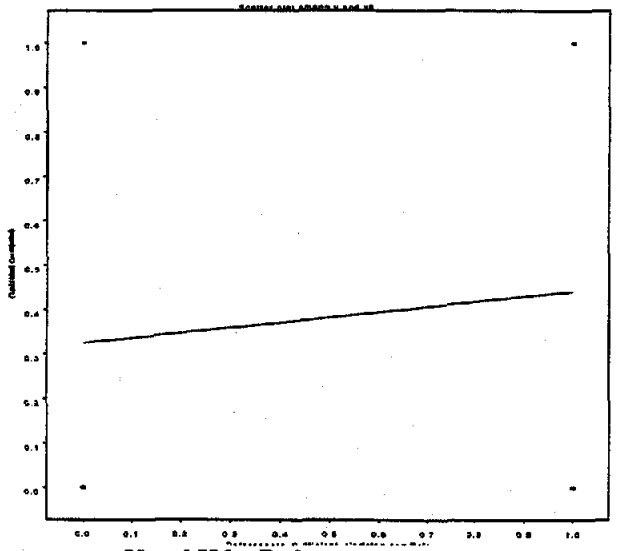
Y and X3 – Writing



Y and X4 – Instrument



Y and X5 – Research Design



Y and X6 – References

SUMMARY

In this paper, we investigated the major problems with research and research articles for an academic journal. Our findings suggest that research quality problems fall into six categories: (a) theme and focus, (b) research design, (c) measurements or instruments, (d) research findings, (e) references, and (f) writing. Among these categories, Research Design was the most influential in determining whether or not an article was published.

Conducting research is always a learning process and represents a good opportunity for evaluation and improvement. The study offers guidance to prospective researchers since those who devote more time to research design, instrumentation, and writing conclusions will increase their chance communicating with their audience.

RFERENCES

Anderson, P. (1980). The need for better research in technical communication. *Journal of Technical Writing and Communication, 10,* 271-282.

Becker, B. J. (1991). The quality and credibility of research reviews: What the editors say. *Personality and Social Psychology Bulletin, 17,* 267-272.

Beebe, L. (Ed.). (1993). Professional writing for the human services. Washington, DC: NASW Press (National Association of Social Workers).

Blyler, N. (1995). Research as ideology in professional communication. *Technical Communication Quarterly, 4,* 285-313.

Charney, D. (1998). From logocentrism to ethnocentrism: Historicizing of writing research. *Technical Communication Quarterly, 7,* 9-33.

Gay, L. R. (1992). *Educational research: Competencies for analysis and application.* New York: Macmillan.

Guralnik, D. B. (1972). Webster's new world dictionary. New York: The World Publishing Company.

Johnson, D. L. (1997). Integrating technology in the classroom: The time has come. *Computers in the Schools, 13*(1-2), 1-5.

Kirt, S. A., & Berger, R. M. (1993). Improving research writing. *Social Work Research and Abstracts, 29*(4). 3-5.

Neter, J., Kurner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear regression models. Chicago, IL: Donnelley & Sons.

Rundall, T. (1996). Conducting and writing research reviews. *Medical Care Research and Review, 53*(1), 132-146.

Sanger, J. (1995). Five easy pieces: The deconstruction of illumination data in research writing. *British Educational Research Journal, 21*(1), 89-98.

Smeltzer, L. R., & Thomas, G. F. (1994). Managers as writers: A meta-analysis of research in context. *Journal of Business and Technical Communication, 8*(2), 186-212.

Stokes, M. E., Davis, C. S., & Koch, G. G. (1996). Categorical data analysis using the SAS system. Cary, NC: SAS Institute.

Suter, W. N. (1998). Primer of educational research. Boston: Allyn and Bacon.

ABOUT THE AUTHORS

Dr. **LEPING LIU** is Assistant Professor, Towson University, Maryland. She has been a SAS user for four years. Her SAS programming areas of expertise include statistical data analysis for educational and psychological research data, application design and development, data presentation, and data management. She currently uses SAS for Windows for statistical analysis and graphics.

Dr. **GEORGE C. J. FERNANDEZ** has ten years of experience teaching introductory statistics, advanced design and analysis, regression models, and SAS programming classes at the University of Nevada, Reno. He has 19 years of SAS experience in BASE, STAT, GRAPH, LAB, INSIGHT, and ETS. He currently uses SAS for Windows for statistical analysis and graphics.