INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University
Microfilms
International
300 N. Zeeb Road
Ann Arbor, MI 48106

Smith-Donaldson, Jacqueline Jill

AN EXPERIMENTAL STUDY TO ASSESS THE ABILITY OF EVIDENCE, ARGUMENT, AND DELIVERY TO DISCRIMINATE FOR WIN/LOSS IN A DEBATE

The University of Arizona PH.D. 1983

AN EXPERIMENTAL STUDY TO ASSESS THE ABILITY OF EVIDENCE, ARGUMENT, AND DELIVERY TO DISCRIMINATE FOR WIN/LOSS IN A DEBATE

by

Jacqueline Jill Smith-Donaldson

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF SPEECH COMMUNICATION

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 8 3

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Jacqueline Jill Smith-Donaldson entitled AN EXPERIMENTAL STUDY TO ASSESS THE ABILITY OF EVIDENCE, ARGUMENT, AND DELIVERY TO DISCRIMINATE FOR WIN/LOSS IN A DEBATE and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

	Date
H. L. Ewbank	6 April 1983
	Date
Ronald J. Matlon	April 6, 1983
	Date
	Date
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

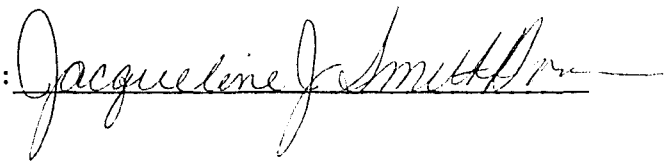
I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

James W. Davis	April 6, 1983
Dissertation Director	Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: 

ACKNOWLEDGMENTS

A dissertation is the culmination of a number of stages in one's graduate work. Consequently, there are several people who I would like to acknowledge. To start, I should thank Dr. Andy King, as the person who guided me through my prelims. I also wish to thank Dr. Dave Nott for his assistance during this same period.

One individual who worked with me from the time I arrived in Arizona, and someone I still turn to for guidance, is Mr. Tim Browning, who is high on my list of thanks. Mr. Browning has always impressed me with his strong support of graduate students and his integrity in forensics. He has supported me a number of times, and for this I will always be grateful. While Mr. Browning was not officially a member of my committee, without his input this dissertation would have been lacking. His assistance was essential from the beginning to the end of this project. Indeed, the conception of this study began through conversations with Mr. Browning.

Dr. Ron Matlon was invaluable for his assistance in reviewing the literature for this study. I will always appreciate his desire for viable research in the area of forensics. I would also like to thank Dr. Henry Ewbank for his support and assistance in writing this dissertation. Additionally, it is important to thank Dr. Barbara Ewbank for her continued support.

Last, I wish to express my gratitude to Dr. James Davis, chair of my committee. Through working with Dr. Davis, I came to appreciate his concept of an elegant experimental design. Dr. Davis is truly committed to good experimental research. I especially appreciate Dr. Davis' supportive attitude throughout this project.

TABLE OF CONTENTS

TABLE OF CONTENTS -- Continued

LIST OF TABLES

ABSTRACT

The purpose of this study was to identify variables which discriminate between winning and losing a debate as measured by judges' responses on semantic differential scales. The dependent variable was membership in either the group "wins" or "losses." The independent variables were measured by semantic differential scales related to Delivery, Argument, and Evidence. The analytical procedure used was discriminant function analysis. Such an analysis discriminates maximally between the win and loss groups.

Four scale items emerged as discriminating for wins and losses in a debate. The most discriminating variable came from the Argument dimension, specifically the scale item Convincing-Unconvincing. The second most discriminating variable was from the Evidence dimension, that is Strong-Weak. The third discriminating variable was from the Delivery dimension, namely Pleasant-Unpleasant. The last significant variable was also from the Evidence dimension, specifically Valuable-Worthless. The final Lambda of .5314 and the canonical correlation of .6845 indicate that the discriminant function produced a fairly high degree of separation between the win and loss groups.

CHAPTER 1

ORIENTATION

Introduction

It is the purpose of this chapter to provide a justification for conducting research in the area of forensics, followed by a rationale for the specific experimental study presented in this dissertation. Further, this chapter will offer a brief overview of the major components of the research undertaken.

Justification for Research

While there exists significant theoretical and quantitative research in argument examining such factors as speaker ethos, speaking position, and primacy and recency of argument, when one seeks to examine argumentation in the area of academic debate, there is not a great deal of experimental research. Walwik (1969, p. 43) characterized forensic research as "approach-avoidance," in that there is much talk about "opportunities" for research, "but, in fact, research in our field has been limited and often of dubious quality." The state of forensic research in 1974 prompted McGlone (p. 144) to comment that his colleagues were not expressing an interest in forensic experimental research and that the research which had been completed was not particularly good.

There would also seem to be a lack of practical research in the field. Anderson characterized this view (1966, p. 113) when he suggested

that current forensic research fails "to test presuppositions," "or to test theory." He goes on to state that experimental research does not address "practical theoretical problems" in the field. The issue seems to be that textbook instructions to debaters as to effective forensic practices have never been adequately subjected to critical quantitative review. Anderson further emphasized the problem with current research when he argued that present empirical research fails to control experimental variables adequately.

Given or based on the apparent need for additional research in forensics, the present study was conducted. The purpose of this study was to determine variables of debater behavior which would discriminate between winning and losing a debate. The goals were to provide practical research such as had been previously sought. That is, coaches would be able to utilize the results of this study when instructing debaters.

Rationale

As stated above, this study attempts to shed practical light on variables which discriminate between wins and losses as measured by semantic differential scales responded to by debate judges. While some previous research has attempted to assess debate variables, generally yielding non-significant results, the current study is measurably different from past studies. First, this study employs discriminant function analysis, a procedure not previously used in forensic research. Earlier studies involving debate evaluation by Burgoon (1975) and Vasilius (1977) used multiple regressions in their statistical analysis.

Evidence would indicate (Huck, Cormier and Bounds 1974, pp. 160-61) that when a dichotomous dependent variable, as used in these earlier studies, is selected, then discriminant function analysis is a proper analytic procedure. Discriminant analysis will determine whether the independent variables will classify if a team won or lost a debate. Given this procedure, one can determine the behaviors, within the pool provided, which will discriminate for winning or losing a debate.

Another important method in which the current study differs from earlier research is in terms of the measuring instrument used. Several researchers relied on the American Forensic Association Form C ballot for information gathering. The problem is simply that evidence exists to suggest that judges do not distinguish between items when marking these scales (Wise 1971). In other words, the problem in some research was that the judges would mark all of the scales high or all of the scales low. It has been argued (Wise 1971, Burgoon 1975) that judges do not make independent assessments of debater behavior in a debate. The tendency thus would be to give winning teams high points and losing teams low points. The present study seeks to avoid this problem by using semantic differential scales and not a standardized American Forensic Association ballot measure. Research conducted by Williams, Clark and Wood (1966) employing a factor analysis of debater behavior offers two conclusions concerning the non-independency issue. First, they suggest that certain behaviors' association with argumentation (as measured by the good-bad scale) are not perceived as independent by judges, thus supporting the subsequent Wise and Burgoon research.

Second, they nevertheless conclude that debaters who do well on one aspect of argument may naturally perform well on other associated behaviors of argument (1966, p. 98). To the extent that non-independency of judgments is related to the measuring instrument used, this study did not use a standard ballot measuring instrument in an attempt to avoid this problem. Rather, this researcher employed the semantic differential as a test instrument for data gathering purposes. Likewise, the scaling was significantly different from that used in the Williams et al. (1966) research. Williams et al. sought to uncover general dimensions of debate and only relied on a single scale measure. On the other hand, this study employs a number of semantic scales to measure a concept, thus seeking a more complete representation of the semantic space involved; hence, more fully defining the concept (Osgood Suci and Tannenbaum, 1961, p. 25).

The semantic differential test instrument used in this study does not presume that judges will fail to make independent assessments of behaviors, as they are expected to do when using the Form C ballot. Indeed, there is no existing research which would indicate that judges would make non-independent evaluations when responding to semantic differential scales. Further, the semantic differential has been shown to be a valid instrument generally (Osgood et al., 1961, p. 195) and specifically in a number of forensic related studies (Stewart and Merchant 1969, Burgoon and Montgomery 1976, King and Phifer 1968), while there is at least some doubt about standardized ballot measures. Additional justification for the test instrument will be made in Chapter 3.

A third manner in which this study differs from earlier research was suggested previously. Specifically, the factors of argument, evidence, and delivery are rated through a number of scale items, thus providing a relatively complete semantic definition of each concept. Clearly, neither the Form C Ballot nor the Williams et al. research does this. The Form C Ballot only requests that a judge respond to the concepts by marking one of the five boxes indicating increasing levels of importance and the Williams et al. study only employed a single scale measure.

Conceptual Considerations

Previous research has established certain variables which discriminate between forensic wins and losses. For that reason in this study the following variables were selected for investigation:

1. Evidence
2. Argument
3. Delivery

The present study starts with the assumption that these dimensions are present in debate. This assumption is based on prior research which generally established that they do, indeed, function in forensic events. In the current study each concept is rated by the judges through a series of related scales.

As will be discussed in greater detail in Chapter 3, specific research hypotheses were not formulated for this study. Given the uniqueness of this research, there was very little on which to base any

assumptions concerning the potential outcome of this study in terms of the specific scale items used. While this study can be seen as a first step in information gathering, two expectations were operating. First, that given an appropriate measuring instrument, judges will make independent assessments of behavior. Second, a set of behaviors exists which characterizes winning debaters and is unique from a set which characterizes losing debaters, as a result of judge responses on semantic differential scales.

Preview of Remaining Chapters

Chapter 2 will be concerned with a review of the pertinent literature in the field. Chapter 3 outlines the methodology used in the present research study. Results of the data analysis are presented in Chapter 4. Chapter 5 discusses these results with Chapter 6 serving as a summary including implications for future research and limitations of the study.

Summary

The emphasis in this chapter has been to establish a need for experimental research in forensics, and further, to indicate how the present study attempts to meet this need.

CHAPTER 2

REVIEW OF LITERATURE

Introduction

The research selected for review in this chapter is that which has a direct bearing on the current study. This chapter principally examines forensic resesarch in the areas of delivery, evidence, and argumentation, as well as more general research investigating criteria used by judges in rendering a decision in a debate. While some literature exists with regard to these variables outside the area of forensics, it was felt that such research would not be specifically applicable to competitive debate. The research reported in this chapter is all empirical in nature, though not necessarily experimental. Specifically, several articles are examined which are based on survey data not treated to statistical procedures. It is hoped that this review will demonstrate first, that the variables selected for treatment in this study are viewed as important by other researchers, and two, that there exist few conclusive statements which can be made concerning how these variables function in a debate.

Sources Consulted

Given the nature of this study, the majority of the reported research comes from The Journal of the American Forensic Association. Since this study covered several debate variables and a statistical

procedure which might not be indexed specifically, each issue of JAFA was consulted individually. Also, issues of The Gavel, The Speaker, and The Speaker and Gavel were consulted for relevant articles. While several journals were reviewed on an individual basis, a number of indices were also used in compiling the final bibliography. These were, the Educational Resources Information Center index, Master's Abstracts, Dissertation Abstracts, the yearly bibliography published in JAFA, The Education Index, the Cumulative Index to Journals in Education, and the Index to Journals in Communication Studies Through 1979. As a final check, footnotes and bibliographies of articles cited in this dissertation were consulted.

General Research on Judging Criteria

A number of earlier studies examined a series of variables used in judging a debate as opposed to centering on a single variable. These studies are discussed below.

Giffin's (1959, pp. 69-71) study which sought to determine if "judgments in tournament debating are related to academic values" provides support for the criteria selected in this study. Though Giffin does not specify whether his figures are statistically significant, nevertheless his data provide insight. The following table (Table 1) represents Giffin's data and that for the McCroskey and Camp (1966) study reported next.

Giffin collected his data from judges at the University of Kansas Heart of America Tournament in 1957. He asked these judges to weight

Table 1. Giffin 1959 and McCroskey and Camp 1966 results.

	Giffin	Camp and McCroskey
1. Selection of logically defensible arguments (case)	19.10%	1
2. Support of arguments with information (evidence)	17.18%	3
3. Perception of irrelevant or irrational arguments (refutation)	17.00%	5
4. Ability to analyze the topic-area (analysis)	14.65%	2
5. Ability to speak well (delivery)	14.65%	7
6. Ability to organize ideas into a structured whole (organization)	8.88%	4
7. Phrasing of concepts clearly and concisely (language)	5.29%	6
8. Other criteria	3.12%	

seven judging criteria. Out of 100% possible, assigned to seven criteria, the ordering emerged as shown in Table 1. Giffin's results suggest the utility of the variables selected for use in the present study. Criteria one, three and four represent the argument variable, whereas the second criterion illustrates that evidence functions in a debate, and five reflects the delivery aspect of the current study. While Giffin supplied his subjects with the criteria they were to weight, he attempted to check these results by having the judges indicate the degree to which they believed they normally used the criteria in rendering a decision. His results indicated that there was a significant relationship between the weights the judges used in the experiment and what they believe they normally did in evaluating a debate.

McCroskey and Camp (1966, pp. 60, 62) conducted a study similar to that done by Giffin in 1959 although the intent was somewhat different. The study was conducted at the Southern Speech Association Tournament in 1965 and included both high school and college debating. The authors had conducted a similar study in 1964, and in 1966 they were seeking to determine the validity of some of the conclusions reached in 1964. College judges' responses in the 1966 study can be found in the previous table. As can be seen, these results tend to support the Giffin study. The authors speculated that judges ranked delivery so low in their 1966 study because delivery was perceived as being relatively unimportant. Yet, they suggested that delivery may be seen by the judge as inseparable from the other elements of debate. So, to rate delivery as a separate variable would actually be rating delivery twice.

In another relatively early study, similar to the previous two, Roever and Giffin (1960, pp. 12-4) attempted to assess the degree to which certain criteria are used by debate judges when rendering decisions. Roever and Giffin gave thirty professional judges one hundred points and asked them to divide the points among the assigned criteria to denote the degree to which each was used when they cast a decision in a debate. The following nine items appeared in order of their perceived importance: Case, Analysis, Refutation, Evidence, Delivery, Organization, Language, Other, and Human Interest.

Unlike the previous three reported studies which did not use standard measuring instruments, Wise (1971) and Burgoon (1975) utilized standard ballot measures to assess debate effectiveness. Wise (1971) analyzed individual speaker means and variables on the standard Form C ballot and his data provide some interesting information. The Form C ballot asks judges to rate debaters in terms of analysis, reasoning, evidence, organization, refutation, and delivery. Wise argued (1971, p. 308) that it was possible that the scales on the Form C ballot did not actually reflect the act of judging a debate. He reported that he "has heard judges assert that they research a decision in a holistic fashion and then 'juggle' these scales . . . to agree with their decision." Wise further maintained that judges may be influenced by a "halo effect." That is, scoring of an individual scale may not reflect an independent assessment. Thus, the rating of one scale may be affected by ratings on other scales, thus indicating that judges may not make independent decisions. It should be noted, though, that since Wise

dealt only with individual speaker means and variances among these means, it would be difficult to draw conclusive results from his data.

Burgoon (1975, pp. 3-4) also conducted research using the Form C ballot. Her research indicated that the six criteria on the ballot were highly correlated with each other and were correlated with the dependent variables (win-loss). Consequently, she concluded that no single factor contributed to success in a debate. Further, she argued that judges do not make independent decisions on these criteria in a debate.

Another study involving tournament success was conducted by Vasilius and De Stephen (1979, pp. 197-204). Data for this study were collected for Vasilius' Master of Arts thesis (1977). She made tape recordings of debate rounds and proceeded to compute the speech rate, the amount of evidence read in a debate round, and the amount of jargon used by debaters. The authors attempted to determine whether increases in these three variables would contribute to winning a debate. They concluded that none of these factors affected either the winning or the losing of a debate.

Burgoon and Montgomery (1976, pp. 171, 175) conducted a study which bears some relevance to the current research project. They attempted to determine elements of source credibility which would be attributed to the ideal debater. They discovered three independent dimensions. The first dimension was labeled "Task Competencies." This dimension involved both message and delivery competencies. They suggested that this dimension involved the general credibility dimension of competence and debate dimensions of argument, vocal correctness, and

overall delivery. The second dimension was labeled "Social Competencies." They maintained that this dimension carried "implications of control: the qualities of poise, organization, perceptiveness, and good-naturedness" in social relationships. The third dimension was labeled "Assertiveness." This dimension is similar to the general credibility dimension of extroversion. Specifically, it reflects the manner in which a debater's "language and delivery style express an energetic and aggressive defense of his/her position." While this study dealt specifically with source credibility, it does indicate the utility of argument and delivery variables within a debate context.

The previous research would tend to indicate that to the extent any one of the variables studied function toward debate success, that variable would be argument. It is important to note two things; first, several of the studies reported no significance in their results; and second, of the studies which resulted in argument being an important variable, none really explained what constitutes argument. Thus, the following section attempts to examine those studies which specifically relate to the argument variable. Several of the following studies give a clearer meaning to the term "argument."

Argument

The majority of research on argument utilizes factor analytic design. One such study, conducted by Williams, Clark and Wood (1966, pp. 95, 100-2), sought to understand if there are "a relatively few 'basic' dimensions of evaluation that underlie judges' assessments of

contest debate." Specifically, the authors had judges respond to a series of semantic differential scales assuming that factors of judging would emerge. The study was completed on college and high school debaters. For the college debaters, the most dominant factor was labeled "argument" including scales such as "analysis," "reasoning," and "case." A second factor was labeled by the authors as "vocal-correctness" with such scales as "articulation," "clarity," and "rate." A third factor was labeled "apparent-character" including the scales of "courtesy," "sportsmanship," and "ethics." The last factor which emerged was termed "overall delivery," including such scales as "facial expression," "eye-contact," and "posture." The authors concluded that a judge renders a decision on a general assessment of argument rather than evaluating specific elements of arguments, such as the "analysis" and "reasoning" elements which were used in this study. Further, they argued that a judge, in rendering a decision, makes an evaluation primarily on two traits: argument and delivery. They also discovered that the better the debaters did on these factors, the more likely it was that they would win the debate. Likewise, debaters were more likely to lose a debate if they did less well on these variables.

In an earlier study involving high school debate, Williams and Webb (1964, pp. 126-27) asked experienced judges to respond to a series of evaluating scales taken from debate texts, journal articles, and debate ballots currently in use. Factor I, "Argument," was the overall best predictor for separate judgments. Argument accounted for 35% of the total variance; high loading scales, or constituents of argument,

were "supporting material," "concreteness," "logic," "relevance of evidence," "analysis," and "refutation." The second factor was "vocal-correctness." Relevant scales were "articulation," "pronunciation," "grammar," and "vocal quality." Factor III was "overall-delivery," including scales such as "eye contact," "rate," and "intelligibility." Several other factors not directly relevant to the present study were isolated.

Williams, Webb and Clark (1966, pp. 15-7) conducted a follow-up study to the one just reported. Their data were obtained from three high school tournaments, which included 138 debates involving 552 debaters. Following a series of factor analyses, "argument" emerged again as the dominant factor. The second highest loading factor was "apparent character." The third, "vocal-correctness," was not as defined as the first two factors. That is, the scales did not as clearly define the factors. "Delivery persuasiveness" was the fourth dimension. Unlike the earlier study, three additional scales emerged which gave delivery this persuasive quality.

An article by Cross and Matlon (1978, pp. 110-23) further attempts to look at types of arguments as they are evaluated by judges. In their study, Cross and Matlon attempted to accomplish three objectives. First, to describe current judging philosophies from information derived from National Debate Tournament judging philosophy booklets; second, to offer explanations to a national questionnaire on judging philosophies; and third, a conceptual analysis of the responses of the judges. Their analysis was specifically concerned with six judging

concepts as they are argued by debaters. Specifically, their concern was with presumption, conditional/hypothetical analysis, contradictions in arguments, probable truth, inherency-causality, comparison of systems, and interjection of theoretical perceptives and their application to debate practice. Their study examined these arguments in light of judging philosophies. Their results indicated that judges are fairly flexible in how they evaluate arguments, but that they do in fact evaluate arguments, thus lending additional support for the inclusion of argument in the current study.

In summary, it would seem that argument plays an important role in the success of a debate team. The majority of the research just reported would tend to confirm this. It should be remembered though that some of the studies reported in the first section of this chapter did not achieve significant results for the argument variable. Since it seems that there is some inconsistency in the literature concerning the importance of argument in a debate, it would seem reasonable to conclude that additional research is justified.

Delivery

This section cites articles addressed to the question of rate of delivery. The majority of the research deals directly with rapid delivery in college debate. Others deal more generally with speech rate, comprehension, and compressed speech. These materials are included at the beginning of this section to give the reader a feeling for what constitutes normal speaking rate. These studies should help to

clarify the debate-related research, by giving one something to which to compare it.

By altering pause time, Diehl, White and Burk (1959, pp. 229-32) attempted to establish whether there is a relationship between rate of speech and listener comprehension. Subjects in the study responded to forty-nine questions dealing with the lecture content and a rating of the delivery. The authors concluded that altering the rate of speech from 126 wpm to 172 wpm through pause time does not affect listener comprehension or "listener's ratings of the quality of a speaker's delivery." They stated that their conclusions support previous research that rate can vary from 125 wpm to 225 wpm without a major change in comprehension. They argued that, based on this, "the listening mechanism is highly adaptable." In all of the variations of pause time, the subjects' ratings of delivery fell between "'good'" and "'very good.'"

A study conducted by Lawrence E. Wheeless (1971, pp. 327-30) attempted to determine the relationship between comprehension and persuasion as a function of time-compressed speech. His results demonstrated that comprehension was significantly higher for a normal speech rate than for the compressed rate. At the same time, his results did not indicate any relationship between comprehension and persuasion. In this case, subjects were no more or no less inclined to purchase a specific product advocated, nor had their attitudes changed toward the purchase.

Several empirical studies have been conducted which seek to determine the importance of delivery skills in debate. At least two of

the previously reported studies would tend to negate its utility (Giffin 1959, McCroskey and Camp 1966).

Jones (1957, p. 93) attempted to assess the most effective word per minute and variety rate, which was defined as to favorably dispose an audience in a debate. The variety rate "was determined by dividing the average deviation of syllables of a four-second segment . . . by the average number of syllables per four-second segment." Audiences were asked to score a speaker on a 200-point scale and then assign a win. The results indicated that a slow rate (125 wpm) with a variety rate of 20 percent was the most effective.

A study conducted by Swinney (1968, pp. 16, 19) revealed a not-too-surprising conclusion. Specifically, speeches by tournament debaters are better comprehended by people who are "technically familiar" with debate theory though not the specific subject matter debated, rather than people who are "technically familiar" with the subject debated but not debate theory, or those individuals unfamiliar with either theory or the subject.

Kline (1953, pp. 59-60) conducted a study to determine the relative influence of argument, evidence, refutation, and analysis on delivery in debate. He concluded "that argument contributes most directly to the skill we describe as delivery." The influence of the remaining factors followed in the same order as presented above. Further, 77.6 percent of skill in delivery is accounted for by these variables. Based on this research, it would seem that delivery does not function alone in a judge's evaluation.

Giffin and Warner (1962, pp. 10-3) attempted to determine whether having an audience would affect the rate of speech of a debater. Specifically, they were interested in whether debaters would exceed 160 wpm when they do not have an audience. They concluded that speaking before an audience does not cause debaters to speak slower.

Specifically related to rate of speech would be the use of the "spread" in debate. The "spread" refers to a technique whereby a debater issues an abundance of arguments and evidence to any given argument by an opponent. Cox's (1974, p. 70) collection of judging philosophies at the National Debate Tournament included remarks on the spread. Concerning this, Cox concluded that all respondents showed a willingness "(though at times reluctantly)" to follow a debate utilizing such practices. However, a few judges stated that they would give low points to teams which used a spread. "So long as debaters met basic requirements for intelligibility, most participants tolerated this form of discourse, 'believing that the ultimate value of competitive debate to be analysis and not oratory.'" This would seem to be consistent with the work of McCroskey and Camp in 1966 and the 1959 research of Giffin, both dealing with factors of debate effectiveness. Cox (1974, p. 70) goes on to indicate that most of the judges indicated that they would not apply arguments which had been dropped by the debaters and "'when final extensions became parenthetical blurbs, . . . teams are playing a kind of roulette game and allowing the judge uncomfortable leeway in selecting the voting issues.'" Further, the majority of judges indicated that while they would try to "'follow' a rapid delivery or multiple extensions

of arguments" they did indicate that there might be some consequences. One critic commented, "'I can only react to speed mouths as I understand them . . . and feel that the burden of understanding is on the debaters, not me to make sense out of "too much in too short a time."'"

Also related to the use of the spread is the research of Olson (1971, pp. 66-69). He sent a questionnaire to sponsors of local chapters of Delta Sigma Rho--Tau Kappa Alpha concerning their opinion on spread techniques. His survey reveals that while a majority of people had negative feelings about the spread, there was no consensus concerning how to solve the problem.

Two additional studies specifically relating speech rate and participants in the National Debate Tournament are relevant. First, Rives (1976, pp. 47-50) assessed that in the final debate round at the 1976 National Debate Tournament the average speaking rate was 245 words per minute. Second, Colbert (1981, pp. 73-6) sought to determine whether the speech rate among National Debate Tournament finalists had increased from 1968 to 1980. He concluded that there was a significant increase in the rate of speech. Specifically, the difference ranged from 100 words per minute to the current rate of 270 words per minute.

There would seem to be two conclusions to be drawn from this research. First, empirically judges would seem not to employ rate of speech as a variable in determining success in a debate. Second, it can be concluded that when asked specifically about speech rate, judges express a concern with the delivery of debaters.

Evidence

Several of the studies reported earlier in this chapter contained data relevant to evidence as it functions in debate. These studies will be summarized below, followed by additional research.

The work of Giffin (1959) and McCroskey and Camp (1966) both suggested that evidence was a factor in judging. The research of Williams and Webb (1964) indicated that while evidence was important, it functioned in conjunction with argumentation. The studies by Vasilius and De Stephen (1979) and Burgoon (1975) tended to suggest that evidence is not a specific factor used in debate evaluation.

The previously reported study by Rives (1976) on the 1976 final round of the National Debate Tournament also made reference to the amount of evidence read in the debate. Specifically, he discovered that approximately two citations per minute were read in the debate.

Benson (1971, pp. 261-62) conducted an extensive study on the use of evidence by three levels of college debaters; novice, varsity, and championship. His research indicated that the amount of evidence read varied according to the amount of experience the debater had. Specifically, championship debaters used 25 percent more evidence than varsity debaters and almost 60 percent more than novice debaters. He also indicated that the increased rate of evidence was consistent for all four speaker positions.

Two conclusions can be made concerning the variable of evidence in a debate. First, given the number of times it has been used in research studies, evidence appears to be a variable with which experi-

menters are concerned. Second, given the diversity of the research results, it would be difficult to isolate its importance in debate success. Both of these conclusions provide support for continued research in this area.

Summary

The previously reported research covers a variety of empirical studies attempting to assess variables of debate success. As can be seen in the summaries of each section of this chapter, there is not a great deal of agreement concerning the relative importance of argument, evidence, and delivery. Generally, it would seem that argument, as variously defined, is an important variable in debate effectiveness. While evidence and delivery are certainly used in a debate, they were not as consistently reported to be important variables. It should be noted though, that in some studies evidence and delivery were viewed as one aspect of argumentation and not as separate variables.

The current research attempts to build on this literature in a number of ways. First, it attempts to define more fully the three variables of argument, evidence, and delivery through the use of semantic differential scales. While some studies attempted to define the variables used, this was not the case in all of the studies. Second, as outlined in Chapter 1, this study's use of semantic differential scales attempts to avoid previous research problems which might have resulted from the measuring instrument. Third, this study employs a statistical procedure not used previously in forensic research.

CHAPTER 3

METHODOLOGY

Introduction

This chapter describes the procedures used in the present experimental study. Further, the chapter provides the justification for these procedures as well as for the test instrument employed and the analytical procedures used.

Hypotheses

As indicated in Chapter 1, due to the uniqueness of the present study, specific research hypotheses were not formulated for testing purposes. While past research has identified some trends, that research has been different from the present study, or so plagued by methodological problems that no hypothetical guidance is provided. The intent of this study is to better understand, through quantification, variables which discriminate winning teams from losing teams, as measured by judges' responses on semantic differential scales. Consequently, the results obtained will be critical to understanding variables related to wins and losses quite apart from confirming or disconfirming research hypotheses.

Procedures

Rationale for Conditions

The decision was made to keep the experimental situation as realistic as possible. While there would be greater control videotaping, or staging a debate, the decision was made not to do so based on several considerations. First, it was felt that if judges believed that their votes were not going to affect the outcome of a tournament they might devote less attention to the task than if they were aware of the importance of their decisions. Second, gathering a group of judges together for the purpose of listening to a debate, staged or not, would no doubt affect their evaluation of the debate. It might, in fact, cause the judges to attempt to second guess the nature of the study. Third, it was felt that the behaviors of the debaters would be affected in an environment other than a real debate.

In order to insure that data on variables selected for study could be obtained, octa-final and quarter-final rounds during the Arizona Debate Institute were used for study. While the assumption was that the variables investigated occur in virtually every debate round, elimination rounds were used since by utilizing the best debates at the tournament, it was assumed there would be greater homogeneity in the behavior of the debaters and thus, all judges would experience more similar experimental conditions. Further, elimination rounds were selected since there would be a panel of three judges in each, increasing the reliability of scores.

Development of Test Instrument

The scales which comprised the test instrument were of the semantic differential type. Judges responded to a series of semantic differential scales related to the debater's delivery, argumentation, and evidence. The focus of the study is on those variables which discriminate for wins and losses as a function of the judges' responses.

A number of considerations influenced the decision to use semantic differential scales in this study. One very important reason was that semantic differential scales had previously been used in forensic research and shown to be effective. Notably, the previously reported studies conducted by Williams, Clark and Wood (1966); Williams and Webb (1964); and Williams, Webb and Clark (1966) would seem to suggest that semantic differential scales can be productively utilized in forensic research. Given the overall lack of empirical research in forensics, it is not surprising that there exists no validated scales similar to McCroskey's (1966) scales for measuring ethos. Consequently, for the majority of the scales, the decision was made to select from the original scales derived by Osgood, Suci and Tannenbaum (1961) for the majority of the scales. Several additional semantic differential scales were drawn from the research of King and Phifer (1968) and one additional scale from the research of Burgoon and Montgomery (1976). (See Appendix A for specific scales and source citations.)

A second important consideration in utilizing semantic differential scales is directly related to the use of generalized scales like Osgood's et al. (1961). Kerlinger (1964, pp. 569, 578) has supported

the general applicability of Osgood's et al. scales for most research. He further maintained that semantic differential scales are "flexible and relatively easy to adapt to varying research demands, quick, and economical to administer and score." Nunnally (1967, p. 535) underscores the flexibility of the semantic differential stating that "it is a very flexible approach to obtaining measures of attitudes and other sentiments. The flexibility of the approach is one of its appealing features." Likewise, Osgood et al. (1961, p. 198) wrote that one of the advantages of the semantic differential "is its flexibility with respect to the nature of the concept judged." Another element influencing the decision to use such scales is as Nunnally (1967, p. 535) notes: the massive quantity of research done by Osgood et al. in testing the scales. Further justification for their use can be found in the research of Stewart and Merchant (1969) where they employed semantic differential scales taken from the research of Osgood et al. and found them a reliable means for assessing attitudes toward debaters. Similarly, research conducted by Burgoon and Montgomery (1976) on debater credibility indicated the utility of semantic differential scales for use in forensic research.

In summary, several factors influenced the decision to use semantic differential scales. First, they are flexible and easy to administer; second, evidence suggests that the generalized scales of Osgood et al. are reliable measures of attitude; and third, the semantic differential and specifically the generalized scales of Osgood et al. have been used for forensic research.

The scales were developed in an initial study during the Wildcat Debate Workshop of 1980. After watching a debate, judges were asked to fill out a series of scales. It was from these scales that the test instrument used in the final study was drawn. Two criteria were used in selecting scales for the final instrument. First, scales were selected which judges in the pretest most often marked on the extreme, which eliminated those scales which did not discriminate. Second, scales were selected which were highly loaded on in previous studies.

The final test instrument was a semantic differential composed of seventeen bi-polar adjectives asking the judges to respond to the debater's delivery, argumentation, and evidence. (See Appendix B for a complete copy of the test instrument.)

As was done with the test instrument, the procedures in the study were pre-tested during the Wildcat Debate Workshop in the summer of 1980. Included on the preliminary instrument were several questions designed to assess the ease with which subjects were able to complete the instrument. Based on the subject's responses, and the experimenter's own evaluation of the execution of procedure pre-test, the final testing procedures were evolved.

Administration of Experiment

Data for the final study were gathered at the Arizona Debate Institute in the summer of 1980. The Institute provided an ideal opportunity for gathering data. Unlike a regional tournament, the Arizona Debate Institute attracts teams from around the country and for this

reason enhances the generalizability of the results of the study. All octa-final and quarter-final rounds were used in the analysis, resulting in a total of twenty-four debates. A total of fifty-eight completed test instruments were returned. Since each judge responded to both a winning and a losing team, there were fifty-eight cases in the win group and the same number in the loss group.

On the third day of the tournament, the elimination rounds, data on the judges' responses to the debater behavior were collected. During the octa-final and quarter-final rounds in Junior and Senior division, judges received packets with the test instrument. Instructions concerning at what point to open the packets were printed on the outside of the packet. (See Appendix C for these instructions.)

Data Analysis

The analytical procedure used in this study was discriminant function analysis. This procedure (Kerlinger and Pedhauzur 1973, p. 337) is appropriately used when a researcher is attempting to classify individuals into two or more groups based on characteristics the groups have in common. Such an analysis discriminates maximally between the two groups and explains "the nature of the discrimination." That is, the procedure will ascertain the "relative efficacies or weights" of the measures used in the discrimination.

In the case of the present study, the groups being classified, the dependent variable, was membership in either the group "wins" or "losses." The characteristics used for classifying independent

variables were measured by the scales of delivery, argument, and evidence. As was indicated in Chapter 1, when a researcher investigates a nominal dependent variable, discriminant function is an appropriate methodology.

The data gathered in the final study were subjected to discriminant function analysis. Additionally, the means and a correlation matrix were calculated. The discriminant function analysis program was that developed at the Computer Center at the University of Arizona, which was based on the discriminant function analysis program in the Statistical Package for the Social Sciences (Nie et al. 1975). A two-group, stepwise procedure, generating a single discriminant function prediction equation was used. The stepwise procedure begins by selecting the single most discriminating variable for win-loss, and continues to select variables which, as a group, discriminate for win-loss. Inclusion of independent variables in the equation was based on Wilk's Lambda. This statistic, according to Klecka (1980, p. 54), uses two criteria in selecting variables comprising the discriminant function. Specifically, the criteria are (a) mean differences between the win and loss groups, and (b) the homogeneity within these two groups. The tolerance level for inclusion of the variables was the default level built into the program since there was no basis on which to justify a change. The default minimum tolerance was thus set at .0001.

Summary

It has been the purpose of this chapter to outline the procedures used in the present research study. A justification has been offered for

the test instrument and research procedures used, and for the methodology employed. Last, the chapter considered the specific statistical program used to generate the results, which will be presented in the following chapter.

CHAPTER 4

RESULTS AND DISCUSSION

Introduction

Presented in this chapter are the results of the discriminant function analysis used to assess the ability of the independent variables to discriminate for wins and losses. Prior to discussing the discriminant function, data relating to the means, discriminant coefficients, and univariate F ratios will be considered. These data provide additional ways of examining wins and losses in a debate, though they do not provide the same conclusive information as generated by discriminant function analysis. When examining the following data, it is important to note that it only provides an indication of discriminating variables and is not the same as the discriminant function. Nevertheless, this information provides for a more thorough understanding of how judges perceive debater behavior. Following this will be an analysis of the strength of the discriminant function for discriminating between wins and losses. Last, possible interpretations of the behavioral impacts of the discriminant function will be discussed.

Preliminary Analysis

Following will be a discussion of the means, discriminant coefficients and univariate F ratios generated by the data.

Means

The means of the ratings of the independent variables provide the first indication of which variables distinguish between wins and losses (Lehmann 1969, p. 507). Intuitively it would seem that those variables with the largest win/loss mean differences would probably contribute most to the distinction between the two groups. Based solely on these means (as indicated in Table 2), Argument (Convincing-Unconvincing) with a win mean of 5.98, a loss mean of 4.34 and a difference of 1.64 would likely be the most important. Further, Evidence (Convincing-Unconvincing) with a win mean of 5.67 and a loss mean of 4.36 yielding a difference of 1.31, might also distinguish between wins and losses. Other variables with comparatively large mean differences are Evidence (Strong-Weak) with a win mean of 5.68, a loss mean of 4.44 and a difference of 1.24; Evidence (Effective-Ineffective) with a win mean of 5.74, a loss mean of 4.56 and a difference of 1.18; Evidence (Valuable-Worthless) generating a win mean of 5.77 and a loss mean of 4.72 resulting in a difference of 1.05; and Argument (Organized-Confused) with a win mean of 5.67, a loss mean of 4.67, and a difference of 1.00.

While these means offer an indication of variables which differ on wins and losses, they may not be the same as those variables in the discriminant function since discriminant function analysis utilizes a number of criteria in selecting discriminating variables and not specifically mean differences.

Table 2. Means.

Variable	Win Mean	Loss Mean	Difference
Evidence			
Valuable-Worthless	5.77	4.72	1.05
Effective-Ineffective	5.74	4.56	1.18
Strong-Weak	5.68	4.44	1.24
Pleasant-Unpleasant	4.67	4.65	.02
Convincing-Unconvincing	5.67	4.36	1.31
Argument			
Good-Bad	5.94	5.00	.94
Clear-Muddled	5.55	4.79	.76
Pleasant-Unpleasant	5.10	4.60	.50
Strong-Weak	5.82	4.60	1.22
Organized-Confused	5.67	4.67	1.00
Convincing-Unconvincing	5.98	4.34	1.64
Effective-Ineffective	6.06	4.56	1.50
Delivery			
Pleasant-Unpleasant	5.34	5.01	.33
Fast-Slow	5.36	4.98	.38
Active-Passive	5.53	5.12	.41
Coherent-Incoherent	5.27	4.93	.34
Good-Bad	5.55	4.98	.57

An important and useful method for examining the data is in the form of the discriminant coefficients. The statistical program used in this study generated both standardized and unstandardized discriminant coefficients. Because the data input for the program was not standardized, the discriminant scores were calculated using unstandardized discriminant coefficients. (Table 7 shows these discriminant scores.) Each resulting score indicates the relative distance in standard deviation units of the case from the grand centroid. This centroid is "that point in space where all of the discriminating variables have their average values over all the points" (Klecka 1980, pp. 22-3). The discriminant score is calculated by multiplying each applicable variable rating by its unstandardized discriminant coefficient, then summing these products with the addition of a constant (Klecka 1980, p. 24). Thus, each unstandardized discriminant coefficient indicates the absolute contribution of a variable, plus the constant, in determining a discriminant score (Klecka 1980, p. 29). It is not valid to compare unstandardized coefficients with each other.

Standardized discriminant coefficients, on the other hand, indicate the relative contribution of each independent variable to the function. Hence, the greater the size of a standardized discriminant coefficient, the greater will be its contribution to the function as compared to the other variables with lesser coefficients (Klecka 1980, pp. 29-30).

Table 3 displays the standardized discriminant coefficients for each of the seventeen independent variables. It is important to note

Table 3. Standardized discriminant coefficients.

Variable	Discriminant Coefficients
Evidence	
Valuable-Worthless	-.523
Effective-Ineffective	.139
Strong-Weak	.726
Pleasant-Unpleasant	-.294
Convincing-Unconvincing	.369
Argument	
Good-Bad	-.179
Clear-Muddled	-.257
Pleasant-Unpleasant	.187
Strong-Weak	-.051
Organized-Confused	.128
Convincing-Unconvincing	.700
Effective-Ineffective	.433
Delivery	
Pleasant-Unpleasant	-.315
Fast-Flow	.044
Active-Passive	.089
Coherent-Incoherent	-.081
Good-Bad	.061

that the magnitude of each standardized discriminant coefficient is not necessarily indicative of the ordering which takes place in the stepwise selection procedure for the discriminant function (Nie et al. 1970, p. 448).

Evidence (Strong-Weak) with a standardized discriminant coefficient of .726 is the greatest contributor to the discriminant function. Argument (Convincing-Unconvincing) with a standardized discriminant coefficient of .700 is the second highest contributor. The third is Evidence (Valuable-Worthless) with a standardized discriminant coefficient of -.523, while the fourth variable which contributes is Argument (Effective-Ineffective) with a standardized discriminant coefficient of .433. Mathematically, these variables would potentially contribute most to the discriminant function. The specific interpretation of the unstandardized discriminant coefficients as they relate to the discriminant function will be considered in the discussion section of this chapter.

Univariate F's

Another indicator of variables which might distinguish between win and loss groups is in the form of the univariate F ratios and their corresponding Wilk's Lambdas. These F's represent a one-way analysis of variance of the group mean for each of the seventeen independent variables (Nie et al. 1975, p. 460). Thus, the analysis here is performed on each variable individually, unlike the multivariate F's and Wilk's Lambdas to be reported later.

Table 4 displays the univariate F's and Lambdas. Considering these data, it would seem that Argument (Convincing-Unconvincing) with the lowest Lambda of .6645, the highest univariate F of 57.56 and significant beyond .000 if one were considering a single variable at a time, would likely distinguish between wins and losses in a debate. Likewise, Argument (Effective-Ineffective) with a Lambda of .7050, a univariate F of 47.69 and significant beyond .000; Argument (Strong-Weak) with a Lambda of .7553, an F of 36.92, also significant beyond .000; would be variables that would distinguish between wins and losses.

On the other hand, it would seem that Evidence (Pleasant-Unpleasant) with a Lambda of .999, a univariabe F of .006 and a significance level of .935 would not be a good predictor variable for wins and losses. The same would probably also be true of Delivery (Pleasant-Unpleasant) with a Lambda of .9892, a univariate F of 1.24, and significant at .267 and Delivery (Coherent-Incoherent) with a Lambda of .9849, a univariate F of 1.74 and significant at .189 and the additional variables in Table 4 with small univariate F's and large Lambdas.

It is important to remember when examining these F's that they are derived individually. That is, they are calculated without consideration of the contribution the other variables could make toward discriminating for wins and losses. Thus, a variable which might have a significant univariate F could quite realistically not be part of the discriminant function because it adds no new information or discriminating ability to variables already included in the function. Nevertheless,

Table 4. Univariate F ratios.

Variable	Wilk's Lambda	Univariate F	Significance
Evidence			
Valuable-Worthless	.8441	21.05	.000
Effective-Ineffective	.7937	29.63	.000
Strong-Weak	.7564	36.71	.000
Pleasant-Unpleasant	.9999	.006	.935
Convincing-Unconvincing	.7694	34.15	.000
Argument			
Good-Bad	.8175	25.44	.000
Clear-Confusing	.9222	9.61	.002
Pleasant-Unpleasant	.9652	4.10	.045
Strong-Weak	.7553	36.92	.000
Organized-Disorganized	.8774	15.92	.000
Convincing-Unconvincing	.6645	57.56	.000
Effective-Ineffective	.7050	47.69	.000
Delivery			
Pleasant-Unpleasant	.9892	1.24	.267
Fast-Slow	.9766	2.72	.101
Active-Passive	.9617	4.54	.035
Coherent-Incoherent	.9849	1.74	.189
Good-Bad	.9369	7.68	.006

With 1 and 114 degrees of freedom

on a univariate level, these F's do provide a useful way of looking at the differences between the win and loss groups.

Discriminating Variables

Discriminant Function

Results. The stepwise procedure selected was Wilk's Lambda. This measure considers the difference between the win and loss groups in addition to the homogeneity within the groups. Wilk's Lambda is an inverse statistic, thus the smallest Lambda is sought for each stepwise procedure. All seventeen variables were included in the discriminant function based on Wilk's Lambda. Had the stepwise procedure been performed with Rao's V, it would appear that only the first four variables in the procedure would have made a significant contribution to the discriminant function. Unlike Wilk's Lambda, Rao's V is concerned with the separation of the group centroids and not with the cohesiveness within the groups (Klecka 1980, p. 54). There is no clear reason why Rao's V and Wilk's Lambda generated different results. One possible reason could be that the study's tolerance levels were less conservative for Wilk's Lambda than for Rao's V. A second explanation probably exists in the nature of the two statistics, specifically that Wilk's Lambda is concerned with the homogeneity of the cases and Rao's V is not. While these may have been factors, they do not provide a precise reason for the differences in the results. Nevertheless, it is apparent that within the confines of this study, Rao's V is a more conservative measure for distancing the win and loss groups.

Although the following report concerning the statistical significance of the discriminant function will be primarily based on Wilk's Lambda and thus include all seventeen variables, it should be noted that it is the first four variables which account for the majority of the discrimination. As presented in Table 5, and discussed previously, this is most clearly indicated by the more conservative measure of Rao's V, which shows that it is the first four variables which make a significant contribution. The change in Rao's V for the first variable, Argument (Convincing-Unconvincing), is 57.56 and significant beyond .000. For the second variable, Evidence (Strong-Weak), the change was 11.74 and significant at .001. The change in the third variable, Delivery (Pleasant-Unpleasant), was 10.94 and significant at .001; and for the fourth variable, Evidence (Valuable-Worthless), had a change of 4.16 and significant at the .041 level. The fourth variable causes the least amount of change in the statistic. Therefore, the first three variables produce the majority of the change in Rao's V and the fourth variable probably does not add a great deal of new information to the discriminant function.

The significance levels for the Wilk's Lambdas would indicate that all seventeen variables are included in the discriminant function. An examination of the multivariate F's to enter or remove would suggest that it is the first four variables which seem to account for the majority of the discriminant function's discriminating ability. For Argument (Convincing-Unconvincing) the F was 57.56 and significant beyond .000, as was the significance of all of the variables. For Evidence (Strong-

Table 5. Summary table.

Dimension	Factor	F to Enter or Remove	Wilk's Lambda	Sig.	Rao's V	Change in Rao's V	Sig.
Argument	Convincing-Unconvincing	57.56	.6644	.000	57.56	57.56	.000
Evidence	Strong-Weak	7.73	.6219	.000	69.30	11.74	.001
Delivery	Pleasant-Unpleasant	6.68	.5868	.000	80.24	10.94	.001
Evidence	Valuable-Worthless	2.37	.5745	.000	84.40	4.16	.041
Evidence	Convincing-Unconvincing	1.63	.5661	.000	87.35	2.94	.086
Evidence	Pleasant-Unpleasant	1.83	.5567	.000	90.75	3.39	.065
Argument	Effective-Ineffective	1.20	.5506	.000	93.04	2.28	.130
Argument	Clear-Muddled	1.16	.5447	.000	95.28	2.24	.134
Delivery	Active-Passive	.76	.5408	.000	96.79	1.51	.218
Argument	Good-Bad	.57	.5378	.000	97.94	1.14	.284
Argument	Pleasant-Unpleasant	.50	.5352	.000	98.97	1.02	.310
Argument	Organized-Confused	.20	.5342	.000	99.39	.41	.517
Delivery	Coherent-Incoherent	.19	.5332	.000	99.80	.40	.523
Evidence	Effective-Ineffective	.18	.5322	.000	100.20	.39	.527
Delivery	Good-Bad	.06	.5318	.000	100.33	.13	.715
Delivery	Fast-Slow	.05	.5315	.000	100.45	.12	.728
Argument	Strong-Weak	.02	.5314	.000	100.51	.05	.809

Final Lambda = .5314 Canonical Correlation = .6845

Weak), the F was 7.73. Delivery (Pleasant-Unpleasant) generated an F of 6.68 and Evidence (Valuable-Worthless) generated an F of 2.37. The combined F of these four variables accounts for 74.34 of the total discriminating ability of the function with only 8.35 remaining, divided among the last thirteen variables. Two cautions should be issued at this point concerning the above statement. First, as will be discussed in more detail later, the change in Rao's V for Evidence (Valuable-Worthless) and the F to enter or remove may be interpreted in such a way that it is not considered a discriminating variable. That is, the change in Rao's V and the F value for this variable seems to be closer in magnitude to the remaining thirteen variables. Second, preliminary analyses, such as those considered in the first section of this chapter, would indicate that Delivery (Pleasant-Unpleasant) might not function as an effective predictor variable.

In terms of the utility of the discriminant function, the final Lambda of .5314 and the canonical correlation of .6845 would indicate that the discriminant function produces a fairly high degree of separation between the win and loss groups. Lambda is measured between 0.0 and 1.0. As Lambda moves toward 1.0, there is less discrimination present. Therefore, the final Lambda, which is approximately midway between these two points, reflects a reasonable degree of separation between the win and loss groups. The canonical correlation is an important measure of the utility of the discriminant function also. Specifically, the canonical correlation indicates how closely related the groups are to the discriminant function (Klecka 1980, p. 36). The

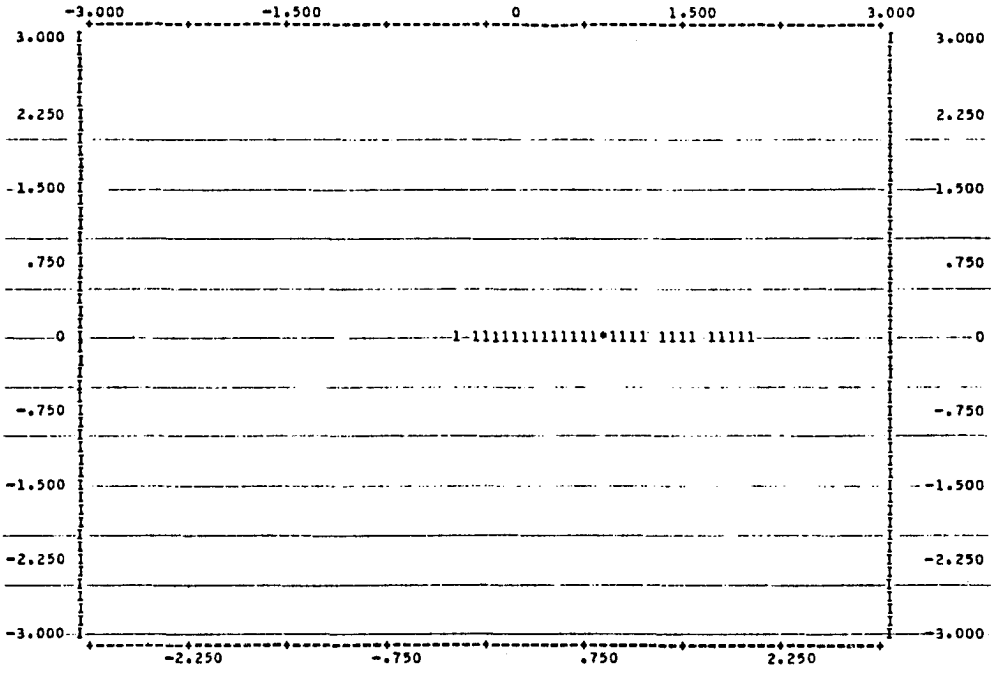
canonical correlation for this study is quite high, which would appear to indicate that the function provides good discrimination between wins and losses.

The group centroids are the central positions for the win and loss groups using the means of each of the seventeen independent variables in the computation. The win group centroid is .93086 and the loss centroid is -.93086, indicating that the win group is positioned on one side of the grand centroid while the loss group is positioned on the opposite side. The degree of separation between the group centroids is 1.86172. While the group centroids indicate that separation between the win and loss groups does exist, this information alone does not provide sufficient analysis. Table 6 displays a plot of the discriminant scores. While the win group is fairly homogeneous, clustering around its group centroid, this is not true for the loss group. An examination of this plot reveals that the loss group overlaps with the win group to some degree. This would tend to indicate that in some cases the discriminant function is less accurate in classifying losses. That is, in some cases debaters lost debates when their scores on the semantic differential would indicate that they should have won, since their scores were similar to other winning teams.

The Classification Matrix, Table 7, provides additional information on this point. The Classification Matrix evaluates the classification ability of the discriminant function for each individual case. Klecka (1980, pp. 42, 49, 50) indicates that a second function of discriminant function analysis is classification. Specifically, this is

Table 6. Plot of discriminant scores. -- (a) Wins; and (b) Losses.

a.



b.

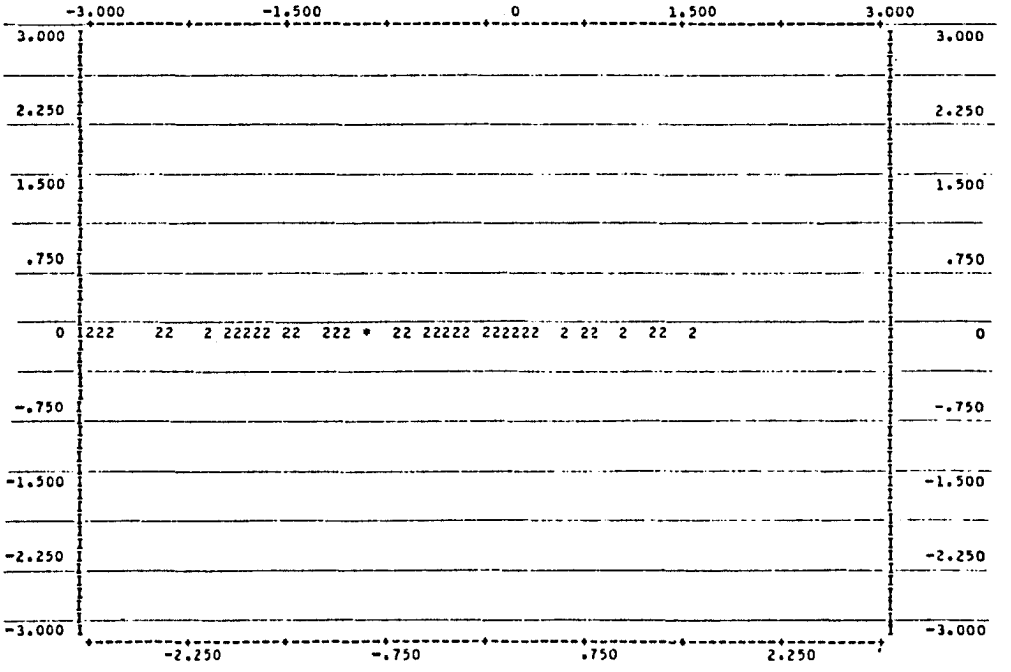


Table 7. Classification matrix. -- (a) Wins; and (b) Losses.

a.

Case	Actual Group	Predicted Group	Discriminant Scores	P(X/G)	P(G/X)
1	win	win	.227	.482	.604
3	win	win	.865	.948	.833
5	win	loss*	-.222	.479	.602
7	win	win	.444	.626	.696
9	win	win	1.002	.943	.866
11	win	win	1.332	.688	.923
13	win	win	1.520	.556	.944
15	win	win	1.027	.923	.871
17	win	win	.905	.980	.844
19	win	win	1.087	.876	.883
21	win	win	1.973	.298	.975
23	win	win	1.717	.432	.961
25	win	win	.870	.952	.835
27	win	win	.520	.681	.725
29	win	win	1.767	.403	.964
31	win	win	.362	.569	.662
33	win	win	.113	.413	.552
35	win	win	1.405	.635	.932
37	win	win	.664	.789	.775
39	win	win	.768	.871	.807
41	win	win	2.023	.275	.977
43	win	win	3.236	.021	.998
45	win	win	.367	.573	.665
47	win	loss*	-.050	.378	.523
49	win	win	.686	.807	.782
51	win	win	.421	.610	.686
53	win	win	.984	.958	.862
55	win	win	1.491	.575	.941
57	win	win	1.815	.377	.967

Table 7.a. -- Continued

Case	Actual Group	Predicted Group	Discriminant Scores	P(X/G)	P(G/X)
59	win	win	1.182	.801	.900
61	win	win	.258	.501	.618
63	win	win	.108	.411	.550
65	win	loss*	-.250	.496	.614
67	win	win	1.027	.923	.871
69	win	win	.286	.519	.630
71	win	win	.719	.832	.792
73	win	win	.322	.543	.645
75	win	win	.787	.885	.812
77	win	win	1.369	.662	.927
79	win	loss*	-.219	.476	.600
81	win	win	1.434	.615	.935
83	win	win	1.871	.347	.970
85	win	win	1.080	.881	.882
87	win	win	1.764	.405	.964
89	win	win	1.825	.371	.968
91	win	win	.853	.938	.830
93	win	win	.177	.451	.582
95	win	win	1.746	.415	.963
97	win	win	1.102	.864	.886
99	win	win	1.537	.544	.946
101	win	win	.325	.544	.647
103	win	win	1.499	.570	.942
105	win	win	.564	.713	.741
107	win	win	.158	.440	.573
109	win	win	.003	.354	.501
111	win	win	.864	.947	.833
113	win	win	1.564	.527	.948
115	win	win	.713	.827	.790

Table 7.b. Losses.

Case	Actual Group	Predicted Group	Discriminant Scores	P(X/G)	P(G/X)
2	loss	loss	-1.067	.892	.879
4	loss	loss	- .241	.491	.610
6	loss	loss	-3.036	.035	.997
8	loss	loss	-1.818	.375	.967
10	loss	loss	-2.369	.150	.988
12	loss	loss	- .378	.581	.669
14	loss	loss	- .652	.781	.771
16	loss	loss	- .477	.650	.708
18	loss	loss	-1.039	.914	.874
20	loss	loss	-2.824	.058	.995
22	loss	loss	-1.855	.356	.969
24	loss	loss	-1.019	.929	.870
26	loss	loss	-2.119	.235	.981
28	loss	loss	- .634	.766	.765
30	loss	loss	-1.427	.620	.934
32	loss	loss	-1.654	.470	.956
34	loss	win*	1.308	.706	.919
36	loss	loss	- .260	.502	.619
38	loss	loss	-1.079	.882	.882
40	loss	loss	-2.128	.231	.981
42	loss	loss	-3.057	.034	.997
44	loss	loss	- .159	.440	.574
46	loss	loss	-2.491	.119	.990
48	loss	loss	-2.410	.139	.989
50	loss	win*	.391	.589	.674
52	loss	win*	.824	.915	.823
54	loss	win*	.723	.835	.793
56	loss	win*	.725	.837	.794
58	loss	win*	.114	.414	.553

Table 7.b. -- Continued

Case	Actual Group	Predicted Group	Discriminant Scores	P(X/G)	P(X/G)
60	loss	loss	- .328	.547	.648
62	loss	win*	.327	.546	.648
64	loss	loss	-2.918	.047	.996
66	loss	loss	- .367	.573	.665
68	loss	loss	-1.500	.569	.942
70	loss	win*	.631	.764	.764
72	loss	loss	-1.880	.343	.971
74	loss	win*	.336	.552	.651
76	loss	loss	-1.491	.575	.941
78	loss	loss	-1.919	.323	.973
80	loss	loss	-2.500	.117	.991
82	loss	win*	.043	.375	.520
84	loss	loss	- .019	.362	.509
86	loss	loss	-1.646	.474	.955
88	loss	win*	1.577	.518	.950
90	loss	win*	1.054	.902	.877
92	loss	loss	- .459	.637	.702
94	loss	loss	- .123	.419	.557
96	loss	loss	-1.614	.494	.953
98	loss	win*	1.320	.698	.921
100	loss	loss	-1.022	.927	.870
102	loss	loss	- .680	.802	.780
104	loss	loss	-1.169	.812	.898
106	loss	win*	.252	.497	.615
108	loss	loss	-1.916	.325	.973
110	loss	loss	-1.720	.430	.961
112	loss	loss	-1.901	.332	.972
114	loss	loss	-3.145	.027	.997
116	loss	loss	-1.101	.865	.886

the ability to predict the group to which a case probably belongs by comparing its discriminant score to the closest group centroid. He goes on to indicate that while classification can be used to predict group membership of unknown cases, it can also function to "test the accuracy of the classification procedure." By applying the discriminant function to known cases, the proportion of cases which were classified correctly as wins or losses indicates the strength of the discriminant function "and indirectly confirms the degree of group separation." Thus, according to Klecka, the percent of known cases which are classified correctly is one measure, along with the overall Wilk's Lambda and the canonical correlation used to determine the degree of discrimination present in the independent variables.

The ratings for each variable in a given case are entered into the discriminant function with a composite discriminant score calculated. This score classifies the case into a win (group 1) or a loss (group 2). The SPSS program at the University of Arizona printed wins as positive integers and losses as negative integers, but this assignment is arbitrary. The asterisks (Table 7) show the cases in which the discriminant function classification is different from the actual classification. While the discriminant function utilizes all seventeen variables to determine the discriminant score, one must realize that the first four and perhaps only the first two account for the majority of the discrimination as indicated earlier.

As Klecka (1980, p. 50) states, the percentage of cases correctly classified will provide information relating to the strength of

the function's discriminating ability. These data would suggest that the discriminant function is more accurate for classifying wins and less accurate for classifying losses. Out of fifty-eight cases, only four cases were classified incorrectly as losses when in actuality they were wins. This corresponds to 93.10% accuracy or strength in classifying wins. On the other hand, fourteen cases were classified as wins when in fact the teams had lost the debate. The classification accuracy for losses is therefore only 75.86%. In examining these figures, it is important to remember that there is a fifty percent chance of classifying cases correctly based purely on chance. Thus for classifying losses, the discriminant function is accurate only twenty-five percent above the chance level. Several reasons could account for this occurrence, which will be considered in the discussion section of this chapter.

The two measures of predicted group membership probability on the Classification Matrix provide additional clarification of the strength of the discriminant function. The first statistic, P(X/G), considers the probability that a case located that far from the group centroid would actually belong to the group. Another way of looking at this statistic is that it is an estimate of the proportion of the cases in the predicted group's population which are further from the centroid than the case is (Klecka 1980, pp. 55-6). If the value is small, there is a chance that the case does not belong to that group even though it is classified as such. The second probability, P(G/X), indicates the probability that a case belongs to the predicted group.

An examination of the four cases predicted to be losses which in reality were wins suggests that there is a moderate probability of these cases actually belonging to the predicted loss group. The P(G/X) for these cases are: case 5, .602; case 47, .523; case 65, .614; and case 79, .600. The P(X/G) indicates that if these cases were losses and matched against the location of real losses, 47.9 percent, 37.8 percent, 49.6 percent, and 47.6 percent, respectively, of real losses are further from the group centroid than each of these cases are. This amounts to a high percent of classification; the odds are better than fifty-fifty that these four cases really are losses. This means that while the teams won the debates, they won based on some reason other than the basic qualities common to the rest of the winning teams.

More interesting are the probabilities for the loss group. The majority of the cases predicted to be wins but which were losses have high probabilities of actually being similar to the scores of other winning teams. Except for those individual cases with really low P(X/G) probabilities, the majority of the cases actually belong to the predicted group. Again it would seem reasonable to conclude that the primary strength of the discriminant function is in its ability to discriminate wins as opposed to losses.

Discussion. The following discussion will seek to interpret the data from two different perspectives. First, an attempt will be made to consider behaviorally the significant variables. Second, the discussion will center on possible explanations concerning the strength of the function for discriminating between wins and losses.

Possible Behavioral Interpretations

The previous section discussed the strength of the discriminant function for distancing the win and loss groups. That discussion reflected all seventeen variables, though as was noted, the first four account for the majority of the function's discriminating ability. The following discussion will consider only the four variables believed to make a significant contribution to the discriminant function.

This section will seek to discuss the results behaviorally and thus develop an understanding for the specific variables that emerge as discriminating variables. This discussion will offer an interpretation of semantic differential data as it is used in discriminant function analysis.

In order that the following analysis be clear, an understanding of the derivation of the unstandardized discriminant function coefficients and the discriminant function is necessary. The discriminant scores, the results of the discriminant function, are the sum of the judge's ratings of the seventeen independent variables multiplied by each weighted unstandardized discriminant coefficient in addition to a constant. Table 8 lists the unstandardized discriminant coefficients of this study. The unstandardized discriminant coefficients are part of the discriminant function which defines the maximum separation between the win and loss groups, considering all of the independent variables.

Since the discriminant function operates as a unit to classify the win and loss groups, an individual discussion of the independent variables' contribution to the discriminant function would not be

Table 8. Unstandardized discriminant function coefficients.

Variable	Discriminant Coefficients
Evidence	
Valuable-Worthless	-.391
Effective-Ineffective	.107
Strong-Weak	.575
Pleasant-Unpleasant	-.259
Convincing-Unconvincing	.269
Argument	
Good-Bad	-.160
Clear-Muddled	-.188
Pleasant-Unpleasant	.139
Strong-Weak	-.411
Organized-Confused	.894
Convincing-Unconvincing	.493
Effective-Ineffective	.312
Delivery	
Pleasant-Unpleasant	-.198
Fast-Slow	.356
Active-Passive	.844
Coherent-Incoherent	-.580
Good-Bad	.545
Constant	-4.43

statistically justified. Thus, it is not specifically correct to attribute certain variables as contributing to a win and other variables as contributing to a loss. Realizing that the positive and negative signs on the SPSS program are arbitrary, in a very general sense though, one can say that specific variables move an individual case toward a win or a loss. That is, for this case, negative coefficients move a case toward a loss while positive coefficients move a case toward a win, since the unstandardized discriminant coefficients cluster around their respective centroids, positive for a win (group 1) and negative for a loss (group 2). Nevertheless, this is only in a very general sense since the unstandardized discriminant coefficients operate simultaneously to discriminate for both wins and losses over all the cases.

It is with caution, though, that one can discuss these unstandardized discriminant coefficients as moving an individual case toward a win or a loss if done so in a hypothetical manner. Within the confines of a hypothetical situation, realizing that positive unstandardized discriminant coefficients move a case toward the win centroid and negative unstandardized discriminant coefficients move a case toward a loss, if a team wanted to win a debate, they would need to minimize the effect of the negative unstandardized discriminant coefficients by achieving small semantic differential scale values for the corresponding independent variables and to maximize the positive unstandardized discriminant coefficients by achieving high scale values on the semantic differential for the corresponding variable.

Mathematically, this occurs since a negative unstandardized discriminant coefficient multiplied by any rating will produce a negative product. Likewise, a positive discriminant coefficient multiplied by a semantic differential value will yield a positive product. If hypothetically, a debate team wanted to lose a debate, the negative unstandardized discriminant coefficients would need to be multiplied by large semantic differential scores. The larger the rating, the more absolutely it contributes to the negative discriminant score and hence the predicted loss. In order to minimize the positive unstandardized discriminant coefficients, they must be multiplied by small semantic differential scale values. Here the reader must recognize that while in theory this is the case, a variety of permutations can occur between the semantic differential scores and the discriminant coefficients which will yield a variety of results.

Discriminating Ability to the Function

In terms of the strength of the discriminant function, several interpretations emerge concerning its ability to discriminate for wins more accurately than losses. Recall from the results section that the function discriminates for wins with 93.19 percent accuracy and losses with 75.86 percent accuracy.

One possible explanation for this occurrence lies in the structure of the normal debate ballot. A ballot asks a judge to determine the winning team and not the losing team. To the extent that this is true, more care could have been taken when responding to the winning

team's scale and less to the losing team's. Another explanation for this is simply that the two teams could be of equal quality, hence both receiving high ratings. Indeed, for the losing teams which were predicted to win the debate this would seem to be the case. Since the debates used in the study were elimination rounds, it would not be unreasonable to expect both teams to do relatively well.

Summary

The primary function of this chapter was to present the results of the discriminant function analysis with an accompanying discussion of these results. The results section considered the means, discriminant coefficients, univariate F's, and the significance of the discriminant function. The discussion section examined the results in terms of behavioral implications, reasons for specific variables' emergence in the discriminant function, and interpretations for the strength of the discriminant function.

CHAPTER 5

CONCLUSIONS

Introduction

This study tested for debate effectiveness in terms of variables which discriminated for wins and losses. It has been a purpose of this study to provide data which would be useful to the intercollegiate debater in improving debate skills. That is, debaters should be able to emphasize those behaviors associated with winning debates and de-emphasize those behaviors associated with losing debates and thereby enhance their chances of doing well. This chapter will focus on conclusions derived from the results reported in the previous chapter.

This chapter will begin by making some initial statements about the research and its methodology. Next will be observations concerning the discriminant function itself, and last will be several conclusions related to the specific components of the discriminant function.

Initial Observations

The first conclusion reflects the measuring instrument used in this study:

1. Semantic differential scales can be used in a meaningful manner so as to discriminate between wins and losses in a debate.

While previous debate-related research has used semantic differential scales (Stewart and Merchant 1969; Burgoon and Montgomery 1976; King and Phifer 1968; Williams, Clark and Woods 1966; Williams and Webb 1964; Williams, Webb and Clark 1966), the intent of these studies was different from the current research. Consequently, the use of semantic differential scales in the current study for discriminating wins and losses makes it unique. The current research would lend additional justification for the continued use of semantic differential scaling techniques and specifically for measuring debate effectiveness.

The second conclusion deals with the use of discriminant function analysis in forensic research:

2. Semantic differential scales can appropriately be treated by discriminant function analysis for discriminating between wins and losses.

Chapter 2 considered, where relevant, the statistical procedures and methods used in previous studies. Although all were valid methods, the present study selected a method significantly different. Based on the results of the current study, discriminant function analysis has proven to be a successful method for discriminating between wins and losses in a debate round.

These two conclusions may appear self-evident. However, it is the very nature of forensic research which makes them warranted. Perhaps more so than other communication research areas, forensics is just now moving into experimental research in a serious way. For this

reason, specifying successes and failures in methodological procedures is valuable to future researchers.

Discriminating Between Wins and Losses

Chapter 4 indicated that four variables made a significant contribution to the discriminant function. Hence, the conclusion that:

3. Convincing-Unconvincing Arguments, Valuable-Worthless and Strong-Weak Evidence, and Pleasant-Unpleasant Delivery are variables which significantly discriminate between winning or losing a debate.

Perhaps one of the more interesting, but unexplained conclusions to be drawn from this study relates to the strength of the discriminant function for discriminating between wins and losses. Based on this is the following conclusion:

4. There seems to be more consistency in the behaviors associated with winning a debate than with losing a debate.

In Chapter 4 several possible reasons were given for this occurrence. Certainly there is no way of determining from this study why this appears to be the case.

While caution must be exercised when separately considering the variables which compose the discriminant function, based on the standardized discriminant coefficients it would appear that:

5. Convincing-Unconvincing Arguments are the single best discriminator for wins and losses.

The emergence of an argument variable would be consistent with the research reported in Chapter 2. That is, the research of Giffin (1959); McCroskey and Camp (1966); Williams and Webb (1964); and Williams, Webb and Clark (1966) all determined in one manner or another that argument is a dominant factor in a debate. The current study differentiates itself from these earlier studies in that it specifies a particular argument variable, specifically Convincing-Unconvincing.

To the extent that the Delivery variable can be interpreted behaviorally, and given the caution issued above concerning interpreting the variable separately, the following conclusion seems justified:

6. In terms of discriminating for wins and losses, specific delivery behaviors (Active-Passive, Coherent-Incoherent, Fast-Slow) seem to be less important than having a pleasant or unpleasant delivery style.

Summary

It has been the purpose of this chapter to identify pertinent conclusions based on the results in the previous chapter. The conclusions reported in this chapter were of three types. First, conclusions were presented which dealt with several of the methodological procedures utilized in the study. Second, were conclusions which dealt with the discriminant function as a unit. Third, and last, were conclusions which reflected two of the specific components of the discriminant function.

CHAPTER 6

IMPLICATIONS, LIMITATIONS, AND SUMMARY

Introduction

The purpose of this chapter is three-fold. First, implications for future research in the forensics area will be discussed. Second, the chapter will consider the limitations of the present study. Third, a summary will be provided of the major points of the research.

Implications for Future Research

Given the overall lack of empirical research conducted in the forensic area, it is not surprising that this study may have raised more questions than it answered. Forensic research can prove to be an exciting area and certainly can be utilitarian. The present study prompted the future research discussed below.

The first implication for future research is clearly linked to one limitation of the study. Specifically, the discipline suffers from a lack of instruments for measuring debater behavior. As was discussed in Chapter 3, there are no scales for measuring debater behaviors as they affect wins and losses in a debate. Scales do exist in debate research, but none specifically related to this research area. The current study adapted scales from other sources, and in doing so may have introduced uncontrolled variables affecting the results of the study since the scales were not originally designed for assessing wins

and losses. While there is probably nothing inherently wrong with each researcher developing individual test instruments, there would seem to be a great deal of utility to be derived from a set of validated scales from which a researcher could select scales appropriate to a given study. Such validated scales could also be used in conjunction with those developed for particular research purposes. It should also be noted that scales used in debate-related research have not been subjected to the kind of rigorous testing done in the McCroskey or Osgood et al. research. In sum, much debate research could benefit from further efforts in developing test instruments.

The present study asked judges to assign a win to either the affirmative or negative team and to respond to the scale items. It may be that the assigning of a win will affect the manner in which scales are marked and in fact there could be a difference between assigning a win and assigning a loss in marking the scales. Consequently, a second research area would involve judges assigning a loss and then responding to a set of scales. As yet there is no empirical evidence concerning the effect assigning a win or a loss may have on rating or ranking of scales, though some researchers have theorized about it. It may be that by the very nature of the debate ballot format, a team receives a loss by default, and perhaps this is reflected in the manner in which judges fill out either a test instrument or ratings on a debate ballot. If indeed this is true, it certainly suggests problems in the assignment of quality rating points for losing teams.

Third, it would be interesting to move beyond a judge's perception of what occurs in a debate round. While it is still the judge's perception of a debate upon which the decision is cast, it could be productive to investigate wins and losses from another perspective. It would be valuable to assess, as closely as possible, the behaviors of debaters but not relying on the judge's perceptions. Vasilius' Master of Arts thesis was an attempt to do this. By recording debates and counting various behaviors on the tapes, she was able to make such an assessment. It is important to note though, that in the Vasilius study and future research which might attempt to accomplish this task, that judgment and measurement are still taking place on the part of the experimenter. Nevertheless, it would seem that more research which is, in effect, speaker-centered could be useful in coaching forensics. The research question then would center around what debaters generally do in a debate, not what debaters generally do as perceived by various judges.

Fourth, given that the variables used in this study did emerge as significant, there is justification to continue research in what may be considered a skills approach to debate. On the other hand, this does not deny the need for research into other areas like the effects of judging paradigms on the outcome of a debate. It does suggest though, that continued research in this particular area is warranted.

As outlined in Chapter 3, the current research chose to utilize actual debate rounds as opposed to a more clinical approach. Future research might seek to use a laboratory situation where some of the

variables, as indicated in the following Limitations of the Study section of this chapter, might be controlled.

Limitations of the Study

Overall, there are five limitations of this study which need to be considered when examining the results. First, while the scales were pre-tested, a factor analysis was not performed on the scales used in the final study. High-loading scales were selected from previous research and used in this study. As indicated in the Implications for Future Research section of this chapter, the validity of the scales may have affected the results of the study.

Second, there may well have been confounding variables as a result of using actual debate rounds instead of utilizing a laboratory approach. The argument was made in Chapter 3 that a real debate situation was more advantageous than a clinical approach since both the debaters and the judges would be less likely to be on guard. Nevertheless, it is possible that variables unaccounted for in the design may well have intervened and affected the final results of the study.

Third, the fact that this study utilized elimination rounds instead of or in addition to preliminary rounds may have affected the results since elimination rounds supposedly exemplify the best debaters in any given tournament. Elimination rounds were selected because it was believed that this would more nearly insure that the judges would experience similar experimental conditions prior to their marking of the scale items. Nevertheless, the results are possibly an indication

of discriminating variables for more experienced debaters and thus may limit the generalizability of the results.

Similar to the issue of generalizability considered above is a fourth limitation to the study. Specifically, that the subjects involved in this study may be unique in several ways. First, the study was conducted during the summer at a debate workshop. Those students interested in attending a workshop in the summer may be different from those who compete during the school year. Interest aside, those students who attend summer workshops may be differentiated from a financial aspect. That is, not all students can afford to attend a workshop due to personal financial reasons or because their school programs are limited financially. The point is simply that when examining these results one would do well to remember that subjects for this study were attending a summer debate workshop and not a regular season tournament.

The final limitation is related to independency of judging assessments. Specifically, the results may be confounded by having the judges respond to both the winning and the losing teams in the debate. Klecka (1980, p. 8) indicates that there may be a problem in having done so. Nevertheless, a precedent had been set for such a methodological format by Klecka (1974). Consequently, while other procedures could possibly have been used in collecting the data for the current study, the research would indicate a justification for the current methodology, even though there may be some confounding of the results.

Summary

The purpose of this study was to identify variables which discriminate for winning and losing a debate as measured by judges' responses on semantic differential scales. It was the intent that the results of this study would provide practical coaching information for use in instructing debaters. This study differed from previous research in several ways. First, the study utilized discriminant function analysis for data evaluation. This procedure had heretofore not been used in forensic research and as indicated in Chapter 5, appears to be a useful statistical tool for forensic analysis. Further, this study generated a different measuring instrument for assessing variables contributing toward wins and losses. Specifically, this study used semantic differential scales.

The variables selected for analysis in this study were Argument, Evidence, and Delivery. Each dimension was defined by a series of semantic differential scales. Subjects for this study were judges and participants at the Arizona Debate Institute held during the summer of 1980. Octa-final and quarter-final rounds at the tournament were used for study.

Four scale items emerged as discriminating for wins and losses in a debate. The most discriminating variable came from the Argument dimension, specifically the scale item Convincing-Unconvincing. The second most discriminating variable was from the Evidence dimension, that is Strong-Weak. The third discriminating variable was from the

Delivery dimension, namely Pleasant-Unpleasant; and the last variable was also from the Evidence dimension, specifically Valuable-Worthless.

APPENDIX A

SCALES USED IN THE STUDY, WITH REFERENCES

The following scales came from the Osgood et al. research of 1961:

Argument

- Good-Bad
- Pleasant-Unpleasant
- Strong-Weak

Delivery

- Pleasant-Unpleasant
- Fast-Slow
- Active-Passive

Evidence

- Valuable-Worthless
- Strong-Weak
- Pleasant-Unpleasant

The following scales came from the King and Phifer research of 1968:

Argument

- Clear-Muddled
- Organized-Confused
- Convincing-Unconvincing

Evidence

- Effective-Ineffective
- Convincing-Unconvincing

The following scale came from the Burgoon and Montgomery research of 1976:

Delivery

- Coherent-Incoherent

APPENDIX B

TEST INSTRUMENT

ARIZONA DEBATE INSTITUTE: JUDGING QUESTIONNAIRE

Doctoral Dissertation
Jacqueline Smith-Donaldson
University of Arizona
August 1980

Currently, I am gathering data for my Doctoral Dissertation. The attached set of scales will form the bases of my empirical investigation. Please take a few minutes and complete the scales.

The purpose of this study is to determine your perception of what the debaters did in this debate round. On the following page you will find a set of instructions explaining the mechanics of marking the scales. The next several pages will contain the specific scales.

Thank you for your cooperation.

Jacque Smith-Donaldson

If you feel that the item (Evidence, Delivery or Argumentation) is very closely related to one end of the scale, you should place your check-mark as follows:

FAIR__X__:____:____:____:____:____:____ UNFAIR

FAIR____:____:____:____:____:____:__X__ UNFAIR

If you feel that the item is quite closely related to one or the other end of the scale (but not extremely), you should place your check-mark as follows:

NICE____:__X__:____:____:____:____:____ AWFUL

NICE____:____:____:____:____:__X__:____ AWFUL

If the item seems only slightly related to one side as opposed to the other side (but not really neutral), then you should check as follows:

FAIR____:____:__X__:____:____:____:____ UNFAIR

FAIR____:____:____:____:__X__:____:____ UNFAIR

The direction toward which you check, of course, depends upon which of the two ends of the scale seem most characteristic of the item you are judging. If you consider the item to be neutral on the scale, both sides of the scale equally associated with the concept, or if the scale is completely irrelevant, unrelated to the item then you should place your check-mark in the middle space:

NICE ____:____:____:__X__:____:____:____ AWFUL

AFFIRMATIVE team's EVIDENCE

valuable _____:_____:_____:_____:_____:_____:_____ worthless

effective _____:_____:_____:_____:_____:_____:_____ ineffective

weak _____:_____:_____:_____:_____:_____:_____ strong

pleasant _____:_____:_____:_____:_____:_____:_____ unpleasant

unconvincing _____:_____:_____:_____:_____:_____:_____ convincing

AFFIRMATIVE team's ARGUMENTATION

bad _____:_____:_____:_____:_____:_____:_____ good

muddled _____:_____:_____:_____:_____:_____:_____ clear

pleasant _____:_____:_____:_____:_____:_____:_____ unpleasant

strong _____:_____:_____:_____:_____:_____:_____ weak

confused _____:_____:_____:_____:_____:_____:_____ organized

unconvincing _____:_____:_____:_____:_____:_____:_____ convincing

ineffective _____:_____:_____:_____:_____:_____:_____ effective

AFFIRMATIVE team's DELIVERY

unpleasant _____:_____:_____:_____:_____:_____:_____ pleasant

slow _____:_____:_____:_____:_____:_____:_____ fast

passive _____:_____:_____:_____:_____:_____:_____ active

coherent _____:_____:_____:_____:_____:_____:_____ incoherent

good _____:_____:_____:_____:_____:_____:_____ bad

NEGATIVE team's EVIDENCE

valuable _____:_____:_____:_____:_____:_____:_____ worthless

effective _____:_____:_____:_____:_____:_____:_____ ineffective

weak _____:_____:_____:_____:_____:_____:_____ strong

pleasant _____:_____:_____:_____:_____:_____:_____ unpleasant

unconvincing _____:_____:_____:_____:_____:_____:_____ convincing

NEGATIVE team's ARGUMENTATION

bad _____:_____:_____:_____:_____:_____:_____ good

muddled _____:_____:_____:_____:_____:_____:_____ clear

pleasant _____:_____:_____:_____:_____:_____:_____ unpleasant

strong _____:_____:_____:_____:_____:_____:_____ weak

confused _____:_____:_____:_____:_____:_____:_____ organized

unconvincing _____:_____:_____:_____:_____:_____:_____ convincing

ineffective _____:_____:_____:_____:_____:_____:_____ effective

NEGATIVE team's DELIVERY

unpleasant _____:_____:_____:_____:_____:_____:_____ pleasant

slow _____:_____:_____:_____:_____:_____:_____ fast

passive _____:_____:_____:_____:_____:_____:_____ active

coherent _____:_____:_____:_____:_____:_____:_____ incoherent

good _____:_____:_____:_____:_____:_____:_____ bad

In this debate I voted for the _____ team.
AFFIRMATIVE or NEGATIVE

Please mark the following:

MALE_____ FEMALE_____

AGE:

20- _____ 36-40 _____

21-25 _____ 41-50 _____

26-30 _____ 50+ _____

31-35 _____

Please circle one of the following judging paradigms you generally follow when judging a debate:

1. policy marker
2. hypothesis tester
3. systems theory
4. other _____

APPENDIX C

INSTRUCTIONS TO JUDGES

PLEASE DO NOT OPEN THIS ENVELOPE UNTIL THE DEBATE IS OVER AND YOU HAVE FILLED OUT YOUR WHITE COPY OF THE DEBATE BALLOT. THANK YOU FOR YOUR COOPERATION.

LIST OF REFERENCES

Anderson, Kenneth E. "Quantitative Research in Debate." The Journal of the American Forensic Association 3 (September 1966): 112-15.

Benson, James A. "The Use of Evidence in Intercollegiate Debate." The Journal of the American Forensic Association 7 (Spring 1971): 260-70.

Burgoon, Judee K. "Evaluation Criteria as Predicators of Debate Success." The Journal of the American Forensic Association 12 (Summer 1975): 1-4.

Burgoon Judee and Montgomery, Charles. "Dimensions of Credibility for the Ideal Debater." The Journal of the American Forensic Association 12 (Spring 1976): 171-77.

Colbert, Kent R. "Speaking Rates of N.D.T. Finalists from 1968-1980." The Journal of the American Forensic Association 18 (Summer 1981): 73-6.

Cox, J. Robert. "A Study of Judging Philosophies of the Participants of the National Debate Tournament." The Journal of the American Forensic Association 11 (Fall 1974): 61-71.

Cross, John D. and Matlon, Ronald J. "An Analysis of Judging Philosophies in Academic Debate." The Journal of the American Forensic Association 15 (Fall 1978): 110-23.

Diehl, Charles; White, Richard C.; and Burk, Kenneth W. "Rate and Communication." Speech Monographs 26 (August 1959): 229-32.

Giffin, Kim. "A Study of the Criteria Employed by Tournament Debate Judges." Speech Monographs 26 (March 1959): 69-71.

Giffin, Kim and Warner, Donald. "A Study of the Influence of an Audience on the Rate of Speech in Tournament Debates." The Speaker 45 (November 1962): 10-13.

Huck, Schuyler W.; Cormier, William H.; and Bounds, William G., Jr. Reading Statistics and Research. New York: Harper and Row, 1974.

Jones, Merritt B. "An Experimental Study of the Effects of Speech Rate on Audience Judgments in Debate Situations." Ph.D. dissertation, University of Southern California, 1956. Abstracted in Speech Monographs 24 (June 1957): 92-3.

Kerlinger, Fred N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1964.

Kerlinger, Fred N., and Pedhauzur, Elazar J. Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1973.

King, Thomas R., and Phifer, Gregg. "The College Debater as Seen by Himself and by His Peers." The Journal of the American Forensic Association (Spring 1968): 48-52.

Klecka, Carol O. "Measurement of Children's Masculinity and Femininity." Ph.D. dissertation, Northwestern University, 1974.

Klecka, William R. Discriminant Analysis. Beverly Hills: Sage Publications, 1980.

Kline, Donald F. "Effect of Debate Skills on Delivery." The Gavel (March 1953): 59-60.

Lehman, Donald R. Market Research and Analysis. Illinois: Richard D. Irwin, 1979.

McCroskey, James C. "Special Reports: Scales for Measurement of Ethos." Speech Monographs 33 (March 1966): 65-72.

McCroskey, James C., and Camp, Leon R. "A Study of Stock Issues, Judging Criteria, and Decisions in Debate." The Southern Speech Journal 30 (Winter 1964): 158-68.

McCroskey, James C., and Camp, Leon R. "Judging Criteria and Judges' Bias in Debate." The Journal of the American Forensic Association 3 (May 1966): 59-62.

McGlone, Edward L. "The Behavioral Effects of Forensics Participation." The Journal of the American Forensic Association 10 (Winter 1974): 140-46.

Nie, Norman H.; Hull, C. Hablai; Jenkins, Jean G.; Steinbrenner, Karen; and Bent, Dale H. Statistical Package for the Social Sciences. New York: McGraw-Hill, 1975.

Nunnally, Jum C. Psychometric Theory. New York: McGraw-Hill, 1967.

Olson, Donald O. "A Survey of Attitudes on the Spread." Speaker and Gavel (March 1971): 66-9.

Osgood, Charles E.; Suci, George J.; and Tannenbaum, Percy H. The Measurement of Meaning. Urbana: University of Illinois Press, 1961.

Rives, Stanley G. "1976 National Debate Tournament Final Debate: Should the Federal Government Adopt a Comprehensive Program to Control Land Use in the United States?" The Journal of the American Forensic Association 13 (Summer 1976): 1-50.

Roever, James and Giffin, Kim. "A Study of the Use of Judging Criteria in Tournament Debate." AFA Register 8 (1960): 12-4.

Stewart, John R., and Merchant, Jerrold J. "Perceived Differences Between Debaters and Non-Debaters." The Journal of the American Forensic Association 6 (Spring 1969): 67-72.

Swinney, James P. "The Relative Comprehension of Contemporary Tournament Debate Speeches." The Journal of the American Forensic Association 5 (Winter 1968): 16-20.

Vasilius, Janet M. "An Investigation of the Relationship Between Tournament Success and Rate, Evidence, and Jargon." M.A. thesis, University of Wyoming, 1977.

Vasilius, Janet M., and De Stephen, Dan. "An Investigation of the Relationship Between Debate Tournament Success and Rate, Evidence, and Jargon." The Journal of the American Forensic Association 15 (Spring 1979): 197-204.

Walwik, Theodore J. "Research in Forensics: An Overview." The Journal of the American Forensic Association 6 (Spring 1969): 43-8.

Wheeless, Lawrence R. "The Effects of Comprehension Loss on Persuasion." Speech Monographs 38 (November 1971): 327-30.

Williams, Frederick; Clark, Ruth Anne; and Wood, Barbara Sundene. "Studies in the Dimensionality of Debate Evaluation." The Journal of the American Forensic Association 3 (September 1966): 95-103.

Williams, Frederick; Webb, Sally Ann; and Clark, Ruth Anne. "Dimensions of Evaluation in High School Debate." Central States Speech Journal 27 (February 1966): 15-21.

Williams, Frederick, and Webb, Sally Ann. "Factors in Debate Evaluation: A Pilot Study." Central States Speech Journal 15 (May 1964): 126-28.

Wise, Charles N. "Relationship Among Certain Rating Judgments on the Form C Ballot." The Journal of the American Forensic Association 7 (Spring 1971): 305-8.