

Eastern Illinois University
The Keep

Masters Theses

Student Theses & Publications

1983

Variables That Affect Success in Debate

Karen S. Shelton

Eastern Illinois University

This research is a product of the graduate program in [Speech Communication](#) at Eastern Illinois University.

[Find out more](#) about the program.

Recommended Citation

Shelton, Karen S., "Variables That Affect Success in Debate" (1983). *Masters Theses*. 2895.

<https://thekeep.eiu.edu/theses/2895>

This is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Masters Theses by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.

THESIS REPRODUCTION CERTIFICATE

TO: Graduate Degree Candidates who have written formal theses.

SUBJECT: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

7/29/83

Date

Author

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because _____

Date

Author

Variables that Affect Success

in Debate

(TITLE)

BY

Karen S. Shelton

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Master of Arts

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

1983

YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

7/29/83

DATE

✓ / ADVISER

7-29-83

DATE

COMMITTEE MEMBER

DATE

COMMITTEE MEMBER

7/29/83

DATE

✓ DEPARTMENT CHAIRPERSON

ACKNOWLEDGMENTS

The author wishes to gratefully acknowledge the assistance of the members of her supervisory committee without whose help the thesis would not have been possible. Professors Douglas G. Bock and Beryl F. McClerren gave freely of their time and expertise which enabled her to complete the thesis.

A special thank-you must be given to her husband, Michael. His expertise, support, and love were also instrumental to the completion of the thesis.

TABLE OF CONTENTS

	Page
Acknowledgments	ii
List of Tables	iv
Abstract	v
Chapter I: Introduction and Statement of the Problem ,	1
Chapter II: Methodology.	27
Chapter III: Results	34
Chapter IV: Conclusions	42
Bibliography	51

LIST OF TABLES

Table I.	Discriminant Analysis for Dependent Measure of Gender	34
Table II.	Discriminant Analysis for Dependent Measure for Decision	35
Table III.	Tests for Significance of Discriminant Analysis for Dependent Measure of Gender	35
Table IV.	Tests for Significance of Discriminant Analysis for Dependent Measure of Decision	36
Table V.	Analysis of Variance on Win/Loss	38
Table VI.	Cell Breakdown for Decision Analysis of Variance	38
Table VII.	Analysis of Variance for Team Rating Points	39
Table VIII.	Cell Breakdown of Points Analysis of Variance	40

Abstract of Thesis Presented to the Graduate
Council of Eastern Illinois University in Partial
Fulfillment of the Requirements for the Degree of
Master of Arts

VARIABLES THAT AFFECT SUCCESS
IN DEBATE

by

Karen S. Shelton

July 1983

Adviser: Douglas G. Bock
Major Department: Speech Communication

This study examines the effect of the six traditional categories of debate evaluation on the A.F.A. Form C and Form W ballots upon win/loss and gender. It also examines the effect of nonperformance variables, such as proximity, gender of the debaters, and gender of the judge, upon the outcome of intercollegiate debates. The data were gathered from the Owen L. Coon Memorial Debate Tournament hosted by Northwestern University in February, 1983. In all, the data pool consisted of 42 debates. The data were submitted to analysis to the SAS computer program at Eastern Illinois University.

The results of the discriminant analysis indicated that winners and losers could be classified according to analysis, reasoning, evidence, organization, refutation, and delivery over 60% of the time. The results of a t-test

indicated that there was no significant difference, however, in relation to scoring on delivery between winners and losers.

The results of the discriminant analysis of the gender dependent measure indicates that the categories predict gender less than 60% of the time. The t-test showed there was no significant difference in regard to the six categories with the exception of delivery in which females scored higher.

An analysis of variance was performed upon the dependent measure decision. The results indicated that there is an interaction of composition of the team and proximity of the team to the judge that affects the outcome in inter-collegiate debate. The analysis of variance performed upon the dependent measure of team rating points found that there was an interaction of composition, proximity, and, in this instance, gender of the judge. This indicates that non-performance variables are a factor in the outcome of a debate, as well as the traditional categories of evaluation.

CHAPTER ONE

Introduction and Statement of the Problem

According to many, to paraphrase Quintilian, an outstanding intercollegiate debater is simply a "good man debating well." Unfortunately, little consensus has emerged as to what combination of elements constitute "debating well." There has been a longstanding controversy in the debate community surrounding this very question. Forensics scholars have long been uncertain as to whether specific, observable elements account for good debating and successful debating, or whether more abstract, non-performance factors play a more central role.

This work will further address the question of what variables affect success in debate. In this first chapter, there will be an introduction and a statement of the problem, as well as a discussion of the hypotheses formulated for study and the dependent and independent variables that are involved. Chapter two will concentrate upon the sampling procedures utilized, the tools involved, and the measurement of data. Chapter three will simply be a reporting of the results from the present study. Chapter four will incorporate a discussion of the results, the limitations of the present study and suggestions for further study.

Much of the intercollegiate debate community has, at least implicitly, endorsed a standard evaluation form which suggests that six factors are of the greatest impor-

tance in debate performance: delivery, reasoning, organization, analysis, refutation, and use of evidence. These six factors are included on the American Forensic Association's Form C debate ballots to facilitate uniform evaluation of debate speakers.

Previous literature has endorsed the validity of utilizing the six factors on the Form C ballot for evaluation. For instance, Professor Burgoon found that a "correlation analysis" computed among the six predictor variables and the criterion variable, "revealed that actually all of the six predictor variables by themselves were significantly related to percentage of wins."¹ She went on to note that "while organization and refutation emerged as being slightly more important, all six factors were relatively equal in their impact."²

Other scholars have also recognized the relevance of the six Form C factors. "The Williams, Clark, and Wood findings suggest that the traditional criteria have a major impact," although they do go on to note that "they are not independent."³ Professor Giffin of Kansas conducted a study which found elements very similar to these traditional six, as constituting the majority of evaluative criteria employed by debate judges. Giffin explained the results in this fashion:

. . . the criteria employed in each debate by each judge were tabulated; of the total consideration given to all different criteria it was found that the judges gave to each of the criteria included in our hypothesis the following weight or consideration:

1.	ability to speak well (delivery). . . .	14.65%
2.	selection of logically defensible arguments (case).	19.10
3.	support of arguments with information (evidence).	17.18
4.	perception of irrelevant or irrational arguments (refutation).	17.00
5.	phrasing of concepts clearly and concisely (language).	5.29
6.	ability to analyze the topic area (analysis).	14.78
7.	ability to organize ideas into a structured whole (organization)	$\frac{8.88}{96.88\%}$ ⁴

Whether each of the six traditional factors independently weigh upon a judge's evaluation and decision is still open to question. For example, Professor Wise has suggested that the "two most difficult skills in academic debate, as measured by mean scores, are 'analysis' and (the use of) 'evidence.'"⁵ The question of whether these two factors, or any of the other four, are actually more important, or whether they function synergistically would seem to warrant examination of each factor individually.

Gerald Sanders has operationally defined reasoning "as the process by which we infer a conclusion from premises."⁶ Although Sanders does not attempt to quantify the relative weight that reasoning plays in a debate judge's evaluation, he does note that one should "emphasize the importance of reasoning in argumentation and the part that it plays in a judge's decision."⁷

Other authorities have suggested that reasoning is at least as important as a debater's use of evidence. Professor Cathcart has noted:

. . . the speaker who skillfully incorporates into his own thinking the evidence gathered, and then weaves it smoothly into his speech, will be just as effective as, if not more so than, the speaker who stops to cite sources for all of his evidence, or the one who documents and qualifies each source.⁸

Again, reasoning is identified as important but the relative weight of such importance is still unclear.

One could surmise that reasoning would obviously be important as a debate skill, but the difficulty in attempting to independently measure its importance is equally obvious. The pervasive nature of reasoning in relation to debate may make it difficult to separate it from other factors.

The great majority of contemporary forensic literature seems to place little value on the independent worth of delivery. Indeed, the conclusion reached by Vasilius and DeStephen seems quite true: "In debate the attitude toward delivery is ambivalent."⁹ Indeed, they went on to note that the "overall lack of significance suggests that a variety of factors contribute to debate success of which delivery, at least in quantitative terms, may be of little importance."¹⁰ Sanders has concurred by noting: "The judge who uses argumentation and logic as his sole criteria for determining the winner of an academic debate sees debate as an intellectual contest with speech being only an incidental element."¹¹

There is actually a solid body of quantitative research which confirms the limited independent value that most debate judges and scholars assign to delivery. An analysis

of judging philosophy statements found that:

Only a few critics indicated they 'generally give low points to spread debaters.' So long as debaters met basic requirements for intelligibility, most participants tolerated this form of discourse, 'believing the ultimate value of competitive debate to be analysis and not oratory.'¹²

Similarly, delivery or "speaking ability" has been ranked extremely low in terms of its importance as an educational by-product of debate. Professor Pearce noted that: "A recent survey of attitudes toward forensics in the U. S. found that members of the American Forensic Association themselves ranked the development of speaking ability last in a list of educational objectives."¹³

Delivery appears to be one factor of evaluation that clearly weighs less heavily than others. The consensus seems to be that it is not sufficient alone to determine the outcome or total performance evaluation of a debate.

There is very little debate-specific literature in relation to the importance of organization. There is general literature concerning organization and speech communication. For example, Elaine Winkelman Butcher has observed:

Results of some previous experimental studies indicated that speech organization did not contribute to message comprehension. Other studies claimed that credibility was not impaired by disorganization and that disorganization did not affect attitude. On the other hand, the majority of the literature as well as speech textbooks acknowledge the importance of speech organization.¹⁴

However, Butcher has also noted that disorganization is not inherently negative or counterproductive. She noted:

Results confirmed the importance of message organization on comprehension, but not on knowledge in some cases. Further, disorganization is detrimental to credibility only on those factors of qualification and safety, but not on warmth. Finally, this study showed no effect of message disorganization on attitude toward the topic.¹⁵

The controversy over the importance of organization in relation to speech generally would seem to be relevant to debate as well. If judges are more concerned simply with the outcome of arguments, organization may not be key. However, good organization may very well affect the outcome of a given argument. Hence, the value of organizational ability as an independent factor in debate evaluation would appear to be open to question.

"Analysis is," according to Sanders, "the arriving at an understanding of the proposition and the discovering of the issues inherent therein."¹⁶ Newman has suggested that deliberative speakers, one would assume this could include the debater, "find that one of their most important tasks is analysis, or breaking a proposition down into its component parts."¹⁷ Professor Rieke has applied the concept more specifically to debaters by noting that "analysis involves essentially two processes: discovering what basic questions must be asked in considering the resolution; and discovering what basic lines of reasoning are appropriate in setting about to answer the questions."¹⁸

Analysis is another factor, like reasoning, that seems to be generally important, but very difficult to isolate and measure against other factors. Indeed, Professor Rieke's comment above clearly draws an interrelationship between analysis and reasoning, further complicating the situation.

Evidence and evidence usage appear to be factors that have stimulated a good deal of debate-related literature. "Evidence is," notes Sanders, "an indispensable element in good debating and the argumentation and logic judge treats it as such."¹⁹ In fact, a concern for evidence use is central to the selection of a debate resolution. Sanders, writing again, has noted: "One of the criteria used for choosing an intercollegiate topic is that adequate evidence should be available on both sides of the proposition."²⁰

According to William Dresser, "contemporary theorists generally agree that the use of carefully selected and tested evidence is important to the advocate . . ."²¹ There are many who feel this is particularly true for the debate advocate. "Championship level debaters," according to Benson, "not only use the greatest amount of evidence but also use a greater portion of their evidence to clash with their opponents by denying arguments or establishing counter contentions."²² Benson has quantified such usage levels: "The championship debaters, /operationally defined as those qualifying for elimination rounds at major tournaments/, use about 25% more evidence than the varsity level debaters, /operationally defined as those with one year or more expe-

rience/, and nearly 60% more evidence than novices."²³

Although "championship" level debaters tend to use more evidence and evidence usage is generally recognized as important, there is no firm consensus on its value or effect. "McCroskey's findings," for instance, "that evidence is the least valuable factor for immediate attitude change" obviously casts doubt upon the inherent value of evidence usage.²⁴ "In debate situations," according to Vasilius and DeStephen, "where the critic must render an immediate decision, the quantity of evidence may be unimportant or at least not as important as other factors."²⁵

Many feel that evidence is interrelated to other factors and debating skills. Some authors have suggested "that evidence is used to support arguments and cannot be considered separate from the arguments."²⁶ Professor Dresser has also suggested that evidence tends to work with, or aid other factors. He has reported that:

This study tends to support the position of those contemporary theorists who hold that the importance of carefully tested evidence in speech making lies not in its contribution to persuasiveness but in its usefulness in helping the speaker to explore his subject intelligently.²⁷

The bottomline of contemporary forensic research seems to be that the value of evidence usage is simply uncertain. Kathy Kellerman, of Northwestern University, summarized the situation rather succinctly:

In contrast to the teachings of most introductory communication courses, theoretical consensus and empirical validation of the usefulness of evidence to a speaker have yet to be established. Indeed, the plethora of empirical research on evidence has produced such inconsistent results that no coherent evidence in argument can be extracted.²⁸

Professor Sanders has defined the last of the six traditional standards in this way: "Refutation is considered to be the attempted destruction of the opponents' argumentation."²⁹ Sanders feels that refutation is one of the key elements that a judge considers in his evaluation of a debater. He has noted:

In this area of the debate, the judge is watchful for a debater's exposure of weaknesses in the opposing case. Such weaknesses could be questionable analysis and interpretation, flaws in evidence, fallacies in structure and argument, and inconsistencies and contradictions in argument.³⁰

There are others who have suggested that refutation is the single most important element for evaluation. "If any single measure could be applied to determine the potency of a debater," writes Professor Faules, "that measure would examine refutation skill."³¹

The results of actual debates seem to validate the relative importance of refutation. Faules noted that "winning debaters were scored superior more frequently for refutation than any other item. Such evidence indicates that refutation skill may be a predictor for debate effectiveness."³² Keeling also found that "the greatest difference in the scores of winning and losing debaters occurred in

the area of refutation. In addition, winning debaters were scored superior more frequently for refutation than any other item."³³

Despite evidence correlating debate success and high scores for refutation, there is still doubt as to whether it is refutation alone that actually accounts for this. In fact, Sanders has gone on to suggest that rebuttal may be equally or more important than simple refutation. He noted: "Rebuttal is the attempted rebuilding of an argument once it has been attacked. It does no good to refute an opponent's argumentation if your own case is in shambles."³⁴ Even Faules has suggested that refutation may be inherently dependent upon other factors. "The presentation," that is delivery, "of refutation will decide its potency."³⁵ He has also noted that the whole process of refutation is "dependent upon a student's ability to examine evidence, reasoning, and the relationship of evidence and inference."³⁶

Apparently, refutation is a critical element relating to debate success, but one dependent upon other factors as well. Faules, for example, has clearly drawn an interrelationship between reasoning, evidence, and refutation. Refutation may well be important, but absent its foundation in these other factors it may well be impotent.

The bulk of contemporary literature tends to endorse the six Form C evaluation factors as important, but it fails to distinguish any one as being uniquely important absent the other five. Indeed, Professor Burgoon has found that:

"Debaters who were rated high on any one dimension were consistently rated high on the other five."³⁷ Vasilius and DeStephen have also found a lack of independent criteria for debate evaluation. They have noted:

Research indicates that debate evaluation is multi-dimensional, that some evaluative dimensions are more important than others, and that the dimensions are not independent, despite "boxes" on a debate ballot indicating evaluative factors.³⁸

Burgoon and Montgomery have gone so far as to suggest that broader, general standards actually account for evaluation rather than the traditional six. They reported:

The collapse of previously discovered dimensions into three in this investigation is a significant finding. It implies that when respondents are asked to reveal their standards for evaluation rather than to rate actual people, a different judgmental structure appears. When evaluating actual people, it seems possible to distinguish among composure, sociability, and character attributes. However, when the ideal is to be rated, all of these attributes seem to be intertwined. The logical extension of this finding is that judges probably only evaluate debaters along these three general lines rather than making six independent judgments, as presumed by the old Form C ballots.³⁹

Hence, these general lines may be more important than the specific criteria suggested by the current debate ballots.

Many judges have taken the option of simply providing a total score for debate performance and ignoring the "boxes" occupied by the six traditional factors. In relation to such action, Professor Burgoon has written:

The failure of judges to discriminate among the six elements implies that either (1) they are only making a gross, global evaluation, (2) they are unable to translate their true evaluation criteria into marking behavior (which reduces the utility of the ballots as feedback to debaters), or (3) other factors are influencing their decisions.⁴⁰

The possibility of "other factors," perhaps nonperformance variables, affecting the outcome or evaluation of a debate is most pronounced. This is, of course, generally true in regard to speech evaluation as Larry Barker has noted:

The many uncontrollable variables present in the evaluation situation, coupled with different concepts of the ideal speech, compound the problem. Evaluations of communication behavior appear to be influenced by a combination of environmental, perceptual, and hereditary factors that influence human judgment.⁴¹

Such factors could obviously influence a judge-evaluator of a debate round.

Debate-specific studies have attempted to measure the effect of nonperformance variables on the outcome of debate rounds. Professor Wise has offered one example:

Although wins over a year's debating will be approximately equally divided, affirmative teams score higher on the average on the six scales than do negatives, particularly on "organization" and "delivery." The first affirmative rebuttal speech and the first negative constructive speeches are "crucial" speeches in a standard format debate.⁴²

The particular variables of "side" and "speaker position,"

however, do not appear to significantly affect the outcome of debates. Sidney Hill found "that the format variables 'side of topic' and 'speaker position' have no significant effect on the overall outcome of intercollegiate debates as measured by the dependent variable index of outcome."⁴³ Any effect associated with topic side would seem to simply reflect pure chance. Halstead concurred by noting:

These figures indicate, then, that there may be a slight advantage for one side on a specific debate question, but that there seems to be no particular advantage for Affirmative per se or Negative per se. Even this advantage may be pure chance, and it is so slight an advantage that it is not likely to influence the decision in a specific debate.⁴⁴

Two other nonperformance variables have produced more controversial findings as to their effect on intercollegiate debates. Those variables are proximity and gender (of debaters and of judges).

"Physical location alone," Brooks has noted, "exerts a powerful influence on amount of interaction. . . . The powerful, almost mechanical, effect of physical distance on friendship patterns is consistently documented."⁴⁵ Brooks has further explained that:

Both the conclusions of debaters and the conclusions of scholars studying debate judging indicate that debate decisions are based on something other than the criteria listed on debate ballots. Hidden criteria, sometimes suggested by debaters, are social distance and geographic distance.⁴⁶

Brooks further reported that "geographical distance was related to debate decisions in a manner not predicted by chance in five of the six tournaments" that he studied.⁴⁷

Hill has also examined the variable of geographical distance, or proximity. Hill noted: "Schools normally do a major portion of their season's debating within their National Debate Tournament district, thus potentially fostering 'friendship through propinquity . . .'"⁴⁸ Hill felt such influence was possibly overstated. He noted: "Because these district lines tend to represent natural lines of travel and traditional rivalries, the effects due to simple geographical proximity might well be over-ridden by the pressures of district loyalty."⁴⁹ Hill further noted that his "model indicated that, within any given N.D.T. district, proximity was a negative influence. Perhaps, in this case, proximity led to the growth of rivalries rather than friendships."⁵⁰

The variable of gender has inspired even greater controversy among forensic scholars. For example, Hayes and McAdoo have found gender to effect speaker rankings beyond simple chance. They reported:

The conclusion is that in debates involving at least one mixed team, the rankings received by both males and females systematically differ from those expected by chance. Under these conditions females receive more "one" and "three" rankings but fewer "twos" and "fours." At the same time males differ from chance in that they receive more "twos" and "fours" but fewer "ones" and "threes."⁵¹

It has further been suggested that gender can affect total outcome (win/loss), not only individual rankings. Rosen, et al found "there is no difference between male and female teams with regard to winning, but mixed teams are more likely to win."⁵²

Some authorities feel that the success of male-female teams actually reflects other factors at work. Hensley and Strother reported:

At least two reasons can be advanced for the advantage of the male-female teams. First, there may be instances when the respective styles of the male and female tend to complement each other better than if members of the same sex were debating as colleagues. Secondly, while in truth, there may be no difference in the abilities of the two sexes, coaches may be reluctant to pair a male and a female.⁵³

Hensley and Strother further suggest that single gender teams are neither more or less successful. The results of their study fails "to give any credence to the superiority of a team composed of two males or to the inferiority of a team composed of two females."⁵⁴ In fact, the success of single gender teams seems to reflect chance alone. Hensley and Strother noted: "By the laws of chance alone, debating teams can be expected to win 50% of their debates and, indeed, teams composed of two males or of two females have records which conform very closely to this expectation."⁵⁵

The gender of those evaluating speech acts may play some part in how those evaluations occur. This has been found to be generally true in the field of speech communi-

cation. According to Barker: "A meaningful relationship was found between instructor's speech ratings and the sex of the communicator."⁵⁶ In relation to debate, Hill found that "female debaters tended to be associated with lower team ratings than did male debaters. Conversely male judges tended to give lower team ratings than female judges."⁵⁷ Hill went on to explain the expected ratings involved in various situations:

This model indicates that the members of mixed teams received lower ratings than either all-male or all-female teams. Before a male judge, the expected speaker rating for the male member of a mixed team was 19.50, as compared to 22.80 for a male debater with a male colleague before a male judge. The expected rating was 19.12. When debating before a female judge, the female in a mixed team had an expected rating of 19.33.⁵⁸

Hill went even further to suggest that:

. . . for any given debate, then these results indicate that all-male teams had a greater expectation of winning before a male than before a female judge. Mixed teams and all-female teams, however, had an expected loss from male judges and an expected win from female judges.⁵⁹

Hence, gender of the judge in relation to gender of the debaters involved may well influence evaluations made by those judges.

The sheer inconsistency and uncertainty associated with the results of previous research would seem to warrant further study. More conclusive data would obviously help resolve the scholarly conflict surrounding the various

issues discussed here, especially given the lack of more recent research. A number of more concrete rationales may also be offered to warrant further investigation.

There is most basically a need for more empirical data relating to the whole of forensic activities. Hill has noted:

Only a small percentage of the research reported each year /in the American Forensics Association bibliographies/ is of a quantitative nature, and only a minor portion of that deals with inter-collegiate forensics.⁶⁰

Benson and Friedly have similarly noted:

Although the forensic community has informally acknowledged a professional responsibility to contribute knowledge by conducting scholarly research and formally acknowledged this commitment at the National Development Conference on Forensics at Sedalia, little research has actually been generated.⁶¹

The need for data specifically relevant to success in debate is even more pronounced than the general need for forensic-related empirical research. Burgoon and Montgomery have noted:

. . . the controversy over what constitutes superior debating has generated much speculation and prescription but very little empirical verification. Debaters and judges alike are still uncertain of the universal standards (if any exist) by which debaters are evaluated during debate competition.⁶²

Burgoon went on to be more specific in advocating further research:

. . . more research using multivariate techniques is needed to obtain a realistic assessment of what factors generate success in debate. Efforts should be made to combine the traditionally identified factors with such variables as geographic biases, sex, reputation of the team's school, source credibility, and refutation forms so that relative influence of each can be determined.⁶³

Williams and Webb have stated that "there is little research evidence that lends insight into the actual bases for judges' decisions."⁶⁴ This confirms what has been indicated all along: there is little knowledge as to what elements actually affect evaluation. The need for such information was underscored by Brooks, who reported:

An integral part of learning is evaluation and feedback. In the educational process we assume that evaluation is a rational act involving systematic analysis and judgment based on relevant criteria, and that the evaluation should be fed back to the learner so that appropriate understandings and behaviors are positively reinforced and erroneous understandings and behaviors are corrected.⁶⁵

The educational necessity for evaluative feedback was confirmed by Professor Burgoon:

Certainly if students are to learn what elements truly contribute to effective argumentation and specifically to successful intercollegiate debate, we must identify those factors that are relevant and those that deserve the most emphasis.⁶⁶

Verderber summarized the concept best by stating: "Intercollegiate debate should be an educational experience; anything that can be done to improve its value is worth the

time and effort."⁶⁷ Hence, if further study were to aid the evaluation and feedback process for debate it would be well worth the effort.

Further research may also aid in the overall process of training debaters. "Training procedures," noted Willmington, "varied widely, and it seemed that the type of training a coach gave to his debaters depended more upon his whims than upon any consensus as to good training procedure."⁶⁸ Whim would certainly seem to be an insufficient approach to debate and argumentation training. Further research into the variables affecting debate success would offer a more reasoned alternative to whim alone.

Hill has concluded that "judges simply don't check the boxes any more."⁶⁹ In other words, the traditional evaluation technique provided on Form C debate ballots is being increasingly ignored. There has even been movement toward abandonment of the "boxes" entirely. Whether this is a wise option or not is a question that also warrants additional research. The risks associated with an abandonment induced by insufficient research were explained by Burgoon:

Abandoning the Form C-type ballot, however, may mean losing valuable information about what factors in reality determine debate success. If, in fact, the six components of evidence, organization, reasoning, analysis, refutation, and delivery are critical factors, we need to know three things: how much of the success they actually account for, what the relative importance of each is, and how independent the judgments are.⁷⁰

Such action as an uniform change in ballot format should only be considered following thorough research and examination.

On a very pragmatic basis, the knowledge of what constitutes successful debating may be extremely important to the very existence of a debate program. In a period of budget-slashing and belt-tightening, few programs that cannot demonstrate their success and worth can avoid becoming the victims of such actions. Benson and Friedley note that "obtaining equitable funding and staff to coach . . . may be intrinsically tied to producing empirical data related to the activity's functions and claimed benefits."⁷¹ Hence, an understanding of what factors actually make up the "good man debating well" may be the key to survival of the debate process itself.

The by-product of further study should be significant to a wide audience in the forensics community. Forensics scholars should be able to benefit from the availability of more precise information relating to performance evaluation. Debate instructors and coaches should gain information that could be utilized in the establishment of training programs for their debaters. Debaters themselves could benefit from a more precise feedback process and a superior understanding of what they should strive for as practitioners.

Hypotheses

A review of the literature left several questions

unanswered. Such questions, if answered, would contribute to the theory and practice of forensics.

There is still controversy as to the usefulness of the boxes on the Form C and Form W ballots. Do these boxes represent the most desirable, independent trait to achieve success, or are all interrelated? And, is there a gender difference with respect to the categories? Does one sex excel in one category or are the sexes equal in their abilities? Although there has been no gender-specific research in relation to the six traditional categories in forensics, other research has shown that some differences do exist in these categories, which leads to two hypotheses:⁷²

- H1: One or more of the six traditional categories of evaluation account for success in debate.
- H2: Females will score lower in the six traditional categories of evaluation.

The six traditional categories of evaluation are defined by the A.F.A. Form C and Form W as delivery, reasoning, organization, analysis, refutation, and evidence. Success is defined as receiving the decision from the judge.

The date of previous studies on gender and proximity also raise questions as to the applicability of their results today. What impact, if any, does the gender of the judge have upon who wins the round? What impact, if any, does the gender of the debaters have upon their success? Do district biases exist? To answer these questions, four additional hypotheses were devised:

- H3: Mixed gender teams win significantly more rounds than same sex teams.
- H4: Mixed gender teams receive higher team rating points.
- H5: Gender of the judge determines the outcome of the debate.
- H6: District bias determines the outcome of the debate.

Team rating points are operationally defined as the total speaker points given to the first and second position speakers of a team for that round. The outcome of the debate is defined in terms of win/loss. District bias is a tendency to vote for the teams that are members of the judge's district as determined by A.F.A. codes.

Independent and Dependent Variables

Hypotheses one and two share the same independent variables: delivery, reasoning, organization, analysis, refutation, and evidence. The dependent variables are decision for hypothesis one and gender for hypothesis two. The discriminant analysis reveals the order of importance that will predict winner and losers; and, the order of importance that will predict maleness and femaleness.

The independent variables for hypotheses three through six are the gender composition of the teams, gender of the judge, and district bias (proximity). The dependent variables are the decision rendered and team rating points.

Endnotes

- 1 Judee K. Burgoon, "Evaluation Criteria as Predictors of Debate Success," Journal of the American Forensic Association, 12 (Summer, 1975), p. 3.
- 2 Ibid., pp. 3-4.
- 3 Ibid., p. 2.
- 4 Kim Giffen, "A Study of Criteria Employed by Tournament Debate Judges," Speech Monographs, 26 (March, 1959), p. 70.
- 5 Charles N. Wise, "Relationships Among Certain Rating Judgments on the Form C Ballot," Journal of the American Forensic Association, 7 (Spring, 1971), p. 308.
- 6 Gerald H. Sanders, "Debate as a Paradigm for Demonstrating Skills in Argumentation and Logic," Paper presented at the Annual Meeting of the SCA (60th, Chicago, December 26-29, 1974), p. 11.
- 7 Ibid., p. 11.
- 8 Robert Cathcart, "An Experimental Study of the Relative Effectiveness of Four Methods of Presenting Evidence," Speech Monographs, 22 (August, 1955), p. 233.
- 9 Janet M. Vasilius and Dan DeStephen, "An Investigation of the Relationship Between Debate Tournament Success and Rate, Evidence, and Jargon," Journal of the American Forensic Association, 15 (Spring, 1979), p. 197.
- 10 Ibid., p. 203.
- 11 Sanders, p. 4.
- 12 J. Robert Cox, "A Study of Judging Philosophies of the Participants of the National Debate Tournaments," Robert J. Branham, ed., The New Debate: Readings in Contemporary Debate Theory (Washington, D.C.: Information Research Associates, 1975), p. 22.
- 13 W. Barnett Pearce, "Attitudes towards Forensics," Journal of the American Forensic Association, 10 (Winter, 1974), p. 136.
- 14 Elaine Winkelman Butcher, "The Effects of Speech Disorganization Upon Information Retention, Credibility, and Attitude," Dissertation Abstracts, 40 (December, 1979), p. 2980-A.

- 15 Ibid., p. 2981-A.
- 16 Sanders, p. 6.
- 17 Robert P. Newman, "Analysis and Issues--A Study of Doctrine," Central States Speech Journal, 13 (Autumn, 1961), p. 43.
- 18 Richard D. Rieke, "Analysis in Debate," Don F. Faules and Richard D. Rieke, eds., Directing Forensics: Debate and Contest Speaking (Scranton, PA: International Textbook Co., 1968), p. 122.
- 19 Sanders, p. 11.
- 20 Ibid., p. 10.
- 21 William R. Dresser, "Effects of 'Satisfactory' and 'Unsatisfactory' Evidence in a Speech of Advocacy," Communication Monographs, 30 (August, 1963), p. 302.
- 22 James A. Benson, "The Use of Evidence in Intercollegiate Debate," Journal of the American Forensic Association, 7 (Spring, 1971), p. 264.
- 23 Ibid., p. 262.
- 24 Vasilius and DeStephen, p. 203.
- 25 Ibid., p. 203.
- 26 Ibid., p. 202.
- 27 Dresser, p. 306.
- 28 Kathy Kellerman, "The Concepts of Evidence: A Critical Review," Journal of the American Forensic Association, 16 (Winter, 1980), p. 159.
- 29 Sanders, p. 13.
- 30 Ibid., p. 13.
- 31 Don F. Faules, "Refutation in Debate," in Don F. Faules and Richard D. Rieke, eds., Directing Forensics: Debate and Contest Speaking (Scranton, PA: International Textbook Co., 1968), p. 190.
- 32 Ibid., p. 47.
- 33 Russell M. Keeling, "An Analysis of Refutation and Rebuttal in Interscholastic Debate," as quoted in Directing Forensics: Debate and Contest Speaking, p. 190.

- 34 Sanders, p. 13.
- 35 Faules, p. 149.
- 36 Ibid., p. 191.
- 37 Burgoon, p. 4.
- 38 Vasilius and DeStephen, p. 198.
- 39 Judee Burgoon and Charles Montgomery, "Dimensions of Credibility for the Ideal Debater," Journal of the American Forensic Association, 12 (Spring, 1976), pp. 175-6.
- 40 Burgoon, p. 4.
- 41 Larry L. Barker, "Irrelevant Factors and Speech Evaluation," Southern Speech Journal, 32 (Fall, 1966), p. 10.
- 42 Wise, p. 308.
- 43 Sidney R. Hill, "A Study of the Effect of Non-Ability Variables on the Outcome of Intercollegiate Debates," Diss. University of Florida, 1973, p. 65.
- 44 William F. Halstead, "Who Wins Debates? A Statistical Study of 1320 Debates," Quarterly Journal of Speech, 26 (April, 1940), pp. 214-5.
- 45 William D. Brooks, "Judging Bias in Intercollegiate Debate," Journal of the American Forensic Association, 7 (Winter, 1971), p. 198.
- 46 Ibid., p. 198.
- 47 Ibid., p. 199.
- 48 Hill, p. 9.
- 49 Ibid., p. 18.
- 50 Ibid., p. 77.
- 51 Michael T. Hayes and Joe McAdoo, "Debate Performance: Differences Between Male and Female Rankings," Journal of the American Forensic Association, 8 (Fall, 1971), p. 236.
- 52 Nathan Rosen, Larry Dean, and Frank Willis, "The Outcome of Debate in Relation to Gender, Side, and Position," Journal of the American Forensic Association, 15 (Summer, 1978), p. 21.

- 53 Wayne E. Hensley and David B. Strother, "Success in Debate," The Speech Teacher, 17 (September, 1968), p. 236.
- 54 Ibid., p. 236.
- 55 Ibid., p. 236.
- 56 Barker, p. 14.
- 57 Hill, p. 67.
- 58 Ibid., p. 67.
- 59 Ibid., p. 67.
- 60 Ibid., p. 3.
- 61 James A. Benson and Sheryl A. Friedly, "An Empirical Analysis of Evaluation Criteria for Persuasive Speaking," Journal of the American Forensic Association, 19 (Spring, 1982), p. 1.
- 62 Burgoon and Montgomery, p. 171.
- 63 Burgoon, p. 4.
- 64 Frederick Williams and Sally Ann Webb, "Factors in Debate Evaluation: A Pilot Study," Central States Speech Journal, 15 (May, 1964), p. 126.
- 65 Brooks, p. 197.
- 66 Burgoon, p. 2.
- 67 Rudolph Verderber, "Judges Criteria and Debate Adaptation: Empirical Evidence," Journal of the American Forensic Association, 5 (Winter, 1968), p. 30.
- 68 S. Clay Willmington, "A Study of the Relationship of Selected Factors to Debate Effectiveness," Central States Speech Journal, 20 (Spring, 1969), p. 3.
- 69 Hill, p. 213.
- 70 Burgoon, p. 2.
- 71 Benson and Friedly, p. 1.
- 72 Eleanor Emmons Maccoby & Carol Nagy Jacklin, The Psychology of Sex Differences (Stanford: Stanford University Press, 1974), pp. 75, 98-9, and 108. Results of studies have shown females superior on verbal tasks, males superior on analytic behavior; & no difference in reasoning.

CHAPTER TWO

Sampling Procedures

The sample was drawn from rounds five and eight of the Owen L. Coon Memorial Debate Tournament sponsored by Northwestern University in February, 1983, for intercollegiate debaters. This provided a possible sample of 82 debate rounds. Ballots from 42 rounds were returned and, thus, used as the basis for the study. This yielded a return rate of 51% which Kerlinger has indicated is common for this type of procedure.¹

The Owen L. Coon Memorial Debate Tournament was chosen because it is one of the largest national tournaments in the country. As such, it provided a fairly representative sample of debate teams competing throughout the country during the year. There were nine teams from District I; three teams from District II; fourteen from District III; fifteen teams from District IV; eighteen teams from District V; thirteen from District VI; nine from District VII; ten teams from District VIII; and three teams from District IX.

The gender composition of the teams was similar to previous tournaments and previous years. Data was obtained from 50 all-male teams, 32 mixed-gender teams, and two all-female teams. Some teams were included twice as in the case of the all-female team; so that there was actually only one all-female team participating in the tournament.

The ratio of male judges to female judges was similar to the ratio of the participants in debate. Thirty six

male judges filled out the ballot, while only six female judges did so. This was not due to the females' lack of concern but simply due to the smaller proportion of females that hold judging (coaching) positions in academic debate.

Tools

Data was obtained from an A.F.A. Form "W" ballot. Examination of the data began with a study of analysis, reasoning, evidence, organization, refutation, delivery, the gender of the participants, the gender of the judge, win-loss decision, team speaker points, and where the participants and judge were from.

The Form W ballot contains a grid that measures the effectiveness of the speaker according to the six categories. The judge evaluates the effectiveness by checking a rating box in the grid: one for poor, two-fair, three-average, four-excellent, and five-superior. This provided the measurement of the effectiveness of the speakers in each of the six categories.

The Form W ballot was chosen as it is a "short form" of the A.F.A. Form C ballot. The Form W ballot utilized the traditional dimensions of evaluation: analysis, reasoning, evidence, organization, refutation, and delivery. "Theoretically, the criteria listed on the debate ballots represent the objectives of debate instruction."²

A.F.A. Form C ballots have been consistently used in many studies.³ As a result, the ballots have been deter-

mined to be useful in the evaluation of debate performance.

Dr. Sanders indicated that:

It is granted that there are other elements of a debate that are considered by a judge. These elements include prima facie case, inherency, presumption, and burden of proof. However, I submit that when these elements are evaluated, the evaluation takes place in terms of . . . /analysis, investigation, evidence, reasoning, refutation, and rebuttal/.⁴

Use of the ballot has generated much useful data:

Widespread usage of the Form C has generated a large amount of data useful in two inter-related fashions: first, as a source of individual feedback for debaters and coaches, and secondly as a conglomerate from which broad trends and patterns may be deduced.⁵

Although the grid provided on these ballots has not been widely utilized, the data generated from this study should yield the same results.

Measurement

Rounds five and eight of the Owen L. Coon Memorial Debate Tournament sponsored by Northwestern University were chosen as the sample for the study. These rounds were chosen due to the employment of power-matching and ease for the tournament hosts. A power-matched round reduced the possibility of prestige inequities between the teams which might influence the judge's decision.

A letter and an A.F.A. Form W ballot were passed out before the beginning of rounds five and eight. The letter

explained the purpose of the study and the directions necessary to fill out the ballot.

The judges were first asked to fill out the ballot putting the first and last names of all participants, including their own first and last names. This was done to determine the gender of all involved without sensitizing the judge as to the purpose. In order to insure that the gender suggested by the names was accurate, volunteers checked the rounds for confirmation.

The judges were then instructed to fill in the grids for each participant after the round was over. Operational definitions for each category of the grid were provided to enhance the uniformity of the evaluations. They were:

- Analysis: to separate into issues or basic principles so as to determine the nature of the proposition.
- Reasoning: the drawing of valid conclusions or inferences from observation, facts, or hypotheses.
- Evidence: the proper use of supporting material that justifies the acceptance or rejection of a claim.
- Organization: the clear arrangement or systematization of the arguments.
- Delivery: manner of speaking. Includes physical behavior, vocal variety, rate variation, enunciation, and fluency.
- Refutation: valid attacks on the oppositions' claims.

The judges returned the completed Form W study ballots with the actual tournament ballots to "ballot collectors" assigned by the tournament director.

To test hypotheses one and two, discriminant analysis

was used. Discriminant analysis was chosen due to its unique suitability. "The function maximally discriminates the members of the group; it tells us to which group each member belongs."⁶ Such analysis would reveal which category of evaluation winners and losers belong and which categories are characteristic of males and females.

Data concerning the effectiveness of each speaker in relation to each category was gathered in terms of whether the speaker won or lost and whether the speaker was male or female. This data was then fed into an equation.

A Factorial Analysis of Variance was performed to test hypotheses three, four, five, and six.⁷ This was done due to the examination of the same three independent variables but different dependent variables: success in terms of win/loss and team speaker points. The analysis was a 2x2x3 design.

The independent variables were gender of the judge, gender make-up of the team, and proximity. The gender of the judge was, of course, subdivided into male and female. The gender make-up of the team was subdivided into all-male and mixed-gender teams. This was due to the inadequate number of all-female teams in this sample and in debate in general. Hence, the ballots involving that one team were excluded from this analysis.

Proximity was determined by district affiliation. Three subdivisions were formed: teams from the same district as the judge, teams from a contiguous district, and others.

It was thought that those teams in the "others" category would have no advantage over other teams in that category in regards to proximity.

Endnotes

¹ Kerlinger is referring to mail questionnaires, of which this method is quite similar to. He indicates: ". . . the researcher must content himself with returns as low as 50 to 60%." Fred Kerlinger, Foundations of Behavioral Research (New York: Holt, Rinehart & Winston, Inc., 1973), p. 414.

² Brooks, p. 197.

³ Some studies that have used the A.F.A. Form C ballot are reported in Wise, pp. 305-308 and Burgoon, pp. 1-4.

⁴ Sanders, p. 5.

⁵ Wise, p. 305.

⁶ Kerlinger, p. 650.

⁷ For an understanding of the applicability and advantages to this approach, see Kerlinger, pp. 255-259.

CHAPTER THREE

Results

The six traditional categories of evaluation were tested first, using discriminant analysis. For gender of team members the categories do not have much predictive potential. The scores will predict males 60.15% of the time and females 54.88% of the time.

TABLE I

Discriminant Analysis for Gender

Gender	Female	Male	Total
Female	18 54.55%	15 45.45%	33 100%
Male	53 39.85%	80 60.15%	133 100%
Total Percent	71 42.77%	95 57.23%	166 100%
Priors	.50	.50	

The scores of winners and losers have more predictive capability. The scores will predict winners 67% of the time and losers 63% of the time. Table II illustrates this percentage breakdown.

TABLE II

Discriminant Analysis for Decision

Decision	Winners	Losers	Total
Winners	52 62.65%	31 37.35%	83 100%
Losers	27 32.53%	56 67.47%	83 100%
Total Percent	79 47.59%	87 52.41%	166 100%
Priors	.50	.50	

To more fully confirm or reject hypotheses one and two, a t-test was performed between the six traditional categories of evaluation for gender and decision. The results indicated there was no significant difference in categories with the exception of delivery when the dependent variable gender was considered. In this instance, females scored significantly higher, inferring that they have significantly better delivery.

TABLE III

t-test for Gender Discriminant Analysis

Category	Male Mean	Female Mean	t-test	Significance
Analysis	4.23	4.09	.88	p < .05
Reasoning	4.17	3.97	1.25	p < .05
Evidence	4.06	4.03	.19	p < .05
Organization	4.04	4.15	.73	p < .05
Refutation	3.96	4.03	.41	p < .05
Delivery	4.18	4.42	1.71	p > .05

The results of the t-test of the decision dependent variable produced results to the contrary. Five categories--analysis, reasoning, evidence, organization, and refutation--were shown to be significant at the .05 level. Whereas delivery was shown not to be significant in determining the winners and losers.

TABLE IV

t-test for Decision Discriminant Analysis

Category	Winners Mean	Losers Mean	t-test	Significance
Analysis	4.43	3.96	3.92	$p > .05$
Reasoning	4.36	3.9	3.54	$p > .05$
Evidence	4.17	3.94	1.77	$p > .05$
Organization	4.17	3.95	1.83	$p > .05$
Refutation	4.2	3.75	3.46	$p > .05$
Delivery	4.19	4.27	.73	$p < .05$

Additionally, the results indicate only a very minimal difference in the two mechanics-oriented categories, evidence and organization. The t-test results, 1.77 for evidence and 1.83 for organization, were just inside the range of significant difference. The thought-oriented categories--analysis, reasoning, and refutation--demonstrated a far larger range of difference between winning debaters and losing debaters.

Based on these results, one can conclude that hypothesis one has been confirmed. The six traditional categories do account for success in debate. However, delivery has been shown not to be a significant factor.

Hypothesis two must be rejected. Females do not score significantly lower on the six traditional categories of evaluation, therefore, the predictive value of gender is somewhat limited. Further, females actually scored significantly higher on delivery.

The data from a factorial analysis of variance was used to reject or confirm hypotheses three through six. A tabulation of the results of the decision dependent variable and team rating points dependent variable display the effect of proximity, gender composition of the teams, and gender of the judge. These results can be found in Table V.

The results indicate that proximity, composition, and the interaction of proximity and composition had a significant effect. However, the gender of the judge did not. Therefore, the interaction of composition and proximity has a strong impact upon the decision, especially given the 16% R-square of variance. See Table 6 for further breakdowns.

The male/female team from the "other" district won significantly less rounds at $p > .05$ than the other three indicated by the asterisk. The all-male team tends to win more often except for those in contiguous districts.

TABLE V
Analysis of Variance for Decision

Source	DF	SS	Mean Square	F	PR F	R ²
Model	10	6.52	0.652	2.90	.0024	0.16
Error	151	33.98	0.225			
Corrected Total	161	40.50				

Independent Variables	DF	ANOVA SS	F-Value	PR F
Composition	1	1.67	7.43	.0072
Proximity	2	1.04	2.31	.1029
Composition/Proximity	2	3.05	6.77	.0015
Judge	1	0.05	.22	.6417
Composition/Judge	1	.00	.00	1.0000
Proximity/Judge	2	.32	.71	.49
Composition/Proximity/ Judge	1	.40	1.80	.18

TABLE VI

Cell Breakdown for Decision Analysis of Variance

Composition	Proximity	N	Mean
Male/Male	Same District	12	.33
Male/Male	Contiguous	46	.48*
Male/Male	Other	42	.76*
Male/Female	Same District	4	.25
Male/Female	Contiguous	42	.48*
Male/Female	Other	16	.13

*Tested to be significantly different from male/female, other district teams.

The analysis of variance performed for the dependent variable of points indicates that there is a significant interaction between the gender composition of the teams, the proximity, and the gender of the judge. Again, the R-square of variance is sufficiently low as to warrant consideration of these results.

TABLE VII

Analysis of Variance for Team Rating Points

Source	DF	SS	Mean Square	F	PR F	R ²
Model	10	531.79	53.18	1.32	.22	.08
Error	151	6066.7	40.18			
Corrected Total	161	6598.49				

Independent Variables	DF	ANOVA SS	F-Value	PR F
Composition	1	40.53	1.01	0.32
Proximity	2	158.68	1.97	0.14
Composition/Proximity	2	0.00	0.00	1.00
Judge	1	27.83	.69	0.41
Composition/Judge	1	73.44	1.83	0.18
Proximity/Judge	2	20.75	.26	0.77
Composition/Proximity/ Judge	1	233.73	5.82	0.02

A cell breakdown displays the interaction effect of the three variables in determining points in Table 8.

TABLE VIII

Cell Breakdown of Points Analysis of Variance

Composition	Proximity	Judge	N	Points Mean
Male/Male	Same	Male	12	50.08
Male/Male	Contiguous	Female	10	50.6
Male/Male	Contiguous	Male	36	48.39
Male/Male	Other	Female	6	54.33*
Male/Male	Other	Male	36	50.42
Male/Female	Same	Female	2	51
Male/Female	Same	Male	2	49
Male/Female	Contiguous	Female	4	49.5
Male/Female	Contiguous	Male	38	48.21
Male/Female	Other	Female	2	40
Male/Female	Other	Male	14	51.43*

* indicates those means compared and found significantly different from the lowest mean.

The highest points were received by all-male teams from the "other" district with a female judge. The female judges were also responsible for the lowest points received. The mixed gender team from the "other" district with a female judge received the lowest points. The male judges gave the second highest point total to a mixed-gender team from the "other" district.

Based on these results several conclusions concerning hypotheses three through six can be made. Hypothesis three is rejected. Mixed-gender teams do not win significantly more rounds than same sex teams. There is no difference in win/loss percentage based on that variable alone.

Hypothesis four is partially rejected. Mixed-gender teams receive higher points from male judges but lower points from female judges. Therefore, one cannot

blanketly say that mixed-gender teams will receive higher speaker points.

Hypothesis five is also rejected. Although gender of the judge interacts with proximity and composition in determining speaker points, it is not shown to be significant when win/loss is considered.

Hypothesis six is also rejected. There was no tendency of teams from the same district as the judge to win more often than those of contiguous or other districts. Again, an interaction between proximity and team composition is more important.

CHAPTER FOUR

Discussion

The confirmation of hypothesis one indicates that the traditional categories for debater evaluation, with the exception of delivery, are valid as evaluative factors. Although the analysis of variables shows other factors to play a part in the evaluation of debaters, five of the six traditional categories still represent important elements in judge evaluation.

The interrelatedness of the variables supports the findings of Burgoon and Williams, Clark, and Wood. It does partially refute Vasilius' and DeStephen's conclusion that there is no independent criteria for debate evaluation. Since delivery differences are not statistically significant, they may not play a part in debate evaluation. However the statistical differences of the other five categories does provide some evidence of the interrelatedness of the variables or an instance of the halo effect on content-related variables.

Confirmation of hypothesis one would seem to have clear implications for the debate coach or forensics educator. It would appear prudent to concentrate training and skills enhancement on the first five categories, excluding or minimizing delivery.

There could very well be implications for scholars and educators outside the forensics community as well. For

those involved in the analysis of rhetoric and public address the confirmation of hypothesis one may also suggest that non-delivery factors are more important in all speech performance situations.

The rejection of hypothesis two means that there is essentially no difference in the performance capabilities of males or females in debate. The only difference noted was in relation to delivery which has been demonstrated to be unimportant to final outcome.

This change could be due to the changing social conditions and the resultant effect it may have towards the attitudes males and females have about their capabilities. Society may have made strides towards reducing the non-physical differences between the sexes.

Given the ability of females to perform as well as males, debate coaches should have fewer worries about pairing female debaters with male colleagues. In other words, gender alone should not dictate team pairings, and it should not significantly affect team success.

Hypotheses three, four, and five were all partially rejected. In each instance, gender was shown not to play a significant, independent role. Hence, there appears to be neither an inherent superiority or inferiority of teams based upon the gender make-up of the teams or the gender of the judge. Other factors, such as proximity, come into play. For instance, given control for proximity, all-male teams tend to win significantly more debates with a female

judge. That result, however, does not arise from gender alone, but from the interaction of gender and proximity.

The rejection of the independent effect of gender composition of the teams refutes the findings of Rosen et al who found that mixed teams were more likely to win; and, supports Hensley's and Strothers' findings of no difference between male and female teams. These studies, however, did not consider the impact of proximity and gender of the judge.

When considering the influence of proximity and gender of the judge in terms of decision and team rating points, the results tend to refute the findings of Hill and Barker. The change in the results could reflect the small sample or changes in the attitudes of those participating in inter-collegiate forensics.

Again, the partial rejection of each of these hypotheses would suggest that gender alone should not play a significant part in a coach's team pairing decisions. Gender only relates to win-loss when combined with other factors which are beyond the control of debate coaches, meaning that one should not isolate gender as a decision basis for pairings.

It is worth noting that the gender of the judge does appear to have some relevance to point assignment, although win-loss decision remains unaffected. Hence, judges should be aware of such an influence and work to divorce themselves from it as best as possible when assigning points.

Hypothesis six was also rejected. There is apparently no district bias. Indeed, all-male teams from the "others" category had a higher percentage win-loss record than either teams from the same or contiguous districts. These results support Hill's findings and refute Brooks'.

As previously noted, proximity appears to interact with other variables, such as gender. Such interaction is obviously beyond the control of a debate coach. Judges and tournament directors could, however, work to enhance the equity of judging if they are made aware of such a relationship.

Limitations

As with most quantitative research, there are limitations associated with the present study. Such limitations relate to sample size, uncontrolled variables, and other factors as well.

The most notable limitation is that of sample size. Only one tournament was employed. Only two debate rounds from that tournament were examined. Such a sampling limitation is not inherently destructive. As noted previously, the tournament that was employed appears fairly representative of the debate community as a whole.

Some may charge that one tournament may simply reflect an atypical or isolated good or bad performance. Although of some validity, there is probably no satisfactory way to avoid such a problem. A larger problem--one that is most

likely insoluble--would become one of trying to magically decide how many tournaments would be sufficient to avoid this change. A team could very well have an unusually good semester or even an unusually good season. Thus, there will likely be some atypical aspect of any performance evaluation.

Also related to sampling was the insufficient number of all-female teams in the study. Although the situation may simply reflect the demographics of intercollegiate debate, it would probably be desirable to include as many all-female teams as possible.

Another sampling limitation would be the limited number of female judges involved in the study. Such a limitation, however, seems largely dependent upon the demographics of debate and subject to little adjustment.

Several other authors have implicitly suggested that a halo effect surrounds the awarding of speaker points on a debate ballot. That is, a speaker who performs very well according to one category also tends to receive high ratings in the other five categories. Given that there are differences demonstrated in relation to the categorical ratings, the halo effect would not seem as pronounced as has been suggested.

The return rate is another problem area. Only about 51 percent of the judges involved actually responded. This should not, however, actually mar the value of the study. So long as the results are representative, such a return

rate should be sufficient.

Variables that were not accounted for may well have influenced the results of the study. Although the use of power-matching should have mitigated the impacts of prestige, it still could have influenced the outcome of some of the debates. However, other than power-matching there appears to be little practical way of controlling for prestige given the subjective nature of the variable.

Non-proximity based friendship between a judge and any given team is another variable that could mitigate against the meaningfulness of the present study. Friendships could develop out of a number of other circumstances that are too numerous to mention, no less test for. Indeed, the total concept of friendship may be beyond the scope of any practical examination instrument.

Although operational definitions were provided for each of the six categories on the debate ballot, there was still ample opportunity for discretion and subjectivity. There is probably a need for greater "fine tuning" of the six categories. For instance, testing different aspects of refutation may be more revealing than examination of a debater's intoto refutation capabilities. Obviously, refutation and each of the other five categories could be divided into types or sub-categories for more precise evaluation.

Further Research

The present study would certainly seem to fulfill the

heuristic duty of research by suggesting several areas for further investigation. Such areas include changes in procedure, approach, and design.

A broad, national survey of coaches and debaters may be in order to ascertain an explanation for the apparent low participation rate of females in intercollegiate debate. Such a survey could shed light on the reasons for the small number of all-female teams and generally low participation rate.

As noted previously, the six traditional categories for debater evaluation still leave much room for judge discretion. A study employing more "finely tuned" categories--different types, sub-categories, etc . . . --might demonstrate why one category taken as whole is more or less important.

Professor Wise has suggested that position can make a difference in speaker evaluation. Given such a possibility, it would be desirable to test for the effect of gender in relation to each of the four speaker positions. Such data might prove useful to coaches when making team assignments.

Further research may also be warranted in relation to team point totals, gender, and decision. For example, one could determine whether there is a difference in the points received by winning female speakers and those received by winning male speakers. In other words, one could investigate the effect or influence of gender at the point comparison level in relation to success.

The whole notion of "point inflation" may be a particularly attractive area of study. Many judges may give a total point score without "checking the boxes" simply because they feel that they would be forced to assign a point total lower than what is normal or expected. Somehow testing the difference between preferred and actual points for a debate would be useful along these lines. Such material may help judges reevaluate their standards for awarding speaker points, or may contribute to a change in the evaluation format.

A longitudinal study, based upon sex differences, would be another area for further study. For example, one could devise a means to measure the progression of teams from their attendance at the Novice National Tournament as freshmen, on through to their attendance at other national tournaments during their career. Such a study would provide not only a long term comparison of the sexes, but other valuable data as well. Such a study could tell educators and coaches much about a student's progress and the factors that influence it.

Another study that would be particularly valuable for educators and coaches would be one that tested different training approaches. For instance, the differences between an instruction-oriented approach (attendance of workshops, classroom training, etc . . .) and an experience-oriented approach (experience in actual debates) might be highly valuable to the educator and coach. How those approaches

affect evaluation as measured by the six traditional categories may be suggestive of the appropriateness of one technique or another.

Bibliography

- Allen, R. R., "The Effects of Interpersonal and Concept Compatibility on the Encoding Behavior and Achievement of Debate Teams," Central States Speech Journal, 14 (February, 1963), pp. 23-26.
- Barker, Larry L., "Irrelevant Factors and Speech Evaluation," Southern Speech Journal, 32 (Fall, 1966), pp. 10-18.
- Benson, James A., "The Use of Evidence in Intercollegiate Debate," Journal of the American Forensic Association, 7 (Spring, 1971), pp. 260-270.
- Benson, James A. and Friedly, Sheryl A., "An Empirical Analysis of Evaluation Criteria for Persuasive Speaking," Journal of the American Forensic Association, 19 (Spring, 1982), pp. 1-13.
- Berryman, Cynthia L., "Perceptions of Male and Female Sex-Appropriate and Sex-Inappropriate Language," Dissertation Abstracts, 39 (January, 1979), p. 3915-A.
- Berthold, Carol A., "A Descriptive Study of Selected Characteristics of Debate Judge's Ballots," Journal of the American Forensic Association, 7 (Winter, 1970), pp. 30-35.
- Bock, Douglas G. and Bock, E. Hope, "The Effects of Sex on the Experimenter, Expectancy Inductions, and Sex of the Rater on Leniency, Halo, and Trait Errors in Speech Rating Behavior," Communication Education, 26 (November, 1977), pp. 298-306.
- Bridges, Dick A. and Reinhard, Jr., John C., "The Effects of Refutational Techniques on Attitude Change," Journal of the American Forensic Association, 10 (Spring, 1974), pp. 203-212.
- Brooks, William D., "Judging Bias in Intercollegiate Debate," Journal of the American Forensic Association, 7 (Winter, 1971), pp. 197-200.
- Burgoon, Judee K., "Evaluation Criteria as Predictors of Debate Success," Journal of the American Forensic Association, 12 (Summer, 1975), pp. 1-4.

- Burgoon, Judee K. and Montgomery, Charles, "Dimensions of Credibility for the Ideal Debater," Journal of the American Forensic Association, 12 (Spring, 1976), pp. 171-177.
- Butcher, Elaine Winkelman, "The Effects of Speech Disorganization Upon Information Retention, Credibility and Attitude," Disseration Abstracts, 40 (December, 1979), pp. 2980-A--2981-A.
- Cathcart, Robert, "An Experimental Study of the Relative Effectiveness of Four Methods of Presenting Evidence," Speech Monographs, 22 (August, 1955), pp. 227-233.
- Cox, J. Robert, "A Study of Judging Philosophies of the Participants of the National Debate Tournament," Robert J. Branham, ed., The New Debate: Readings in Contemporary Debate Theory, Washington, D.C.: Information Research Associates, 1975.
- Cronen, Vernon E., "Responding to Weaker Opponent: A Study of Likeability and Refutation in Public Debate," Western Journal of Speech Communication, 38 (Spring, 1974), pp. 108-116.
- Dresser, William R., "Effects of 'Satisfactory' and 'Unsatisfactory' Evidence in a Speech of Advocacy," Speech Monographs, 30 (August, 1963), pp. 302-306.
- Dunne, Dennis P.; Mack, Herschel L.; and Pruett, Robert, "Empirical Evidence on the 'Logical'--'Proficiency' Dichotomy in Debate Judging," Journal of the American Forensic Association, 7 (Winter, 1971), pp. 201-207.
- Faules, Don F. and Rieke, Richard D., Directing Forensics: Debate and Contest Speaking, Scranton, PA: International Textbook Co., 1968.
- Faules, Don, "Measuring Refutation Skill: An Exploratory Study," Journal of the American Forensic Association, 4 (Spring, 1967), pp. 47-52.
- Giffin, Kim, "A Study of the Criteria Employed by Tournament Debate Judges," Speech Monographs, 26 (March, 1959), pp. 69-71.
- Halstead, William P., "Who Wins Debates? A Statistical Study of 1320 Debates," Quarterly Journal of Speech, 26 (April, 1940), pp. 213-221.
- Hayes, Michael T. and McAdoo, Joe, "Debate Performance: Differences Between Male and Female Rankings," Journal of the American Forensic Association, 8 (Fall, 1971), pp. 127-131.

- Hensley, Wayne E. and Strother, David B., "Success in Debate," The Speech Teacher, 17 (September, 1968), pp. 235-237.
- Hill, Sidney R., "A Study of the Effect of Non-Ability Variables on the Outcome of Intercollegiate Debates." Ph.D. dissertation, University of Florida, 1973.
- Hill, Sidney, "Scoring Differences in the Use of Alternate Debate Ballots," Journal of the American Forensic Association, 10 (Spring, 1974), pp. 213-216.
- Kellerman, Kathy, "The Concept of Evidence: A Critical Review," Journal of the American Forensic Association, 16 (Winter, 1980), pp. 159-172.
- Maccoby, Eleanor Emmons and Jacklin, Carol Nagy, The Psychology of Sex Differences, Stanford, CA: Stanford University Press, 1974.
- Newman, Robert P., "Analysis and Issues--A Study of Doctrine," Central States Speech Journal, 13 (Autumn, 1961), pp. 43-54.
- Pearce, W. Barnett, "Attitudes toward Forensics," Journal of the American Forensic Association, 10 (Winter, 1974), pp. 134-139.
- Rosen, Nathan; Dean, Larry; and Willis, Frank, "The Outcome of Debate in Relation to Gender, Side, and Position," Journal of the American Forensic Association, 15 (Summer, 1978), pp. 17-21.
- Sanders, Gerald H., "Debate as a Paradigm for Demonstrating Skills in Argumentation and Logic," Paper presented at the Annual Meeting of the SCA (60th, Chicago, December 26-29, 1974).
- Vasilius, Janet M. and DeStephen, Dan, "An Investigation of the Relationship Between Debate Tournament Success and Rate, Evidence, and Jargon," Journal of the American Forensic Association, 15 (Spring, 1979), pp. 197-204.
- Verderber, Rudolph, "Judges' Criteria and Debater Adaptation: Empirical Evidence," Journal of the American Forensic Association, 5 (Winter, 1968), pp. 28-30.
- Williams, Frederick and Webb, Sally Ann, "Factors in Debate Evaluation: A Pilot Study," Central States Speech Journal, 15 (May, 1964), pp. 126-128.

Willmington, S. Clay, "A Study of the Relationship of Selected Factors to Debate Effectiveness," Central States Speech Journal, 20 (Spring, 1969), pp. 36-39.

Wise, Charles, "Relationships Among Certain Rating Judgments on the Form C Ballot," Journal of the American Forensic Association, 7 (Spring, 1971), pp. 305-308.