THE EFFECTS OF SEX ON EVALUATIONS IN INTERCOLLEGIATE DEBATE

Sidney R. Hill, Jr.

In the first half of this century, the organization of intercollegiate forensics left little doubt that women were seen as incapable of competing with men on equal terms. The exceptional woman might enter men's contests, but the more usual procedure was to employ separate divisions for the sexes.[1] Men's and women's divisions in oratory, extemperaneous speaking, oral interpretation, and debate were the rule rather than the exception.[2]

In the abandonment of these distinctions, however, forensics has moved ahead of the contemporary concern for the eqaulity of women. With the growth of coeducation in institutions of higher learning, women have participated more and more in debate against men. The tournament with separate men's and women's divisions is rapidly disappearing.[3] Women have participated in the National Debate Tournament and have compiled outstanding records. Some of the most respected coaches in the nation are women. Women have served as officers of the major forensic honoraries and play an active part in the professional societies. The format of contemporary debate tournaments suggests that sex is not seen as a variable which in any way affects the outcome of intercollegiate debate. Yet, recently published research tends to challenge that view.

PREVIOUS RESEARCH

Two studies are of immediate interest. Hensley and Strother in 1968 reported results indicating that sex does affect win-loss decisions.[4] They discovered that a mixed (i.e. one male, one female) team stood a "greater than random chance of winning" any given debate round.

Stimulated by the work of Hensley and Strother, Hayes and McAdoo in 1972 pursued the examination of sex as an influence on evaluations in debate.[5] These researchers utilized data generated by speaker rankings rather than the win-loss results employed by Hensley and Strother. A Chi-square test found significant deviation from expected results at a .01 level of significance, and the direction of the results indicated that female debaters were evaluated more highly than were male debaters.

There are a number of limiting factors in the research reported above. In the case of Hensley and Strother, the study deals only with team win-loss results. While this is of obvious importance, it does not focus directly on the evaluation of the individual debater as that evaluation is affected by sex.

The sampling procedure employed by Hayes and McAdoo raises questions as to their results. The ballots studied included all of those accumulated by three different college debate programs over a three-year span. Obviously this resulted in multiple measures of the same female and male debaters. While this procedure may give conclusive evidence of the superiority of female debaters at the schools involved, there is no basis for generalization to the entire population of female debaters.

Hayes and McAdoo also excluded from their data pool all ballots in which the competition consisted of all men or all women. Thus the possibility exists that their results were contaminated by interaction between male and female debaters.

Neither study cited made any distinction as to the sex of the judge in the round. Thus the possibility of yet another contaminating variable exists; i.e. interaction between the sex of the judge and the sex of the debater evaluated.

This study proposes to examine the influence of sex on evaluation in debate while controlling for interaction effects from the sex of the debater's colleague and the sex of the critic judge.

METHODOLOGY

Sampling procedures

The data for the study consisted of 2170 debate ballots collected from three different tournaments for five successive years.[6] In order to eliminate possible effects due to a specific debate topic, the test samples were constructed to provide equal representation for each debate year.

To avoid any possible effects of time within a tournament, the test samples were also stratified across time for early, late, and elimination round debates. It was decided to limit possible errors in estimation to no more than one point on the 5-30 point scale for speaker ratings in the estimation of sample means.[7]

After calculation of the variance of each of the strata, the appropriate sample size was computed to be 75 measurements. Samples for the study were drawn from the data pool by a "1 in 30" systematic sampling procedure until each stratum was filled. The allocation of the total sample across each stratum was derived by the formula $n_i = nw_i$.

Dependent measures

The A.F.A. Form "C" ballot provided four different measures of outcome for each debate: win-loss decision, speaker rating, team rating, and speaker ranking. The study was replicated using each of the ballot measures as the dependent variable. The specific hypotheses tested were:

1. There is no effect due to sex in the win-loss decisions in debate.
2. There is no effect due to sex in speaker ratings in debate.
3. There is no effect due to sex in rating debate teams.
4. There is no effect due to sex in the rankings assigned to debaters.

Predictor variables

The method chosen for the testing of the research hypotheses was an adaptation of multivariate regression analysis, using "dummy" variables to introduce nominal data into the regression equation. Since the variable tested lay outside the range of ability, any significant apportioning of variance in the dependent measure to sex was taken as evidence of a non-ability effect.

Regression analysis was chosen as the appropriate statistical procedure because the goal of the experiment was the construction of a prediction equation for the outcome of intercollegiate debates based on sex.

Actual calculations were performed by the University of Florida Computing Center using the BMDX63 program "Multivariate General Linear Hypothesis" developed by the U.C.L.A. Health Sciences Computing Facility. The output from this program includes regression coefficients for each of the predictor variables in the model, various cross-product matrices, and appropriate "F" statistics with associated degrees of freedom for hypotheses selected by the user.[8] BMDX63 tested a regression model of the general form:[9]

$$\hat{y} = \beta_o + \beta_1\chi_1 + \beta_2\chi_2 + \beta_3\chi_3 + \beta_4\chi_4 + \beta_5\chi_5 + \beta_6\chi_6 + \beta_7\chi_7 + \varepsilon,$$

where $\hat{y}$ was the estimated value of the dependent measure;

$\beta_o \ldots \beta_7$ were the weights associated with each of the eight possible combinations of sex of the debater, colleague, and judge;

$x_1 \ldots x_7$ were dummy variables (interpreted as either 1 or 0) representing the various sex combinations; and

ε represented unexplained variance in the model.

RESULTS

When considering win-loss as the dependent measure, there were a number of significant differences revealed by the data. Sex affected win-loss both for debaters and for judges. These results are shown in Table I.

TABLE I

Regression Of Sex On Win-Loss For Groups

Group	F-score	d.f.	Interpretation
Male debaters	10.37	3,92	$p < .01$
Female debaters	10.25	4,92	$p < .01$
Male judges	16.13	3,92	$p < .01$
Female judges	5.22	4,92	$p < .01$

Interaction between the sex of the debaters and the sex of the judge was also examined. In all cases, the calculated regression weights were found to be significant with $p < .05$. These are the results shown in Table II.

TABLE II

Regression Of Sex On Win-Loss

Variable (debater/colleague/judge)		Weight	F-score
X_1	male/male/female	-0.54	18.47
X_2	male/female/male	-0.88	24.66
X_3	male/female/female	-0.44	6.91
X_4	female/male/male	-0.88	24.66
X_5	female/male/female	-0.44	6.91
X_6	female/female/male	-0.83	23.36
X_7	female/female/female	-0.50	8.05

When considering the effects of sex on the dependent measure speaker rating, analysis again revealed the presence of significant regression. No significant effects were found for groups in speaker rating. However, this was not the case when considering interaction among the sex of the debater, the sex of the colleague, and the sex of the judge. These results are shown in Table III.

TABLE III

Regression of Sex on Speaker Ratings

Variable	Weight	F-score	Interpretation
X_1	-0.30	0.08	$p > .05$
X_2	-3.30	4.94	$p < .05$
X_3	-1.47	1.06	$p > .05$
X_4	-3.68	6.12	$p < .05$
X_5	-3.47	5.92	$p < .05$
X_6	0.28	0.05	$p > .05$
X_7	-2.30	2.40	$p > .05$

Sex affected team ratings both for debaters and by judges. Regression of sex on team ratings by groups is summarized in Table IV.

TABLE IV

Regression Of Sex On Team Ratings For Groups

Group	F-score	d.f.	Interpretation
Male debaters	2.34	3,92	$p > .05$
Female debaters	2.67	4,92	$p < .05$
Male Judges	3.44	3,92	$p < .05$
Female Judges	1.82	4,92	$p > .05$

Once again, significant interaction was found to exist among the sex of the debater, the sex of the colleague, and the sex of the judge. The regression weights for various sex combinations and the results of the tests for significance are shown in Table V.

TABLE V

Regression Of Sex On Team Ratings

Variable	Weight	F-score	Interpretation
X_1	-0.09	0.19	$p > .05$
X_2	-0.58	4.53	$p < .05$
X_3	-0.48	3.40	$p > .05$
X_4	-0.58	4.53	$p < .05$
X_5	-0.48	3.40	$p > .05$
X_6	0.13	0.32	$p > .05$
X_7	-0.45	2.77	$p > .05$

When considering the effects of sex on the dependent measure speaker ranking, analysis again revealed the presence of significant regression. The differences between groups were tested, and the results are shown in Table VI.

TABLE VI

Regression Of Sex On Speaker Rankings
For Groups

Group	F-score	d.f.	Interpretation
Male debaters	4.09	3,92	$p < .05$
Female debaters	4.26	4,92	$p < .01$
Male judges	6.60	3,92	$p < .01$
Female Judges	2.19	4,92	$p > .05$

As was the case for the other three dependent measures, speaker ranking revealed interaction effects between the sex of the debaters and the sex of the judge. The calculated regression weights, with their associated "F" statistic and interpretations, are shown in Table VII.

TABLE VII

Regression of Sex on Speaker Rankings

Variable	Weight	F-score	Interpretation
X_1	0.80	6.49	$p < .05$
X_2	1.43	10.42	$p < .01$
X_3	0.63	2.24	$p > .05$
X_4	1.43	10.42	$p < .01$
X_5	0.96	5.21	$p < .05$
X_6	1.30	11.38	$p < .01$
X_7	0.80	3.28	$p > .05$

CONCLUSIONS

The rater-bound variable "sex" appears to have a significant effect on some measures of the outcome of debates. Each of the ballot measures revealed some significant regression effects.

Tests of the dependent measure win-loss challenged the conclusions of Hensley and Strother. The finding of significant differences allows for prediction of win-loss results by sex as follows.

$$\hat{y} = 1.0 - .54x_1 - .88x_2 - .44x_3 - .88_4 - .44x_5 - .83x_6 - .50x_7$$

where x_1 represents a male team before a female judge;

x_2, x_4 represent mixed teams before a male judge;

x_3, x_5 represent mixed teams before a female judge;

x_6 represents a female team before a male judge; and

x_7 represents a female team before a female judge.

The presence of interaction effects invalidates any general statements as to the comparative expectations of winning among male, female, and mixed teams. If $\hat{y} \geq .5$ is defined as an expected win, and $\hat{y} < .5$ as an expected loss for any given debate, then these results indicate that all-male teams had a greater expectation of winning before a male than a female judge. Mixed teams and all-female teams, however, lose more frequently with male judges and may expect to win with female judges.

When considering the dependent measure speaker rating, the corrected model for predicting outcome on the basis of sex was as follows.

$$\hat{y} = 22.8 - 3.3x_2 - 3.68x_4 - 3.47x_5$$

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 predicted speaker rating for the male member of a mixed team was 19.5, as compared to 22.8 for a male debater with a male colleague before a male judge. The expected rating for the female member of a mixed team before a male judge was 19.12. When debating before a female judge, the female in a mixed team had an expected rating of 19.33.

The dependent measure team rating also revealed significant effects due to sex. In general, 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. The corrected model for predicting team ratings on the basis of sex was as follows.

$$\hat{y} = 3.70 - .58x,$$

where x represents mixed teams before a male judge.

The data also provided sufficient evidence to reject hypothesis #4. The corrected model for predicting speaker rank by sex was as follows.

$$\hat{y} = 1.7 + .8x_1 + 1.43x_2 + 1.43x_4 + .96x_5 + 1.3x_6$$

These findings tend to conflict with the results reported by Hayes and McAdoo. The model indicates that the members of mixed teams were ranked lower by male than by female judges. The expected rankings were 3.13 as compared to 1.70. For all-male teams, higher rankings came from male judges (1.7 as compared to 2.5 for female judges). For all-female teams, better rankings were received from female than from male judges. Hayes and McAdoo suggest the possibility of a leniency error by male judges in favor of female debaters. These results point in the opposite direction.

The presence of significant interactions between the sex of the debaters and the sex of the judge represents a serious challenge to the integrity of intercollegiate debate. Immediate research is needed to discover means of compensating for the biases revealed by this study. In the absence of such compensatory measures, debate judges can only strive individually to purge their decisions of bias.

It should be noted in conclusion that this researcher did not accept the results reported here as indicative of a difference in performance between the sexes. Such a difference, if it did exist, might well represent the source of the regression effects discussed above. An experimental design which could hold ability constant while manipulating sex as the independent variable would shed more light on the issue, and would be of great value to the field of forensics.

FOOTNOTES

Mr. Hill (Ph.D., University of Florida, 1973), is Assistant Professor of Communication and Director of Forensics at Mississippi State University.

[1] Documentation of this can be found in a number of places. Nichols pointed it out in 1937, and, as late as 1952, Emery advocated it as the most desirable procedure. See Egbert R. Nichols, "A Historical Sketch of Intercollegiate Debating: III," Quarterly Journal of Speech, XXIII (April, 1937), 259-278; and Emogene Emery, "Rehabilitating Women's Debate," Southern Speech Journal, XVII (March, 1952), 186-191.

[2] Berry reported in 1928 that forty-four out of fifty-six schools surveyed maintained separate men's and women's teams. Mildred F. Berry, "A Survey of Intercollegiate Debate in the Mid-West Debate Conference," Quarterly Journal of Speech, XIV (February, 1928), 86-94.

[3] The 1973-73 A.F.A. Calendar of Tournaments reports only 12 which have a separate women's division in debate. Of these, seven are located in the Pacific Northwest. Jack Howe, (ed.), "A.F.A. Calendar 1973-74," Journal of the American Forensic Association, IX (Spring, 1973), 413-425.

[4] Wayne E. Hensley and David B. Strother, "Success in Debate," The Speech Teacher, XVII (September, 1968), 235-237.

[5] Michael T. Hayes and Joe McAdoo, "Debate Performances: Differences Between Male and Female Rankings," Journal of the American Forensic Association, VIII (Winter, 1972), 127-131.

[6] The years covered were the academic years 1967-68, 1968-69, 1969-70, 1970-71, and 1971-72. The tournaments involved were the Peachtree Debate Tournament, hosted by Emory University in Atlanta, Georgia; the Birmingham Invitational Debate Tournament, hosted by Samford University in Birmingham, Alabama; and the Gator Invitational Debate Tournament, hosted by the University of Florida in Gainesville, Florida.

[7] In order to achieve this limit, the following formula for determining the size of the test sample was used:

$$n = \frac{\sum_{i=1}^{L} \frac{N_i^2 \sigma_i^2}{w_i}}{N^2 \frac{B^2}{4} + \sum_{i=1}^{L} N_i^2 \sigma_i^2}$$

Since this formula requires some estimate of population variance for each of the sample strata, a "1 in 20" systematic sample with N = 200 was drawn from the data pool to provide those estimates. For a discussion and mathematical validation of this formula, see William Mendenhall, Lyman Ott, and Richard Schaefer, Elementary Survey Sampling (Belmont, California: Wadsworth Publishing Company, Inc., 1971), p. 61.

[8]For a description of this program, see W. J. Dixon, BMD Biomedical Computer Programs: X-Series Supplement. (University of California Publications in Automatic Computation No. 3, 1973), pp. 23-33.

[9]Using the values calculated by BMDX63, various null hypotheses of regression effect were tested as follows. The general hypothesis of some effect due to sex was of the form

$$Ho: \beta_1 = \beta_2 = \ldots = \beta_K = 0$$

$$H_\alpha: \text{at least one } \beta_i \neq 0$$

In order to isolate the specific items in which an effect due to sex was to be found, each of the terms in the model was tested under the general form

$$Ho: \beta_i = 0$$

$$H_\alpha: \beta_i \neq 0$$

All tests were made using a preset value of = .05, and the test statistic

$$F = \frac{MSR}{MSE} \quad \Omega \quad F_{\nu_1 \nu_2}$$

In all cases, actual calculation of the values was performed by the University of Florida Computing Center through the use of the BMDX63 program.