DOCUMENT RESUME

ED 354 570 CS 508 085

AUTHOR Dudczak, Craig A.; And Others
TITLE A Comparative Analysis of the Predictive Validity of Questionnaires and Philosophy Statements in CEDA Debate.
PUB DATE Nov 92
NOTE 23p.; Paper presented at the Annual Meeting of the Speech Communication Association (78th, Chicago, IL, October 29-November 1, 1992).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *Debate; *Evaluation Methods; Higher Education; *Judges; National Surveys; *Predictive Validity; Predictor Variables; Questionnaires
IDENTIFIERS Cross Examination Debate Association; Debate Tournaments

ABSTRACT

A study assessed the predictive validity of two substantially different instruments (a questionnaire or a philosophy statement) which may be used to predict critics' ballot behavior in Cross Examination Debate Association (CEDA) debate. Questionnaires were distributed to 29 debate tournaments across the United States for completion by critics judging at those tournaments. Judge philosophy statements were retrieved from among those solicited by the CEDA national tournaments. A total of 87 subjects completed the questionnaire with 34 having a minimum of 6 or more written ballots. Usable philosophy statements for 24 of these respondents were gathered. Hence, 34 sets of subjects were used in analysis of questionnaire-ballot correlations and 26 sets of subjects were used to assess philosophy-ballot correlations. Results indicated that: (1) when critics were the unit of analysis, "new arguments" had high negative predictive validity and "inherency" had high predictive validity; and (2) if ballots were taken as the unit of analysis, philosophies were substantially better predictors, but if ballots by critics are combined to make the critic the unit of analysis, the effect disappeared. Findings are limited by the small number of discriminants which emerged as significant and reliable. (A figure representing the construct and technique matrix of tools and eight tables of data are included. (Contains 16 references.) (RS)

A Comparative Analysis of the Predictive Validity of Questionnaires and Philosophy Statements in CEDA Debate

Craig Dudczak*, Donald Day and Tina Hartwell

Syracuse University

Speech Communication Association Conference

Chicago, Illinois

November 1992

Address Correspondence to: Craig A. Dudczak
Speech Communication Department
400 Sims Hall
Syracuse University
Syracuse, NY 13244-1230

(315) 443-5144

Running Head: Predictive Validity in CEDA Debate

A Comparative Analysis of the Predictive Validity of Questionnaires and Philosophy Statements in CEDA Debate

Predictive validity is at the heart of applied science. Although there is undeniable benefit to pure research, the ability to at least correlate observations (if not establish causal relationships) is the ultimate test of understanding in dealing with real world phenomena. If measurable results cannot be predicted based upon an understanding of underlying principles, the utility of any avenue of research may be called into question. The main problem addressed in the current study is an assessment of the predictive validity of two substantially different instruments which may be used to predict critics' ballot behavior in CEDA debate. The two instruments, a survey questionnaire and a structured philosophy statement, are characterized by major differences in how they guide critics' reporting of principles which underlie their debate decisions. The characteristics of measurement instruments mold the responses of subjects. The problem is simply to assess which instrument is better at predicting critics' ballot behavior, and to explain why.

A previous study (Day & Dudczak, 1991) sought to establish the degree to which questionnaires and philosophy statements map to each other (i.e., the extent to which their metrics vary consistently in response to similar real world situations). The current study takes this objective a step further in seeking to establish which of the two instruments better predicts ballot behavior. Credible satisfaction of this goal would contribute significantly to the discipline of argumentation in that it would validate a methodology for assessment of critics' views which could be used first to establish a taxonomy of paradigms

applied in debate decisions, and second to train collegiate debate teams in the intricacies of argumentation as perceived by relevant experts: judges in tournament rounds.

If the two instruments vary widely vis-a-vis one another (as reported in Day & Dudczak, 1991), it is likely that (a) one has a higher level of predictive validity than the other, (b) both are equally predictive for varying reasons, or (c) both are equally non-predictive for varying reasons. The main goal of the current study is to establish which of these cases is most probably true.

In addressing the problem of instruments' predictive validity, this study was guided by Dudczak & Day's regional pilot study (1989a), which indicated that judge philosophy statements have substantially higher predictive power than do survey questionnaires.

Brief Description of the Study, Including Hypothesis

The current study addresses the question of which instrument (a questionnaire or a philosophy statement) is most effective in predicting actual ballot behavior by a pool of critics active in tournaments in various CEDA regions during the Fall 1989 debate season. "Predictive validity" is considered coincident with the correlation between individual critics' relatively abstract assertions regarding decision criteria and their actual behavior as evidenced in debate ballots.

In order to field the current study, questionnaires were distributed to 29 tournaments across the U.S. for completion by critics judging at those tournaments. (However, only 11 tournaments returned questionnaires.) Judge philosophy statements were retrieved first from among

those solicited by the 1990 CEDA National Tournament, then (if necessary) from statements completed for other CEDA national tournaments or from the 1989 Syracuse Debate Tournament.[1]

Each instrument was compared with actual ballot behavior for corresponding critics to determine the instrument's predictive validity.

Hypothesis

The following hypothesis was tested in the current study. It attempts to extend findings of the Dudczak & Day (1989a) regional pilot study to a non-regional population.

H1. Judge philosophy statements are not better predictors of ballot behavior than are survey questionnaires.

Relationship of Current Study to Pilot and Non-regional Studies

This paper reports results of the final experiment of four conducted using a non-regional sample of CEDA debate critics. The first (Dudczak & Day, 1991a) in part replicated the earlier pilot study, which had examined the broader issue of whether debate critics' espoused decision criteria are in fact implemented in actual ballot behavior. Survey questionnaires and judge philosophy statements were matched against corresponding ballots to determine the consistency of professed criteria to decision criteria (Experiment #1). This first non-regional study also compared selected portions of critic questionnaires against the top, more easily quantified, portions of debate ballots (Experiment #2). The third experiment in the series was reported in Day & Dudczak (1991). This effort compared attributes on survey questionnaires to their corresponding items on judge philosophy statements, to ascertain the degree to which the two instruments measured similar underlying

principles and attitudes.

The final experiment reported in the current paper matches survey questionnaires and judge philosophy statements independently against ballot behavior.

The non-regional study of four experiments was preceded by a pilot study based upon questionnaires, philosophies and ballots from tournaments in the Northeast during the fall of 1989 (Dudczak & Day 1989a; 1989b). That study not only attempted to match professed criteria against actual behavior, but also sought to establish a taxonomy of CEDA debate paradigms. A further extension of this line of inquiry (Dudczak, Day, & Hartwell 1992) attempts to ascertain criterion validity of the paradigms employed by critics.

Literature Review

While a number of studies have evaluated critics' paradigm preferences in NDT (Cox 1974; Cross & Matlon 1978; Thomas 1977) and in CEDA (Buckley 1983; Lee, Lee & Seeger 1983), these surveys have not established whether expressed preferences actually are used in judging debates. Unless confirmed by decision criteria actually employed in debate rounds, the utility of judge philosophy statements in academic debate is open to question.

The current study is justified by the scarcity of research regarding debate critic decision criteria. Early investigations (cited above) surveyed critic paradigm preferences through self-report instruments. These surveys were limited to indicating "professed" beliefs, since they were not intended to validate the extent to which preferences actually were applied. More recent work by Gaske, Kugler and Theobald (1985)

attempted to discriminate among CEDA judging paradigms, but relied upon unequal cell sizes (therefore, they may have been flawed methodologically). Brey (1989; 1990) analyzed CEDA philosophy statements to discover the elements of judge preference, but his analysis did not indicate whether paradigm preferences correlated with discernible patterns of judging behavior.

Even less research has focused upon the artifacts of debate evaluation. Bryant (1983) conducted a content analysis of NDT and CEDA debate transcripts to compare evidence use within each format. Hollihan, Riley, and Austin (1983) used content analysis of NDT and CEDA ballots to determine thematic "visions" embraced respectively within these two debate formats. While their analysis of ballots suggested that different visions are held by NDT critics versus CEDA critics, without knowledge of the critics' prior attitudes one cannot know whether ballot comments reflected critic preference or circumstances unique to debate rounds.

There have been five research reports that compared judge philosophy statements with ballot artifacts. Henderson and Boman (1983) reported high consistency (83.5%) between a set of NDT judge philosophy statements and corresponding ballot comments, although their analytic procedures make their findings suspect. Dudczak and Day (1989a) found lower consistency (54.9%) in a pilot regional study of CEDA critics. They also reported that several clusters of paradigms were correlated with decision criteria cited in critics' ballots. A secondary analysis of Dudczak and Day's pilot data (1989b) sought to isolate differences among traditional paradigms. Paradigm boundaries were found to be

porous and unreliable. Unlike the earlier work by Dudczak and Day (which included only data from the Northeast), their 1990 (Dudczak and Day 1991a) non-regional study included tournaments from across the U.S. Their first two experiments replicated the previous pilot effort, investigating three research questions and nine hypotheses. Results showed little reliability for questionnaires as predictors of critics' ballot behavior (thus the current paper, comparing questionnaires to philosophies as they predict ballot behavior). The 1990 experiments by Dudczak and Day showed limited association between professed paradigms and subsequent ballot behavior, and indicated that the components assigned by critics to traditional paradigms largely overlap one another. In fact, the non-regional study indicated less consistency between professed beliefs and actual ballot behavior than had been observed with purely regional data.

The latest experiment by Day and Dudczak (1991) compared variables in questionnaires to corresponding variables in philosophies, to evaluate the degree to which the instruments measure similar aspects of critic preference. That experiment showed little similarity between the two instruments. It also demonstrated that inconsistencies between professed and actual behavior noted in earlier work were not an artifact of intrasample cancellation (due to data aggregation). Critics were inconsistent individually from ballot to ballot, not merely as a group.[2]

Methodology

Materials

The work products and instrument examined in this study included (a) judging philosophies, (b) ballots completed during competition at

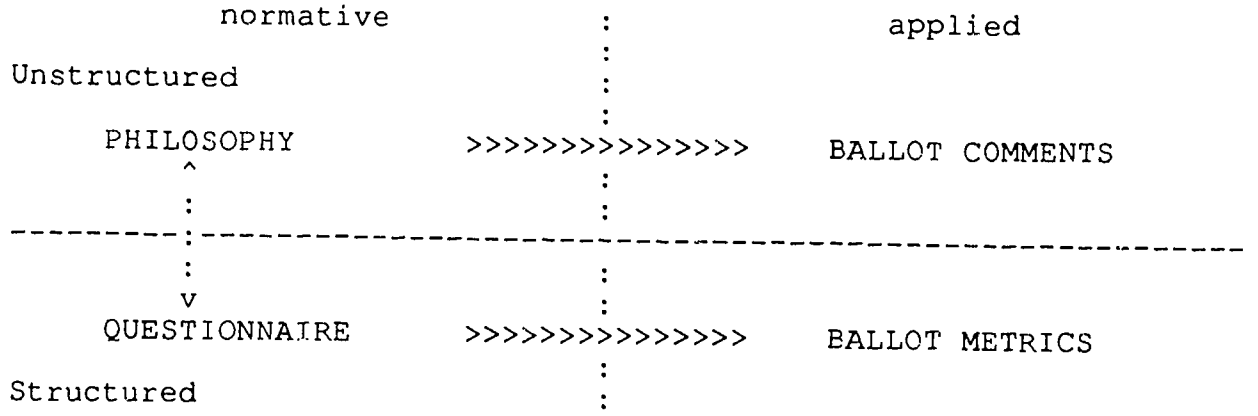
tournaments, and (c) a structured questionnaire administered at tournaments (following a majority of the rounds).

Coding forms used for Dudczak and Day's first two non-regional experiments (1991a) were expanded further to include new discriminants; the coding category description form developed for the earlier experiments also was revised, to minimize ambiguity in and overlap among discriminants.

The one instrument and two work products used in the study may be visualized in a two-by-two table. Both the philosophy and questionnaire are normative --"ought"-- documents; the ballots are applied documents. The philosophy and comment portions of ballots are unstructured; the questionnaire and template (top) portions of ballots are structured. Using these distinctions, the current study examines the predictive validity of the questionnaire and philosophy statement. A future study may examine the construct validity of these documents.

FIGURE 1

Construct and technique matrix of tools in the study



The two-page questionnaire incorporated 32 Likert scale items, five yes/no selections, five multiple option questions, two single selection

choices, one 10-item value assessment ranking question, and two 3-item proportional weighting scales. Twenty-eight of the Likert scale items also asked whether the operation of an element in a round would help or hurt the team involved.

Of the 42 items on the judge philosophy coding form, 10 were binary, 30 were category choices, and two were 10-category choices.

Subjects

Subjects used in this study were debate critics who judged debate rounds at CEDA tournaments during the Fall 1989 season. For a subject's work products and instrument to be included in the current study, s/he must have completed either a judge philosophy statement and/or a survey questionnaire, plus a minimum of six ballots written for the Fall 1989 CEDA topic.[3]

Eighty-seven subjects completed the questionnaire with 34 having the minimum of six or more written ballots. Usable philosophy statements for 24 of these respondents were gathered from the CEDA Judge Philosophy Handbooks or solicited at one tournament.[4] Two additional philosophy statements from critics with sufficient ballots (but who had not answered the questionnaire) were obtained from the CEDA Judge Philosophy Handbooks. Hence, 34 sets of subjects were used in analysis of questionnaire-ballot correlations and 26 sets of subjects were used to assess philosophy-ballot correlations.

Procedures

Twenty-nine tournament directors who had hosted CEDA tournaments during the Fall 1989 season were asked to administer the questionnaire to judges at their tournaments. Sixty-nine questionnaires were returned

from eleven tournaments; two additional questionnaires were returned directly by respondents. A follow-up solicitation mailed to critics yielded an additional 16 questionnaires. A total of eighty-seven completed surveys were obtained.

Official ballots submitted by judges at 11 or the 29 CEDA tournaments comprised the second source of data. For the bulk of the study, each round was considered an unique case for purposes of statistical analysis, and critic response patterns were considered in the aggregate. However, in one analysis (composite critic response to key discriminants) all remarks by one critic on any of his or her ballots were combined, to determine whether the critic ever cited key discriminants in those work products. Of the 1653 ballots returned, 1519 were usable.[5] Only the usable ballots for the 34 subjects who had completed a questionnaire were included in the questionnaire-ballot portion of the study (N = 307); only the usable ballots for the 26 subjects who had completed a judge philosophy statement were included in the philosophy-ballot part of the study (N = 236). Two coders were trained to code the ballots. Ballot comments were recorded on a standardized coding form independently by the two coders.

The third source of data was judge philosophy statements, which also were rated independently by two coders. Intercoder reliability was disappointing in the first two experiments of this study. Therefcre, for this final experiment two coders performed pretest coding of a small sample of philosophies and ballots. After discussion of differences in interpretation of source documents, changes were made in the discriminator reference sheet used in coding. Additional discussion and

mutual training ensued before coding of the actual sample for this experiment began. As a result, an improved intercoder reliability of r = .613 was achieved.[6] Table 1A reports the discriminants for which coders experienced relatively high levels of reliability for the Philosophy coding task while Table 1B reports the relatively high levels of reliability for the Ballot coding task.

Table 1A
Discriminants Revealing High Intercoder Reliability
Philosophy Statements

DISCRIMINANT	RELIABILITY
Paradigms	
Judicial	1.000
Value Comparison	1.000
Hypothesis Testing	.799
Argument Skills	.693
Tabula Rasa	.652
Substantive Elements	
Affirmative Burden of Proof	1.000
Coverage	1.000
New Arguments in Rebuttal	1.000
Turnarounds	1.000
Uniqueness	1.000
Debate Philosophy Arguments	.946
Counter-Warrants	.943
Cross-Examination	.931
Topicality	.928
Obnoxious Behavior	.854
Burden of Rejoinder	.838
Inherency	.828
Ethics	.817
Prima Facie	.785
Justification	.716
Organization	.710
Counter-Intuitive Arguments	.708

Note#1: High Inter-Coder Reliability was operationalized as those exceeding the overall reliability for Philosophy Statements (r = .705).

Table 1B

Discriminants Revealing High Intercoder Reliability
Ballot Comments

DISCRIMINANT	RELIABILITY
Debate Theory Arguments	.847
Uniqueness	.842
Topicality	.807
Delivery	.791
Organization	.790
New Arguments in Rebuttal	.731
Dropped Arguments	.721
Cross-Application	.621
Turnarounds	.588
Inherency	.576

Note#1: High Inter-Coder Reliability were operationalized as those exceeding the average reliability for Ballot Comments ($r = .605$).

Note#2: No explicit paradigm identification was made by critics on the ballots (N=307).

Data processing for the study was performed on an IBM-PS/2 using PC-FILE+ (a database program) and on an IBM 3090 mainframe using SAS (a statistical package). Data were entered via PC-FILE, converted to standard data format (SDF), manipulated using BASIC programs written for this study, then uploaded to the mainframe for SAS correlation runs.

Results

Three sets of correlations between philosophy statements and ballots were compared with corresponding sets of correlations between questionnaires and ballots, to test which instrument was superior in predicting critics' ballot behavior. After comparing the predictive

ability of each instrument for each set of correlations, an aggregate score could be calculated for each instrument. Disconfirmation of the research hypothesis could occur if philosophies were superior to questionnaires either because of a higher aggregate score because of higher predictability on one or more individual sets of correlations.

H1. Judge philosophy statements are not better predictors of ballot behavior than are survey questionnaires.

Acceptable levels of intercoder reliability were experienced for only some ballot and philosophy discriminants. Those discriminants which exceeded a reliability threshold (r = .700) were included and used in comparing the predictive validity of the questionnaire and philosophy (Table 2).[7]

Table 2

Philosophy/Ballot Discriminants with Acceptable Reliability

DISCRIMINANT	COMBINED RELIABILITY
* Uniqueness	.921
Debate Theory Arguments	.897
Topicality	.868
New Arguments	.866
* Turnarounds	.789
* Organization	.750
Inherency	.702

Note#1: Intercoder reliability for these items (r = .838)

Note#2: Discriminants for which there were no equivalent items on the questionnaire are indicated by an asterisk (*).

Not all of the seven discriminants with high reliability could be used in comparisons, since the questionnaire lacked corresponding items for three discriminants. Consequently, only four discriminants were used to determine the comparative predictive validity of philosophy

statements versus questionnaires.

For each discriminant, three sets of comparisons were made between philosophy statements and questionnaires. The first comparison determined the correlation between a discriminant's presence on a ballot and its occurrence on the predictive instrument.[8] For this first analysis, each ballot was treated as a separate case. Table 3 reports the correlation of discriminant by instrument type.

Table 3

Correlation of Discriminant by Instrument Type:
Predicted Use of Discriminant Using Ballot As Case

DISCRIMINANT	INSTRUMENT TYPE Questionnaire	Philosophy
Debate Theory Arguments	-.069	.136*
New Arguments	-.018	-.155*
Topicality	.074	.002
Inherency	.077	.138*

Note#1: * ($p < .05$)

A second comparison of discriminants by instrument type investigated the valence of discriminants on ballots versus their valence on predictive instruments. Valence in this sense constitutes an opinion by the critic that the occurrence of a discriminant would/did help or hurt the team in question. Table 4 reports the correlation between discriminant valence and instrument type, still treating each ballot as a separate case.

Table 4

Correlation of Discriminant by Instrument Type:
Predicted Valence of Discriminant Using Ballot As Case

DISCRIMINANT	INSTRUMENT TYPE Questionnaire	Philosophy
Debate Theory Arguments	-.029	.105
New Arguments	.021	-.030
Topicality	-.014	.063
Inherency	.043	.130*

Note#1: * ($p < .05$)

A final comparison was made treating the critic rather than the ballot as the unit of analysis. All ballots from an individual critic were combined to create a single case for that critic. This approach allowed us to ask, in effect, whether the critic ever applied the discriminant. It also reduced the degree to which the objective presence or absence of a discriminant in an individual round could affect results by providing greater opportunity for the discriminant to occur. Table 5 reports the correlation of discriminant by instrument type when treating the critic as the unit of analysis.

Table 5

Correlation of Discriminant by Instrument Type:
Predicted Use of Discriminant using Critic as Case

DISCRIMINANT	INSTRUMENT TYPE Questionnaire	Philosophy
Debate Theory Arguments	-.184	.192
New Arguments	-.278	-.296
Topicality	.034	.319
Inherency	.318**	.247

Note#1: ** approached significance ($p = .067$)
Note#2: Ballot x = 9.03 per critic.

Discussion

It is a curiosity of strict predictive validity that relatively high negative correlations as well as relatively high positive correlations are considered to be desirable in terms of predictive power. Although most correlations reported here are too low to be influenced by this consideration, when critics are the unit of analysis, "new arguments" appear to have high negative predictive validity and "inherency" appears to have high predictive validity. In other words, critics can be presumed to behave in ballot remarks regarding new arguments in rebuttal in a manner opposite to that claimed in either questionnaires or philosophy statements, predictably.

In our introduction. we noted that this study was guided by Dudczak & Day's regional pilot study (1989a), which indicated that judge philosophy statements have substantially higher predictive power than do survey questionnaires. The current study both replicates and refutes this finding. If ballots are taken as the unit of analysis (as they were in Dudczak & Day, 1989a), philosophies are substantially better predictors. However, if ballots by critics are combined to make the critic the unit of analysis, this effect disappears.

Table 6 presents this unexpected effect of the two differing treatments. Both "use" and "valence" analyses show that philosophies are three times as predictive as questionnaires when individual ballots are the unit of analysis. However, except for the topicality discriminant, when critics are the unit of analysis there is essentially no difference between the predictive validity of philosophies versus that of questionnaires.

Table 6

Ratio of Correlations By Discriminant,
Philosophy: Questionnaire,
Strict Predictive Validity

TREATMENT	DISCRIMINANT	RATIO
Ballot As Case		
(Use)	Debate Theory Arguments	1.97:1
	New Arguments	8.61:1
	Topicality	0.03:1
	Inherency	1.79:1
	(Use Average)	**(3.10:1)**
(Valence)	Debate Theory Arguments	3.62:1
	New Arguments	1.43:1
	Topicality	4.50:1
	Inherency	3.02:1
	(Valence Average)	**(3.14:1)**
(Ballot As Case Average)		**(3.12:1)**
Critic As Case		
(Use)	Debate Theory Arguments	1.04:1
	New Arguments	1.06:1
	Topicality	9.38:1
	Inherency	0.78:1
(Critic As Case Average)		**(3.07:1)**

Note#1: Strict predictive validity occurs whenever a questionnaire or philosophy correlates to a ballot or critic, regardless of direction (i.e., a high negative correlation would be considered a positive sign of predictive validity despite the fact that it would mean that a critic frequently professes one position, but in fact acts exactly the opposite).

The apparent impact of shifting the unit of analysis from individual ballots to critics may have profound implications for the future study of professed preferences versus ballot behavior. It might be considered incidental only if one assumes that critics will feel free to

cite favorite discriminants in ballot remarks even if the discriminant figures only slightly in the round, reducing the influence that variability between rounds may have upon apparent critic consistency when ballots are the unit of analysis. We feel this assumption would be highly suspect.

We make the case in our review of literature that studies which examine only critics' professed positions are of limited value because there is no behavioral standard against which to measure results. Clearly, critics' claims of preference are meaningless if they are not implemented in ballot behavior. However, even the five studies which have examined ballots can be questioned if they did not use combined ballots for a given critic as the unit of analysis, despite the fact they were targeted at assessing critic consistency rather than predictive validity.[9] Only the fact that few of the correlations in the current study achieved statistical significance blunts the potential implications of the unit of analysis issue.

Two other discriminants emerged as significant when critics were used as the unit of analysis (albeit with low intercoder reliability). Both discriminants emerged from the philosophy statements with high correlations and are reported in Table 7.

Table 7
Discriminants Predicted by Philosophy Statements
With High Correlations But Inadequate Reliability

DISCRIMINANT	CORRELATION	PROBABILITY
Justification Arguments	.497	p <.05
Evidence Quality	.389	p =.06

The small number of discriminants which emerged as significant and reliable limits the conclusions which can be derived from this study. The relatively small number of critics treated as cases when used as the unit of analysis for Questionnaires (N = 34) and Philosophy Statements (N = 24) contributed to the lack of significant results. Further, the intercoder reliability quotient further limited the number discriminants which emerged.

Although to some the issues raised by predictive validity may seem a methodological labyrinth of questionable value, we feel the difference in findings seen with one treatment versus the other should act as a warning to researchers in the field to give such issues serious consideration in future studies. If nothing else, future research should focus upon the critic as the appropriate unit of analysis utilizing a sufficient number of ballots for each critic.

References

Brey, J. (1989). "A Descriptive Analysis of CEDA Judging Philosophi s Part I: Definitive Acceptance of Rejection of Certain Tactics and Arguments." CEDA YEARBOOK 10: 67-77.

Brey, J. (1990). "An Analysis of CEDA Judging Philosophies - Part II: Accepting Certain Tactics and Arguments." CEDA YEARBOOK 11: 72-79.

Bryant, M. (1983). "A Content Analysis of Selected NDT and CEDA Transcripts." Unpublished. Paper presented at the Third SCA/AFA Summer Conference on Argumentation. Alta, Utah.

Buckley, D. (1983). "A Comparison of Judging Paradigms." ARGUMENT IN TRANSITION: PROCEEDINGS OF THE THIRD SUMMER CONFERENCE ON ARGUMENTATION (hereinafter, ARGUMENT IN TRANSITION). Eds. Zarefsky, D., Sillars, M. and Rhodes, J. Annandale, VA: Speech Communication Association (SCA), 858-870.

Cox, J. (1974). "A Study of Judging Philosophies of the Participants of the National Debate Tournament." JOURNAL OF THE AMERICAN FORENSIC ASSOCIATION (JAFA) 11: 61-71.

Cross, J., and R. Matlon (1978). "An Analysis of Judging Philosophies in Academic Debate." JAFA 15: 110-123.

Day, D., and Dudczak, C. (1991). "Debate Philosophy Statements As Predictors of Critical Attitudes". ARGUMENT IN CONTROVERSY: PROCEEDINGS OF THE SEVENTH SCA/AFA CONFERENCE ON ARGUMENTATION. D. Parson (ed.) Annandale, VA: SCA, 441-445.

Dudczak, C., D. Day & T. Hartwell (May 1992). "The Elements of CEDA Debate Paradigms: An Investigation of Paradigm Accuracy," Unpublished. Eastern Communication Association Conference, Portland, Maine.

Dudczak, C., and D. Day (1991a). "The Impact of Paradigm Consistency on Taxonomic Boundaries in CEDA Debate". CEDA YEARBOOK 12, 27-47. An earlier version was presented to the 1990 SCA Conference, Chicago. (ERIC # ED 328 925).

Dudczak, C. and Day, D. (1991b). "Debate Philosophy Statements as Predictors of Critic Attitudes: A Summary and Direction of Research." Unpublished. SCA Conference, Atlanta. (ERIC # ED 341 103).

Dudczak, C., and D. Day (1989a). "A Profile of CEDA Debate Critics," SPHERES OF ARGUMENT: PROCEEDINGS OF THE SIXTH SCA/AFA SUMMER CONFERENCE ON ARGUMENTATION. Ed. Gronbeck, B. Annandale, VA: SCA, 427-433.

Dudczak, C., and D. Day (1989b). "A Taxonomy of CEDA Debate Critics".

Unpublished. Paper presented to the Speech Communication Association Conference, San Francisco. (ERIC # ED 318 071).

Gaske, P., D. Kugler, and J. Theobald (1985). "Judging Attitudes and Paradigmatic Preferences in CEDA Debate: A Cumulative and Construct Validity Investigation." CEDA Yearbook 6: 57-66.

Henderson, B., and D. Boman (1983). A Study to Determine If Debate Judges' Philosophy Statements are Consistent with Their Later Related Ballot Statements." JAFA 19: 191-198.

Lee, R., Lee, K., and Seeger, M. (1983). "Comparison of CEDA and NDT: Differences in Program Characteristics and Forensic Director Attitudes". ARGUMENT IN TRANSITION. Eds. Zarefsky, D., Sillars, M., and Rhodes, J. Annandale, VA: SCA, 745-857.

Thomas, D. (1977). "Analysis of Judges and Debaters at NDT." Unpublished. Data collected from the 1977 National Debate Tournament. Used with permission.

Endnotes

1. Nineteen of twenty-six philosophies used were from the 1990 CEDA Tournament Booklet. Of the additional seven philosophy statements, three were from 1989 booklet, one each from the 1988 and 1987 booklets, and the remaining two were free response philosophies solicited for the 1989 Syracuse Debate Invitational Tournament. We assumed that judging philosophies are relatively stable, allowing us to use older forms.

2. For a complete review and critique of the research methodology see Dudczak and Day (1991b).

3. One hundred and twenty potential subjects wrote six or more ballots.

4. Of these twenty-five with philosophy statements and the requisite six ballots, one was unused because his philosophy statement consisted of a statement rejecting the use of philosophy statements.

5. The unusable ballots included 68 blank ballots, 13 illegible ballots, 21 round forfeits, 22 judge disqualified (i.e., a member of the research team), 6 "oral critiques", 5 "useless" comments, and 2 duplicate ballots.

6. The overall inter-coder reliability represents the average of the reliability of each coding task, weighted for the number of documents coded. The intercoder reliability for the philosophy statements was $r = .708$ while the reliability for coding the ballots was $r = .605$.

7. No intercoder reliability for questionnaire responses was required since respondents' answers simply were recorded as provided.

8. The Likert scale values on the questionnaire were recoded to binary values to create an important/unimportant dichotomy for each item.

9. Of the five studies cited (Henderson & Boman 1983; Day & Dudczak 1991; Dudczak & Day 1991a; 1989a; 1989b), only Day & Dudczak (1991) included combined ballots by critic as a unit of analysis. We believe Henderson & Boman's reported high level of consistency is further called into question by their de facto use of ballots as cases (most critics used in their study used a single ballot per critic).