Using the Vanderbilt Television Abstracts to Track Broadcast News Content: Possibilities and Pitfalls

Scott L. Althaus, Jill A. Edy, and Patricia F. Phalen

The Vanderbilt Television News Archive supplies written abstracts for its video collection of news programs. Researchers from many disciplines use the abstracts to locate stories, track specific topics, and measure the evaluative tone of news. This study examines the validity of using abstracts as substitutes for full-text transcripts. Drawing on an analysis of the abstract writing process, we highlight potential sources of error and analyze the correspondence between transcripts and abstracts. Results of a quantitative content analysis suggest abstracts can reflect important elements of news when used at high levels of aggregation but may be unreliable as substitutes for news content.

In the last 20 years, new and powerful research tools have become available to scholars who study both print and electronic media texts. Some of these tools, such as software for conducting content analysis, come with extensive documentation, but other commonly used tools may be employed with little consideration of how they work or what they are intended to do. Among the most important resources for scholars interested in broadcast news is the Vanderbilt Television News Archive. In the early 1970s, the archive began offering indexes and abstracts to help researchers locate items in its extensive collection of videotaped news broadcasts. While some

---

Scott L. Althaus (Ph.D., Northwestern University) is an Assistant Professor of Speech Communication and Assistant Professor of Political Science at the University of Illinois Urbana-Champaign. His research examines opinion surveys and political behavior as channels for mass communication, the construction of policy discourse in the news, and patterns of information acquisition among users of Internet and traditional news sources.

Jill A. Edy (Ph.D., Northwestern University) is an Assistant Professor in the Department of Communication at the University of Oklahoma. Her research interests include political communication and public opinion, with a focus on the role of collective memory in public life.

Patricia F. Phalen (Ph.D., Northwestern University) is an Assistant Professor and Director of Graduate Studies in the School of Media and Public Affairs at the George Washington University. Her research interests include the economics of information and the socio-economic history of media industries.

Earlier versions of this study were presented at the annual meeting of the International Communication Association, June 1-5, 2000, Acapulco, Mexico, and at the annual meeting of the American Political Science Association, September 2-5, 1999, Atlanta, GA. The authors thank John Lynch, director of the Vanderbilt Television News Archive, for his help with this project, as well as the Research Board of the University of Illinois Urbana-Champaign, which funded the collection of content analysis data used in this paper.

researchers use the abstracts for just this purpose, others use them as substitutes for the videotapes. To date, no systematic research has been conducted to determine whether abstracts represent television news stories in the ways that researchers seem to expect.

Although full transcripts of broadcast news stories are available through the Nexis database, scholars turn to the Vanderbilt Abstracts for several reasons. First, Vanderbilt's record of broadcast news stretches back to 1968, while Nexis only goes back to 1990 for CBS and ABC nightly news, and 1997 for NBC. Second, the Vanderbilt Abstracts is available free of charge to anyone with Internet access, whereas Nexis is a subscription service. Third, the Academic Universe search engine, common in academia due to its lower cost, is not as powerful as full-scale Nexis or Vanderbilt Abstracts searches. The standard keyword search of Academic Universe covers only headlines and lead paragraphs, which increases the likelihood of missing stories. And finally, abstracts are short, so researchers can tackle larger data sets than are practical using transcripts or videos.

This study offers guidelines to scholars using the Vanderbilt Abstracts. First, we review the academic studies that rely on this resource and analyze the abstract writing process and its potential effects on research. Second, we report the results of quantitative comparisons between abstracts and transcripts of news broadcasts about the 1990-1 Persian Gulf Crisis. These comparisons do not capture the visual elements so important to television news, and the findings of our modest case study do not generalize to all uses of Abstract content. However, they do illustrate some potential problems that can arise in research using abstracts to represent more complete renderings of news content.

**Literature Review**

The Vanderbilt Abstracts, most fundamentally, should make it easier for scholars to locate relevant news content archived in the collection. Kaid, Tedesco, and McKinnon (1996), for example, used the Abstracts to locate adwatches aired during the 1988 and 1992 election seasons. They watched the adwatch stories and concluded that adwatches are becoming a more prominent part of campaign coverage. The validity of these conclusions depends in part on the Abstracts' reliability in locating all relevant stories.

Other scholars have used abstract entries as proxies for videorecorded news in the Vanderbilt Archive. Scholars seeking to identify broad trends in television news coverage over long time periods have used the Abstracts as a substitute for viewing a large number of stories. Norris (1995) used them to identify countries, themes, and issues covered in foreign news reporting before and after the cold war. Kuklinski and Sigelman (1992) and Ragsdale and Cook (1987) used the Abstracts to measure the visibility of individual members of Congress by searching for their names. In
addition to using the Abstracts to count the number of mentions of a particular topic, issue, or person, some scholars have used the amount of time devoted to stories as recorded in the Abstracts to measure the salience of a person or issue (see Bailey, 1990; Edwards, Mitchell & Welch, 1995; Erfle & McMillan, 1990; Harrington, 1989).

Others have relied even more heavily on the Abstracts, counting on them not only to successfully record the "factual" characteristics, but also to gauge the more subtle and interpretive evaluative dimensions of network news. It is with the latter that we are most concerned in this analysis. Iyengar and Simon (1993), for example, used the Abstracts to distinguish "episodic" from "thematic" frames in television news stories (see Iyengar, 1991). Page, Shapiro, and Dempsey (1987) and Page and Shapiro (1992) used the Vanderbilt Abstracts to measure the direction and balance of media discourse over decades of policy making. They coded statements made by a news source as favoring or opposing a particular policy option, thus relying on abstracts to both capture the discourse and accurately depict its direction.  

Given the number of important theoretical developments that have relied on Vanderbilt Abstract data for empirical validation, a critical question arises: How well do abstracts serve the requirements of scholarly research? In this study, we answer two research questions that address this issue:

RQ1: Are there idiosyncrasies in the Vanderbilt Abstracts that might introduce error in studies that use them?

RQ2: When do the Vanderbilt Abstracts provide a reliable and accurate depiction of broadcast news content?

The Abstract Creation Process

Our analysis of the abstract creation process is based on an interview with John Lynch, director of the Vanderbilt Television News Archive (personal communication, June 11, 1999). Abstracts are written to help locate news stories, not to replace actual television programs. According to Lynch, abstractors "let you know what you are going to be able to see and hear—but [don't] replace what you are going to see and hear." Researchers who use these summaries to replace video should be aware of potential problems with their methodology. In fact, any form of transcription can fail to convey important information, such as a speaker's tone of voice or onscreen manner. Perhaps more importantly, the judgments abstractors make when they write summaries can introduce systematic bias into a research project. In this section, we review abstract writing procedures and highlight ways in which this process can affect research results.

As with the New York Times Index (Althaus, Edy, & Phalen, 2001), precision and continuity are essential to Vanderbilt's abstracting process. In Lynch's words, abstractors have to avoid the "tendency to want to do it a little different this time than you did it for the last [program]." They follow a pattern, first watching the story to
gain a general understanding and to record names of people quoted, and then watching it again to fill in details. While subject terms can be subjective, writers aim to consistently reflect six story characteristics in an abstract: (a) who reported the story and from where, (b) story length, (c) what it was about, in general terms, (d) who was mentioned, (e) who was quoted, and (f) what video or audio clips were used. Currently, a single editor checks all entries for consistency and quality.

Each index of news content is unique. At the New York Times Index, for example, entries are created and catalogued using a list of subject headings, called a thesaurus, created by the publication (Althaus et al., 2001). The Vanderbilt Abstracts is not based on a thesaurus because the archive does not have the staff resources to develop one. Instead, abstractors use the New York Times and Facts on File as guides to improve spelling and terminology, and they rely heavily on institutional memory. As the archive’s director explains, “Our thesaurus is the past.” Indeed, the Vanderbilt Abstracts is much more idiosyncratic than the New York Times Index in connecting similarly themed stories, for the abstractors are more likely to use terms developed in-house to catalogue stories than the language of the news itself.

Many highly context-dependent rules for abstracting are not written down—employees learn them on the job. Abstractors match current stories with similar entries from the past, trying to use terminology that will link the two. This skill takes a long time to develop, which means that a writer’s abstracts improve over time. This learning curve accounts for some inconsistencies in the way entries are written. However, as is the case with the New York Times Index (see Althaus et al., 2001), employees have long tenures with the Vanderbilt Archive. At the time of our interview, the director and senior abstractor had 25 and 23 years of service, respectively, and the video services specialist had worked at the archive for 15 years. Employees with the shortest tenures were the client services representative (10 years) and the second abstractor (2 years).

The archive does have precise rules about the inclusion of names in Abstract entries. Names are included if (a) the person is a significant part of the story or (b) the person speaks in the story. In the first case, abstractors exercise some discretion. They may, for example, group individuals into a single category if they judge that the group name conveys the necessary information. A phrase like “members of Congress against listed” would mean that several congresspersons who are opposed to a policy were named briefly in a newscast. The second case is a more patent rule: When someone is quoted, his or her name appears in brackets with the last name capitalized, followed by a summary of what was said. This means that researchers interested in identifying people who speak in a newscast are likely to find all relevant stories. They would not, however, be able to locate all references to a person’s policy positions.

Significantly, abstractors deliberately avoid reproducing the precise language of news reports. Instead, they try to reflect what the reporter is saying. In order to emphasize the difference between “reflecting” and “quoting,” Lynch instituted a stylistic change to the passive voice in the mid-1970s. He describes the abstractor’s
job as “translating what the reporter is doing or saying into our sort of formatted language." This is true for both the “synopsis sentence” at the beginning of an abstract and the description that follows it. Scholars need to be aware of these professional practices so that they do not mistakenly use abstracts to represent the language of the news or of newsmakers.

The system for creating abstracts has evolved over the last 30 years. Researchers conducting longitudinal studies should identify times when changes in procedures might affect content. In 1973, for example, Vanderbilt University students were hired on a temporary basis to write index and abstract entries for the 1968-1972 backlog of news. Although full-time employees edited the students’ work, these early entries are likely to be less consistent and precise than are those done in later years by the professional staff. Greater care should be taken when searching these entries.

Another important change occurred in 1989. Until then, indexing and abstracting were separate; a staff person either catalogued stories based on the abstracts or wrote the abstracts themselves. Indexers often required the use of specific terms in abstracts to facilitate consistent indexing. They also edited abstracts that “didn’t make sense” in order to guard against the unintentional omission of important information. Overall, the system was a double check on the work of abstractors. When it changed, staff had to learn how to do cataloguing as well as abstracting. Coincidentally, this change facilitated the 1996 move to online entries. As Lynch put it, “Our abstractors are expected to think like indexers . . . to apply a terminology that describes the story, but which may not be a word that was actually used in the story.”

The advent of online databases compounded some problems of terminology. When the Abstracts were a print-only resource, writers could watch news develop over several weeks, adjusting terms to account for new information. Now abstractors have little time to craft online entries—typically about four days from the date of broadcast—and the constant flow of news stories does not allow time for major adjustments. Staff members make minor spelling or formatting revisions routinely, but they only make major changes when the development of a story indicates that important subject terms or other information should be added to early entries. Even then, changes are only applied in the short term—as Lynch explained, “We can’t remember far enough back to do that over a long period of time.”

Until well into the 1980s, abstractors abbreviated words to save time for their typists. Therefore, older abstracts contain abbreviations of key terms that are spelled out completely in later entries. While the search engine used by the archive is sophisticated enough to find both the complete spelling and the standard abbreviation of words, scholars concerned about this should find out what the abbreviation was and do a test search to make sure they are retrieving all relevant stories over time.

Knowledge of the way proper names are handled in a database is also important for researchers. According to Lynch, video sources often include either inconsistent information or no information at all on how to spell personal or place names, and sometimes the accepted spelling of proper names changes over time. For example,
since the 1960s Vanderbilt has consistently spelled the name of Libya’s leader “Muammar el-Kadhafi.” They also maintained the spelling of Mao Tse-Tung, although they spelled the names of people who came to power after him according to the new way of transliterating. Our test showed that the search engine picked up both the archive spelling and the current standard spelling (Qaddāfī and Zedong, respectively), no matter which version was entered as a search term. However, we encourage scholars to test variations on a name so there is no doubt about the results.

Researchers should be aware that special reports are treated differently than the daily news. The focus is on cataloguing more than describing program content in detail. In fact, prior to 1989, entries for these specials do not have any content information other than the title, the date, the length, and perhaps a very short description. A list of participants appears at the bottom of the entry. This means that scholars cannot use the same search terms to locate stories among specials and daily news programs. While such use of keywords may satisfy expectations that quantitative research be systematic and reliable, we believe that validity should never be sacrificed to consistency. It is more important that the tool fit the job.

Tracking the amount of coverage given to a story can be tricky. The archive uses a 10-second clock, so the estimated time can be off by 10 seconds at either end. This means that error is minimized only when differences are large. Additionally, there is no consistent relationship between the length of stories on screen and the length of abstract entries. Even though longer news stories tend to have longer abstracts, there is no ratio that can predict the relationship. Abstractors write very little relative to what the reporter says in long stories, but they may write more than what the reporter says in short stories.

In summary, information provided to us by the director of the Vanderbilt Archive suggests that while great effort is made to construct abstracts according to consistent rules, they can be expected to deviate in systematic ways from full-text news transcripts. These are not the result of mistakes, but rather of conscious professional decisions made by the archive staff to produce consistent and timely abstracts. Their routines and decision rules shape information content in ways that are often quite predictable.

In the following section we analyze policy statements made about U.S. strategy in the 1990-1 Gulf Crisis to show how abstracting procedures can affect the degree to which abstracts mirror transcripts. We are not testing the reliability of the Vanderbilt Abstracts in locating stories but rather the relationship between the content of abstracts and transcripts. While the findings of this quantitative analysis are in many ways linked to the context of this particular case, its coding scheme reflects common uses of abstract data in communication, sociological, and political science research. Rather than offering a definitive statement about the validity of abstract content, we present this case study to illustrate some appropriate uses of abstract data as well as some ways that the use of proxy content can trap the unwary scholar.
Methodology

Our comparative content analysis is based on data from a case study of the policy debate prior to the Persian Gulf War. Transcripts of every ABC World News Tonight (WNT) broadcast during this pre-war period were obtained from Nexis; corresponding abstracts were taken from the Vanderbilt Television News Archive's Web site. A trained coder read every transcript and abstract, identifying stories containing content relevant to the Gulf Crisis. The resulting data set included 1,204 stories that appeared in both transcripts and abstracts.

All transcripts and abstracts were coded for statements supporting or opposing policy options. The identity of the person making the statement (the “source”) was also recorded. Sources were divided into nine categories for analysis: administration officials, Iraqi officials, other foreign officials, members of Congress, U.S. citizens, foreign citizens, Iraqi citizens, experts, and journalists or “unattributed” sources. The four major policy options in the debate were: military force; economic or diplomatic sanctions; negotiations; and a catchall “other policy” category, which included assassination, energy conservation, and vague recommendations (e.g., “do something”). Policy statements were coded only once per source per transcript paragraph, and once per abstract. However, if the same policy position was articulated by two different sources in a paragraph or abstract, each source was recorded with that policy statement.

Intercoder reliability was measured by both agreement and Brennan and Prediger's (1981) kappa, which subtracts a chance agreement term from the initial agreement. “Agreement” meant coders had to code the same statements in the same paragraphs rather than simply reach equivalent totals in the aggregate. For this test, two coders independently analyzed 101 full-text transcripts of WNT stories. Intercoder agreement was .943 for coding sources (Brennan & Prediger's kappa = .936), and .917 for coding policy options (Brennan & Prediger's kappa = .912).

Comparing Abstracts to Transcripts

Our assessment of the abstracts as proxies for full-text transcripts of television news focused on three areas (following Althaus et al., 2001). First, we compared the “density” of information found in transcripts to that found in abstracts: How many stories in each data set contained at least some codeable content, how many policy statements were identified in each data set, and what was the average number of codeable items per story? For obvious reasons, we expected that the abstracts would contain fewer coded stories and fewer coded units than the transcripts. The precise degree of difference should help clarify the limitations of abstract entries relative to transcripts as proxies for television news coverage.

Second, we tested whether the proportional mix of sources and policy statements in abstracts corresponded to that found in transcripts. We predicted that compilation
procedures should lead to systematic differences between the two. Specifically, we expected that "minor" sources and their policy statements would be numerically overrepresented in abstract data due to the tendency of abstractors to name sources once in an entry regardless of how frequently or prominently they are quoted in the story.

Finally, we assessed the correspondence between transcript and abstract data for policy statements made by various sources in order to test the accuracy of abstracts in representing the evaluative tone of news. Because abstractors simplify the content of stories to capture broad contours, we expected sources with consistently low levels of variance in support for administration policy to appear more extreme in the abstracts than in the transcripts. For example, a source that is 90% opposed to administration policy in the transcript data might appear 100% opposed in the simplified abstract entries. We also expected that this type of distortion would be related to the frequency with which a source appears in WNT coverage. Fewer policy statements attributed to a source in WNT would mean greater distortion in the abstracts.

Information Density

The density of information in a content proxy is one way to assess its validity. We examined density of information in two ways: (a) the total number of policy statements coded for each data set, and (b) the average number of policy statements in stories containing at least some policy content. The closer abstracts come to matching the number and density of policy statements in the full-text transcripts, the more valid they should be as a proxy for the content of news broadcasts.

The abstract data set contained 29% of the number of policy statements found in the corresponding transcripts (243 compared to 836). Most of the 1,204 World News Tonight stories relevant to the Gulf Crisis contained no policy content at all, with only 30.6% of transcripted stories and 14.7% of abstract entries containing at least one codeable policy statement. The average number of policy statements in stories with at least some policy content was 1.37 statements for abstract entries and 2.27 for transcripts. While on this measure the abstracts contain only about half as much information as the transcripts, they are relatively more information-dense than New York Times Index entries, which had an average of 1.50 policy statements compared to the 5.09 policy statements in the average full-text Times story containing at least some policy content (Althaus et al., 2001). Moreover, the New York Times Index contained only 15% of the number of coded units found in full-text stories (Althaus, et al. 2001).

On this measure, the Abstracts would appear to represent broadcast news content better than the New York Times Index represents the New York Times. However, the superiority of television abstracts stems in large part from the fact that newspapers carry a great deal more substantive content than television newscasts. It should be
easier for an abstract to convey the content of a 750-word television story than a 2,000-word newspaper article.

Consistency in Overall Distribution of Sources and Policy Statements

Because the abstracts contain only one third the number of policy statements found in the transcripts, we have reason to question how well they represent news content. Given the common uses of abstract data by social scientists, the performance of the abstracts in three areas is of special concern: reproducing the overall distribution of sources appearing in news broadcasts, reproducing the overall distribution of supportive and opposing statements made about particular policies, and reproducing the net support or opposition to a policy articulated by different categories of sources (following Althaus et al., 2001). Chi-square tests were conducted to determine whether these observed differences between the abstract and transcript data were statistically significant. It is important to point out that because our analysis considers the entire population of stories relevant to the Gulf Crisis, we can draw inferences about the validity of proxy data in this particular case from all of the observed differences reported below, regardless of their statistical significance. We nonetheless report the significance of these differences as a measure of the confidence we have in generalizing from this particular case to similar uses of abstract data in other contexts.

We found that the abstracts provided a fairly accurate representation of the aggregate distribution of sources in the transcripts (see Table 1). The greatest difference in proportions between data sets was for foreign officials, where the abstracts overrepresented the percentage of foreign officials making substantive policy statements by 6 points. The remaining differences are small and of little consequence: correlating the rank order of source categories between the data sets produces a Spearman’s rho of .95 (p < .001), indicating almost perfect consistency in the ordering of most- to least-mentioned sources. These findings confirm that the special attention given by abstractors to mentioning the names of sources produces an accurate record (at least in the aggregate) of who speaks in the news.

These results are similar to those we documented in our study of the New York Times Index and the Libya crisis (Althaus et al., 2001). The Index also reproduced the distribution of sources quite accurately, aside from a slight tendency to overrepresent foreign officials. As in the case of the New York Times Index, overrepresentation of foreign sources in the Vanderbilt Abstracts may be explained by the abstractors’ practice of making sure all directly quoted sources appear in the summary. While a variety of foreign officials are quoted briefly, a small number of administration sources are permitted to speak at length in broadcast news. Abstractors include the statements of each without regard to length, and as a result, the abstracts, like the Index, tend to exaggerate the prominence of those who speak briefly.

The Abstracts fares somewhat less well in reproducing the distribution of aggregate support for various policy alternatives reported in the news. Support for a specific
policy was calculated by dividing the number of pro-statements by the total number of statements—pro, con, or neutral—that were made about it. We also constructed a 3-point measure of support for the administration by coding every policy statement in the transcripts and abstracts as supportive, neutral, or opposed to positions taken by the administration. Support for the administration was calculated by dividing the number of supportive statements by the total number of statements about administration stances. This allowed us to assess general levels of reported support for President Bush’s three-pronged approach to dealing with the Gulf Crisis (relying on sanctions and shunning negotiations while threatening and ultimately using force).  

Table 2 shows that the abstract data reflect the aggregate levels of support for force and negotiations recorded in the transcripts more accurately than for sanctions and other policies. The largest discrepancy (which achieves marginal levels of statistical significance) is in support for sanctions—81% of statements about this policy in the abstracts supported it, while only 63% of sanctions statements made in the transcripts were supportive. When individual policy statements were transformed into a global measure of support or opposition to administration policy, differences between abstracts and transcripts decreased. Despite the sometimes large discrepancies on specific policies, aggregate support for the administration differed only 4 percentage points between the two data sets. The pattern of differences in Table 2 suggests that the greater the number of individual statements made about a specific policy, the more likely that the level of support in the abstracts will approach the level of support in the transcripts. Once again, these results are similar to our finding
Table 2
Consistency in Overall Distribution of Support for Policy Alternatives

<table>
<thead>
<tr>
<th>Policy</th>
<th>% Support in Transcripts (No.)</th>
<th>% Support in Abstracts (No.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support for Administration Policies</td>
<td>45.8 (836)</td>
<td>50.2 (243)</td>
<td>+4.4</td>
</tr>
<tr>
<td>(All Policies Combined)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support for Individual Policies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Force</td>
<td>53.1 (416)</td>
<td>52.5 (120)</td>
<td>−0.6</td>
</tr>
<tr>
<td>Negotiations</td>
<td>58.0 (212)</td>
<td>53.2 (62)</td>
<td>−4.8</td>
</tr>
<tr>
<td>Sanctions</td>
<td>63.0 (92)</td>
<td>80.6 (31)</td>
<td>+17.6†</td>
</tr>
<tr>
<td>Other Policies</td>
<td>71.6 (116)</td>
<td>80.0 (30)</td>
<td>+8.4</td>
</tr>
</tbody>
</table>

Note: Cells contain the percentage of all statements about a policy that are supportive of that policy. Parentheses contain the total number of statements made about a policy.
†Two-tailed p < .10.

that more data points improved estimates of policy support in proxies for the New York Times (Althaus et al., 2001).

In addition to representing the overall distribution of sources and policy statements in the news, abstracts are often used to identify variations among the different news sources in support for a policy or an administration (e.g., Bennett, 1990; Bennett & Manheim, 1993; Page et al., 1987). Table 3 displays the results of this type of analysis for the entire seven-month time period, using five source categories especially relevant to the American policy debate in the Gulf Crisis: administration officials (including President Bush), American citizens (including “person on the street” interviews, poll results, and demonstrations), foreign government officials (excluding Iraqi government representatives), members of the U.S. Congress, and nonpartisan experts brought in to provide analysis and perspective. Two questions are of particular importance in such an analysis: how accurately the abstracts reproduce each source’s level of support when compared to the transcripts, and whether differences between the abstracts and the transcripts in reported levels of support are systematic or random.

Our findings reveal that the positions attributed to various sources in the abstracts are often dramatically and consistently different from the levels of support recorded for those sources in full-text transcripts. Of the 25 sets of measures presented in Table 3, 15 have differences in levels of support between abstracts and transcripts that are greater than 10 percentage points; 7 show differences greater than 20 points, and 4 have differences of 30 points or more. As we might expect, larger errors occur when fewer data points are available, but even prominent sources can be misrepresented to a surprising degree. For example, the transcript data show that foreign officials tended toward neutrality or opposition toward administration policies, voicing high levels of support for negotiations and sanctions while remaining almost evenly split in their support for the use of military force. On balance, this mix of views caused the
## Table 3
Consistency in Reproducing Levels of Support for Administration’s Policies, by Source Category

<table>
<thead>
<tr>
<th>Source Category</th>
<th>% Support in Transcripts (No.)</th>
<th>% Support in Abstracts (No.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administration Officials</td>
<td>69.6 (313)</td>
<td>70.2 (94)</td>
<td>+0.6</td>
</tr>
<tr>
<td>U.S. Citizens</td>
<td>43.1 (181)</td>
<td>38.6 (44)</td>
<td>-4.5</td>
</tr>
<tr>
<td>Foreign Officials</td>
<td>42.3 (97)</td>
<td>55.8 (43)</td>
<td>+13.5</td>
</tr>
<tr>
<td>Members of Congress</td>
<td>36.4 (66)</td>
<td>43.8 (16)</td>
<td>+7.4</td>
</tr>
<tr>
<td>Experts</td>
<td>25.9 (27)</td>
<td>41.2 (17)</td>
<td>+15.3</td>
</tr>
<tr>
<td>Force</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administration Officials</td>
<td>88.2 (127)</td>
<td>91.9 (37)</td>
<td>+3.7</td>
</tr>
<tr>
<td>U.S. Citizens</td>
<td>48.2 (137)</td>
<td>34.3 (35)</td>
<td>-13.9</td>
</tr>
<tr>
<td>Foreign Officials</td>
<td>53.8 (39)</td>
<td>58.8 (17)</td>
<td>+5.0</td>
</tr>
<tr>
<td>Members of Congress</td>
<td>34.4 (32)</td>
<td>40.0 (10)</td>
<td>+5.6</td>
</tr>
<tr>
<td>Experts</td>
<td>0.0 (4)</td>
<td>50.0 (4)</td>
<td>+50.0</td>
</tr>
<tr>
<td>Negotiations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administration Officials</td>
<td>29.0 (107)</td>
<td>41.2 (34)</td>
<td>+12.2</td>
</tr>
<tr>
<td>U.S. Citizens</td>
<td>85.0 (20)</td>
<td>100.0 (2)</td>
<td>+15.0</td>
</tr>
<tr>
<td>Foreign Officials</td>
<td>84.6 (26)</td>
<td>44.4 (9)</td>
<td>-40.2*</td>
</tr>
<tr>
<td>Members of Congress</td>
<td>100.0 (7)</td>
<td>100.0 (1)</td>
<td>0.0</td>
</tr>
<tr>
<td>Experts</td>
<td>20.0 (5)</td>
<td>42.9 (7)</td>
<td>+22.9</td>
</tr>
<tr>
<td>Sanctions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administration Officials</td>
<td>77.8 (36)</td>
<td>92.3 (13)</td>
<td>+14.5</td>
</tr>
<tr>
<td>U.S. Citizens</td>
<td>50.0 (2)</td>
<td>100.0 (2)</td>
<td>+50.0</td>
</tr>
<tr>
<td>Foreign Officials</td>
<td>83.3 (18)</td>
<td>80.0 (10)</td>
<td>-3.3</td>
</tr>
<tr>
<td>Members of Congress</td>
<td>81.8 (11)</td>
<td>66.7 (3)</td>
<td>-15.1</td>
</tr>
<tr>
<td>Experts</td>
<td>75.0 (4)</td>
<td>100.0 (1)</td>
<td>+25.0</td>
</tr>
<tr>
<td>Other Policies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administration Officials</td>
<td>72.1 (43)</td>
<td>70.0 (10)</td>
<td>-2.1</td>
</tr>
<tr>
<td>U.S. Citizens</td>
<td>77.3 (22)</td>
<td>100.0 (5)</td>
<td>+22.7</td>
</tr>
<tr>
<td>Foreign Officials</td>
<td>64.3 (14)</td>
<td>71.4 (7)</td>
<td>+7.1</td>
</tr>
<tr>
<td>Members of Congress</td>
<td>81.3 (16)</td>
<td>100.0 (2)</td>
<td>+18.7</td>
</tr>
<tr>
<td>Experts</td>
<td>50.0 (14)</td>
<td>80.0 (5)</td>
<td>+30.0</td>
</tr>
</tbody>
</table>

Note: Cells contain the percentage of all statements made by a source about a policy that are supportive of the administration’s position on that policy. Parentheses contain the total number of statements made by a source about a policy.

*Two-tailed p < .05.
global measure of administration support among foreign officials to reach only 42%. Yet the portrayal of foreign officials in the abstracts suggests a net approval for administration policies of 56%—a nearly 14-point difference that misrepresents the majority position found in the transcripts. Moreover, the abstracts suggest that foreign officials opposed a diplomatic solution, while the transcripts record near-unanimous support for negotiations—a 40-point difference in apparent levels of support for the second-most-frequently discussed policy alternative in these data. Similarly, the abstracts greatly exaggerate opposition to the use of force among American citizens, who appear evenly divided on this issue in the transcripts but opposed to force by a two-thirds majority in the abstracts.8

In 17 of the 25 categories considered in Table 3, abstracts overstate reported support for administration policies. This tendency is especially apparent in administration support scores, where four of the five measures of difference between abstracts and transcripts are positively signed. While nearly all observed differences across categories fall short of statistical significance due to the small number of data points, the magnitude and consistency of patterns are of great substantive importance.

Consistency in Reproducing Levels of Support over Time

Our final set of validity tests focused on the ability of abstracts to reproduce trends in news content over time, a common use of content proxies in communication and public opinion research. Given that more data points in the abstracts produce better estimates, the validity of proxies should decline as data are disaggregated into smaller units for time series analysis. As in the previous discussion, our concern is with the accuracy of abstracts in reproducing trends found in transcripts and the possibility that trend differences are systematic rather than random.

We divided the seven months of content data into two-week periods beginning three days before the invasion of Kuwait and ending on the day that the ground war started. We recognize that in any trend study of news content, the time periods of interest will be determined by the focus of a particular research agenda. In the case of the Gulf Crisis, dividing news coverage into discrete two-week periods seemed a good way to capture the ebb and flow of support for administration policy over time while ensuring that each time period contained a reasonable number of data points. Trends in the percentage of policy statements supporting administration positions are especially informative for analyzing the debate over American policy during the Gulf Crisis, and they have the added advantage of minimizing the amount of fluctuation due merely to small numbers of data points. The mean number of data points across the 15 time periods is 16.2 statements for the abstracts and 55.7 for the transcripts.

When examining changes over time in the level of support for administration policy, both absolute levels of support and trends in support are important. Both are widely divergent in comparisons of the abstracts and the transcripts. Figure 1 shows that the levels of support for administration policy that appeared in the abstracts frequently diverged from those that appeared in the transcripts. For instance, in the
first month of the Gulf Crisis (two-week periods beginning 7/29 and 8/12), the abstracts record a policy debate with more than 60% support for administration policy, while the transcripts reveal a debate with only 45% support in the 7/29 period and 51% in the 8/12 period. At the end of September and beginning of October, the abstracts portray a policy debate that was 78% supportive, while the transcripts recorded only 53% support, a 25-point difference. Between early November and late January, the critical period of deliberation leading up to and including the start of offensive action against Iraq, the abstracts portray a strikingly different policy debate than that found in the transcripts. The abstracts suggest that support over this period rose steadily from 47% in early November to a high of 75% in the latter half of December, before plummeting 17 points in early January and 26 points in late January to a low of 32% after the start of the air war. In contrast, the transcripts during this same period show a much more stable debate over administration policy, hovering just above 50% support in November and December before declining gradually to a low of 33% support in early February.

Figure 1
Trends in Overall Support for Administration Policies

While the movement between these trends is roughly parallel over time, with a correlation of .74 (p < .01), abstract data tend to overstate levels of support relative to transcript data by an average of 6.5 points over the 15 time periods. In 7 of the 15 time periods, levels of support in the abstracts exceeded those in the transcripts by more than 10 points. Just as importantly, the bivariate regression slopes of the two series are signed in opposite directions, taking a value of −.64 in the abstract data
but .25 in the transcript data, a difference that approaches conventional levels of significance ($t_{26} = -1.56, p = .13$). A researcher using the abstracts might conclude that the news gradually became more negative toward the administration’s policies, while someone using the transcripts might surmise just the opposite.

These findings sharply diverge from those of our *New York Times Index* study (Althaus et al., 2001). Although the overall correlation of the comparable distribution (p. 719) in that earlier study is similar, about .76 when including only those weeks of coverage that contain at least one data point, analysis of the *Index* alone would not lead to false conclusions about the general levels or trends in administration support contained in the full text of the *New York Times*. This may be because discrepancies in the *Vanderbilt Abstracts* appear to be systematic, overestimating administration support in nightly news reports, while the errors in the *New York Times Index* appear to be relatively more random.

Disaggregating these trends in broadcast news further by examining net support for administration policies among various sources reveals extreme differences between abstract and transcript data (not shown). While reported levels of support over time correlate between abstract and transcript data fairly well for administration sources ($r = .63, p = .01$), the two trends in support for administration policy for the next largest source categories diverge so much that they are statistically unrelated to one another: The correlation between trends for American citizens was only .14 ($p = .74$); for foreign officials it was -.13 ($p = .73$). The abstracts understated levels of support among American citizens by 20 points in an average time period, and they overstated levels of support among foreign officials by 15 points in an average time period. It would seem that abstract data are particularly unsuited to this type of analysis. Here again, these findings dovetail with those of our earlier study. The *New York Times Index* is also unable to reasonably represent the content of the *New York Times* at such low levels of aggregation.

**Discussion**

Our findings suggest many guidelines for researchers using the *Vanderbilt Television Abstracts* to represent the content of TV news. Abstracts are constructed with specific purposes in mind, and these purposes may or may not fit the needs of particular research projects. For example, abstractors do not intend for the abstracts to be used as a surrogate for videotapes or transcripts, and therefore they do not write them to meet such research needs. Nevertheless, as a practical matter, many researchers do use the *Vanderbilt Abstracts* instead of the videotapes to answer questions that are of great importance to the social sciences. They should be aware of several issues raised in this study, and they are advised to modify their research designs and data analyses to account for the likely pitfalls in abstract data.

Researchers must be aware that abstractors do not attempt to reproduce the language used by journalists or their sources. In fact, wording may be entirely that of
the abstractor. This means that scholars can use abstract entries to advantage if they want to capture the broad themes and general ideas conveyed in stories. However, if they are studying the exact words used in broadcast reports, they would not be well-served by using the Abstracts. Researchers should also understand that optimum search terms for locating stories might never appear in the stories themselves. To make best use of the Abstracts, they should use a snowball approach, developing and testing keywords to make sure all relevant content is retrieved. Scholars can also use the abstracts to track people who are quoted in the news, but they can't rely on these summaries to pick up every mention of a person's name.

Over-time discrepancies in abstract content should be expected due to changes in archive staff and to the learning curve effect. We believe that the adherence to professional guidelines in crafting abstract entries and the availability of word-search tools to locate stories overcome the worst problems of inconsistency. However, while it is true that the current archive staff is quite senior, employees have moved from one job to another over the years, and some abstractors have served comparatively short tenures. Development of search terms, and potential problems that may result, should be documented in discussions of methodology. We would point out the careful documentation of Gilens (1996) and Paletz, Short, Campbell, Cooper, and Oeslander (1980) as exceptional examples of how this can be done.

Our quantitative analysis confirms that abstract data can accurately reflect some dimensions of full-text transcripts, but it also sounds a warning about leaning too heavily on proxy measures of news content. While the abstracts do a good job of reproducing the aggregate distribution of news sources and topics, they provide, at best, imprecise representations of the evaluative tone of policy statements. At worst, they can produce decidedly inaccurate and misleading portrayals of what many sources actually say in the news. When the picture of news content in the abstracts differed substantially from that in full-text transcripts, their tendency was to overstate levels of support regardless of policy or source category. However, levels of opposition can also be substantially inflated. We cannot be sure if these patterns would hold for domestic policy cases or even other foreign policy cases, but the tendency toward misrepresenting levels of support seems to be of a nature that is not easily amenable to statistical or other correction. Studies attempting to discover whether the media are “watchdogs” or “lapdogs” with regard to the administration may be especially vulnerable to these kinds of problems.

As was the case with the New York Times Index (Althaus et al., 2001), the more numerous the data points for any source, the more accurate the measures of support provided by the Vanderbilt Abstracts tend to be. Because the Abstracts, like the Index, contain relatively small numbers of policy statements compared to full text, breaking them further into individual source categories for the purpose of ascertaining differences in support levels invites misrepresentation of the full text data. Yet our findings suggest no firm threshold in number of data points beyond which reasonable correspondence to transcript data should be expected. On the contrary, it would seem that the only way to ensure that the Abstracts provide a reasonable approxi-
formation of policy statements found in full-text content is to compare the abstracts directly to transcript data on the research topics of interest in particular studies, as illustrated by Steele's (1997) careful approach to confirming the validity of abstract data.

Our comparisons between the *New York Times Index* and the *Vanderbilt Abstracts* help define the key aspects of a successful proxy. Taken at face value, the Abstracts might appear to be the better indicator of full-text content. Charged with abstracting the relatively simple stories of television news, this resource captures relatively more of the information contained in the original source documents as measured by information density. In contrast, the *Index*, charged with capturing the longer and more complex stories of the *New York Times*, captures relatively little of the original documents. However, our analyses suggest that the Abstracts may fare worse as a proxy for television news coverage than the *Index* does as a proxy for *New York Times* coverage. It appears that the absolute number of data points provided by the proxy is more important than the relative density of information in the proxy.

The key is not how much of the original source material is captured by the proxy, but whether there are meaningful patterns in the elements that are captured and missed. In the *New York Times Index*, the errors are mostly random. In the *Vanderbilt Abstracts*, the errors are more systematic, and these errors may well lead unwary researchers to incorrect conclusions about the content of broadcast news.

**Notes**

1 For other examples of this use of the Abstracts, see Bailey, 1999; Behr and Iyengar, 1985; Bollen and Phillips, 1982; Broh, 1983; Burden and Mughan, 1999; Dearing, 1989; Erfile, McMillan, and Grofman, 1990; Gilens, 1996; Harrington, 1989; Holden, 1986; Kuklinski and Sigelman, 1992; Paletz, Short, Campbell, Cooper and Eslander, 1980; Pasadino, 1990; Robinson and Appel, 1979; Singer and Ludwig, 1987; Singer, Rogers, and Corcoran, 1987; Smith and Hogan, 1987; and Steele, 1997.


3 See Danielian and Page, 1994; Durr, Gilmour and Wollbrecht, 1997; and Walters and Hornig, 1993.

4 The case study spans the period between August 2, 1991, the date that Iraq invaded Kuwait, and February 24, 1991, the day that the Allied ground assault into Kuwait was reported to American audiences.

5 The selection process identified 1,297 relevant stories from the transcript data and 1,221 relevant entries from the abstract data. Ninety-three stories in the transcripts had no corresponding abstract entries due apparently to program preemptions at the recording site used by the Vanderbilt archives. Similarly, 17 abstract entries had no corresponding story in the transcript due to missing transcripts in the Nexis database. Further analysis revealed that most of the missing data came from weekend broadcasts that were apparently preempted or bumped to a later broadcast time by sporting events. The other main source of discrepancies was an expansion of the time allotted to WNT broadcasts at the start of the air war in mid-January.
While standard abstract entries were recorded for many of the stories appearing in these special newscasts, the Vanderbilt archive staff shifted to a format for abstracting these broadcasts that collapsed an hour or more of broadcast time into a single paragraph. These entries could not be compared in a meaningful way to the transcript data and were therefore counted as “missing.” The 1,204 stories in this analysis represent all of the stories that could be parallel coded in both sources.

These tests were conducted separately for each row in the two columns by constructing a series of $2 \times 2$ tables in which the number of statements made by a particular source and the number of statements made by all other sources were compared for abstract data and the full text data. Similar chi-square tests were run on the row comparisons in tables 2 and 3. Because $2 \times 2$ tables containing cells with low frequencies tend to deviate from the chi-square distribution, in accordance with standard practice, Fisher’s exact test was used to determine the significance of the observed differences (Blalock 1979: chapter 15; Bohrnstedt & Knoke 1988: chapter 9).

To construct this variable, pro-force, con-negotiation, and pro-sanctions statements were given a value of +1 (supportive of administration positions), con-force, pro-negotiation, and con-sanctions statements were given a value of −1 (opposed to administration positions), and any pro or con statement about other policy options was given a value of 0 (neutral toward administration positions). The administration’s position on negotiation changed somewhat over time, becoming at least publicly open to the idea as a “last ditch” effort to forestall a bombing campaign and eventually agreeing to hold talks. Yet the administration consistently signaled its unwillingness to settle for anything less than complete withdrawal of Iraqi forces from Kuwait, and the private memoirs of key administration officials (e.g., Baker, 1996, pp. 350-364) clarify that the talks were primarily held to remove a potential objection to the use of force. No serious consideration of entering into formal negotiations with Hussein was ever undertaken or suggested by the administration.

These are precisely the kinds of distortions that we observed between the New York Times Index and the New York Times itself (Althaus et al., 2001). The Index overstated foreign support and understated citizen support for administration policy in the 1986 Libya crisis.

References


