Using Substitutes for Full-Text News Stories in Content Analysis: Which Text Is Best?

Scott L. Althaus; Jill A. Edy; Patricia F. Phalen


Stable URL: http://links.jstor.org/sici?sici=0092-5853%28200107%2945%3A3C707%3AUSFFNS%3E2.0.CO%3B2-2

American Journal of Political Science is currently published by Midwest Political Science Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/mpsa.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
Using Substitutes for Full-Text News Stories in Content Analysis: Which Text Is Best?

Scott L. Althaus  University of Illinois at Urbana-Champaign
Jill A. Edy  Middle Tennessee State University
Patricia F. Phalen  The George Washington University

This study examines the validity of newspaper indexes, lead paragraphs, and headlines as representations of full-text media content. We analyze the effects of production decisions on content and categorization in the New York Times Index, based on interviews with its senior editor. We then compare the content of three proxies with that of full-text articles by conducting a parallel content analysis of New York Times stories covering the 1986 Libya crisis and their corresponding Index entries. The study suggests that proxy data can be used to roughly estimate the broad contours of Times coverage but do not reliably represent several key aspects of New York Times reporting.

In a recent Workshop article, Woolley (2000) examined the use of media indexes for measuring media attention and counting various types of events. The central concern of Woolley’s study was the degree of correspondence between the occurrence of real-world events and media reports of those events, and he found that published indexes of media content suffer from several validity problems when used in event count research (see also White 1993). Another important concern is whether media indexes adequately represent the actual content of the news itself. In this article, we test a common practice employed by political scientists to analyze news content: the use of proxies, such as index entries, headlines, and lead paragraphs, as surrogates for the actual content of the news.

Because of time, cost, and access constraints, many researchers code proxies rather than the full content of news texts. Even scholars who ultimately code full text often rely on indirect indicators of news content, such as subject headings in printed indexes and keywords in news databases, to locate that text. Thus, at some level, virtually all content analysis relies on surrogates for full-text content in one form or another. Many important theoretical studies rely on evidence from news proxies, including work on intrastate political conflict (Davenport 1995; Jackman and Boyd 1979; White 1993), the development of social movements (McAdam 1982, 1983), the indexing hypothesis used to explain elite influence in news content (Bennett 1990; Bennett and Klockner 1996; Bennett and Manheim...
Using Proxies to Locate and Represent Media Content

News indexes are critical research tools for locating news content because of the limited usefulness and inherent difficulty of drawing true random samples of media content. Where scholars are interested in a particular sort of coverage, such as foreign or campaign news, sampling by days often produces a great deal of irrelevant material. Further, random sampling can be labor-intensive, time-consuming, and expensive. While it may seem that electronic databases like Lexis/Nexis could streamline this process, these databases are themselves limited as research tools (see Snider and Janda 1998; Kaufman, Dykers, and Caldwell 1994). One obvious problem is that they only contain relatively recent material and are therefore incomplete records of media content (in the case of The New York Times, articles in the Nexis database go back only to June 1980). For scholars interested in studying long-term trends, historic events, or alternative newspapers, indexes remain a primary means of locating coverage.

Researchers who use indexes to locate media content often assume that index writers use consistent subject headings and reliable systems for categorizing news stories. Deficiencies in either area will introduce error into the research (see Zollars 1994). For example, Hertog, Finnegan, and Kahn concede that their findings regarding the relative prominence of various diseases in public health news might in part be an artifact of the newspaper indexes they used to obtain their data: “Another explanation . . . is that the indexing format for stories in one or more of the newspapers may have changed at some point in time . . . resulting in an increase or decrease in the number of entries . . . even in the absence of a correlation among actual story numbers in the newspapers themselves” (1994, 302). When the volume of news coverage or the number of real world events is tracked through subject headings, samples and event counts may not reflect changes over time in the wording of subject headings or in the ways subject categories were defined by index writers. Results could then reflect methodological artifacts rather than real differences. An understanding of the process of index production is crucial to ensuring the validity of research results, even when scholars merely use indicators to select stories that later will be analyzed in depth.

Identifying texts is only one important use of index data; a second is as a proxy for complete articles. This use is motivated by the human and monetary resources needed to complete any extensive coding of media con-
tent. Most content analysis protocols are designed with very specific and narrowly defined research questions in mind, so it is difficult to amortize the cost of collecting data across many projects. While enormous improvements have been made recently in computer software for content analysis, many scholars continue to rely on human coders for nuanced, though reliable, readings of text. Large databases of precoded news content, the textual equivalent of the large databases of survey results now available, have yet to become widely popular. This makes it imperative for researchers to seek ways of making the data collection process more efficient.

Some scholars apply content analysis to relatively narrow case studies (e.g., Althaus et al. 1996; Entman and Page 1994; Entman 1991; Page 1996; Zaller and Chiu 1996). This allows them to examine all the news relevant to an issue that they can locate. An alternative tactic is to create a relatively narrow time limit for the study, as Entman and Rojecki (1993) did in their study of the nuclear freeze movement during the Reagan years. The reliability and validity of this type of coding is likely to be better than that of proxy coding, as we demonstrate below, but case studies limit the generalizability of results.

Sampling texts by day or within subject categories makes it practical to study broader issues over longer time frames than would be possible with a census of news coverage (Entman 1992; Patterson 1993). Yet despite the advantages of such an approach, sampling makes it difficult to answer the “why” questions so critical to theoretical development. The researcher can document the changes that appear in the coverage but will often have a difficult time explaining them, for a single, critical story can dramatically alter the shape of the coverage. If that article is not captured in the sample, the scholar may see the symptoms but miss the cause. For example, the Reagan administration changed its preferred policy position several times throughout the Libya crisis. A census of articles reveals journalists and political elites reacting to the administration’s shifts. Had we been unaware of the administration’s changes, we would be at a loss to explain the changed behavior of other actors.

Some scholars, attempting to maintain continuity while at the same time examining lengthy policy debates or broad aspects of political culture, have coded proxies in place of actual news content. The New York Times Index has been used to represent coverage of war (Bennett and Manheim 1993), foreign policy (Bennett 1990), and terrorism (Crelinsten 1989), and to select articles for a series of audience experiments (Neuman, Just, and Crigler 1992). Headlines have also been used as substitutes for full text (Bennett and Klockner 1996; Hughes 1995; Johnson and Wanta 1995), as have lead paragraphs (Bennett and Lawrence 1995; Johnson and Wanta 1995).

However reasonable and practical such proxies for full-text news content may be, our study suggests that their use can be problematic. Some limitations have been noted by the researchers themselves. Hughes cautions that his findings about press portrayals of the Clinton and Carter “honeymoon” periods “may be little more than a partisan skew in the construction of Times headlines” (1995, 848). Yet despite the concerns raised by researchers about the validity of content proxies, to date no one has examined the extent to which they accurately depict the evaluative tone and topical content of full-text coverage.

To our knowledge, the present study represents the first systematic attempt to test the validity of news proxies for these applications in political science research.

Methodology

The New York Times data used here were originally collected to study media coverage of the policy debate leading up to the U.S. air strikes against Libya in 1986 (see Althaus et al. 1996). The Libya bombing was typical in two important ways of the many small-scale military interventions that have become more common since the end of the Second World War: the Libya crisis erupted and then subsided relatively quickly, and the president’s intention to use force was telegraphed well in advance of the April bombing. We collected the content data for the original study in order to test and refine a theoretical approach in the political communication literature known as the “indexing hypothesis” (Bennett 1990) that is used to predict the breadth of policy discourse in the news.

The indexing hypothesis specifies conditions under which news discourse will be relatively more supportive or critical of an administration’s preferred policies.

For the present study we conducted a parallel coding of New York Times Index entries corresponding to the full-text stories gathered for the original study. The coding protocol used in this analysis is comparable to protocols used in studies of indexing behavior in political news coverage (e.g., Bennett 1990; Bennett and Manheim 1993) as well as more generally to simple valenced coding schemes that capture evaluative tone in news stories (positive, negative, or neutral) and have been widely used in public opinion and political communication research (e.g., Caldeira 1987; Dalton, Beck, and Huckfeldt 1998; Hofstetter 1976; Just et al. 1996; Page, Shapiro, and Dempsey 1987; Patterson 1993; Patterson and McClure...
two teams of researchers. In order to count as "agreement," coders had to code the same sources and statements in the same paragraphs rather than simply reach equivalent totals in the aggregate. Intercoder agreement for the Times was .847 (Brennan and Prediger's kappa = .829), and for the Index was .854 (Brennan and Prediger's kappa = .841).

In the sections that follow, we first discuss the procedures used to compile the New York Times Index, paying special attention to how these practices may affect political science research that relies on the Index. Our description of the procedures used by the editors of the New York Times Index to catalogue the content of the Times is based upon interviews Patricia Phalen conducted with Harvey Holmes, senior editor of the Index, in June 1997 and August 1998. Mr. Holmes has been on staff with the Index since 1967. We use these data to illuminate the potential pitfalls of using the Index both as a means of locating media content and as a proxy for that content. Then, we turn our attention to a quantitative analysis of the validity of various proxies for representing the full text of the Times.

Potential Sources of Bias in the New York Times Index

As Woolley (2000, 158) points out, an understanding of media practices leads to better interpretation of media content. These media practices, such as the indexing procedures reviewed in this section, are a potential source of bias in political science research. At the New York Times, a staff of twelve creates index entries for on-line and print distribution. They work directly from the printed newspaper, and each typically indexes one section of the paper on a consistent basis. Indexers write abstracts for stories in their section and choose subject terms for major headings as well as cross-referenced headings. Abstracts are released daily to the online system, and bi-monthly, quarterly, and annually in print publications.

Omitted Stories

Nearly all of the Times stories are indexed. The only items that are not included in the Index are society news, letters appearing elsewhere than the editorial pages, and "question and answer" columns. Before 1997, the Index did not include most letters to the editor. Letters from average citizens were excluded; letters from influential people were indexed. Influential people included, for example, senators, foreign ambassadors, experts on a given topic
(such as Libya), or prime ministers. In 1997 the Index began a transition, completed in 1998, to including all letters to the editor that appear on the editorial page. About fifteen letters were indexed per day in 1998, compared to the three to four letters per week that were done previously. Thus, Index entries for letters prior to 1997 are systematically biased. Even now, although letters on the editorial page are indexed, letters found elsewhere in the paper are not. Because other newspaper indexes may not be as complete as the Times's, we would recommend that any researcher using a media index either to locate coverage or as a proxy for media content should first find out what types of stories are routinely left out of the index. It might also be significant to a given project whether the types of stories included in an index have changed during the time period being studied.

Changes in Subject Headings

Without guidelines, different individuals would index certain stories quite differently. Indexers avoid some potential problems by writing down categorization rules as stories develop. For example, a story covering the Clinton-Lewinsky scandal would be indexed under "ethics" and "Lewinsky" for every article, but "Kenneth Starr," "special prosecutors," and "perjury" would only be used when they are actually mentioned in the article. There could be any number of rules that make sense for a given story, and codifying the criteria minimizes confusion caused by inconsistencies.

Indexers work from a comprehensive list of subjects called a thesaurus. At present the thesaurus contains approximately 3,500 subjects, and editors ensure that subject terms are applied consistently. Every Index entry is assigned to a major subject heading and is often cross-referenced to other subject terms. Prior to 1947, the full text of each Index entry could be included under each of the appropriate subject headings, but when the volumes became unmanageably large, the cross-referencing scheme was substituted. In this scheme, the complete entry appears under one subject heading and is cited by date under the other relevant headings. Research based on the pre-1947 Index might be subject to errors of double-counting, and research in later years might undervalue content if cross-referenced articles are overlooked.

Prior to 1983, when the subject thesaurus at the Times was rewritten, indexers used up to 25,000 subject terms. Up to that point, each writer specialized in only a few subject areas because an indexer could only memorize terms corresponding to five or ten major topics. Under this system, one person had to read through the entire paper and assign every story to a subject expert. With the shorter thesaurus it became possible for indexers to handle more subject areas, and this new efficiency prompted the Index to assign work by newspaper section rather than topical specialty. Although sections generally include similar topics, overall the indexers today are generalists rather than specialists. One consequence of this change is that researchers studying a given topic from, say, 1980 to 1990 might find articles indexed under different headings before and after 1983. The results of some studies (particularly agenda-setting research) could be dramatically affected by this type of change. Further, as Holmes explained, with the older system of subdivisions "you never got the whole story anywhere," so one cannot assume that finding all the stories on a given topic in 1980 will be as easy as finding similar stories in 1990.

Researchers should also be aware that headings may change from year to year, even within the new set of subject terms. As stories develop, editors may decide to index them differently than they did at the start. Within a given calendar year they make every effort to be consistent with the headings, but the same consistency is not necessarily maintained from year to year. For example, the Watergate affair started out as a story under "Democratic Party" because it was "about" a break-in at Democratic headquarters. It was only later that the story became "Watergate." Similarly, a certain type of story might fall under "Vietnam" in 1963 and "Vietnam War" in later years. While it would be consistently indexed under the first heading in 1963, and under the second in, say, 1983, the headings would not necessarily be consistent across years. These changes in subject headings are similar to those noted by Zollars (1994) in her analysis of sample construction from The Readers' Guide to Periodical Literature.

We conclude from these findings that researchers must take into account a certain amount of dynamism in subject headings. Those scholars who rigidly pursue a set list of subject headings to locate media content over several years of coverage are likely to be misled about the universe of media coverage in which elites and citizens debated and decided important public issues.

Procedures for Writing Index Entries

The Index is written according to a philosophy that is quite close to that of professional journalists. Indexers seek to identify key persons, places, and events appearing in a story while at the same time offering a balanced view of the subject. Indexers are trained specifically to pick up names and key quotes. Notions of objectivity and timeliness drive the selection of content to be included in the Index, and the effort to include the "who, what, when,
where, and why” defines a “good” entry. The professionals who write these index entries are, essentially, writing “stories about stories.” Like the journalists who write *Times* articles, indexers exercise news judgment when selecting specific words and phrases and when choosing specific subject terms under which to index the story. They consider “fairness” when constructing the entry, making sure to include conflicting quotes whenever possible so that opposite sides are represented. This affects both the content and the length of an index entry. For example, when Clinton was inaugurated in 1992, an indexer wrote a very good five-line entry that happened to be seven lines shorter than the entry for the 1988 Bush inauguration. Editors required the writer to expand the summary to avoid the appearance of bias. Researchers should take note that the length of an index entry may have little to do with the length or importance of a story. There are, however, certain rules of thumb that relate a story’s importance to the length of an entry. For example, when specific names appear in the article, indexers are instructed to include them in the summary. However, there are practical limitations to the comprehensiveness of written rules. It seems that the most important rules guiding indexing practices are the body of unwritten guidelines learned through apprenticeship and experience, similar to the kind of “news judgment” cultivated among professional journalists.

One factor that enhances consistency in indexing judgments at the *Index* is the extremely low turnover among employees. In fact, at the time of this writing, the nonclerical employee with the shortest tenure has been at the *Index* since 1973. This suggests not only a consistency in the way stories are indexed, but a very long institutional memory as well. The job of an indexer is difficult to learn because it is predicated on experiential knowledge gained through handling many different situations. It would be virtually impossible to write down every decision criterion because, as the editor interviewed for this study explained, “for every rule there are a thousand exceptions.” Consequently, ways of handling particular stories are handed down verbally and become part of an indexer’s specialized knowledge. This can only be developed over time, and once employees acquire the skills necessary to complete their work accurately and quickly it is very difficult to replace them.

Consistency is also maintained with a process of editorial review. Three editors review the work of the indexers. Editors check the abstracts to make sure they are grammatically and factually correct, and they check the index terms for consistency as well as accuracy. Yet despite the best efforts to ensure uniformity of information content over time, changes in systems and personnel in-

---

3Indexers also make the initial choice of which important stories or facts should appear in boldface type in the *Index* Entries corresponding to front-page stories typically contain bold type, and stories from inside pages are bolded if they are particularly important for the subject. Usually the first sentence of the entry is bolded, but in a very long entry a “multipart boldface” style is used in which one or more additional sentences are set in bold. Editors may later change the allocation of boldface.
produce inconsistencies into the process. This could affect research projects that cover long periods of time.

There are important differences among various versions of the New York Times Index. Planned differences in on-line vs. print texts are a function of distribution demands and the fact that stories evolve over time. In many cases, the primary difference will be in subject headings. For example, regional election results will be reported in a single on-line abstract, but will be reported by state in the printed Index. The text of entries also differs between on-line and print. On-line abstracts frequently include more repetition of the details concerning an ongoing story than their print counterparts because users see the abstracts in isolation, without the benefit of a full page of entries that provide more background.

Once an abstract goes on-line it is almost never altered in any way. In contrast, the print indexes are compiled and edited for bi-monthly, quarterly and annual publication. Because the size of the printed book is limited by practical considerations, entries have to be edited to make the product manageable. The printed Index benefits from a longer term perspective on any given story. As Holmes explained, articles about a particular story might be indexed differently because “maybe the indexer who did it on June 5th thought a little bit differently than the person who did it on September 5th.” Consequently, a story may show up under one subject category in the quarterly publication and another in the annual. The annual could also include more information than the earlier publications, such as cross-references under the main subject heading to the names of people involved with the story. The need to edit the larger annual publication could mean that fewer print entries reiterate all the facts of a story. This would be the case for a story such as the OJ Simpson murder trial, for example. As Holmes explained, with the printed index “sometimes you take a second look at something and you say, ‘I don’t like this. This is not as complete as it should have been or this is way overdone.'”

For all of these reasons, although researchers often work hard to collect timely data, we would recommend caution in combining data gathered from more than one version of the New York Times Index. The information provided to us by the senior editor of the Index suggests that although its entries conform to the standards of journalism, and while great effort is made to ensure that they are constructed according to consistent rules, the Index can be expected to deviate in systematic ways from the full-text content of the Times. These are not the result of “errors,” but rather of the conscious professional decisions made by indexers to ensure that Index entries are produced consistently and in a timely manner. The following section presents quantitative findings from a parallel analysis of Times articles and Index entries that illustrate some of the validity problems that can arise when researchers use proxies for full-text coverage.

### Comparing Proxies to Full Text

Our assessment of the validity of various proxies for full-text Times coverage focused on two areas that represent common uses of proxy data by political scientists. First, we examined how accurately these proxies represent the distribution of sources and policy statements found in the full text of the Times. We predicted that Index entries should diverge systematically from full-text data due to the procedures under which entries are compiled. In particular, we predicted that overall levels of support for administration policy should be higher in the Index than in the Times because previous studies have found that opposition voices tend to be located toward the end of Times stories while administration voices tend to be located toward the beginning of stories (Althaus et al. 1996; Entman and Page 1994). Since indexers are trained to read only the top paragraphs thoroughly while skimming later paragraphs, we expected that these reading practices would make the Index more likely to miss the oppositional content that tends to cluster in later paragraphs. While we were less certain about how the practices of headline writers or the literary demands of the lead paragraph would affect these proxies’ representation of full text, we suspected that they would misrepresent the content of the newspaper to some degree. In particular, our previous research (Althaus et. al. 1996) suggested that headlines tend to emphasize conflict, which might make oppositional voices more prominent in headlines than in the full text of the Times.

Second, we assessed whether it is possible for researchers to improve the quality of proxy data by constructing scales for content variables in ways that maximize the aggregation of individual data points. Scaling content data is a common technique used by political scientists to reduce measurement noise and identify subtle but consistent patterns in content data without undue damage to the precision and validity of the raw measures of news content. Some of the more common scales include measures of relative liberalism or conservatism in news content and overall levels of support or opposition for the administration (see Bennett 1990; Bennett and Manheim 1993; Page, Shapiro, and Dempsey 1987 for examples of the support/opposition scale). Here, we test this latter approach in the context of the Libya Crisis.
One possibility is that combining individual policy statements into valenced scales reduces random error and improves the quality of our proxy data as a representation of actual news content. However, it is also possible that aggregation will multiply directional error in the raw measures and make our proxy data even more unrepresentative of the New York Times.

For reasons just discussed, we expected systematic error in Index entries to generally overstate levels of support for administration policy voiced by different sources. But because indexers necessarily simplify full-text content in an attempt to capture the broad contours of a story, we also expected Index entries to oversimplify the policy positions attributed to various sources in ways that systematically privilege neither pro- nor con-administration positions. Rather, sources with consistently low levels of variance in their support for administration policy should be portrayed in the Index as even more extreme in their positions than they appear in the Times. For example, if a source were 90 percent opposed to administration policy in Times coverage, we might expect the simplified Index coverage to show that source as 100 percent opposed. We also expected that this type of distortion would be related to the frequency with which a source appears in Times coverage. The fewer the policy statements attributed to a source, the greater the distortion should be from the Index's simplification process. Once again, we were uncertain about how alternative proxies would be affected by the professional practices that govern their creation, but we believed that they, too, would misrepresent the Times to some degree. Our goal here is to measure the patterns of distortion in order to offer researchers some guidance about how validly to use proxies.

### Overall Distribution of Sources and Policy Statements

The drastically lower numbers of data points for the proxies suggest that these indicators should do a poor job of reflecting the mix of sources and policy statements found in the full-text stories. Contrary to our expectations, we found that each of the content indicators reflected fairly

### Information Density

Before examining how well each proxy represents full-text content, it will be helpful to first compare the density of information found in the New York Times to that found in each of our proxies. We can think of information density in two ways: (1) the number of stories from each source of content data containing at least one coded policy statement and (2) the number of policy statements coded from each source of content data. It seems plausible that the closer a proxy comes to matching the number of policy statements and coded stories found in the full-text data, the more valid it should be as a stand-in for full text.

Table 1 shows that the three proxy indicators contain far fewer stories with at least some codeable content than the full-text data. The discrepancy among these content indicators was even more pronounced when comparing numbers of policy statements captured by our coders. The full-text stories contained a total of 1,283 policy statements. In contrast, the corresponding Index entries taken together contained a total of 195 policy statements, amounting to only 15 percent of the number of coded units found in full-text stories. Lead paragraphs and headlines had even less policy content, respectively containing only 8 percent and 5 percent of the number of attributed policy statements found in the full-text data. The third row of Table 1, which shows the ratio of policy statements to coded stories, reveals that proxies contained between a third and a fifth the amount of content information found in full-text stories. Dividing the number of policy statements by the total number of stories analyzed in each of the four data sets (N=403) reveals even larger discrepancies. Relative to full-text content, the Index is more information-dense than lead paragraphs and headlines, but all three proxies capture only a small portion of the attributed policy statements found in the full text.
### Table 2 Percentage of All Policy Statements Made by Each Source Across Content Indicators

<table>
<thead>
<tr>
<th>Source</th>
<th>Full Text</th>
<th>Index</th>
<th>Lead</th>
<th>Headline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreign Officials</td>
<td>47.6</td>
<td>49.2</td>
<td>(+1.6)</td>
<td>53.7</td>
</tr>
<tr>
<td>U.S. Administration</td>
<td>611</td>
<td>96</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>U.S. Congress</td>
<td>32.5</td>
<td>24.1</td>
<td>(-8.4*)</td>
<td>25.0</td>
</tr>
<tr>
<td>U.S. Citizens</td>
<td>417</td>
<td>47</td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>Foreign Citizens</td>
<td>5.8</td>
<td>8.7</td>
<td>(+2.9)</td>
<td>5.6</td>
</tr>
<tr>
<td>U.S. Citizens</td>
<td>75</td>
<td>17</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Journalists</td>
<td>4.9</td>
<td>9.7</td>
<td>(+4.8*)</td>
<td>6.5</td>
</tr>
<tr>
<td>Experts</td>
<td>63</td>
<td>19</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>U.S. Citizens</td>
<td>3.7</td>
<td>2.1</td>
<td>(-1.6)</td>
<td>4.6</td>
</tr>
<tr>
<td>Journalists</td>
<td>48</td>
<td>4</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Experts</td>
<td>45</td>
<td>8</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Total All Sources</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
<td>100.0</td>
</tr>
<tr>
<td>N</td>
<td>1283</td>
<td>195</td>
<td></td>
<td>108</td>
</tr>
</tbody>
</table>

† two-tailed p < .10 * two-tailed p < .05

Note: Cells contain the percentage of all sources within the content indicator, such that percentages in each column add up to approximately 100%. Parentheses contain the percentage-point differences between the proxy indicator and full-text data.

accurately the aggregate distribution of sources and statements found in the full-text articles. Table 2 displays the frequency distribution of all coded sources appearing in the full-text and proxy data. In this table, each cell contains the total number of policy statements made by a given source as well as the percentage of all statements contained in the content indicator that were attributed to a given source. For instance, a total of 611 policy statements were attributed to foreign officials in the full-text data, which comprised 47.6 percent of all policy statements found in the full-text data. In addition, cells for each proxy indicator contain the percentage-point difference between the proxy indicator and the full-text data for each source category. For instance, 49.2 percent of policy statements appearing in the Index data were attributed to foreign officials, which differed by only 1.6 points from the percentage found in the full-text data. Chi-square tests were conducted to determine whether these observed differences between the full text and each of the three content proxies were statistically significant. It is important to point out that because our analysis considers the entire population of full-text stories and proxy data relevant to the Libya 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 can have in generalizing from this particular case to similar uses of proxy data in other contexts.

The picture presented in Table 2 is quite consistent: when sources are aggregated across the entire five months of Libya crisis coverage, each of the content proxies presents a fairly accurate estimate of source proportions found in the full-text stories. Chi-square tests between the full-text data and each of the three content proxies reveal that the only three of the observed differences are statistically significant. Two of these are found in the Index data, which significantly underrepresented administration officials and significantly overrepresented foreign citizens relative to the full-text data. In addition, the headline data underrepresented policy statements made by administration officials to an even greater degree than the Index data, though this relationship is only marginally significant. It is important to point out that if the number of policy statements in the headline and lead paragraph data were larger, some of the larger observed differences in source percentages between those content indicators and the full-text data would eventually attain statistical significance. Another way of assessing the degree of misrepresentation...
afforded by each proxy is to calculate Spearman rank-order correlations between the percentages of sources in each content indicator and those recorded in the full-text data. This comparison produces a Spearman’s rho of .90 between Index data and full-text data, .96 between lead paragraph and full-text data, and .88 between headline and full-text data. This finding suggests that while the three proxies may vary somewhat in the percentages assigned to each category, they all tend to agree on the relative prominence of different types of sources in the news.

Nonetheless, examining the data in Table 2 within rather than across columns confirms that even small differences in proportions for each source category can create substantial distortions within a given content indicator. For instance, the Index and lead paragraph data suggest that statements by foreign officials are twice as numerous as those made by administration officials, and the headline data suggest foreign officials spoke nearly three times as much as administration officials. In contrast, the full-text data show that foreign officials made only about one and a half policy statements for every one made by an administration official. All three proxies overestimate the proportion of foreign officials and foreign citizens while they underestimate the proportion of administration officials who spoke out on U.S. policy. If we had used the New York Times Index to study the policy debate surrounding the Libya crisis, we would have mistakenly concluded that sources from the Reagan administration were even more marginalized in news coverage than was actually the case, and that members of Congress were more prominent in news coverage than the full-text data show (Althaus et al. 1996), although in neither case would these errors have affected the larger conclusions of our earlier study.

Some differences between full-text and Index entries can be explained by the journalistic practices of Times indexers, who are concerned not with reproducing the relative proportions of statements made by various sources but rather with making sure that key named sources are mentioned, and that conflicting quotes are represented. This rule had little effect on the proportion allotted to foreign officials, as no single official dominated this category of sources. But because a smaller number of administration officials spoke on behalf of the U.S. government (particularly the “big three”: President Reagan, Secretary of State Shultz, and Secretary of Defense Weinberger), repeated statements made by one of those officials in any given story might appear as a single summary statement in an Index entry. As shown in Table 2, a consequence of this rule is that the Index should tend to underrepresent the proportions of source categories containing a small number of named individuals who speak frequently or at length in a story. The Index should do a better job of representing the proportions of source categories containing a larger number of named individuals who each make brief statements or appear only once in a story.

The findings presented above refer only to the accuracy of proxies in representing when a member of some group appears in the coverage. We also want to know whether proxies accurately reproduce the policy statements made by these sources. Table 3 shows the overall distribution of policy statements in the full-text and proxy data. In this table, each cell contains the total number of statements made about a given policy as well as the percentage of those statements that supported or advocated the policy. This comparison suggests that all three proxies tended to provide a fairly accurate picture of policy content at this high level of aggregation. Chi-square tests between the full text and each of the three content indicators confirm that there are no statistically significant differences across proxies, even though many of the observed differences are quite large, most notably in the case of the minor policy options (negotiations, espionage, War Powers, and other), in which proxy data contain only one or a handful of observations. When comparing across individual policies, each of the content proxies seems to give a statistically valid estimate of the relative amount of support for different policies recorded in the full text of the New York Times.

That said, it is important to note that many of the differences which fail to attain conventional levels of significance (due to the small number of observations in the proxy data) are substantively important. For example, the contents of headlines suggest that con-force and con-sanctions statements were, respectively, more than twice as prevalent as pro-force and pro-sanctions statements during the five months of the Libya Crisis. In contrast, the full text shows that con-force and con-sanctions voices were only slightly more prevalent than pro-force or pro-sanctions voices. While this pattern can be explained by the tendency among editors to emphasize conflict rather than consensus in headline content (Althaus et al. 1996), a similar imbalance is found in the Index data, where con-force statements outnumber pro-force voices by a factor of nearly two to one. These distortions in the Index can be explained by the sourcing patterns discussed above, which underreported the statements of Reagan administration officials, who tended to be the strongest and most vocal supporters of military force during the crisis period. As foreign officials were consistently anti-administration in their views while administration officials toed the line, the underrepresentation of American officials caused by indexing practices created an anti-force tilt in the mix of policy statements.
TABLE 3 Support for Each Policy, by Content Indicator

<table>
<thead>
<tr>
<th>Policy</th>
<th>Full Text</th>
<th>Index</th>
<th>Lead ¶</th>
<th>Headline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Force</td>
<td>43.3</td>
<td>36.8</td>
<td>(-6.5)</td>
<td>43.1</td>
</tr>
<tr>
<td>N =</td>
<td>894</td>
<td>133</td>
<td>72</td>
<td>17</td>
</tr>
<tr>
<td>Sanctions</td>
<td>43.5</td>
<td>52.4</td>
<td>(+8.9)</td>
<td>42.9</td>
</tr>
<tr>
<td>N =</td>
<td>292</td>
<td>42</td>
<td>28</td>
<td>17</td>
</tr>
<tr>
<td>Negotiations</td>
<td>73.7</td>
<td>80.0</td>
<td>(+6.3)</td>
<td>50.0</td>
</tr>
<tr>
<td>N =</td>
<td>19</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Espionage</td>
<td>88.2</td>
<td>100.0</td>
<td>(+11.8)</td>
<td>100.0</td>
</tr>
<tr>
<td>N =</td>
<td>17</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>War Powers</td>
<td>75.0</td>
<td>66.7</td>
<td>(-8.3)</td>
<td>100.0</td>
</tr>
<tr>
<td>N =</td>
<td>24</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>97.3</td>
<td>90.9</td>
<td>(-6.4)</td>
<td>100.0</td>
</tr>
<tr>
<td>N =</td>
<td>37</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Neg., Esp., War Powers and Other Combined</td>
<td>85.6</td>
<td>85.0</td>
<td>(-0.6)</td>
<td>87.5</td>
</tr>
<tr>
<td>N =</td>
<td>97</td>
<td>20</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

Note: Cells contain the percentage of supportive statements regarding each policy, and the total number of statements made about each policy. Parentheses contain the percentage-point differences between the proxy indicator and full text. None of the observed differences in this table was found to be statistically significant.

Not only was the level of opposition to the use of force overstated, but relative levels of support for economic and diplomatic sanctions against Libya were also overrepresented in Index data. While researchers using the full text of the Times would conclude that force and sanctions had roughly equal levels of support in the news, researchers using Index entries would mistakenly presume that sanctions received much more favorable news coverage than the force option, and that there was majority support among news sources against force but in support of sanctions. Both conclusions based on Index data would have been wrong.

This analysis suggests that researchers should exercise caution in attributing significance to the precise frequencies of sources and substantive statements found in proxy data. Our analysis suggests that the production practices of the New York Times Index lead to systematic underrepresentation of administration sources and systematic overrepresentation of opposition to administration policies. However, our case involves foreign officials arrayed in opposition to American officials. We cannot be sure if these patterns would hold for domestic policy cases or even other foreign policy cases.

Impact of Data Aggregation on Proxy Validity

We also tested the validity of aggregating our proxy data into a summary scale. Could scaling proxy data reduce measurement noise and simplify data interpretation without sacrificing validity? First, we examined whether the proxy data accurately represent the reported overall level of support or opposition to administration policy for each type of source. Can the proxies accurately tell us who supported the president? Second, we examined the changes in levels of support for administration policy over time. Scaling policy statements in terms of whether they support or oppose administration positions allows us to account for the changing terms of the debate as it unfolded, to see whether the larger contours of the debate were accurately represented in proxy data. Finally, we tested whether the proxies could accurately capture three dimensions of the debate at once: the level of support by each source over time.

Table 4 presents the reported levels of support for administration policy among various categories of news sources for the full-text data and each of the three content proxies. In keeping with the standard use of such data in the scholarly literature, the aggregate measure of administration support includes all six policy options. Individual pro- and con-policy statements were recoded as supportive, opposed, or neutral toward U.S. administration policy. Throughout the crisis, the administration opposed a negotiated settlement, favored covert action against the Libyan regime, advocated diplomatic and economic sanctions, and (at least publicly) favored force as a last resort. Although Reagan initially spoke out against the use of force, arguing that military strikes would endanger the lives of Americans working in Libya,
Throughout the crisis period his administration publicly endorsed the use of air strikes if sanctions failed to sway Qaddafi. To produce a valenced summary measure of reported support for the administration's policies, policy statements in the "other" category were recoded as neutral, and all substantive policy statements were recoded as supporting or opposing administration positions. Reported levels of support were measured as the percentage of all statements made by a given source that were in line with administration policy positions.

Data presented in Table 4 show that, although the Index tended to overrepresent support for the administration among foreign officials, the level of support attributed to the two most common news sources—foreign officials and administration officials—was fairly close to what actually appeared in the Times. Support among other groups was less accurately represented. For example, the full-text data show that members of Congress appearing as news sources were about evenly split in their support for administration policy. Yet the picture of Congressional opinion provided by the Index suggests strong approval for the administration, while that provided by lead paragraph and headline data suggests strong opposition. Such variance is typical of the source categories with relatively smaller numbers of coded statements, which makes lead paragraph and headline content especially prone to bias. But even the Index often proves to be a poor indicator of full-text content: besides misconstruing Congressional opinion, it greatly exaggerates levels of opposition to administration policy among experts and foreign citizens as reported in the full text of the New York Times. Both could be artifacts of indexing procedures. As noted earlier, indexers are trained to pick up "key" names as they scan the full text of articles. If members of Congress whom indexers consider "key" are more likely to support the administration, then mentioning them in the Index will lead to a bias towards the pro side of the debate. Furthermore, once the pro side is represented in an entry, indexers have no reason to pick up supportive statements by members of other categories, such as experts and foreign citizens. Instead, they will choose quotes that provide a conflicting view to the pro-administration sentiment already covered.

Contrary to our expectations, the proxies did not consistently portray sources with low levels of variance in their support for administration policy as more extreme in their positions than they appear in the Times. The expected patterns were observed in some cases, notably statements made by foreign citizens and experts. However, the Index also reported higher levels of support among foreign officials and lower levels of support among American citizens than found in the Times. While we expected the proxies to inflate overall levels of support for administration policy, the bottom row of Table 4 shows that the net effect of these differences is to exaggerate overall levels of opposition toward administration policy. Headlines are the worst offenders in this regard.

A general pattern seems apparent in Table 4: the more numerous the data points for any source, the more accurate the measure of administration support in the news will tend to be. Because each of the three content proxies contains such relatively small numbers of policy statements, breaking them further into individual source categories for the purpose of ascertaining differences in support levels invites misrepresentation of the full-text data. Coupled with differences introduced by the indexers' decision rules, such disaggregation can lead to extreme and systematic inaccuracies when proxy data are taken to represent actual news content.

Another common application of news content data is to capture trends in the reported level of administration or policy support. The simplest way to approach this is to chart developments in support over time, without regard to differences among the various sources in the news. 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 1986 Libya Crisis, dividing the data into discrete
weeks of news coverage seemed the best way to capture Times reporting of the ebb and flow of support for administration policy over time. We compare weekly changes in administration support in full text and each of the three proxy indicators.

Figure 1 shows that levels of support calculated from the Index tended to be reliable estimates of week-to-week changes in levels of support appearing in the full-text data.\(^5\) The correlation between the full text and Index data is an impressive .89. The major exception to this trend in the Index data comes in the week beginning February 4\(^{th}\), where the Index was more than 30 percentage points off of the support level in the full-text data. But on the whole, aggregate support in the Index tended to track changes in support from the full-text data. The average percentage point deviation across all fourteen weeks in which at least one policy statement was recorded was 11.2 for the Index data.\(^6\) In other words, the aggregate support level from the Index data tended to be off by around 11 percentage points in any given week from the level found in the full-text data, although once we control for the weeks of February 4\(^{th}\), February 11\(^{th}\), and March 18\(^{th}\)—each of which contained fewer than five policy statements in the full-text data—this average degree of error drops to 6.4 percentage points. Controlling for these three weeks also reduces the correlation between these trends to .76. Taken together, the high degree of similarity in the trends produced by Index and full-text data is in many ways a remarkable finding given that the mean weekly number of policy statements in the Index data (not counting the week of the April 15\(^{th}\) bombing of Tripoli, which contained seventy-two) was only eight.\(^7\)

The parallel analysis of lead paragraphs and headlines in Figure 1 shows that they are less accurate indicators of aggregate support over time than Index entries. While these proxy data sets produce a correlation of .87

\(^5\)No coverage of the Libya Crisis appeared in the Times in the weeks from February 18\(^{th}\) through March 11\(^{th}\).

\(^6\)By "average percentage point deviation" we mean the square root of the mean square error, where error is calculated for each week as the percentage point deviation from levels of administration support found in the full-text data. Our calculation of mean error therefore indicates only the absolute value of the percentage-point deviation rather than the direction of that deviation relative to the full-text value.

\(^7\)In contrast, the full-text data contained an average of sixty-seven policy statements per week, not counting the week of April 15\(^{th}\), which contained 409 policy statements.
between full text and lead paragraph data, and .54 between full text and headline data, the mean error in weekly support levels for the lead paragraph data was 17.8 percentage points, while that for headline data was 26.1 percentage points. Even after controlling for the three weeks mentioned earlier, these average weekly errors drop to just 16.1 points in the lead paragraph data (the headline data contain no observations for these weeks). As was the case with Table 4, these indicators tend to exaggerate the levels of support or opposition found in the full-text data. Moreover, levels of support within these indicators often swing wildly from week to week, a tendency produced by the small number of policy statements per week (averaging 5.2 policy statements per week for the lead paragraph data and 3.1 for the headline data). Our original analysis of full-text data (Althaus et al. 1996) led us to conclude that news coverage given to the Libya crisis was surprisingly balanced between critics and supporters of the Reagan administration’s policies, and we would have arrived at the same conclusion using Index data. But had we used headline or lead paragraph data instead, we would have erroneously concluded that news coverage swung cyclically between overwhelming support and deep opposition to the administration’s actions. Such a fluctuating pattern would have suggested that the degree of support in news coverage was largely determined by the appearance of critical events over the course of the crisis period, whereas the more balanced patterns in the full-text data suggest that news coverage was determined primarily by journalistic norms of news construction, such as the desire to produce “objective” reporting by seeking out critical sources to counterbalance supportive sources.

It would seem that the Index can be a reasonable proxy for capturing broad, over-time changes in full-text content. A more stringent test of Index data is to track reported levels of administration support over time for individual source categories. At this level, the emphasis is on tracking changes over time among groups with an eye to comparing differences in support levels. This analysis, which is summarized in Table 5, suggests that the Index is an inconsistent proxy for full-text content when there are so few data points. Index coverage of administration sources deviated from full-text data by an average of 11 percentage points per week. Moreover, as expected the Index consistently portrayed these sources as more extreme in their levels of support than they appeared in full-text Times articles. In ten of the twelve weeks in which administration sources made policy statements, the Index portrayed administration officials as 100 percent supportive of administration policies. In only three of those weeks did the Times also report 100 percent support among administration sources. Larger inconsistencies were found for the most common source of policy statements, with the degree of support among foreign officials misrepresented by an average of 30 percentage points in a given week. Likewise, support for administration policies among foreign citizens and journalists was off by an average of more than 40 percentage points per week. The average errors reported in Table 5 show that none of the proxy indicators can be counted on to stand for full-text data at this level of analysis, although a longer time unit (such as months instead of weeks) or a more intensely covered issue might produce different results.

Table 5 further suggests that at least two factors are potentially damaging to the validity of proxy indicators when the data are disaggregated for trend analysis. First, the roughly linear relationship found in Table 4 between source prevalence and the accuracy with which the source’s statements are captured in the Index disappears at this lower level of aggregation. Foreign officials, which constitute the largest source category, have one of the higher average errors in the Index data, while experts, the least common news source, have the lowest. In these time-series data, there does not seem to be a linear relationship between the number of statements recorded for

### Table 5  Average Weekly Deviation from Full Text Data in Percentage Supporting Administration Policy, by Source and Content Indicator

<table>
<thead>
<tr>
<th>Source</th>
<th>Index Avg. Error</th>
<th>% of Missing Weeks (N)</th>
<th>Lead % Avg. Error</th>
<th>% of Missing Weeks (N)</th>
<th>Headline Avg. Error</th>
<th>% of Missing Weeks (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreign Officials</td>
<td>30.4</td>
<td>0.0 (0)</td>
<td>22.3</td>
<td>16.7 (2)</td>
<td>28.4</td>
<td>16.7 (2)</td>
</tr>
<tr>
<td>U.S. Administration</td>
<td>11.3</td>
<td>14.3 (2)</td>
<td>19.2</td>
<td>28.5 (4)</td>
<td>14.1</td>
<td>42.9 (6)</td>
</tr>
<tr>
<td>U.S. Congress</td>
<td>12.5</td>
<td>66.7 (4)</td>
<td>17.0</td>
<td>66.7 (4)</td>
<td>33.7</td>
<td>66.7 (4)</td>
</tr>
<tr>
<td>Foreign Citizens</td>
<td>41.5</td>
<td>50.0 (3)</td>
<td>16.5</td>
<td>66.7 (4)</td>
<td>16.7</td>
<td>66.7 (4)</td>
</tr>
<tr>
<td>U.S. Citizens</td>
<td>13.6</td>
<td>60.0 (3)</td>
<td>14.2</td>
<td>40.0 (2)</td>
<td>61.3</td>
<td>60.0 (3)</td>
</tr>
<tr>
<td>Journalists</td>
<td>40.3</td>
<td>62.5 (5)</td>
<td>49.3</td>
<td>75.0 (6)</td>
<td>24.4</td>
<td>62.5 (5)</td>
</tr>
<tr>
<td>Experts</td>
<td>7.8</td>
<td>75.0 (6)</td>
<td>24.8</td>
<td>75.0 (6)</td>
<td>11.1</td>
<td>87.5 (7)</td>
</tr>
</tbody>
</table>
a source and the accuracy with which the Index reproduces those statements, which means that the errors produced within proxy data are difficult to anticipate and perhaps impossible to remedy without conducting a detailed comparison of proxy data to full-text data. A similar lack of linearity is found in the headline and lead paragraph data.

A second pattern in Table 5 is the presence of a linear relationship between the number of statements made by a given source and the frequency of “missing” weeks: weeks in which statements made by the source were captured in the full-text data, but not in the proxy data. The general pattern suggested in an analysis of missing data seems clear: the more prevalent the source, the fewer the missing data points. For instance, the Index data contain at least one statement by a foreign official for each of the weeks that the full text contains such a statement. With this most numerous source category, there are no missing weeks in the Index data. But for the least prevalent source category—experts—the Index recorded no codeable statements for six of the eight weeks in which such statements were captured in the full-text data. The prevalence of sources is undoubtedly confounded with their “importance,” as judged by Index coders. Foreign officials were not only the most numerous group, they were also well-known people whom Index protocol would rank worthy of inclusion.

The conclusion we draw from this analysis is that none of the content proxies provides a valid picture of full-text content at such lower levels of aggregation. Research using proxy data to track the changing levels of support for administration policy among even the most prominent news sources would severely misrepresent the actual levels of support recorded for those sources in the actual pages of the New York Times. In our study, the validity of Index data was compromised by the small number of policy statements contained in any one week and by the growing number of weeks with missing data for all but the largest source categories. Although the sort of analysis represented in Table 4, Table 5, and Figure 1 is fairly common in political communication and public opinion studies, our findings suggest that proxy data are not always well-suited for such purposes.

**Conclusion**

The findings presented here should help researchers make more informed use of proxies to both locate and represent media content. Scholars using indexes to find stories should verify that the type of content they are researching was consistently indexed over the time frame they are studying. They should also be aware of potential inconsistencies in categorization. Because of concerns about reliability in research protocols, researchers sensitive to the demands of quantitative research may be inclined to use the same subject headings to locate coverage on a given issue over a long period of time. Yet while the editors and indexers strive for consistency, and while the Index staff remains constant for many years at a time, the process of assigning index entries to subject headings can vary from year to year as long-standing issues develop and change. Annual volumes may contain important variations in classification schemes. We would encourage scholars using the Index not to sacrifice validity for what at first glance resembles reliability. The evidence presented here suggests that a variety of subject headings in the paper index should be examined for each year of interest and that the search terms should change in accordance with changes in the relevant subject headings. In fact, the difference in subject headings over time may itself constitute important data about news content. The longer the time frame of the study, the more important this precaution becomes. This finding also suggests that important changes should be made in reporting content analyses. Currently, published results do not consistently report the subject headings used to locate stories. We believe that this information is an important element in evaluating studies and that editors should insist on it.

Our comparison of Index entries, lead paragraphs, headlines, and the full text of The New York Times during the 1986 Libya crisis suggests that these proxies can adequately represent original content when researchers are working at a high level of aggregation, whether this aggregation takes the form of scaling “raw” content data into a smaller number of categories or amassing a large number of “raw” data points. Our proxy data performed most effectively when indicating highly aggregated trends, measuring overall levels of support, and ordinarily ranking the prominence of sources and policy statements. However, drawing conclusions based upon the fine-grained analysis of the marginals, such as interval-level estimates of the relative prominence of sources or estimates of the level of administration support in a given week were more problematic, in part because some categories of the variables had few data points to work with. This problem was most dramatically depicted in our attempts to assess the relative levels of support for specific policy options using proxy data. The results from this analysis suggests that even with a relatively large number of data points, estimates of relative prominence based on proxy data can be misleading.

Scaling proxy data does not seem to magnify the effects of directional errors found in the raw measures of news content. To the contrary, by increasing the number
of data points available for analysis, this approach may tend to reduce such error. Among our most accurate measures are week-by-week changes in overall levels of support for administration policies. But as shown in the preceding analyses, the use of such scales does not necessarily improve data quality. Our work further demonstrates that as proxy data are divided into smaller subsets, they no longer accurately reflect the content of the Times, and more importantly, that the direction of the distortions is unpredictable. For example, while the overall level of support for the administration indicated in the Index only slightly underestimates the level of support in the Times as a whole, the level of support among members of Congress is dramatically overestimated in the Index relative to the level of support described in the Times itself. Here again, our proxies fail because they are subjected to a fine-grained analysis that relies on too few data points, and mathematical aggregation does not save us from wrong conclusions.

Interviews with the senior editor of the New York Times Index persuade us that the reason the Index can work reasonably well in some situations as a proxy for full text is because it is produced in a journalistic culture similar to that of the newsroom, one that emphasizes fairness and balance. It would not, however, be appropriate to say that the Index is a summary of the content of the Times, and some of the practices of the indexers do seem to produce systematic differences between the Index and the Times. One of the most important is the indexers’ emphasis on including key public figures in an Index entry. This tendency may explain the underestimation of support for the administration in the Index relative to the support reported in the Times full text, since support for the administration came from relatively few sources. Foreign sources opposed to the administration were quoted much more briefly in the Times, but there were many more of them. Thus, specific characteristics of the coverage may affect the validity of the Index as a proxy for full text.

While our quantitative analysis of the validity of New York Times Index entries as a proxy for New York Times newspaper content is based on a case study, it uses measures of the text often applied in communication and political science research. We consider ours to be a relatively lenient test of the legitimacy of proxy-based findings. Our coding scheme may overestimate the quality of the Index especially since, like the indexers themselves, we sought to link individual news sources to the policy statements they made. Our data demonstrate that proxies can be used in the broadest ways to represent highly aggregated media content, but that proxy data are likely to misrepresent full text in more fine-grained analyses of media content. However, because these findings are the product of a case study, these findings cannot suggest how much coverage is enough to make a proxy measure valid or how much disaggregation is required to compromise that validity. More research is needed to establish these kinds of guidelines. For now, we can only sound a warning and suggest study-by-study validation procedures. Our findings lead us to recommend that researchers who wish to use any proxy for full-text content should first test the assumption that proxy data provide a valid indicator of full-text content. This could be done by drawing a sample of stories from the full-text population to test the suitability of the proxy-coding scheme as a substitute for full-text coverage. Bennett (1990) employs a similar technique in his study of Times coverage about U.S. policy toward Nicaragua, and while he does not report the results of his comparison in detail, his careful approach to the use of Index data is an example well worth following.

Ultimately, whether the New York Times Index provides an adequate representation of the actual content of the New York Times depends upon the specific research question posed. However, this study suggests that we should not expect too much of the Index, for its use is fraught with pitfalls that may trap the unwary. Additional research, focusing on the critical issues we have raised here, is needed to establish the validity of other media indexes as a way to locate stories and represent full-text data. Studies of this kind will enable researchers to make informed decisions about the use of proxy data, enhance the rigor of content analysis methods, and thus make important contributions to a broad range of social research.

References


