6-2005

The Effect of Content Mix on Circulation Penetration for U.S. Daily Newspapers

Jack Rosenberry
St. John Fisher University, jrosenberry@sjf.edu

Follow this and additional works at: https://fisherpub.sj.edu/commj_facpub

Part of the Communication Commons

Custom Citation

This document is posted at https://fisherpub.sj.edu/commj_facpub/12 and is brought to you for free and open access by Fisher Digital Publications at . For more information, please contact fisherpub@sjf.edu.
The Effect of Content Mix on Circulation Penetration for U.S. Daily Newspapers

Abstract
For a generation, the U.S. newspaper industry has suffered from stagnant circulation performance and declining penetration, but some papers manage to maintain healthy penetration. This project investigated whether content mix could account for the variance in penetration among papers in a national sample, but found content does not have a significant impact on circulation variance after controlling for the influence of market characteristics.

Disciplines
Communication

Comments
This version is pre-publication.

The final copy of this article is © SAGE Journals and is available through the publisher's website here: http://dx.doi.org/10.1177/107769900508200209

This article is available at Fisher Digital Publications: https://fisherpub.sjfc.edu/commj_facpub/12
Manuscript Title:

The Effect of Content Mix on Circulation Penetration for U.S. Daily Newspapers

Prepared for re-submission to

Journalism and Mass Communication Quarterly

May 2005 (revision of a draft originally submitted November 2004)

Submitted by:
Jack Rosenberry, Assistant Professor of Communication/Journalism
St. John Fisher College, Rochester, N.Y.

Author bio/acknowledgements:

Jack Rosenberry is an assistant professor of communication/journalism at St. John Fisher College in Rochester, N.Y. He would like to acknowledge and thank Robyn Kraft, Tom McDonald and Stacey MacDonald for their assistance, and also thank the anonymous reviewers for several helpful suggestions.

Word counts:
Abstract: 82
Main paper: 5,530 (including tables but exclusive of references)
Paper Title:

The Effect of Content Mix on Circulation Penetration for U.S. Daily Newspapers

Abstract

For a generation, the U.S. newspaper industry has suffered from stagnant circulation performance and declining penetration, but some papers manage to maintain healthy penetration. This project investigated whether content mix could account for the variance in penetration among papers in a national sample, but found content does not have a significant impact on circulation variance after controlling for the influence of market characteristics. The study thus offers no prescriptive ideas for editors about improving circulation penetration performance by adjusting content mix.
Introduction

When University of Wisconsin researchers Bruce Westley and Werner Severin did their groundbreaking research on newspaper readership in the early 1960s, they could say with some accuracy: “Reading the daily newspaper is doubtless one of the most thoroughly institutionalized behaviors of Americans.”¹ Indeed, what made their research into non-readers interesting and valuable was that, in their time, non-readers were the unusual case rather than the common one.

But today, the situation is vastly different. One survey found that only 42 percent of the U.S. population reads a paper every day.² The sense within the U.S. newspaper industry is that a generation of “flat line” circulation – stagnation in the total number of newspapers sold – is a major problem. This decline is well established and documented. Picard ³ discusses a nearly 50 percent decline in penetration rates from 1950 to 1990. Putting it another way, about as many daily newspapers are sold now in the United States as in the mid-1950s, even though the national population has grown by 64 percent since then.⁴ Recent statistics from Newspaper Association of America put weekday circulation at 55.6 million and Sunday circulation at 59 million.⁵

An interesting and important question for the industry is: what, if anything, can journalists do about it? Can circulation be affected by decisions that reporters and editors make about newspaper operations, such as the amount or types of content that compose the newspaper? Or have demographics, competition and other external factors made declining penetration inevitable and irreversible?

The conventional wisdom in the industry is that local news and lifestyle coverage are thought to build circulation by giving readers what they want and what they need but
cannot get from other sources. Recent academic research has focused on the connections of newspaper business success and editorial quality, operationalized by content measurements such as newshole. The goal of this project is to empirically and systematically evaluate content determinants of circulation to see whether the conventional wisdom is correct or not, and also to contribute to the ongoing conversation about content-based quality measures and business success with a detailed analysis of the relationship of content to circulation performance. The key question in this study will be whether content variables, defined as space devoted to specific types of news, can account for the variance in circulation penetration across a broad sample of newspapers.

**Theoretical basis**

Traditional economic theory says that supply and demand in a market combine to create an equilibrium point that matches quantity produced with quantity consumed and also sets a price. But newspaper circulation performance doesn’t neatly follow this model. A half-century ago Ray demonstrated that non-price competition was a key factor in newspaper demand because of imperfect competition. A few years later Landau and Davenport determined that cost-price theories did not apply to newspaper supply-demand models. Picard found newspaper circulation demand is inelastic with respect to price. Lacy has repeatedly demonstrated that content offers a form of product differentiation for newspapers that operate in markets where different media are imperfect substitutes for each other, echoing what Ray and Landau/Davenport had said a generation earlier. Lacy and Simon elaborate on that, saying that “Traditional utility theory is not entirely satisfactory in explaining reader demand ... circulation is a function
of quality, which is dependent on the newspaper’s content.”  

Attempts to explain how factors beyond economic ones affect newspaper circulation have proceeded from two main perspectives. One says that circulation performance rests largely, perhaps exclusively, on structural factors such as market demographics. The other view says that while structural factors matter, content also plays a role. The current study seeks to shed some light on these competing paradigms.

**Literature review**

Research regarding circulation and readership determinants is broad but not deep. Inquiry in the field also has been complicated by a wide variety of dependent and independent variables. Some published studies focus on circulation, but more delve into the related (but not identical) construct of readership. Some researchers focus on what Blankenburg has called the “structural” determinants of circulation, such as demographic characteristics of the market, while others explore the “discretionary” side of the issue, meaning things over which journalists can exercise discretion or control, such as content and format. Some studies make individual readers the unit of analysis – e.g. why do individuals choose to read a paper – while other projects focus on communities or markets. Many studies in this latter category are single-market projects; only a handful of studies encompass a wide range of newspapers.

A review of literature on determinants of newspaper readership and circulation was conducted that covered more than 50 years of work from the 1950s to 2004. Many of the studies came from the middle portion of that time period (1977 to 1989) including a large body of work sponsored by the American Newspaper Publishers Association under
the auspices of the ANPA Research Institute. Not as much work in the field, especially work of a quantitative nature, has been done since then, although recently the question of whether quality journalism can contribute to newspapers’ economic success has been a focus of some researchers’ attention\textsuperscript{13} and the Readership Institute at Northwestern University surveyed 37,000 readers and content-analyzed newspapers from 100 markets nationwide in an effort to measure how content influences readership. \textsuperscript{14}

\textit{Demographic description as the dominant research paradigm}

Blankenburg summed up the focus on demographics when he wrote that “Published research in this area tends heavily toward structural effects on readership, yielding repetitious findings that readership is strongly associated with age, education and income.”\textsuperscript{15} He further noted that structural market factors were the best predictors of circulation and that quality of the paper as expressed through features such as format and content had “nothing to do with circulation change.”\textsuperscript{16}

\textbf{Readers vs. non-readers} The earliest research into demographic impacts on readership concerned itself with differing characteristics of newspaper readers vs. non-readers, such as Westley and Severin’s,\textsuperscript{17} ground-breaking study in the early 1960s. At that time, non-readers made for an interesting research topic because they were atypical. In other studies of readers vs. nonreaders, Poindexter\textsuperscript{18} tried to create a composite profile of non-readers, as did Sobal and Jackson-Beeck.\textsuperscript{19} Schweitzer\textsuperscript{20} noted that differences between subscribers and non-subscribers among younger readers (ages 18 to 24) could be found in demographics but not so-called “lifestyle” variables. Stevenson\textsuperscript{21} and Peiser\textsuperscript{22} used cohort analysis to determine that succeeding generations of readers each were reading less than their older counterparts.
Households and markets. The focus on individual reader characteristics was accompanied by investigations of the impact of household and market characteristics on readership and subscribing. Stone, for example, noted that while demographic characteristics such as income and education were good predictors of readership at the individual level, they were not so effective at the market level. Rarick concluded that household characteristics (such as income) made for good predictors but only if combined into an index consisting of several individual demographic variables. Denbow replicated Rarick’s work in a Kentucky market a few years later in the interest of demonstrating whether the index was generalizable, and concluded that it was. Eberhard and Turpin both related circulation strictly with population trends while Guthrie, Ludwin and Jacob created a statistical model using purely market data to predict circulation in outlying areas of a metro newspaper’s market. Stone also created a regression-based formula that used home ownership, total community population and single-family dwellings in a community to explain 64 percent of the variance among 200 markets in total circulation numbers.

Competitive concerns. Content as a performance determinant was the focus of research into competitive market situations, especially in the 1960s and ’70s when competitive markets were more common than they are today and the decline of competition within individual markets was a primary focus of newspaper research. But this line of research is complementary with the economic theories mentioned earlier; namely, that newspaper performance is based on product attributes such as content rather than solely on price. Lacy and Bernstein investigated the key product differentiator of cycle (AM or PM distribution) and found a relationship of cycle to both content and
circulation, which they measured by overall numbers (not penetration). Morning and evening papers differed in 10 of 22 content variables measured; large papers differed from smaller ones in content measures such as total news space available and proportion of news space devoted to local copy.

Looking specifically at competitive differences, Rarick and Hartman\textsuperscript{31} discovered that papers in competitive situations carried more local news and more sensational news. Weaver and Mullins,\textsuperscript{32} on the other hand, discovered few significant differences among content or format in papers in a competitive situation. McCombs\textsuperscript{33} likewise found little difference in the content of competitive papers vs. those in a monopoly market. It may be worth noting that the changes in the industry over a 20-year time frame, especially the decline of two-newspaper markets, could account for the differences between McCombs’s findings in the mid-'80s and Rarick and Hartman’s from the mid '60s.

Weaver, Schweitzer and Stone\textsuperscript{34} reported that similarities and differences of content and appearance in dual-newspaper markets correlated with differences in penetration, although not strongly. But Lacy, Sohn and Stephens\textsuperscript{35} in a study of two markets (Detroit and Denver) said that differences in content were not correlated with differences in penetration. In another study, Lacy and Sohn also reported that content of competing metro and suburban dailies in the same cities had little similarity. But in the same article they advocated more focus on content as a determinant of circulation, saying “One of the underlying limitations of much of the existing research in this area is the absence of actual content examination.”\textsuperscript{36} One goal of this project is to address this concern.

**Issues of quality.** A corollary to whether content differences affect competitive
performance is whether the general quality of a newspaper has an impact on its market performance, and much research has sought to demonstrate that it does. Stone, Stone and Trotter\(^\text{37}\) concluded that while market characteristics could be used as predictors of circulation, newspaper quality could explain some of the circulation variance among papers. The quality assessment they used was not empirical, however, but rather a subjective judgment by newspaper editors about other papers in their states. Becker, Beam and Russial\(^\text{38}\) created a quality-performance index (also based on a survey of editors, not on content) and said the best-performing papers among 109 in a sample also had the best circulation “saturation” performance, their term for penetration. Lacy and Fico\(^\text{39}\) attributed about 22 percent of circulation variance among 114 papers in a sample to a “quality index” that was operationalized by certain amounts and types of content. Reiterating a finding from some other studies, they determined that overall population in a market accounted for most of the variability in circulation (again using total numbers, not penetration). Meyer and Kim\(^\text{40}\) used a similar content-based approach to quantify quality in an objective way. Lacy\(^\text{41}\) said newspaper demand curves were “kinked” around a quality point and theorized that in general, higher quality would be associated with more subscribers.

**Local content.** Direct evaluations of content’s impact on circulation have tended to focus on local news. Alperstein\(^\text{42}\) found a relationship between local news content and circulation penetration in a study of 12 Canadian markets. Grotta, Larkin and DePlois\(^\text{43}\) likewise noted that local information is a key point of interest for readers of smaller newspapers, while Stone and Morrison\(^\text{44}\) found circulation was related to significant differences in local news content among weekly newspapers. Anecdotal a similar
finding was reported by Callahan,\textsuperscript{45} who wrote about a small North Carolina newspaper that emphasizes local news to the virtual exclusion of other content – and had the highest circulation penetration among audited U.S. papers in 2000 to show for it.

\textit{Discussion of the literature}

None of the studies that incorporate content variables provides a clear answer to the question of how much circulation variance, if any, can be attributed to content differences on a broad-scale basis. Some researchers, notably Blankenburg, take the bleak view that none of it can be thusly accounted for – circulation determination rests entirely in the market (structural) forces. Bogart\textsuperscript{46} is dubious as well, noting that a survey of content preferences of frequent vs. infrequent readers found few differences. On the other side, Lacy and Martin conclude “the preponderance of large-scale U.S. studies support the connection between newspaper content and circulation and penetration.”\textsuperscript{47}

Yet this extensive review of the literature did not find any investigation comparable to the one undertaken here, with a detailed analysis of content mix across a broad range of U.S. daily newspapers directly correlated with penetration as a dependent variable. The bulk of past work in which a wide range of papers were evaluated has been focused on demographics rather than content, or has related content to overall (total number) circulation results. Penetration studies that have used content analyses have generally investigated single markets or small groupings. Large-scale investigations of penetration as a variable have not used detailed content analysis of individual papers.

Alperstein\textsuperscript{48} did relate local content of Canadian newspapers to their market penetrations. But that was a study of strictly local information (both news and advertising) in just 12 markets, outside the United States, with a simple item-count
content analysis and analysis through descriptive (rather than inferential) statistics. Another study with some similarities to the current one was Stone’s 1976 project that related content to circulation of weekly newspapers. But this project differs in that Stone used total circulation rather than penetration as the dependent variable and studied weeklies rather than dailies. Some of Lacy’s work also has investigated correlation of penetration and content variables, but generally in single-market, highly specific circumstances (the Denver-Detroit studies). The one study that did assess multiple markets used total circulation rather than penetration as the dependent variable and found (not surprisingly) that market population accounted for most of the circulation variance. Becker, Beam and Russial used penetration as a dependent variable but related it to a quality evaluation by editors rather than relating it to direct content measurements.

**Methods**

Newspaper circulation is a function of many variables, starting with market size. Population differences have been documented as the largest single source of variance in overall circulation numbers in various studies. But the industry’s benchmark for performance is not raw numbers but circulation penetration, defined as the ratio of the number of copies sold to market size. Thus, the major purpose of this study is to use circulation penetration as a dependent variable with various measures of news content as independent variables, applied to a cross-section of U.S. newspapers.

**Sampling**

The roster of newspapers in the 2002 edition of the annual *Editor and Publisher Yearbook* was used as a sampling frame. A random starting point and nth–entry selection
process was used to select 59 newspapers from among the approximately 1,500 listed in the publication. The Yearbook and accompanying Editor and Publisher Market Guide also contain reliable data regarding circulation and market size.

Papers selected for sampling were asked to provide a “constructed week” sample as described by Riffe, Aust and Lacy. The constructed week consisted of the following issues: Friday Sept. 12, 2003; Tuesday Sept. 16; Thursday Sept. 18; Sunday Sept. 21; Monday Sept. 22; Wednesday Oct. 1; and Saturday Oct. 4. Papers in the sample were first contacted by mail in the last week of September 2003 asking them to supply papers from each of those dates. Those that did not immediately reply received a follow-up letter a couple of weeks later and were contacted by telephone as well with a request that they participate. Of the 59 newspapers that were contacted, 41 ultimately supplied papers, for a response rate of 69.5 percent and a total of 258 individual issues to be analyzed.

The resulting data set ended up being highly representative of the overall U.S. newspaper population with respect to penetration, circulation-size stratification and other measurements. The mean penetration rate among the 41 papers was .533, nearly identical to the U.S. national rate of .530. The national rate was calculated from the Newspaper Association of America’s circulation statistic of 55.6 million daily circulation divided by 105 million U.S. households, according to the 2000 Census, a methodology also used by Meyer. The median paper size in the data set was 14,695, again nearly identical to the national value of 14,359. Finally, U.S. newspaper circulation is distributed according to a power curve, with a few very large newspapers, a smaller number of mid-sized papers, and a large number of smaller-circulation community dailies; this natural skew can be used to test the similarity of the sample to the population. The correlation (Pearson’s r) of
the percentages in each circulation stratum for the total U.S. newspaper population vs. the sample is .904. (See Table 1).

**Table 1: Comparison of Sample to Newspaper Population**

<table>
<thead>
<tr>
<th>Circulation Category</th>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percentage in population</td>
<td>Percentage in data set</td>
</tr>
<tr>
<td>n = 1468</td>
<td>n = 41</td>
<td></td>
</tr>
<tr>
<td>250,001+</td>
<td>2.7% (39 papers)</td>
<td>2.4% (1 paper)</td>
</tr>
<tr>
<td>100,001 - 250,000</td>
<td>4.3% (64 papers)</td>
<td>7.3% (3 papers)</td>
</tr>
<tr>
<td>50,001 - 100,000</td>
<td>8.6% (128 papers)</td>
<td>9.8% (4 papers)</td>
</tr>
<tr>
<td>25,001 - 50,000</td>
<td>13.6% (203 papers)</td>
<td>12.2% (5 papers)</td>
</tr>
<tr>
<td>10,001 - 25,000</td>
<td>29.3% (437 papers)</td>
<td>36.6% (15 papers)</td>
</tr>
<tr>
<td>5,001 - 10,000</td>
<td>24.4% (364 papers)</td>
<td>24.4% (10 papers)</td>
</tr>
<tr>
<td>5,000 or less</td>
<td>17.1% (254 papers)</td>
<td>7.3% (3 papers)</td>
</tr>
</tbody>
</table>

Correlation (Pearson’s r) of percentages from columns A and B = .904

**Definition of variables: Penetration**

For purposes of the analysis, circulation was defined as the newspaper’s weighted average daily-Sunday circulation \[ \frac{([\text{Mon-Sat} \times 6] + \text{Sunday})}{7} \]. For most of the papers, the penetration rate was calculated by taking this circulation and dividing by the number of households in the newspaper’s home county. Exceptions to this rule were made for five papers out of the 41 that had unusually small circulations relative to the county’s size because they were small community papers in what is commonly called an “umbrella” market situation, where they operate in the shadow of a nearby major metro daily. For these cases, penetration was figured on the basis of their home-community households (the market they can more realistically be said to serve) rather than the entire county. Because many papers circulate across multiple counties, a check on the efficacy of using just the home county as a surrogate for the market was done by comparing it to
the number of households in the newspapers’ Designated Market (or City Zone or Retail Trade Zone) as reported by Editor and Publisher. Not all papers report these statistics, but for the 31 that did list an NDM, CZ or RTZ figure, the correlation (Pearson’s r) of these values with home-county households was .890.

**Definition of variables: Content**

All of the material in each issue was categorized in a variety of ways, starting with the most fundamental division of newspaper content: that of advertising vs. editorial (news) material. Editorial content then was further classified along several dimensions, including local vs. non-local and text vs. graphic, and by topic such as general news, sports news, lifestyle/entertainment news and business news. These dimensions combine to yield a total of 38 categories of content. (See Appendix A.)

The primary reason for breaking down the content along these multiple dimensions was so that the measurements later could be recombined into larger units, but with enough information captured about each particular element that it could be counted in all the relevant subsets. For example, a photograph of a high-school football game would be classified into Category 25 (local sports graphic), contrasted with an NFL game photo in Category 26 (national sports graphic). Both photos naturally would be counted as part of sports coverage. But since the coding captures the local dimension, the high school photo also could be tallied as part of a paper’s overall local news component. Finally, both also could be counted in a tally of all graphics (as opposed to text).

This categorization/coding was done by the principal researcher and a team of undergraduate students who were trained in the methodology. Inter-coder reliability was established by having all coders categorize all content elements – about 200 items in all –
from one edition of a medium-sized metro newspaper that was not a part of the study. These category ratings were then evaluated with the PRAM computer program\textsuperscript{54}, which calculated an average value of .797 for both Cohen’s Kappa and Krippendorf’s Alpha as to category assignments. (Reliabilities for individual coder pairs ranged from .755 to .846 in each statistic.) As a second reliability check, a correlation measure (Pearson’s \( r \)) was used to compare the total space in each of the 38 categories assigned by the project director and one of the student coders for one issue of the same paper, with a resulting value of .981. The unit of analysis for the overall project was the entire newspaper, while the unit of measurement was square-inches of material, a common way of measuring content to account for differences in column-widths and page sizes.\textsuperscript{55}

The 38 specific categories in the coding system then were used to create more general attributes based on typical organizational strategies for news content, such as local, national, sports or business news. For example, all of the various sports measures (local and non-local sports text [Categories 9 and 10], local and non-local graphics [Categories 25 and 26] and sports tabular material [Category 37]) were added together for each paper to create a “sports news” attribute. Ten of these combination categories were created: total news space (newshole), local news, political news, sports news, business news, opinion, graphics, a “hard news” attribute representing mostly world and national general and political news, and entertainment news. (A complete breakdown of which detailed categories comprised each combined attribute is in Appendix B.)

\textit{Definition of variables: Demographics}

Because prior research has demonstrated the significance of reader demographics and market characteristics on circulation performance, it was necessary to control for
in the analytic models that were used to assess the impact of content. U.S. Census figures\textsuperscript{56} were used to determine the following characteristics for each county (or, in the case of those five “umbrella” papers, home community): household and per-capita income, proportion of adults who are high school graduates, proportion of adults who are college graduates, percentage of residents who had lived in the same place for five years or longer and percentage of residents who owned their home. These characteristics were selected to control for market characteristics investigated in prior research, including income level, educational level and community stability.

\textit{Analysis procedure}

Measurements for the market characteristics and various content categories were used as independent variables in a regression analysis to identify which types of content tend to account for differences in circulation performance. This methodology is similar to that used by Weaver, Schweitzer and Stone\textsuperscript{57} in their comparison of content and subscribership to both the morning and evening papers in a single market. Lacy and Sohn\textsuperscript{58} also demonstrated the use of regression based on content measurements within single markets. One thing that makes this study different from those, however, is its application of this method to a national cross-section of papers.

The wide range of circulations and market sizes created a problem in the analysis, however, because the sample was not distributed normally. Rather, it followed the exponential curve described earlier with a handful of very large papers, a few large to middle-sized ones and a large number of small ones. Market sizes were similarly skewed. As noted earlier, this distribution represents the population of U.S. newspapers well, but poses problems in analysis, particularly for regression models based on linear
relationships. To account for this, a logarithmic transformation was applied to circulation figures, market sizes and all final text measurements, which were compiled as the average daily total for a given content category. The log transformation normalized the data to prepare it for linear regression analysis. A least-squares line fitted through the raw circulation sizes had an $r^2$ value of 0.339, compared with $r^2 = 0.899$ for the log-transformed data. The daily average amounts of space (also log-transformed) devoted to each of the combined categories were used as content variables in the regression against penetration.

A hierarchical regression analysis was used to control for the impact of the market variables. The control part of the model included newspaper cycle as well as Census data for high school and college completion rates, homeownership percentage, and length of residence (proportion of people living in the same place for five years or longer). These variables were selected in line with Lacy and Bernstein’s findings\(^5^9\) about the significance of cycle, as well as Stone’s work\(^6^0\) showing the variance attributable to community stability, especially home ownership. Income was not part of the control block because trial models that included it lacked statistical significance. This, too, is consistent with prior research, especially another study by Stone\(^6^1\) which said that demographic characteristics such as income and education were good predictors of readership at the individual level, but not so effective at the market level.

**Findings**

*Content measurements*

The content analysis measured widely varied results across the data set. On average, papers devoted nearly 60 percent of their total space to news. About 20 percent
of that newshole, on average, was devoted to local news of all types (including sports and business). Similarly, “hard” news (state, world and international events) took about 20 percent of newshole as well. Sports and entertainment news each occupied about 13 percent of news space, while between 3 and 4 percent was devoted to business and opinion coverage. Individual papers varied widely around these means, however. Total newshole percentage, for example, ranged from 73 to 43 percent. Local news ranged from 31 percent of newshole down to 7 percent; sports ranged from nearly 20 percent down to 8 percent. (See Table 2.)

### Table 2: Summary of Content Analysis – Space Percentages by Category

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newshole*</td>
<td>58.9</td>
<td>73.1</td>
<td>42.9</td>
</tr>
<tr>
<td>Local</td>
<td>19.2</td>
<td>31.4</td>
<td>7.1</td>
</tr>
<tr>
<td>Hard</td>
<td>20.7</td>
<td>34.4</td>
<td>8.7</td>
</tr>
<tr>
<td>Entertainment</td>
<td>13.5</td>
<td>25.0</td>
<td>5.3</td>
</tr>
<tr>
<td>Sports</td>
<td>12.7</td>
<td>19.6</td>
<td>8.1</td>
</tr>
<tr>
<td>Business</td>
<td>3.7</td>
<td>9.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Opinion</td>
<td>3.6</td>
<td>6.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Graphics</td>
<td>13.5</td>
<td>22.3</td>
<td>7.2</td>
</tr>
<tr>
<td>Tabular</td>
<td>8.6</td>
<td>15.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Political</td>
<td>7.8</td>
<td>18.0</td>
<td>4.2</td>
</tr>
</tbody>
</table>

*Newshole is defined as news space compared to total space available (news plus advertising). All other percentages are based on news content only.*

In general, larger circulation papers had more pages and therefore more total space devoted to each content measure. Circulation and total newshole were correlated at .856. This finding is congruent with The Readership Institute’s Impact Study, which found that larger papers publish more pages. In that study, papers with circulations between 10,000 and 25,000 averaged 32 pages and 72 stories in a typical weekday issue, while those larger than 200,000 circulation average 104 pages with 162 stories.62
However, larger papers on average tended to devote less space proportionally to local news and hard news; circulation was negatively correlated with each of these content categories. As with the total space devoted to various content measures, these proportional figures make logical sense. Because they have more total space available, larger papers can afford to devote more space to coverage of certain topics, such as entertainment and business, that many of the smallest papers virtually ignore. This reduces the proportion of space devoted to, say, national or world news.

**Regression results**

The regression analysis started with entry of a block of structural variables describing market characteristics. These were home ownership (M=68.4%; s.d. = 9.2), residential stability (percentage living at same address for five years or longer; M=54.4%; s.d. = 7.2) and education (two variables: percentage of high school graduates, M=81.0%; s.d. = 6.2; percentage of college graduates M=20.9%; s.d.=8.2). This structural block also included newspaper cycle, coded as a dummy variable (AM=1; PM=2). Although cycle was the only individual variable that was statistically significant, this model overall explained 32.4 percent of the variance at a statistically significant level. \( r^2 = .324; F = 3.348; p = .014 \). However, when the 10 content variables were added as a second block, none of the individual variables nor the model itself was statistically significant. (Table 3)

With respect to cycle, the negative coefficient indicates that AM papers (which were coded “1”; PMs were coded “2”) have significantly better circulation performance. This was confirmed by comparing the mean penetrations (AM = .624; PM = .487), which were found to be significantly different according to a nonparametric test (Mann-Whitney U = 128; \( p = .035 \)) that was used because they were not normally distributed.
A second set of regressions was conducted, in which each of the 10 content variables was added independently as a second block consisting of just a single variable. These models resulted in miniscule beta weights for each variable, with only small changes in additional variance ($r^2$) explained and no statistical significance, with p values ranging from .21 to .92 (Table 4).

Table 3: Regression Results – Content Categories as Single Block

Block 1: Demographics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle</td>
<td>-0.408**</td>
</tr>
<tr>
<td>Home Stability</td>
<td>-0.058</td>
</tr>
<tr>
<td>High School grad</td>
<td>0.224</td>
</tr>
<tr>
<td>College grad</td>
<td>-0.431*</td>
</tr>
<tr>
<td>Owner%</td>
<td>-0.287</td>
</tr>
</tbody>
</table>

Increase in $r^2$ for block = 32.4%; $F = 3.348; p = 0.014**

Block 2: Content

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newshole</td>
<td>-1.389</td>
</tr>
<tr>
<td>Local</td>
<td>-0.021</td>
</tr>
<tr>
<td>Hard</td>
<td>0.207</td>
</tr>
<tr>
<td>Entertainment</td>
<td>0.573</td>
</tr>
<tr>
<td>Sports</td>
<td>-0.381</td>
</tr>
<tr>
<td>Business</td>
<td>0.642</td>
</tr>
<tr>
<td>Opinion</td>
<td>0.318</td>
</tr>
<tr>
<td>Graphics</td>
<td>0.188</td>
</tr>
<tr>
<td>Tabular</td>
<td>0.137</td>
</tr>
<tr>
<td>Politics</td>
<td>0.187</td>
</tr>
</tbody>
</table>

Increase in $r^2$ for block = 13.6%; $F = 0.63; p = 0.774

* = $p < .10$; ** = $p < .05$
Table 4 Regression Results – Content Categories as Individual Variables

Block 1: Demographics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle</td>
<td>-0.408**</td>
</tr>
<tr>
<td>Home Stability</td>
<td>-0.058</td>
</tr>
<tr>
<td>High School grad</td>
<td>0.224</td>
</tr>
<tr>
<td>College grad</td>
<td>-0.431*</td>
</tr>
<tr>
<td>Owner%</td>
<td>-0.287</td>
</tr>
</tbody>
</table>

Increase in $r^2 = 32.4\%$; $F = 3.348$; $p = 0.014**$

Block 2: Content

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>$r^2$ # increase</th>
<th>$\ p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newshole</td>
<td>0.134</td>
<td>1.1%</td>
<td>.46</td>
</tr>
<tr>
<td>Local</td>
<td>-0.034</td>
<td>0.1%</td>
<td>.83</td>
</tr>
<tr>
<td>Hard</td>
<td>0.015</td>
<td>0.4%</td>
<td>.66</td>
</tr>
<tr>
<td>Entertainment</td>
<td>0.024</td>
<td>2.8%</td>
<td>.23</td>
</tr>
<tr>
<td>Sports</td>
<td>0.002</td>
<td>0.0%</td>
<td>.92</td>
</tr>
<tr>
<td>Business</td>
<td>0.016</td>
<td>3.2%</td>
<td>.21</td>
</tr>
<tr>
<td>Opinion</td>
<td>0.024</td>
<td>2.0%</td>
<td>.32</td>
</tr>
<tr>
<td>Graphics</td>
<td>0.017</td>
<td>0.8%</td>
<td>.54</td>
</tr>
<tr>
<td>Tabular</td>
<td>0.013</td>
<td>1.3%</td>
<td>.42</td>
</tr>
<tr>
<td>Politics</td>
<td>0.022</td>
<td>1.6%</td>
<td>.37</td>
</tr>
</tbody>
</table>

# Increase in $r^2$ based on each variable’s addition to model as a single-variable “block”

* = $p<.10$; ** = $p<.05$

Discussion

This study was conducted against a backdrop of research in the newspaper industry (and among academic researchers interested in newspapers) about what it will take to keep readers attracted to newspapers. Trade journals report on topics such as “Bad News About Newspaper Circulation” while the relationship of a quality news product to overall business success is an area of particular interest among academic researchers.
these days. Bogart, Lacy and Martin, Meyer and others have approached this topic from angles such as whether competition helps to spur higher product quality and whether a dis-investment strategy (i.e. cost-cutting) in resources such as newshole and staffing, while beneficial for the short-run bottom line, is a losing business strategy for the longer term because lower quality drives away readers, leading to a downward circulation and revenue spiral.

This study contributes a couple of significant findings to this ongoing discussion. For one, it looks at daily penetration and content mix on a broad-scale basis, something that has not really been done before. Previous broad-based studies have looked at total circulation numbers, while penetration studies have generally been of single markets or small groupings, rather than attempting to investigate both variables on a national basis. The way this data set so closely represents the industry lends a measure of external validity to the findings, especially with respect to demographic variables. The lack of statistically significant relationships with respect to content and the large variance attributable to cycle and market characteristics can be most reasonably interpreted as meaning that these really are the controlling aspects of circulation performance. There is no “magic bullet” for reporters and editors to fall back upon with regard to using content strategies to boost circulation performance.

By failing to find that the level of coverage devoted to different types of news has any impact on circulation variance, the study falls short of the goal of offering prescriptive ideas for editors about what they could do to adjust content “mix” in the interest of improving circulation penetration performance. This is unfortunate because a positive finding in this regard would have helped in the ongoing effort to define and
operationalize “quality” news coverage by relating circulation specifically to certain
types and amounts of content, paralleling the effort scholars such as Lacy and Fico\textsuperscript{69} and
Meyer and Kim\textsuperscript{70} have made with relating content quality to general business
performance.

But while the results of this study don’t provide unqualified support for the idea
that quality of coverage is related to circulation success, it’s not certain that they
contradict it either. One limitation of this study is that both because of the sheer volume
of material to be coded (several thousand pages, tens of thousands of individual items,
and only four coders), and because of an interest in avoiding subjective judgments in an
effort to enhance statistical validity, throughout the content analysis items that in all
likelihood would be seen as having varying news quality were coded identically. In other
words, an investigative story about City Hall corruption and the routine police beat would
have been categorized the same (Category 3: Local government, police and court news).
Likewise, a dramatic local news photo and a “grip and grin” both would be classified the
same (Category 20, local news photo or graphic).

Whether these category definitions and the way they capture a wide variety of
news presentations tainted results is impossible to say. Any sort of content analysis that
would seek to measure amounts of “high quality” material vs. more mundane material
still has the problem of conceptually defining “quality,” which at the item-coding level
would come down to a subjective judgment such as whether investigative reporting
constitutes quality journalism and grip-and-grin photos do not. Such judgments were
deliberately avoided here, but with careful definition could be made a part of future
research that builds upon the findings and methods developed here.
This study also did not investigate the impact of marketing or promotional variables, such as how much effort papers put into acquiring and retaining subscribers. Clearly, this could affect penetration, yet it was not accounted for here. Neither was competition, in any comprehensive way (aside from the adjustments of the penetration rate that were made for those three papers in umbrella markets). Future studies might find it productive to control for these variables in addition to the market-level “structural” variables controlled for here.

Circulation performance is an important issue for the industry and one that academic researchers must continue to investigate in the interest of better understanding its determinants.
## Appendix A: Coding Categories for Content Analysis

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Non-editorial material</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Advertising</td>
</tr>
<tr>
<td>2</td>
<td>Administrative (e.g. promos, page flags and house ads)</td>
</tr>
<tr>
<td><strong>News Text</strong></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Local government, political, police and court news text</td>
</tr>
<tr>
<td>4</td>
<td>Local general news text</td>
</tr>
<tr>
<td>5</td>
<td>State and regional news text</td>
</tr>
<tr>
<td>6</td>
<td>National government, political and court news text</td>
</tr>
<tr>
<td>7</td>
<td>National general news text</td>
</tr>
<tr>
<td>8</td>
<td>International news text</td>
</tr>
<tr>
<td>9</td>
<td>Local sports news (including college and pro teams in paper's market) text</td>
</tr>
<tr>
<td>10</td>
<td>Non-local sports news (mostly professional and out-of-town college) text</td>
</tr>
<tr>
<td>11</td>
<td>Local business news text</td>
</tr>
<tr>
<td>12</td>
<td>Non-local business news text</td>
</tr>
<tr>
<td>13</td>
<td>General lifestyle information (local and other) text</td>
</tr>
<tr>
<td>14</td>
<td>Entertainment/celebrity news text (including local arts and entertainment)</td>
</tr>
<tr>
<td>15</td>
<td>Staff generated opinion text (including unsigned editorials)</td>
</tr>
<tr>
<td>16</td>
<td>National/syndicated opinion text</td>
</tr>
<tr>
<td>17</td>
<td>Reader-generated opinion text (primarily letters to the editor)</td>
</tr>
<tr>
<td>18</td>
<td>Other text (not classifiable elsewhere)</td>
</tr>
<tr>
<td><strong>News Graphics</strong></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Local government, political, police and court photos/graphics</td>
</tr>
<tr>
<td>20</td>
<td>Local general news photos/graphics</td>
</tr>
<tr>
<td>21</td>
<td>State and regional news photos/graphics</td>
</tr>
<tr>
<td>22</td>
<td>National government, political and court news photos/graphics</td>
</tr>
<tr>
<td>23</td>
<td>National general news photos/graphics</td>
</tr>
<tr>
<td>24</td>
<td>International news photos/graphics</td>
</tr>
<tr>
<td>25</td>
<td>Local sports photos/graphics</td>
</tr>
<tr>
<td>26</td>
<td>National/professional sports photos/graphics</td>
</tr>
<tr>
<td>27</td>
<td>Local business news photos/graphics</td>
</tr>
<tr>
<td>28</td>
<td>Non-local business news photos/graphics</td>
</tr>
<tr>
<td>29</td>
<td>General lifestyle information photos/graphics (incl. comics and puzzles)</td>
</tr>
<tr>
<td>30</td>
<td>Entertainment/celebrity news photos/graphics</td>
</tr>
<tr>
<td>31</td>
<td>Staff generated opinion photos/graphics (e.g. editorial cartoons)</td>
</tr>
<tr>
<td>32</td>
<td>National/syndicated opinion photos/graphics</td>
</tr>
<tr>
<td>33</td>
<td>Reader-generated opinion photos/graphics</td>
</tr>
<tr>
<td>34</td>
<td>Other photos/graphics not classifiable elsewhere</td>
</tr>
<tr>
<td><strong>Tabular material</strong></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>General tabular (e.g. real-estate transactions and lottery results)</td>
</tr>
<tr>
<td>36</td>
<td>Business tabular (e.g. stock listings)</td>
</tr>
<tr>
<td>37</td>
<td>Sports tabular (e.g. standings, box scores)</td>
</tr>
<tr>
<td>38</td>
<td>Lifestyle tabular (e.g. TV listings, calendars of entertainment events)</td>
</tr>
</tbody>
</table>
Appendix B: Content attribute components

Defined according to categories as delineated in Appendix A

Newshole – sum of:
Categories
2-38 Everything except advertising (category 1)

Local news – sum of:
3 Local government, political, police and court news text
4 Local general news text
9 Local sports news text (including college and pro teams in paper’s market)
11 Local business news text
15 Staff generated opinion text (including unsigned editorials)
19 Local government, political, police and court photos/graphics
20 Local general news photos/graphics
25 Local sports photos/graphics
27 Local business news photos/graphics
35 General tabular (e.g. real-estate transactions and lottery results)

Sports – sum of:
9 Local sports news text (including college and pro teams in paper’s market)
10 Non-local sports news text (mostly professional and out-of-town college)
25 Local sports photos/graphics
26 National/professional sports photos/graphics
37 Sports tabular (e.g. standings, box scores)

Business – sum of:
11 Local business news text
12 Non-local business news text
27 Local business news photos/graphics
28 Non-local business news photos/graphics
36 Business tabular (e.g. stock listings)

Opinion – sum of:
15 Staff generated opinion text (including unsigned editorials)
16 National/syndicated opinion text
17 Reader-generated opinion text (including letters to the editor)
31 Staff generated opinion photos/graphics (e.g. editorial cartoons)
32 National/syndicated opinion photos/graphics
33 Reader-generated opinion photos/graphics

Graphics – sum of:
19-34 Reader-generated opinion photos/graphics
### Appendix B: Content attribute components (continued)

**Tabular – sum of:**
- 35 General tabular (e.g. real-estate transactions and lottery results)
- 36 Business tabular (e.g. stock listings)
- 37 Sports tabular (e.g. standings, box scores)
- 38 Lifestyle tabular (e.g. TV listings, calendars of entertainment events)

**Politics – sum of:**
- 3 Local government, political, police and court news text
- 5 State and regional news text
- 6 National government, political and court news text
- 8 International news text
- 19 Local government, political, police and court photos/graphics
- 21 State and regional news photos/graphics
- 22 National government, political and court news photos/graphics
- 24 International news photos/graphics

**“Hard” news (non-local) sum of:**
- 5 State and regional news text
- 6 National government, political and court news text
- 7 National general news text
- 8 International news text
- 21 State and regional news photos/graphics
- 22 National government, political and court news photos/graphics
- 23 National general news photos/graphics
- 24 International news photos/graphics

**Entertainment – sum of:**
- 14 Celebrity news and entertainment text
- 30 Celebrity news and entertainment photos/graphics
Notes


14 *Impact Study*. Chicago: Readership Institute/Media Management Center at Northwestern University, 2001: 3

15 Blankenburg, “Structural Determination of Circulation.”

16 Blankenburg, “Structural Determination of Circulation.”

17 Westley and Severin, “A Profile of the Daily Newspaper Non-reader.”


40 Meyer and Kim, “Quantifying newspaper quality.”


Content mix and penetration – page 31


47 Lacy and Martin, “Competition, Circulation and Advertising.”

48 Alperstein, “The Influence of Local Information on Daily Newspaper Household Penetration.”

49 Lacy and Fico, “Link Between Newspaper Content Quality and Circulation.”

50 Becker, Beam and Russial, “Correlates of Daily Newspaper Performance in New England.”


54 Skymeg Software, Program for Reliability Assessment with Multiple Coders (PRAM) computer program release 0.4.5 Available: http://www.geocities.com/skymegsoftware/pram.html

55 Stephen Lacy, “The Effects of Intracity Competition on Daily Newspaper Content,” Journalism Quarterly 64 (Summer-Fall 1987): 281-90.

57 Weaver, Schweitzer and Stone, “Content, Appearance and Circulation.”

58 Lacy and Sohn, “Correlations of Newspaper Content with Circulation in the Suburbs.”

59 Lacy and Bernstein, “Daily Newspaper Content’s Relationship to Publication Cycle.”

60 Stone, “Using Community Characteristics.”

61 Stone, “Community Commitment.”


65 Lacy and Martin, “Competition, Circulation and Advertising.”


69 Lacy and Fico, “Link Between Newspaper Content Quality and Circulation.”

70 Meyer and Kim, “Quantifying newspaper quality.”