A Key Word History of Marketing Science

Carl F. Mela    Jason Roos    Yiting Deng

November 5, 2012

1Carl F. Mela is the T. Austin Finch Foundation Professor of Business Administration at the Fuqua School of Business, Duke University (email: mela@duke.edu, phone 919-660-7767). Jason Roos is an Assistant Professor at the Erasmus School of Business, University of Rotterdam (email: roos@rsm.nl, phone 206-317-1713). Yiting Deng is Ph.D. candidate at the Fuqua School of Business, Duke University (email: yiting.deng@duke.edu, phone 510-735-6255). The authors would like to thank seminar participants at the 2012 ISMS and AMA Doctoral Consortiums, the University of Houston, and the University of Frankfurt, as well as Bart Bronnenberg, Preyas Desai, Rick Staelin, and an anonymous reviewer for their feedback.
Abstract: A Key Word History of Marketing Science

This paper considers the history of key words used in Marketing Science to develop insights regarding the evolution of marketing science. Several findings emerge. First, pricing and game theory are the most ubiquitous words. More generally, the 3C’s and 4P’s predominate, suggesting key words and common practical frameworks align. Various trends exist. Some words, like pricing, remain popular over time. Others, like game theory and hierarchical Bayes have become more popular. Finally, some words are superseded by others, like diffusion and social networking. Second, the overall rate of new keyword introductions has increased but the likelihood they will stick around has decreased. This suggests a maturation of the discipline or a long tail effect. Third, a correspondence analysis indicates three distinct eras of marketing modeling, comporting roughly with each of the past three decades. These eras are driven by the emergence of new data and business problems, suggesting a fluid field responsive to practical problems. Fourth, we consider author publication survival rates which increase up to six papers and then decline - possibly due to changes in ability or motivation. Fifth, survival rates vary with the recency and nature of words. We conclude by discussing the implications for additional journal space and the utility of standardized classification codes.

Keywords: Key Words; History; Marketing; Data Visualization.
1 Introduction

We characterize the history of marketing science using the key word choice of authors in *Marketing Science*. Authors’ choices of key words not only afford a sense of what topics were once, have always been, or are becoming central to marketing thought; they also reflect factors that drive innovation in the field. That is, the collective efforts of the field, as expressed by the key word choices of its authors, provide a sense of the field’s interests and the potential for subsequent innovations. This research also complements a strategic review of the journal, appearing in this issue of *Marketing Science*, regarding stakeholders’ perspectives about the journal’s goals and focus Chintagunta et al. (2013). Our findings about the state of the field, and how it has evolved over time, can be viewed as a potential input to such a strategic assessment. For example, we consider i) the interplay between methods and topics and between research and practice and ii) the rate at which seminal topics have been introduced over time - issues central in that strategic review.

Other research exists with similar aims. These studies can be characterized as either survey based (using the authors’ review of the literature) or citation based (using citation analyses). For example, Leeflang et al. (2000) characterize five eras of marketing science, the latter three of which are spanned by the publication of *Marketing Science* (see Lilien (1994), for another summary). Our analysis also suggests three eras of marketing science that loosely comport with their characterizations. Related, Buzzell (1968) and Lilien et al. (1992) developed definitive texts outlining conceptual foundations for the field of marketing science. In contrast, our work is more empirically grounded. Along these lines, there also exists a broad literature on citation analyses in marketing (e.g., Baumgartner and Pieters, 2003; Tellis et al., 1999). By assessing author cross-citations, these papers assess the similarity between journals. More recently, Moussa and Touzani (2010) use Google scholar citations to assess the relative impact of journals. In contrast, we focus our attention on the impact of topics and assess how this varies over time.
We are aware of one marketing paper that also analyzes key words. Kevork and Vrechopoulos (2009) consider key words used in customer relationship management research, using expert judgment to classify these key words into topics. Our approach extends this work by i) considering the broader domain of marketing models, ii) applying multivariate analyses for classification, and iii) considering dynamics in word usage.

Several findings emerge. First, the modal use of a key word is just once, perhaps reflecting a long-tail for key words. Second, regarding the words that are most prevalent, there is a correspondence between the common marketing vernacular of the “3C’s” and “4P’s” and the most predominant words used. Third, key words tend to cluster into several themes, including promotions, channels, and new products. Of note, “game theory” is atop this hierarchical clustering, being the word most likely to connect these various threads. Fourth, patterns exist in the prevalence of key words over time, with some words in ascendance (like “game theory”), some in maturity (like “promotions”) and some being substituted with newer concepts (like “diffusion” and “social networks”). Fifth, while the rate at which new key words appear is accelerating, the rate at which they become enduring is decelerating, possibly suggesting a long-tail phenomenon wherein the additional inventory of journal space leads to more variety in key words. Finally, we conclude with an analysis of author productivity rates, in which we observe that about 3/5 of all authors who publish in Marketing Science do so just once. Moreover, the probability of continuing to publish peaks around the fifth paper, indicating longevity per se does not explain survival rates in publication. Of note, these survival rates are related to the choice of words (methods or topics), their novelty, and authors’ tendencies to use different words on new papers.

The remainder of this note proceeds as follows. First, we detail the key word and authorship data used in our analysis. Next, we outline our approach for analyzing these data. We conclude by summarizing key insights and their potential implications.
2 Data

Data used in the analysis comprise all articles published in *Marketing Science* from its inception in 1982 through 2011 - 30 volumes in total. For each article, the data include the authors, volume, issue and date of publication, title, and the sets of key words chosen by the article’s author(s). There are no restrictions used on key words that authors can select in *Marketing Science* nor are there any ex ante classifications of key words from which to choose (c.f. the JEL Classification Codes used by many economics journals, which we discuss below). All totaled, there are 4749 key word choices, of which 2415 are unique, describing 1085 articles by 1051 authors (or co-authors). Given our interest lies with the organic evolution of the field, rather than policy, all 35 editorials appearing in the data are excluded, leaving 1050 articles, 1050 (co-)authors, and 4654 key words (of which 2357 are unique).

One obstacle to analyzing these data arises from conceptual overlap in key words with slightly different spellings, e.g., “auction” and “auctions;” “word of mouth” and “word-of-mouth;” and “variety seeking,” “variety-seeking,” and “variety-seeking behavior.” Accordingly, highly similar terms are merged. Likewise, we combine synonyms such as “store brands” and “private labels;” and “movies” and “motion pictures” using our collective judgment for these decisions. These substitutions are listed in an online technical appendix. After combining words, the data reduce further to 1975 unique key words.

Remarkably, most of the key words appearing in *Marketing Science* have been used just once, and 97% of papers include at least one key word never before used. The mean and median number of key word appearances over more than three decades are 2.3 and 1 respectively. This suggests that most key words never “stick.”
3 Analysis and Historical Insights

3.1 Top Words

Table 1 reports the frequency count of the top 25 words used in *Marketing Science* over the past three decades. “Pricing” appears most often. Of note, elements from the so-called “4P’s” - “pricing;” “advertising” and “promotion;” “channel” and “retailing” (place); and “new products” - occupy six of the top eleven slots, demonstrating the centrality of this framework to the field and providing evidence for a strong alignment between marketing science and marketing practice. Likewise, aspects of the so-called 3C’s are also heavily represented in the leading words, with “choice models” (customer) and “competitive strategy” in the top six. The impact of methodological innovation is evidenced in *Marketing Science*; over a quarter of the top 20 key words are method oriented.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Frequency</th>
<th>Keyword</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>▲ Pricing</td>
<td>122</td>
<td>Diffusion</td>
<td>37</td>
</tr>
<tr>
<td>Game theory</td>
<td>89</td>
<td>Hierarchical Bayes</td>
<td>37</td>
</tr>
<tr>
<td>▲ Advertising</td>
<td>61</td>
<td>Buyer behavior</td>
<td>27</td>
</tr>
<tr>
<td>▲ Choice models</td>
<td>59</td>
<td>Forecasting</td>
<td>27</td>
</tr>
<tr>
<td>▲ Channel</td>
<td>58</td>
<td>Conjoint analysis</td>
<td>26</td>
</tr>
<tr>
<td>▲ Competitive strategy</td>
<td>52</td>
<td>Customer satisfaction</td>
<td>25</td>
</tr>
<tr>
<td>Econometric models</td>
<td>50</td>
<td>Price discrimination</td>
<td>23</td>
</tr>
<tr>
<td>▲ Promotion</td>
<td>46</td>
<td>Market structure</td>
<td>22</td>
</tr>
<tr>
<td>▲ New products</td>
<td>46</td>
<td>Marketing mix</td>
<td>22</td>
</tr>
<tr>
<td>▲ Brand choice</td>
<td>45</td>
<td>Marketing strategy</td>
<td>22</td>
</tr>
<tr>
<td>▲ Retailing</td>
<td>45</td>
<td>Structural models</td>
<td>22</td>
</tr>
<tr>
<td>▲ Competition</td>
<td>43</td>
<td>Internet marketing</td>
<td>21</td>
</tr>
<tr>
<td>Bayesian analysis</td>
<td>38</td>
<td>Motion picture</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 1: Most frequently used key words in *Marketing Science*, 1982 - 2011. Among the words used more than 40 times, the 4P’s are marked with a closed triangle (▲) and the 3C’s with an open triangle (△).

The appearances of individual key words are not independent. To understand how these words relate to each other, we perform a hierarchical cluster analysis using the Jaccard (1901) distance metric, defined as the ratio i) of the number of times both keywords appear in the
same article to ii) the total number of articles featuring either keyword, \(\#(A \cap B)/\#(A \cup B)\).

This metric classifies words as more similar if they appear alongside many of the same key words. The dendrogram in Figure 1 depicts the resulting clusters for the top 20 key words. Distances between leafs in the dendrogram, i.e., the sum of the lengths of the stems separating any two leafs, depict the (dis)similarity of words. For example, “Bayesian analysis” and “hierarchical Bayes” are closely related because they tend to appear with the same sets of other words.

![Dendrogram](image)

Figure 1: Dendrogram depicting a cluster analysis of the 20 most prevalent keywords. Groups of similar key words are contained in the same boxes.

Four key groupings are suggested by Figure 1. The words in the lowest box comprise the most specialized cluster, which represents words not as frequently used. In the lower level of this cluster, we observe “Bayesian methods,” “price discrimination,” “customer satisfaction,” and “advertising,” which may reflect customer-level targeting. At this cluster’s highest level, “diffusion,” “new products,” “forecasting,” and “conjoint analysis,” are focused on new product demand, and the upper level’s link to the lower level of the cluster may lie in the concept
of demand assessment. The next box up includes “econometric models,” “brand choice,” and “choice models,” possibly reflecting the scanner data era of research. The second box from the top includes “channel,” “competition,” and “pricing,” and appears to reflect research pertaining to channel strategy. Finally, of special interest is the key word “game theory” (at the top of Figure 1), which is not closer to any one topic cluster than another, but rather spans them all. Applied initially in the context of analytic models, “game theory” can be construed as a unifying lens through which researchers look in order to study a plethora of issues in marketing science. This trend has been extended to empirical research in the context of structural models, which have also adopted game theoretic techniques. The theme of how keyword use evolves over time is a point we consider next.

3.2 Key Word Lifecycles

Figure 2 reports word clouds for all keywords occurring over three 10 year periods of our data, aligning roughly with the 1980’s, 1990’s and 2000’s. In the word cloud, the font size of the word is proportional to its frequency of use over that duration. For example, a dominant word in each decade is “pricing,” consistent with Table’s 1 result that “pricing” is the most heavily used keyword.
Figure 2 suggests that the use of key words is not static; some have always been popular (e.g., “advertising,” “pricing,” “new products,” and “promotion”), some are in growth (e.g., “game theory” and “hierarchical Bayes”) and some are in substitution (“diffusion” was popular in the 1980’s, but “social networks” and “Internet marketing” are increasingly popular in the 2000’s). The figure also suggests the concentration of the leading key words by decade; there are a smaller percentage of very popular words in *Marketing Science*’s first decade, perhaps indicating that the field is in the growth phase of its product life cycle (Levitt (1965)) and becoming more fragmented over time (in terms of words used).

Figure 3 plots the share of selected key words over time in order to illustrate some of the more prominent trends in the data: growth, maturity and substitution. In panel (b) and (d) on the right, one observes the rapid ascendance of game theory and Bayesian methods as prominent paradigms in marketing. In panel (a) of Figure 3, we see that “pricing” and “promotion” were especially popular in the 1990’s, likely a consequence of the ubiquity of
scanner data research at the time and suggestive of maturity. The substitution of “social networks” and “Internet marketing” for “diffusion,” which we noted earlier, is also evident in panel (c).

Figure 3: Key word popularity over time. The four panels illustrate patterns of growth and substitution.

3.3 Innovation and What Sticks

Figure 4 depicts a scatter plot of key words by time. The vertical axis lists key words in the order of their first appearance in the data (earlier appearing words are at the bottom). The words used are suppressed to conserve space. The horizontal axis represents the years in which these keywords first appeared. Each circle on the plot represents the appearance of a word in a year, and the size of the circle is proportional to a key word’s share in that same year. Editor names are listed across the top of the figure for reference.
Figure 4: Emergence and popularity of new key words by year. Rows represent individual key words. Circle size is proportional to key word share in each year. Colored bands group key words together by the year they first appeared.

One insight emerging from the figure is that some words have been quite enduring. Many words appearing in 1992, for example, continue to appear today. Key words from other years (e.g., 1989 or 2004) have been less enduring. Denser horizontal “bands” reflect eras during which new, enduring key words were established (it would be of interest to assess what aspects of these eras led to their more enduring word use). Another insight is that the pace of key word introduction, as evidenced by the slope of the key word frontier, is accelerating. Much of this acceleration can be ascribed to an increase in journal space in 2004. The slope of the frontier is roughly linear both before and after this increase (though may have leveled some in the mid 1990’s). This acceleration might indicate the field has entered a growth phase, suggesting it has never been easier to innovate in terms of key words (and topics to the extent these equate with words). Another potential explanation for the patterns
observed is that there is a long-tail phenomenon with key word choice, wherein a core set of papers continues to have impact, but most of the growth in the field arises from papers that are marginally less impactful (Anderson, 2006). If so, one might expect - as we see in the data - that the increase in the number of new words would not be accompanied by an increase in the number that become widely used.

A similar scatter plot is produced in Figure 5, but now focuses solely on the top ten most popular key words in each of ten three year periods (we use three year increments in lieu of one year increments to facilitate interpretation, and include the top 11 words when there is a tie). It is important to note that the first appearance of a word on this plot marks the time at which a word first became popular enough to enter the top 10 in market share, and

---

### Figure 5: Emergence, persistence, and decline of highly popular key words, grouped in three-year periods. Words are ordered by their popularity in the year they first broke into the top 10. Circle size represents popularity within each year.
not the first time the word appeared in the literature. For example, even though Coughlan (1985) is the first paper published in *Marketing Science* to use the key word “game theory,” the topic did not become popular until the mid 1990’s. That is, there can be a response latency between a word’s first appearance and when it becomes popular. Of note, Figure 5 reinforces the notion that some words become enduring (e.g., “game theory,” “channel,” and “competitive strategy”), whereas some words are more ephemeral (e.g., “variety seeking” and “market segmentation”).

Figures 4 and 5 present an informative contrast: Figure 4, which includes all key words, is convex, whereas Figure 5, which includes only the most popular key words, is concave. Apparently, while it has become easier to innovate, it has also become more difficult to become (and remain) popular. This is consistent with the product life cycle notion that the field may be transitioning from growth to maturity (Klepper, 1996). This result is also suggestive of a long tail aspect to the increase in journal space. Though it is impossible to answer this question with these data (for example, words might not equate to topics and there is a latency between their first appearance and when they become popular), the issue merits additional consideration.

Regarding the concern that words might not align well with topics, one way to assess this is the preceding cluster analysis. Another approach, which further considers dynamics, is correspondence analysis (Hoffman and Franke, 1986), which we discuss next.

### 3.4 Eras of Marketing

In our application of correspondence analysis, years and keywords are collocated based on the frequency with which the words appeared, as depicted in Figure 6. One can think of a year being placed in the center of the words that were used that year, and a key word being placed in the center of the years in which it appeared. In order to keep the figure interpretable, we focus on words occurring at least 10 times. Results are reported in Figure 6.
Figure 6: Correspondence analysis showing key words and years. Key words are placed closest to the years in which they were the most popular.

As one might expect, terms that are used in all years, like “pricing,” appear closer to the center of the plot. These terms are less informative about the evolution of the field than those located closer to the periphery, which are more representative of words used in specific eras. Three such eras emerge, roughly corresponding to the decades of analysis, 1980’s (located on the right), 1990’s (top), and 2000’s (bottom). In the 1980’s, “consumer choice” and “logit” were especially popular, reflecting the promotional tools developed in that era (e.g., Guadagni and Little, 1983). In the 1990’s, scanner data became widely available and accordingly emphasis shifted toward work developed around “brand choice” and “economet-
ric models,” as further evidenced in 3. In the 2000’s, we see the emergence of “Bayesian methods,” “Internet,” “CRM/targeted marketing,” and more recently, “social networks.” The evolution from aggregate models of brand choice to individual level models of social engagement suggests that the field is reacting strongly to the informational environment in which it operates. As new industries and data (e.g., log files of customer interactions from fields such as the Internet, telecom and banking) become increasingly available, the field responds in terms of tools and topics. Arguably, 6 stands in contrast to the perspective of some that the field is becoming less relevant (Lehmann et al. (2011)).

3.5 Author Publications

In this section we consider author productivity and mortality at *Marketing Science*, and the potentially moderating role played by key words.

3.5.1 Author Productivity

The publication data are indicative of authors’ research productivity, allowing one to assess norms within the field. Figure 7 plots the distribution of publications per author in panel (a) on the left and rate of publications in panel (b) on the right. The latter is defined as the total number of publications divided by the number of years between the first and last publication observed. These figures are conditioned on having published in *Marketing Science*, a challenging task to begin with. Panel (a) indicates that authors who publish more than once are in the top 40% of the field in their productivity, reflecting the challenging nature of publishing in an elite journal. Panel (b) suggests that those who publish two or more papers do so at a rate of about two papers per year (with the caveat that these observations are skewed towards authors with fewer than five papers). Overall, the message is that it is difficult to publish, but that once this is done, publication rates become somewhat more regular.
Figure 7: Distribution of research productivity. The histogram in panel (a) shows the proportion of authors publishing a certain number times in Marketing Science. Panel (b) shows the proportion of authors achieving a given publication rate.

Exploring this idea in greater detail, Figure 8 depicts the author survival rate, defined as $\Pr \left[ \sum_{k+1}^{K} n_k > 0 | n_k \right]$ where $n_k = 1$ if an author publishes $k$ papers and 0 otherwise. One insight is that it becomes easier to publish up until 6 papers, at which point the likelihood of publishing an additional paper reaches 80%. This increase might reflect increasing skill or possibly selection bias. The likelihood of publishing then declines slowly, potentially owing to either ability (as skills become obsolete) or motivation (as tenured researchers pursue alternative endeavors). One noteworthy aspect of these figures is that it is difficult to publish, and remains so over the course of a career; there is no evidence of an “inside club.”
Figure 8: Author survival rates. Bars indicate the probability of publishing an additional paper, given a certain number have already been published.

### 3.5.2 State Dependence and Author Survival

Systematic differences exist among survival rates for authors depending on the types of key words used (e.g., methods vs. topics, growth vs. decline), and authors’ tendencies to repeat the same key words across many papers. Accordingly, we integrate our author survival and key word analyses. To achieve this aim, we first compute, for each author-paper-key word observation, the following set of dichotomous variables: i) whether the author used the same key word on their next paper (repeat), ii) whether they did not use the same key word on their next paper (switch), or iii) whether the author who used that key word published another paper (survival). Then, for each key word, we computed an empirical probability (across author-papers) of repeating the key word, switching to a new key word, or never publishing again. Note that these three state variables sum to one, and that the probability of survival after using a key word is the sum of that word’s repeat and switch probabilities.
Figure 9: State Dependence in Key Word Usage. Circles indicate the relative usage of key words. Larger circles denote greater use. Solid circles indicate methods key words, open circles indicate topics key words. The dashed lines are iso-curves indicating regions with constant survival rates. The lowest iso-curve corresponds to a 35% survival rate, the highest to 60%. Observations in the lower left (upper right) indicate lowest (highest) survival.

We compare the resulting repeat, switch, and survival probabilities for the top 25 key words in Figure 9. Several insights emerge. First, authors using methods words tend to have higher repeat likelihoods and lower switch likelihoods. This finding might be ascribed to the use of these tools as over-arching paradigms to study a variegated mixture of substantive topics that change from paper to paper. Greater inertia in methods may be indicative of higher barriers to entry for methods relative to topics (e.g., learning costs). Second, more common words, as evidenced by circle size, tend to be repeated more often - as one might expect owing to their increased prevalence. Third, words that are in their ascendancy (“Game Theory,” “Bayesian Analysis”, and “Internet Marketing”) relative to those that are not (“Forecasting” and “Market Structure”) are also words that are high in repeat and low in
switching.\textsuperscript{1} Such a pattern might arise if authors were to focus on “hot” issues by publishing runs of paper on these topics. In contrast, older words are more likely to have switches because it is harder to publish in declining areas.\textsuperscript{2}

Finally, it is possible to link research topics to survival. If all topics had the same survival rate, but differed only on the trade-off between repeat and switch, then all 25 words would perfectly align along a negative 45 degree slope, as indicated by the red dashed lines. In fact, a least square error fit of a line to the observations yields an intercept of 0.58 and a slope of -0.94. This result suggests that the slope of the observations in Figure 9 is indeed very close to -1, and that key words lie on an iso-survival curve of 60%. However, there is substantial dispersion from this line, as indicated by an $R^2$ of 0.29. This dispersion is greatest for the observations on the left hand side of the Figure 9 where repeat rates are low. As discussed earlier, words in the left tend to describe older topics. Two strategies are thus evidenced. First, above the 60% survival iso-contour, words such as “market structure” and “price discrimination” indicate higher survival rates. This pattern is consistent with survival by switching to new topics. Second, for those older words below the 60% iso-survival contour (such as “marketing strategy”) lower switching corresponded with lower survival rates. Interestingly, there is less dispersion of key words around the 60% survival iso-survival curve in the right hand side of the figure. These words evidence higher repeat and lower switching. Given that many of these words are method-oriented, one possibility is that authors use a method as a base, switching topics as time passes.

4 Conclusion

This paper identifies trends in marketing science through the lens of key words used in \textit{Marketing Science}. The analysis first and foremost helps to document the history of the

\textsuperscript{1}Note the similarity of the north-west to south-east orientation of word order in Figure 9 to the north-east to south-west orientation of word order in Figure 6). This result is consistent with the interpretation of words in growth and decline over the decades.

\textsuperscript{2}This finding pertaining to new words could also relate to censoring in that new words have yet to enter a decline that would incent authors to switch. To explore this possibility we created the same plot while omitting observations from the past five years. The results appear qualitatively similar.
journal, and in that sense may be a useful input in constructing a situation analysis for assessing the state of the journal. Moreover, the history of *Marketing Science* is potentially of interest in its own right to the authors who helped develop it.

The analysis generates a number of insights. First, the foundational aspects of marketing, as construed via the 3C’s and 4P’s, comport with the most frequent words used, suggesting a tight link between commonly used frameworks for defining the field and the choice of research questions considered. Second, the field is dynamic, with patterns of growth (e.g., “game theory”) and substitution (e.g., “diffusion” and “social networks”). This is an encouraging sign that the field is actively evolving. In the case of “game theory,” we further find that it does not cluster with other words; in this sense it is more of a unifying theme/paradigm that enables researchers to find common ground across topics and methods, like the “one ring to rule them all” (Tolkien, 1954). Third, there is a direction to this evolution that mirrors the evolution of data and issues faced in practice, indicating that the field does not appear to be veering from its applied roots. More data would be useful to better understand how trends in practice drive trends in research - and whether applied topics lead research, or research leads applied topics. Fourth, inspection of publication rates reveals that while it is challenging to publish, it becomes somewhat easier up to a point - and that the choice of key words seems related to the likelihood of survival. Last, survival rates may be affected by the types of words chosen and the authors likelihood of repeating them.

In addition, there are a number of relevant policy questions that emerge from this analysis. Owing in part to the addition of more journal space in 2004, it has become easier to innovate with key words. This is a good thing, in the sense that it is potentially enabling new ideas and might reflect a field in growth. Yet this growth in innovation has been matched by a deceleration in the rate with which new words stick. This might indicate a field in maturation, or perhaps reflect that more space is coincident with a longer tail in which the marginal papers published are not likely to be as influential as the infra-marginal papers. The latter might suggest that there are diminishing returns to adding additional journal pages. The decrease
in rates of new popular words should also be recognized by tenure committees looking for “home run” papers. The data suggest that the likelihood of publishing such papers over time might be declining.

A second policy question pertains to the use of classification codes such as the Journal of Economic Literature (JEL) codes used by the American Economic Association. In *Marketing Science*, it has been possible to use any key words, and in many years, there has been no code list. There are several advantages to using such a list. First, standardizing key words makes it easier to link words to topics. In our analysis, because words are often similar, words provide a noisy measure of topics, making it difficult to track the evolution of the field in terms of topics and the relative impact of ideas. Second, key words make it easier for authors to find similar works, and therefore encourage relevant early papers to diffuse into subsequent papers. For example, we aggregated several words into “motion picture,” including “film production,” “films,” “motion picture distribution and exhibition,” “motion picture industry,” “movie industry,” “movie theaters,” and “movies.” Papers using “films” might be harder to find by those using “movies” or “motion picture” even though these are synonyms. Third, key words enable editors to match reviewers and papers more easily, as one can consider the overlap between the papers and the authors that publish those papers. On the negative side, if the field is moving quickly, it might be somewhat constraining to limit organic growth of words as early indicators of new topics - i.e., it would be difficult for a committee tasked with key word classification to foresee all new topics as they first emerge. Hence a hybrid system might be of value (e.g., requiring most key words be chosen from a list, while allowing a limited number of new words). It seems sensible to explore these tradeoffs in greater depth and assess the suitability of classification codes for *Marketing Science*.

In sum, it is our hope that this analysis of the history of *Marketing Science* key words provides a portrait of the field that is of historical value and has implications going forward - and that this approach might be more generally applied in the context of other journals to obtain a broader understanding of this and other fields.
References


