



# Know your audience: how language complexity affects impact in entrepreneurship science

Hannes W. Lampe<sup>1</sup> · Jan Reerink<sup>1</sup>

Accepted: 28 December 2020 / Published online: 10 February 2021  
© The Author(s) 2021

## Abstract

This article addresses the importance of tailoring publications to expectations of the intended scientific sub-community it addresses. But what does this mean when writing an article and adopting community specific jargon? This article disentangles the effects of articles' language complexity on their impact. In the domain of entrepreneurship science, we show that language uniqueness (in form of aligning jargon uniquely to one community) has a positive effect on article's impact. An article's novelty (in form of novel recombination of community jargon) has an inverted U-shape relationship with impact. We further show that the optimal level of novelty decreases with increasing uniqueness, yielding higher overall impact. These findings have implications not only for authors of scientific articles but also for their audience.

**Keywords** Entrepreneurship science · Language complexity · Language uniqueness · Novelty · Natural language processing · Bibliometric analysis

**JEL Classification** L26 · Z13

*'If you talk to a man in a language he understands, that goes to his head. If you talk to him in his own language, that goes to his heart.'* Nelson Mandela.

## 1 Introduction

Researchers in all disciplines strive to increase their research impact in form of citations which have become the default measure of scientific success (Martin and Irvine 1983; Wang 2014). According to Merton (1968) citations serve as a major building

---

✉ Hannes W. Lampe  
hannes.lampe@tuhh.de

<sup>1</sup> Institute of Entrepreneurship, Hamburg University of Technology, Am Irrgarten 3-9, 21073 Hamburg, Germany

block of the scientific reward system. Even in the absence of citations, an article's publication in a peer-reviewed journal is considered as a signal of its original contribution and a measure of quality. Accumulating citations indicates additional peer recognition of an article's value and its impact for science (DeBellis 2009; Merton 1973; Simonton 2004). Previous research shows that the majority of Nobel laureates were among the top 0.1 percent cited authors (Garfield 1973). Furthermore, Cole and Cole (1967) show that citations are more important than the amount of publications in receiving awards and being widely known in the scientific community.

Recent research analyzing drivers of scientific impact focused their analysis on several dimensions: authors (Podsakoff et al. 2008; Judge et al. 2007; Nerkar and Lahiri 2014), journals (Podsakoff et al. 2005; Rafols et al. 2012), references (Mingers and Xu 2010; Pehlan et al. 2002) or linguistic attributes (Judge et al. 2007; Antons et al. 2019). We expand research on the latter with more sophisticated measures to detail the use of language and study the corresponding effect on article impact. Existing research uses basic linguistic attributes, e.g. an article's length or its number of keywords. More recently, Antons et al. (2019) analyzed the content of articles and their effect on impact by extracting underlying topics and their structure.

However, little is known on how articles use the variations of language in scientific communities that typically follow disciplinary conventions regarding language complexity (Stremersch et al. 2007; Locke and Golden-Biddle 1997). This is surprising as the choice of language seems to be likely to modify the article's reception and thus may provide a barrier for the author's research. Understanding these effects not only enables recommendations for authors but further explains how readers, reviewers as well as the evolution of scientific disciplines is affected by language complexity. We seek to address this research gap by asking: *How does language complexity in communicating science affect impact?* Contrary to previous research we take on a different perspective. We assume that scientific communities have their own languages and jargon (Kramsch 1998). We therefore unpack language complexity in two forms: first, we theorize and test how uniquely aligned an article's jargon should be to one scientific community (further called *uniqueness*) to gain legitimacy and maximize impact. Second, we disentangle the effect of novel recombination of community specific jargon (*novelty*) and propose theoretical arguments for the effect on impact. Furthermore, are we interested in uncovering the trade-off scientists face when choosing an article's level of uniqueness and novelty. Our results reveal that an article's uniqueness increases its perceived legitimacy and thus enables impact in the form of citations. We also show that the relationship between novel recombination of community specific jargon and impact follows an inverted U-shape. Furthermore, we show that for articles with a high level of uniqueness the optimal level of novelty is lower and the overall effect on impact is higher.

We answer our research questions based on a comprehensive sample of entrepreneurship articles. The context of entrepreneurship was chosen for multiple reasons: earlier articles from this discipline often focus on definitions of the field (Gartner 1990) or the potential of the field for research (Shane and Venkataraman 2000; Ireland et al. 2005). Meanwhile entrepreneurship has grown into a mature research field, as indicated by several systematic structuring approaches for the field (Cornelius et al. 2006; Reader and Watkins 2006; Schildt et al. 2006) as well as literature

analyzing its evolution (Busenitz et al. 2014; Grégoire et al. 2006). However, the effects of novelty and thus the underlying speed of the evolution has not yet been analyzed in the domain of entrepreneurship science. To do so, we combine traditional bibliometric techniques (Lampe and Hilgers 2015; Schildt et al. 2006) with measures from natural language processing (Robertson 2004). The combination of these techniques allows to detect and analyze the similarity and divergence of articles to scientific communities and thus their specific language or jargon. As entrepreneurship science is regarded as cross-disciplinary in nature it provides a good context for our study as it likely entails sub-communities with disparate language structures. We contribute to recent literature in several ways. First, we contribute to the quickly expanding body of research into the domain of science of science, unpacking potential effects of language complexity on impact (Judge et al. 2007). Second, we add to literature on legitimacy (Garud et al. 2014; Gurses and Ozcan 2015; Tauscher et al. 2020) by analyzing how scientific articles gain legitimacy and thus impact by uniquely aligning to one community language as well as recombining community language in a novel way. We further add to literature on novelty and its varying effects on research impact (Boudreau et al. 2016; Lee et al. 2015; Trapido 2015; Uzzi et al. 2013). Fourth, our findings improve our understanding of how citation behavior emerges and how it can influence the evolution and convergence of a scientific field such as entrepreneurship (Busenitz et al. 2014; Grégoire et al. 2006). Our findings also provide several practical implications for scientists seeking to maximize their research impact by showing the effects of using community specific language in combination with novelty.

## 2 Theoretical background

### 2.1 Language, jargon, and communities

Previous research has suggested that linguistic characteristics affect an article's impact. For example, Stremersch et al. (2007) show that an article's length has a positive effect on impact. The authors further find that the number of keywords has a negative effect on impact. Diving more deeply into article attributes, previous research shows that better readability, associated with a greater writing clarity, positively affects an article's impact (Judge et al. 2007). Surprisingly others find the opposite effect, leading to the implication that some scientists do not necessarily find more readable research more legitimate (Stremersch et al. 2007).

A closely related research domain focuses on the rhetoric of scientific texts (Gephart 1988; Gross 1990; McCloskey 1994; Simons 1990). Rhetoric is most broadly construed in the Aristotelian tradition, as honest argument intended for an audience (McCloskey 1994). This definition implies that as soon as scientists frame ideas for presentation to an identified audience, they are engaging in rhetoric. Locke and Golden-Biddle (1997) identify rhetorical practices that award credibility of contributions. These so-called markers are depicted by expressing inclusiveness (e.g. "both", "and", "not only") or exclusiveness (e.g. "but", "else", "nor") (Tausczik and

Pennebaker 2010; Pennebaker et al. 2015). Readers then may find articles presenting a blend of these rhetorical markers credible and cite their contributions in future work.

However previous literature on potential effects on scientific impact have mostly neglected that communication, especially language and rhetoric, is suggested as a symbol of social identity (Kramsch 1998; Vilhena et al. 2014). The use of a certain language might be understood as affiliation or belonging to a certain community (Kramsch 1998). Thus, language reflects a community's matters of focus, expertise and special concern, often depicted via artificial words, enunciated via compressed terms—jargon—frequently used as synonyms for more complex constructs. These artificial words or constructs are used to refer in the most efficient manner to familiar as well as common concepts. This linguistic compression, via jargon, for efficient communication between peers of a common community might likewise occur in criminal argot, a subcultural lingo or in regional dialects (Vilhena et al. 2014).

Due to its epistemic cultures, language and rhetoric are of special importance in science (Knorr-Cetina 1999). Scientific jargon allows a more precise and efficient communication with peers within the same scientific community. For example, when the term 'fitness landscape' is used by an evolutionary biologist, a comparison of expected relative reproductive successes across multiple genotypes is implied (Vilhena et al. 2014). It is very likely that a scientist from the domain of entrepreneurship would need a bit more of an explanation to understand the full context of this artificial term. Every scientific field has its own ideas, constructs and measures, often expressed via specialized jargon, which might not overlap with those close to other disciplines or communities.

Based on prior research we find that language or jargon is important when distinguishing scientific disciplines. We further argue that this finding applies not only to the distinction of scientific disciplines which are remote from each other, but also in sub-research fields of scientific domains and their underlying communities'. Especially entrepreneurship science, due to its heterogeneous nature, caused by its cross-disciplinarity, represents an ideal context for a more fine-grained analysis of the effects of community-based language or jargon. According to Kuhn's (1996) argument communities' underlying differences of language are like proponents of different theories. Thus, analyzing community specific rhetoric does not superficially refer to the language per se, but rather shows deeper layers of content as theories, norms and constructs as well as measures, utilized in a scientific community.

We now derive our hypotheses, elaborating the potential relationship between language complexity in form of *uniqueness* and *novelty* with impact, as well as the potential moderating role of uniqueness on the relationship between novelty and impact.

## 2.2 Uniqueness—choice of community

As described above, language is a powerful instrument of identity and belonging. As the introductory quote points out, addressing somebody in his or her own language might have a profound effect on that person. Thus, we assume that jargon

used in scientific publications is highly relevant for an article's perceived legitimacy in a certain community, followed by impact. Previous research has shown that narrative strategies are important in making meaning of opportunities, allowing them to contextualize innovations and make content meaningful in order to become legitimate (Garud et al. 2014; Gurses and Ozcan 2015). Institutional scientists argue that the deviation from a categorical prototype reduces the comprehensibility of a proposed new venture because it prevents audiences from linking the unknown to a familiar cognitive template (Navis and Glynn 2011). McKnight and Zietsma (2018) suggest that the notion of analogies to situate one's own approach into the current thinking to create a common ground from which to further separate one's own ideas. A different stream of research, focusing on category spanning, has shown that audience members refer to established categories to make sense of products (Hsu et al. 2009; Kovács and Johnson 2014). This research has also shown that spanning multiple categories has negative effects on audience appreciation and thus legitimacy.

In science, generally, it is well understood that knowing your audience is important for tailoring communication to the expectations of the recipients. This implicitly suggests that the jargon of the target audience is to be used in an article. Hence the question arises to what extent an article should be committed to a community's language. Should an article serve several communities in terms of jargon or focus on one community only? We therefore focus on articles' language uniqueness in terms of how clearly an article is assigned to one community (compared to other communities). On the one hand, an article might be assigned to several communities, thus using the jargon of several communities in terms of a more uniform distribution. On the other hand, an article might be focused on a certain community thus using more of that community's jargon relative to the jargons associated with other communities.

Taking up the example from the previous section, the term 'fitness landscape', we argue that when an article uses this jargon associated with a certain community, for scholars e.g. from the domain of psychology it is probably hard to understand, whereas scholars from the associate community—familiar with the term—are more likely to ad hoc understand and see this article as legitimate. This leads to the assumption that talking in a community's language yields higher impact. Whereas when intermingling jargon this would be more likely to drive away several audiences in form of not being legitimate for them. This again would result in less impact.

We conclude that if an increase in language uniqueness is observed it should increase an article's understanding (Boudreau et al. 2016; Garud et al. 2014; Gurses and Ozcan 2015) and its legitimacy resulting in higher impact. This leads to our first hypothesis:

*Hypothesis 1. An article's language uniqueness has a positive effect on the article's impact.*

### 2.3 Article's novelty

Novelty in science might be understood in terms of Schumpeter's (1939) concept of a recombinant nature of innovation: explaining innovation as a novel

recombination of existing knowledge, which would likely result in a mix of jargons from different communities. The construct of novel recombinations has been central in recent studies on scientific impact (Lee et al. 2015; Trapido 2015). The literature argues that scientific papers that draw on unusual or novel combinations of journals in their references can be thought of as representing relatively more novel knowledge (Uzzi et al. 2013). Among others, this idea has diffused into research areas such as technology (Kaplan and Vakili 2015; Valentini 2012) and science (Boudreau et al. 2016; Trapido 2015; Uzzi et al. 2013). However, previous research often used the constructs of novelty and impact interchangeably (Lee et al. 2015). Uzzi et al. (2013) as well as Lee et al. (2015) are among the first to disentangle the concepts of novelty and impact in the context of science. Although research has begun to understand the relationship between novelty and impact, mixed and contrary results are presented in previous literature, emphasizing the need for a better understanding of this relationship. Lee et al. (2015) argue that the relationship between novelty and impact is positive linear. Boudreau et al. (2016), in the context of research proposals, show that the novelty of proposals has a negative effect on evaluations. As an exploratory part of their article they allow novelty to take on a more flexible relationship with evaluations. Trapido (2015) moves closer to a curvilinear relationship, showing that lower-novelty work is associated with higher citation counts, while higher-novelty work has a negative effect on impact.

We are taking these different perspectives into consideration to derive two hypotheses about the potential effect of novelty on impact. Previous literature has shown that higher levels of novelty make research more interesting and moves it into new unknown territory which may lead to higher impact (Newman and Cooper 1993; Schoenmakers and Duysters 2010). Aldrich et al. (1994) argue that research, especially in the social sciences, is driven by novelty, surprise, controversy and interest. A closely related finding from previous research shows that interdisciplinary research has higher impact in the long run, by combining references from different disciplines (Van Noorden 2015). The argumentation behind this finding is that accessing and combining unusual knowledge domains or relying on a high variety of knowledge increases impact. Thus, a positive effect of an article's novelty on impact might be assumed (Lee et al. 2015). In this article's context, the novel recombination of community specific jargon is assumed to have a linear positive effect on article's impact. Thus, we propose our second hypothesis:

*Hypothesis 2. An article's novelty has a positive effect on the article's impact.*

Despite the assumed positive relationship between novelty and impact, novelty might also have contrasting effects. For example, research in psychology suggests a bias against novelty, arguing that more novel ideas might be more difficult to process (Miller 1986; Mueller et al. 2012). In management science, recent research points toward the negative effects of novelty as well: Boudreau et al. (2016) find that the evaluation of research proposals is negatively biased if the proposal's content is novel. A closely related finding by Uzzi et al. (2013) shows that high impact science derives for the most part from conventional (common and existing) recombination of

knowledge. Furthermore, Van Noorden (2015) shows that interdisciplinary research has lower impact in the short run.

In line with these arguments we argue that adopting more novel recombinations of community jargon, might confuse the reader of such an article and thus make an article's content more difficult to process. This difficulty in processing articles' content, might be due to non-understanding of the introduced artificial constructs and terms or a perceived distance and hence lack of interest in the topic. This would lead readers to reject the content of an article and would thus result in less impact. We further argue that this negative relationship increases with higher novelty values. For example, having solely a few artificial terms from different communities to deal with, a reader might be willing to investigate the meaning of these few terms. For higher values of novelty, more and more relatively new combinations of jargon—previously not combined—might be included. This would lead to a more excessive amount of artificial terms (not combined previously), which is most likely to frustrate any reader. We therefore propose that the downsides of novelty tend to increase as novelty increases, resulting in a convex or exponential negative effect on impact.

While there are several benefits of high novelty, with novelty also come escalating disadvantages. After a certain point, these costs start to dominate the linearly increasing benefits of novelty (see Hypothesis 2). An inverted U-shaped relationship between articles' novelty and impact is therefore predicted resulting in the following hypothesis:

*Hypothesis 3. The relationship between an article's novelty and its impact follows an inverted U-shape.*

## 2.4 The moderating effect of uniqueness on novelty

Having discussed the anticipated effects of uniqueness and novelty on an article's impact, we now turn towards the potential moderation effect of uniqueness on the relationship between novelty and impact. How important the use of language is when proposing new ideas is exemplified in the case of Isaac Newton, who wrote his revolutionary *Principia* in Latin. One reason was that his approaches would have sparked too much resistance when written in the English language (Hall 1980; Honig et al. 2014)—as English would have been far more difficult for the audience to understand. Thus, scientists are prone to use not only different languages but rather have to be aware how rhetorical nuances might affect legitimacy when writing about novel ideas. Similarly Uzzi et al. (2013) bring up the example of Darwin's scientific manifest *The Origin of Species*, arguing that the combination of convenient domain-level thinking was critical for the link between innovativeness and impact. We argue that it is just this convenient domain-level thinking which is formalized in community specific jargon and thus gaining a community's legitimacy.

These examples elucidate the important relationship between uniqueness and novelty. As proposed in Hypothesis 3 we expect novelty to have two opposing effects on impact, a linear positive one and a negative convex one (combined resulting in the proposed inverted U-shaped relationship). We thus will now elaborate on how uniqueness is likely to affect both of these effects.

Research into entrepreneurship has identified organizational needs to use narratives to allow them to contextualize novel and innovative content (Garud et al. 2014) and thus make their technology or invention meaningful and legitimate to others (Gurses and Ozcan 2015). Taking these considerations into account, it seems obvious for scientists to be aware of their article's uniqueness in order to propose novel ideas. We argue that uniqueness, in form of an article's unique alignment towards one scientific community in form of rhetoric and jargon, seems of vital importance to 'sell' novel ideas and thus to ensure the legitimacy of these novel ideas. This is further in line with Uzzi et al. (2013) who argue that conventional knowledge, here the unique alignment to one community, is critical to the link between novelty and impact. We thus argue that an article needs to have a high uniqueness to make a novel contribution legitimate and thus yielding higher impact. In technical terms, a higher uniqueness decreases the negative convex effect of novelty on impact, again due to higher uniqueness, less negative effects are prone due to novelty. This would result in the steepening of the curvilinear effect and thus a stronger mechanism of the inverted U-shaped relationship, followed by a higher effect of novelty on impact in its optimal point (Haans et al. 2016). Thus, we propose Hypothesis 4a:

*Hypothesis 4a. An article's higher level of uniqueness steepens the inverted U-shaped relationship between novelty and impact.*

Even though increasing novelty's legitimacy and impact, uniqueness is likely to affect the positive effect of novelty on impact in a negative manner. Again, we refer to Uzzi et al. (2013) who found that "the highest-impact science is primarily grounded in exceptionally conventional combinations of prior work yet simultaneously features an intrusion of unusual combinations" (p. 468). Thus, it is rather a nuance of novelty which is increasing impact. An alternative reasoning is coming from literature into the Not-Invented-Here (NIH) syndrome (Katz and Allen 1982; Antons and Piller 2015). NIH is defined as the tendency of a stable group (here a scientific community's jargon) to reject new ideas from outsiders (the recombination with jargon from other communities). Following this argument, we assume that higher uniqueness levels and thus a clearer affiliation to a certain community lead to lower acceptance of novel ideas—lower values of novelty seem to be optimal when an article holds high uniqueness values. High uniqueness is weakening the positive effect of novelty on impact, resulting in a turning point shift of the inverted U-shaped relationship between novelty and impact to a lower optimal level of novelty (for more detailed technical elaboration of a turning point shift see Haans et al. 2016), resulting in the following hypothesis:

*Hypothesis 4b. An article's higher level of uniqueness leads to a turning point shift towards lower optimal values for novelty.*

Figure 1 depicts our proposed theoretical model.

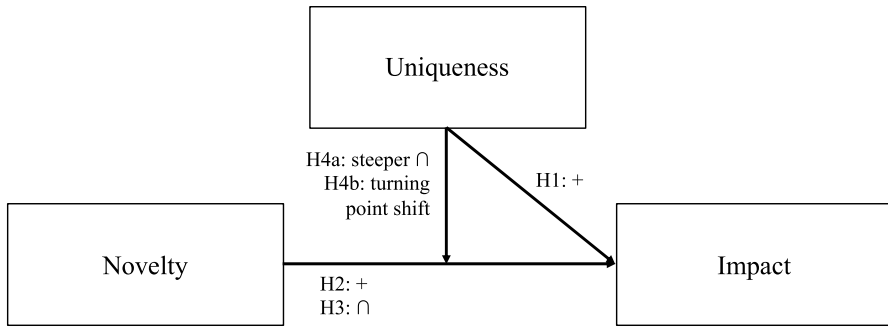


Fig. 1 Know your audience

### 3 Data and method

In this section we discuss our data and the methods applied. First, we describe the process of obtaining and preparing data for the analysis. In the second sub-section, we give an overview of the deployed methods. The procedure in this analysis is based on two steps: first, we detect scientific communities, and second, we detect the content similarity between all articles in the sample and the prior detected communities. For the first step we use document co-citation analysis to define sub-research streams of entrepreneurship science, we then match articles not included in the clusters to these clusters based on natural language processing. Lastly, we refer to our regression analysis elucidating details about used variables and model specifications.

#### 3.1 Data

To analyze entrepreneurship research, we use the Thomson Reuters Web of Science (WOS) to retrieve bibliometric data on corresponding publications. WOS is a prominent citation database, covering over 10,000 high impact journals and 120,000 international conference proceedings. In order to capture a broad selection of potentially relevant articles we used the search term ‘entrepre\*’ (with \* as wildcard).<sup>1</sup> The query was applied to paper titles, abstracts as well as keywords (both original keywords and keywords generated by WOS). The search was conducted in August 2014 including a timespan from 1945 to August 2014 resulting in 21,973 unique WOS records. Excluding all non-articles (such as book chapters or conference

<sup>1</sup> We use such a general search string even though it is likely that *entrepre\** is mentioned in several articles only loosely related to the research into entrepreneurship. By using a citation-based technique (see Sect. 3.2.1), and thus a backward oriented approach, to detect research clusters of entrepreneurship we assume that clusters emerging in our DCA analysis solely focus on entrepreneurship. Others, only loosely related are unlikely to yield high co-citation metrics with respect to articles citing these are all to some extent associated with entrepreneurship. Indeed, all our clusters are strongly related to the topic of entrepreneurship, as can be seen in Table 7.

proceedings) resulted in 16,683 records; leaving out all non-English articles and articles with missing values left us with 14,028 documents.

## 3.2 Methods

Dividing entrepreneurship research into communities (e.g. clusters or sub-research fields) and studying these clusters' development over time requires a combination of diverse methods. Hence, the following sections give a brief introduction to co-citation analysis and natural language processing. Furthermore, we explain how we linked these two methods to increase the quality of our results.

### 3.2.1 Delineating scientific communities in entrepreneurship science

In order to delineate scientific communities within the scientific field of entrepreneurship, and thus to be able to assess the uniqueness of each article, we perform a document co-citation analysis (DCA). Such an analysis is particularly relevant for this purpose, because it measures paper relatedness based on the frequency with which two documents are cited together by other documents (Cawkell 1976; Garfield et al. 1978; Small 1973), overcoming subjectivity due to its quantitative analysis of citations (Lampe and Hilgers 2015; Schildt et al. 2006). Due to their strong relatedness, these detected sub-research fields or clusters of a scientific domain might be equated to a scholarly community (Schildt et al. 2006). The cleaning of the data was conducted following Lampe and Hilgers (2015).<sup>2</sup>

We first excluded all papers with less than four references to only include research articles, resulting in 14,657 papers. In a second step we only kept articles which received 15 or more citations to ascertain the analyzed citation behavior to be validated by specialists in this research domain (resulting in 3358 articles). Hence our findings built upon a wide range of expert opinions (scholars' citations) and thus accepted principles. After building the DCA-network, we deleted isolates (i.e. articles not linked to any other articles) resulting in the final DCA dataset of 2117 articles with 62,511 co-citation links. These steps enable a robust citation analysis, minimizing the possible effect of noise (Lampe and Hilgers 2015).

Following earlier research, we adopt the Jaccard index (Jaccard 1901) as a normalized measure for the connectivity of co-cited articles (Small and Greenlee 1980). This index gives the ratio of the number of co-citations to the total citations of A and B less their common co-citations (Gmür 2003). The value of the Jaccard index (S) ranges from 0 (no co-citations) to 1 (representing perfect co-citation) and is defined as follows:

---

<sup>2</sup> This process can be distinguished into four steps. First, normalizing all letters end thus enabling case sensitive algorithms to detect similarity in author names. Second, we merged identical authors using the Jaro-Winkler metric (Jaro 1989, 1995; Winkler 1999). As authors are not the restrictive character of our analyses, we kept the underlying threshold quite small. We manually corrected for errors between a similarity threshold of 40% and 60% and automatically merged similar authors above 60%. Third, based on an 'Authoritative Journal Merging List' provided by the Sci2 Team (2009), we merged identical journals to account for misspellings in the references of articles. Fourth, citations were matched to documents.

$$S = \frac{\text{number of common citations to articles } A \text{ and } B}{(\text{Total citations to } A + \text{Total citations to } B - \text{Co-citations of } A \text{ and } B)} \quad (1)$$

When defining co-citation clusters and distinguish them from each other we opted for the straight-forward approach of removing weak links. We exclude all links with a Jaccard value lower than 0.2. The cut off value of 0.2 results from a comparison of various cut-off values and the resulting number of disconnected components in the network. We tried to find a value where the number of clusters would not change with a slight change of the threshold (Lampe and Hilgers 2015). Compared to previous research, this cut off value is quite small (Schildt et al. 2006), a necessity following the larger number of articles considered in this dataset. The issue of false positives showing up in the dataset due to the basic search query is mitigated by this step: papers that do not belong to the field of entrepreneurship are unlikely to have been highly co-cited by those papers that do belong to the field. Overall, we identified 35 different sub-research fields of entrepreneurship science (stated with description and metrics in Table 8).

### 3.2.2 Variables

The dependent variable, impact, is operationalized by citation counts (average per year) in Web of Science. This measure is commonly used by scholars when analyzing patents or publications (Lee et al. 2015; Martin and Irvine 1983; Moed 2005; Wang 2014). We further use the average yearly citation count to allow older articles to be more cited.

Two variables, concerning our hypotheses are included: article uniqueness and novelty. In order to determine the uniqueness and novelty of articles, with respect to their affiliation to a cluster, we need to identify each article’s similarity to each cluster first. We therefore use a widely-accepted method in data analysis for weighting the importance of words in text collections, namely tf-idf (term frequency – inverse document frequency) (Robertson 2004).<sup>3</sup> Given a collection of texts in a corpus ( $d \in D$ ) the tf-idf weight of a word for one text can be calculated as the product of the frequency of that word in the current text ( $f_{w,d}$ ) and the inverse document frequency. The inverse document frequency is the logarithm of the number of texts in a corpus ( $|D|$ ) divided by the number of texts containing the word to be weighted ( $f_{w,D}$ ):

$$tf - idf = f_{w,D} * \log \left( \frac{|D|}{f_{w,D}} \right) \quad (2)$$

Defining the content similarity and thus the similarity/deviation of articles to each cluster (previously determined by DCA), available abstracts and titles for one

<sup>3</sup> A matching based on keywords was excluded due to the quality of the underlying data. Both, ‘new ISI keywords’ (keywords generated by Thompson Reuters WOS) and ‘original keywords’ (keywords supplied by the papers’ authors) are not standardized whereas the ISI keywords fail to capture the information relevant for identifying research clusters, they seem to be more suitable for a more abstract classification.

cluster are added up to a new document. After appropriate pre-processing,<sup>4</sup> the term frequencies are calculated per cluster and over the collection of clusters. Subsequently, tf-idf scores are obtained and a list of tf-idf-weighted words is created for each article and cluster.<sup>5</sup> We used this information to obtain similarities between articles and clusters. We therefore transformed article abstracts into a vector representation allowing us to use the inverse cosine similarity measure to obtain distances between articles and clusters:

$$distance = 1 - \cos(\theta) = 1 - \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (3)$$

As we are interested in the deviation between articles and their corresponding research sub-fields we only calculate the inverse similarity measures for instances where the article has been published after the beginning of a cluster (earliest publication date of articles defining a cluster). We also ignore articles for which no data (e.g. abstracts) are available. The resulting dataset consists of 9846 articles.

Our first focal independent variable, article's uniqueness, is measured in terms of the deviation between the highest similarity measure (compared to communities/clusters) and the average of the other similarities between the focal article and communities. This measure allows to detect how unique an article is associated to a certain community. This measure therefore enables the measurement of the uniqueness of jargon and thus the focus on only one community to use jargon from.

Second, we measure an article's novelty in terms of its inverse cosine similarity of the similarity distribution of an article and the detected scientific communities in entrepreneurship science (detected using DCA). We therefore compare each article's distribution of community jargon (the similarity between an article and a community's language) to all distributions of articles published in the same or previous years. Each article is therefore defined as its similarity to each of the 35 communities represented by a vector of length 35 for each article (using tf-idf as explained above). This approach allows to analyze which article is a relatively new recombination of different jargons and thus community languages (defined by DCA clusters). To allow for the emergence of new communities, a similarity of a cluster might only arise when the first publication of clusters' defining articles is published in the same year as or before the focal one. Furthermore, restricting the comparison to previous articles allows our novelty measure to detect novelty with a changing definition over time. For instance, the same article's recombination in 1995 and 2005 might be novel for the earlier one whereas the later one is not as novel.

<sup>4</sup> Pre-processing includes stop-word removal, i.e. the filtering of words that add little meaning to a text such as articles or pronouns. Remaining words are stemmed using the Python library Natural Language Toolkit (NLTK) (Bird et al. 2009). This allows for aggregating different forms of words (e.g. plural and singular of a word). Furthermore, numbers, very short words (less than 3 characters) and non-alphabetic characters are removed, and capital letters are replaced by their underscore equivalents.

<sup>5</sup> We obtain tf-idf scores using the Python package Gensim (Rehurek and Sojka 2010).

**Table 1** Descriptive statistics and correlations

Variable	Mean	SD	1	2	3	4	5	6	7	8
Citations per year	1.48	3.09								
Novelty	0.09	0.04	0.10							
Novelty squared	0.01	0.02	0.08	0.92						
Uniqueness	0.04	0.01	0.09	-0.07	-0.04					
Seminal inactivity	4.67	6.93	-0.13	-0.08	-0.08	0.34				
Age of paper	6.62	5.69	0.17	0.03	0.04	-0.44	-0.80			
Number of authors	2.08	1.33	0.05	-0.01	0.00	0.17	0.12	-0.17		
Number of pages	18.02	9.07	0.11	0.00	-0.01	0.08	0.05	-0.04	-0.01	
References	52.24	32.53	0.014	-0.03	-0.02	0.27	0.26	-0.29	0.10	0.39

On the one hand the resulting vector assumes relatively high values when the similarity is relatively low and thus an article is rather novel in its nature. On the other hand, relatively low values represent a high similarity between articles. In accordance to Uzzi et al. (2013) we then used the 1 percent and as a robustness test, the 10 percent quantile value of each vector in comparison to all similarity values as an indicator for novelty.

Furthermore, several control variables are incorporated into the model. Our first control variable is seminal inactivity. More precisely an article's associated community's stagnation. This variable expresses the distance (in years) between the publication of the focal article and the newest article of the community with the highest similarity. This measure allows to detect how long no seminal article (detected by DCA) emerged in the community, compared to the publication date of the focal article. Further control variables are the age of the paper (in years compared to 2015), the number of authors, the amount of included references in a paper and the number of pages. As proxy for an article's quality we include 2237 journal dummies. Furthermore, we control for the corresponding sub-research field by including cluster dummies in accordance to the identified 35 clusters.<sup>6</sup> The descriptive statistics are shown in Table 1.

### 3.2.3 Analysis

Citations are discrete and typically have a broad distribution, with some articles receiving very high citation counts. Furthermore, our dependent variable cannot assume values smaller than 0 and corresponds to count data. The obvious approach would be a Poisson model (Hausman et al. 1984). However, the citation distribution is over-dispersed, thus many more highly cited articles occur than would be the case for Poisson-distributed data. Therefore, the relationship between citations and our

<sup>6</sup> Due to missing abstracts in two very small cluster we were not able to detect similarities between articles and these cluster text corpora, leading to neglect 2 clusters.

independent variables might be estimated using negative binomial regression, a generalization of the Poisson model that accounts for over-dispersion in the data.<sup>7</sup>

Negative binomial models, like Poisson models (Hausman et al. 1984), assume that the logarithm of the expected value of the dependent variable can be modeled by a linear combination of known predictors. In this sense, it is similar to estimating a regular linear regression with the logarithm of citations as the dependent variable. We therefore follow Foster et al.'s (2015) approach and assume the following:

$$Citations_a \sim NegativeBinomial_{\mu} \quad (4)$$

where  $Citations_a$  depicts the average amount of yearly citations received by article  $a$ . We use ordinary least square regression to test for our two hypotheses, using the natural logarithm of the number of citations (plus one) relative to the article's age as dependent variable. To challenge the robustness of our findings, and in order to correct for an excessive number of zeros in our data we also used zero-inflated negative binomial regression (Long 1997).

## 4 Results

In order to test our hypotheses, we first had to detect scientific communities in the domain of entrepreneurship research. To determine a quantitative categorization of entrepreneurship sub-research fields or communities, we conduct a document co-citation analysis akin to that of Schildt et al. (2006) to reveal the different clusters of these research areas. Given that we are interested exclusively in the most cited and coherent groups of articles, some of the highly-cited articles will be excluded from this analysis due to their lacking affiliation to a cluster. In total, we found 35 clusters. As expected, top clusters are represented mostly by papers published in the last two decades. Very recent papers may not be available in Web of Science or may not have been cited often enough for co-citation patterns to emerge. But it is surprising that older papers do not seem to be part of these clusters. Intuitively, papers that had more time to be cited and that are upstream in a field of research should receive many co-citations and therefore show up in clusters. A possible explanation is that this intuitive reasoning applies but is moderated by the small yearly publication numbers before 1990, which may in turn be influenced by data coverage of the Web of Science database.

Given the importance of scientific impact, Table 2 shows the results of the regression analysis including the effect of article characteristics on their citations per year (in natural logarithm).

Model 1 is the basic model, including all control variables. All of our control variables have a significant positive effect on article impact. The age of a paper,

<sup>7</sup> Negative binomial models contain an extra parameter to capture over-dispersion (i.e., a thick right-hand tail); this parameter (alpha) equal 0 in the case of a Poisson. In all our cases, the over-dispersion parameter is significantly different from zero according to a likelihood ratio test comparing negative binomial model to a Poisson mode,  $p < 0.001$ , so our use of the negative binomial is justified throughout.

**Table 2** Regression results of ordinary least squares estimations on articles' citation per year

	DV: natural logarithm of citations per year (plus 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Age of paper	0.176*** (0.024)	0.154*** (0.023)	0.212*** (0.023)	0.222*** (0.023)	0.221*** (0.023)	0.221*** (0.023)
Number of authors	0.051*** (0.004)	0.042*** (0.004)	0.042*** (0.004)	0.043*** (0.004)	0.043*** (0.004)	0.043*** (0.004)
Number of pages	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
References	0.005*** (0.0002)	0.004*** (0.0002)	0.004*** (0.0002)	0.004*** (0.0002)	0.004*** (0.0002)	0.004*** (0.0002)
Seminal inactivity	0.139*** (0.024)	0.106*** (0.023)	0.164*** (0.023)	0.174*** (0.023)	0.172*** (0.023)	0.173*** (0.023)
Uniqueness		10.880*** (0.397)	11.491*** (0.390)	11.733*** (0.389)	9.308*** (0.786)	5.100*** (1.369)
A: Novelty			3.274*** (0.142)	6.972*** (0.336)	5.961*** (0.441)	3.324*** (0.829)
B: Novelty squared				-11.382*** (0.941)	-11.357*** (0.940)	-2.907 (2.439)
Uniqueness * A					28.285*** (7.966)	102.431*** (21.294)
Uniqueness * B						-235.263*** (62.663)
Constant	-1.161*** (0.196)	-1.326*** (0.191)	-2.169*** (0.191)	-2.499*** (0.192)	-2.395*** (0.194)	-2.251*** (0.198)
Cluster dummies	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.147	0.191	0.221	0.229	0.230	0.230
Adjusted R <sup>2</sup>	0.145	0.189	0.219	0.227	0.227	0.228
Residual Std. error	0.601	0.585	0.574	0.572	0.571	0.571
F statistic	67.190***	89.209***	104.268***	106.405***	104.147***	102.045***

Standard errors are in parentheses. 14,028 observations (10,408 non-zero and 3620 zero); 31 cluster dummies are included (due to quasi-complete separation three dummies were neglected)

\*p < 0.05; \*\*p < 0.01; \*\*\*p < 0.001

the number of authors, the number of pages as well as the number of references all affect an article's impact positively. Seminal inactivity, the distance in years between an article and its associated clusters' latest article, also has a positive effect on article impact. If a paper associates itself to a community where seminal articles are quite old its impact is likely to be higher than associating a paper to a community where the last seminal article is rather recent.

Our first hypothesis, that an article's uniqueness has a positive effect on articles' impact is tested in Model 2. The focal variable of Hypothesis 1, uniqueness

**Table 3** Test of an inverted U-shaped relationship between novelty and impact

	Cultural novelty
Test of joint significance of focal variables (standard and squared) (p-value)	0.000
Sasabuchi-test of inverse U-shape (p-value)	0.000
Slope (low)	4.985***
Slope (high)	-5.961***
Estimated extreme point	0.27
95% Confidence interval—Fieller method	(0.251, 0.295)
Test of joint significance of control variables (p-value)	0.000
Test of joint significance of all variables in the model (p-value)	0.000

is positive significant ( $\beta = 10.880$ ;  $p < 0.001$ ) and thus lends support for Hypothesis 1. An article's uniqueness has a positive effect on its impact.

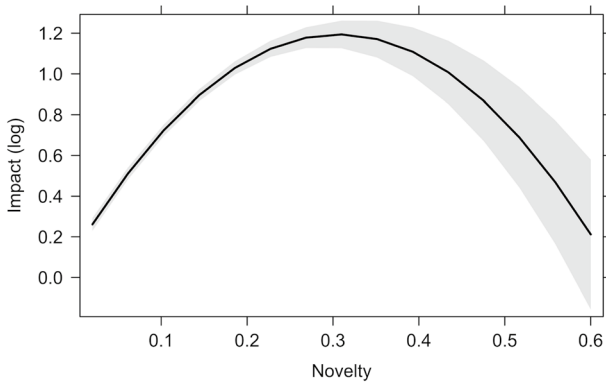
Hypothesis 2 predicted a positive linear effect of an article's novelty on impact. The results, presented in model 3 in Table 2 support this hypothesis showing a significant positive effect ( $\beta = 3.274$ ;  $p < 0.001$ ). Hypothesis 3 predicted an inverted U-shaped relationship between novelty and impact. The results found support for this relationship ( $\beta = -11.382$ ;  $p < 0.001$ ). To ensure the correct interpretation of our results we follow the three-step procedure to test an inverted U-shaped relationship, proposed by Lind and Mehlum (2010). First, as stated above,  $\beta$  of the squared term needs to be significant and of the expected sign. Furthermore, we test the joint significance of the direct and squared terms of novelty, following Sasabuchi's (1980) test for an inverted U-shaped relationship for novelty.

Second, the slope must be sufficiently steep at both ends of the data range. Table 3 shows the directions of the slopes at low and high values of novelty. If the slope at the low value of novelty is positive and significant ( $\beta = 4.985$ ;  $p < 0.001$ ) and if the slope at the high value of novelty is negative significant ( $\beta = -5.961$ ;  $p < 0.001$ ), then preliminary evidence of an inverted U-shape relationship is present.

Following Lind and Mehlum (2010), the third step to test for U-shaped relationships is to assess whether the turning point is located well within the data range. We therefore estimated the extreme point of the effect of novelty and calculated confidence intervals based on Fieller's standard error (Lind and Mehlum 2010). In addition, the confidence intervals for the Fieller standard error indicate that the novelty values were within the limits of the data (0.251, 0.295). As shown in Table 3 the inverted U-shaped relationship is significant.

To round out the robustness of our findings, Fig. 1 shows the predicted U-shaped relationship between novelty and impact based on our estimates of Model 4 in Table 2. Overall these findings lend support for Hypothesis 3, an inverted U-shaped effect of novelty on articles' impact.

Model 5 and 6 take the proposed moderation effect into account. Hypothesis 4a assumed that higher values of uniqueness steepen the inverted U-shaped relationship between novelty and impact. The interaction term between uniqueness and the squared novelty term (B) in model 6 is significant negative ( $\beta = -235.263$ ;



**Fig. 2** Effect of novelty on impact (95% confidence intervals are displayed)

$p < 0.001$ ), lending support for Hypothesis 4a. Figure 2 graphically displays the effect of novelty on impact with respect of low, medium and high values of uniqueness.<sup>8</sup> The observed steepening of the inverted U-shape as well as the upwards movement of the optimal point of novelty lend additional support for Hypothesis 4a. Hypothesis 4b assumed decreasing levels of the optimal novelty for increasing uniqueness values of an article, namely a turning point shift. The proposed effect is observable (highlighted graphically), lending support for Hypothesis 4b. Even though the interaction term with the linear term is significant, this is neither a necessary nor a sufficient condition (Haans et al. 2016) for the proposed effect of Hypothesis 4b. Thus, we use a formal test for a turning point shift proposed by Haans et al. (2016). The following equation states the full model specification, including all interaction:

$$Y = \beta_0 + \beta_1X + \beta_2X^2 + \beta_3XZ + \beta_4X^2Z + \beta_5Z \tag{5}$$

Haans et al. (2016) set the first derivative of the regression equation with respect to novelty to zero to derive the turning point of the inverted U-shaped effect of novelty on impact. The authors further take the derivative of this equation with respect to the moderator resulting in:

$$\frac{\delta X^*}{\delta Z} = \frac{\beta_1\beta_4 - \beta_2\beta_3}{2(\beta_2 + \beta_4Z)^2} \tag{6}$$

Evident from the above equation is that a potential turning point shift does not only depend on the first order interaction but also on the second order interaction term ( $\beta$ ) (Haans et al. 2016) supporting the use of the full model specification (Model 6). As suggested by Haans et al. (2016), we assess whether above equation

<sup>8</sup> The three levels depict the mean value of cultural uniqueness and one standard deviation above (high) and below (low) the mean.

as a whole is significantly different from zero, for specific meaningful values for the moderator. We deploy values of our moderator variable uniqueness ( $Z$ ) ranging from  $\min=0.000$  to  $\max=0.095$  with 0.01 steps. For all of these values the equation is negative and significantly different from zero ( $p<0.001$ ) (Table 7). These results lend support for Hypothesis 4b, a moderated turning point shift of the optimal level of novelty with respect to uniqueness. As hypothesized, the moderation effects the turning point shift in a negative manner, thus decreasing the level of optimal novelty with increasing uniqueness of an article.

## 5 Robustness test

To challenge the robustness of our results we replicate the results adding journal dummies (2237 dummies) to account for the potential effects of being published in different journals. The journal an article is published in might not only affect its impact directly but further indirectly via a quality approval of an article being published in a highly accredited journal. Imitating the results from Table 2 and further adding journal dummies are presented in Table 4. The results mostly stay the same. Hypothesis 1 to 3 again are affirmed by significant effects with the proposed signs. Uniqueness (Model 2) positive significantly affects article's impact ( $\beta=8.003$ ;  $p<0.001$ ) lending support for Hypothesis 1. Hypothesis 2 proposed a positive linear relationship between novelty and impact. The effect in Model 3 of Table 4 supports this hypothesis ( $\beta=2.982$ ;  $p<0.001$ ). Hypothesis 3, assumed an inverted U-shaped relationship between novelty and impact. The results are significant ( $\beta=6.590$ ;  $p<0.001$ ) further lending support for Hypothesis 3. A difference compared to the above results might be observed for the interaction effect of uniqueness and novelty in model 5 and 6. Whereas the single interaction effect in model 5 is not significant anymore, the double interaction effect in model 6 is slightly significant ( $\beta=46.541$ ;  $p<0.05$ ). The negative significant effect of uniqueness on the squared novelty term again lends support for Hypothesis 4a ( $\beta=-127.786$ ;  $p<0.05$ ), the steepening of the curvilinear relationship. As the significance of the interaction term is neither a necessary nor a sufficient condition (Haans et al. 2016) for the proposed turning point shift in H4b, we again test this Hypothesis following the formal test of Haans et al. (2016) (Eq. 6). Again, the results lend support for Hypothesis 4b.

To further test the robustness of our results we make use of the zero inflated negative binomial model (Tables 5 and 6). We therefore use the rounded publications per year as the dependent variable. The z-value of the Vuong test (Vuong 1989) is significant in all models and thus supports the model (Model 5:  $z=10.41$ ;  $p<0.000$ ). The results are similar to the ones using ordinary least squares in Table 2. Table 5 states the main part of the zero inflated negative binomial estimation and Table 6 states the zero-inflated part of the zero inflated negative binomial model. Similar to the results including journal dummies (Table 4) Hypothesis 1 ( $\beta=22.473$ ;  $p<0.001$ ), Hypothesis 2 ( $\beta=6.612$ ;  $p<0.001$ ) and Hypothesis 3 ( $\beta=14.878$ ;  $p<0.001$ ) are supported. The significance for the interaction effect between novelty and uniqueness is vanished in the zero inflated

**Table 4** Regression results of ordinary least squares estimations on articles' citation per year (including journal dummies)

	DV: natural logarithm of citations per year (plus 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Age of paper	0.118*** (0.023)	0.118*** (0.023)	0.167*** (0.023)	0.178*** (0.023)	0.178*** (0.023)	0.178*** (0.023)
Number of authors	0.022*** (0.004)	0.020*** (0.004)	0.020*** (0.004)	0.021*** (0.004)	0.021*** (0.004)	0.021*** (0.004)
Number of pages	0.005*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.001)
References	0.001*** (0.0002)	0.001*** (0.0002)	0.001*** (0.0002)	0.001*** (0.0002)	0.001*** (0.0002)	0.001*** (0.0002)
Seminal inactivity	0.083*** (0.023)	0.076** (0.023)	0.125*** (0.023)	0.135*** (0.023)	0.135*** (0.023)	0.135*** (0.023)
Uniqueness		8.033*** (0.444)	8.708*** (0.437)	8.913*** (0.435)	8.468*** (0.807)	6.112*** (1.373)
A: novelty			2.982*** (0.141)	6.590*** (0.335)	6.385*** (0.459)	4.893*** (0.840)
B: novelty squared				-10.759*** (0.909)	-10.723*** (0.910)	-6.095** (2.365)
Uniqueness * A					5.242 (8.007)	46.541* (21.062)
Uniqueness * B						-127.786* (60.279)
Constant	-0.955 (0.567)	-1.129* (0.560)	-1.689** (0.550)	-1.984*** (0.547)	-1.966*** (0.548)	-1.884*** (0.549)
Journal dummies	Yes	Yes	Yes	Yes	Yes	Yes
Cluster dummies	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.445	0.460	0.480	0.486	0.486	0.486
Adjusted R <sup>2</sup>	0.338	0.355	0.379	0.386	0.386	0.387
Residual Std. error	0.529	0.522	0.512	0.509	0.509	0.509
F statistic	4.149***	4.406***	4.770***	4.886***	4.884***	4.885***

Standard errors in parentheses; 14,028 observations (10,408 non-zero and 3620 zero); 2237 Journal dummies are included; 31 cluster dummies are included (due to quasi-complete separation three dummies were neglected)

\*p < 0.05; \*\*p < 0.01; \*\*\*p < 0.001

negative binomial model specification. For H4b this is neither a necessary nor a sufficient condition (Haans et al. 2016) again. Thus, we again conducted the formal test proposed by Haans et al. (2016). The results lend support for Hypothesis 4b (p < 0.009). In this nonlinear model specification, the significance of the interaction term between uniqueness and the squared novelty term (B) is also neither a

**Table 5** Main part of the zero-inflated negative binomial estimations

	DV: natural logarithm of citations per year (plus 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Age of paper	0.587*** (0.063)	0.530*** (0.063)	0.625*** (0.064)	0.652*** (0.063)	0.574*** (0.065)	0.573*** (0.065)
Number of authors	0.139*** (0.012)	0.122*** (0.012)	0.120*** (0.011)	0.155*** (0.013)	0.164*** (0.012)	0.164*** (0.012)
Number of pages	0.005** (0.002)	0.006*** (0.002)	0.004** (0.002)	0.007*** (0.002)	0.010*** (0.002)	0.010*** (0.002)
References	0.010*** (0.001)	0.009*** (0.001)	0.009*** (0.001)	0.008*** (0.001)	0.012*** (0.001)	0.012*** (0.001)
Seminal inactivity	0.450*** (0.063)	0.376*** (0.063)	0.479*** (0.064)	0.510*** (0.064)	0.478*** (0.065)	0.478*** (0.065)
Uniqueness		22.473*** (1.018)	23.204*** (1.003)	23.691*** (1.001)	25.150*** (2.088)	20.982*** (3.571)
A: novelty			6.612*** (0.369)	14.878*** (0.841)	13.737*** (1.153)	11.302*** (2.050)
B: novelty squared				-24.898*** (2.254)	-21.834*** (2.192)	-14.443** (5.599)
Uniqueness * A					-12.491 (20.643)	58.302 (53.414)
Uniqueness * B						-215.738 (149.965)
Constant	-5.240*** (0.539)	-5.609*** (0.539)	-7.130*** (0.547)	-8.000*** (0.550)	-7.585*** (0.573)	-7.439*** (0.583)
Cluster dummies	Yes	Yes	Yes	Yes	Yes	Yes
Log likelihood	-20,983	-20,737	-20,561	-20,522	-20,351	-20,350

Standard errors in parentheses; 14,028 observations (10,408 non-zero and 3620 zero); 31 cluster dummies are included (due to quasi-complete separation three dummies were neglected); Vuong test of zero inflated negative binomial model vs. standard negative binomial mode:  $z = 10.17$ ,  $p < 0.000$

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

necessary nor a sufficient condition (Haans et al. 2016) to test our Hypothesis 4a. Thus, we rely on our previous results (from Tables 3 and 4) to confirm H4a.

## 6 Discussion, limitations and future research

In this chapter we discuss the study's results and their contribution to existing literature. Moreover, we address the limitations of this study and provide suggestions for future research. Overall our findings show that scientists should be aware of their language complexity in form of community specific jargon when writing scientific articles.

Our first result (H1) shows that uniqueness in terms of adopting one scientific communities' language—and focusing on this jargon with respect to other jargon—is

**Table 6** Zero-inflated part of the zero-inflated negative binomial estimations

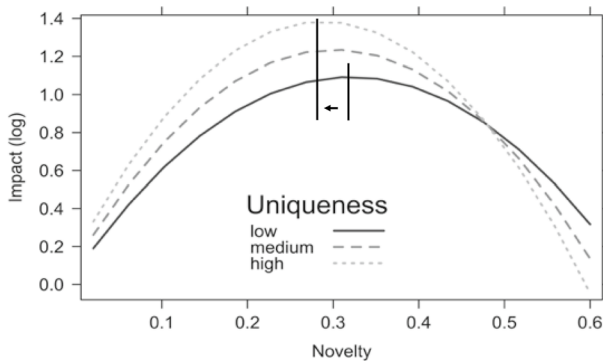
	DV: natural logarithm of citations per year (plus 1)					
	(1)	(2)	(3)	(4)	(5)	(6)
Age of paper	0.229*** (0.033)	0.234*** (0.035)	0.170*** (0.034)	0.120*** (0.029)	-1.381*** (0.075)	-1.380*** (0.075)
Number of authors	-0.644*** (0.117)	-0.641*** (0.124)	-0.573*** (0.127)	0.065 (0.068)	0.011 (0.028)	0.011 (0.028)
Number of pages	-0.074*** (0.015)	-0.075*** (0.016)	-0.077*** (0.018)	-0.014 (0.009)	0.0001 (0.008)	0.0002 (0.008)
References	-0.063*** (0.007)	-0.065*** (0.007)	-0.077*** (0.008)	-0.107*** (0.008)	-0.004 (0.002)	-0.004 (0.002)
Seminal inactivity	0.020 (0.016)	0.021 (0.017)	0.013 (0.018)	0.021 (0.018)	0.025 (0.018)	0.025 (0.018)
Constant	-0.754 (0.560)	-0.885 (0.591)	0.226 (0.526)	-0.423 (0.456)	1.527*** (0.287)	1.525*** (0.287)
Log likelihood	-20,983.590	-20,737.400	-20,561.720	-20,522.310	-20,351.530	-20,350.500

Standard errors in parentheses; 14,028 observations (10,408 non-zero and 3620 zero)

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

shown to have a positive effect on articles' impact. This finding is manifested in science by statements as *Know Your Audience* or evident by our starting quote, suggesting to speak your audience's or community's language to increase impact. For authors of scientific articles in the field of entrepreneurship research this further implies that aligning and thus focusing on just one audience in terms of unique jargon is likely to increase impact. This effect is rooted in articles becoming legitimate when talking in one community language uniquely. This is in line with research into legitimacy (Garud et al. 2014; Gurses and Ozcan 2015; Navis and Glynn 2011) and further adds to this literature in the form that epistemic cultures, scientific communities, also underlie the need for legitimacy due to rhetorical strategies as well. For researchers the recommendation is clear: focus uniquely on one community language to increase impact. Thus, authors seeking higher levels of uniqueness in terms of the alignment to one community should increase the use of highly definitional term for a sub-research field's topic. For example, an author wishing to uniquely align to the sub-research field of *entrepreneurship in family firms* (Cluster 10 in Table 8) should use terms associated with the clusters topic as "(non)family", "owner(ship)", "culture" or "altruism".<sup>9</sup> Again, uniqueness is achieved when an article is more aligned to one community/cluster compared to others. Also, in order to link an article to the cluster of entrepreneurship and family firms it may be advisable to avoid terms highly associated with other sub-research fields such as "university", "spinoff", "transfer", "technology", "academic" or "licensing" (associated with the sub-research field *university-industry relations and*

<sup>9</sup> These terms represent the highest tf-idf weighted words for a sub-research field.



**Fig. 3** Interaction effect of uniqueness and novelty on article's impact

*entrepreneurship*—Cluster 2 in Table 8). Again, the unique affiliation to just one sub-research field should be aimed for.

We also show that recombining different scientific communities' jargon in a novel way, affects articles' impact in form of an inverted U-shape (H2 and H3). This finding helps to overcome mixed results from previous research on novelty and its effects (Boudreau et al. 2016; Lee et al. 2015; Trapido 2015; Uzzi et al. 2013). On the one hand, for low novelty levels, an increase in an article's novelty (from low to medium) has a positive effect on impact (Lee et al. 2015; Newman and Cooper 1993; Schoenmakers and Duysters 2010). On the other hand, for high levels of novelty, an increase of novelty has a negative effect on impact (from medium to high) (Boudreau et al. 2016; Mueller et al. 2012). Hence, the level of novelty defines the degree of the effect. Work which is 'too' novel might be too distant to the audience's knowledge and thus does not lead to high impact. A possible explanation for this relation is the view that science is required to advance gradually in order to re-test hypothesis before they can be accepted. Authors should be aware of that, when writing their articles in order to secure impact. As a result, research that does not follow established principles, and thus is assumed to be very novel, may not get a high level of impact (Uzzi et al. 2013; Antons and Piller 2015). It would be of further interest if this finding holds for research in more general management communities (e.g. Academy of Management Journal, Administrative Science Quarterly, etc.) or other disciplines, in particular those which are not characterized by a large degree of heterogeneity. It may also be interesting to compare disciplines or sub-communities with different degrees of maturity to test whether new communities differ in their acceptance of novelty relative to well established clusters.

Additionally, our article presents results taking two different moderating effects of an article's uniqueness on the relationship between novelty and impact into account. First, we show that higher levels of an article's uniqueness steepen the inverted U-shaped relationship between novelty and impact (H4a). Thus, an article's higher level of uniqueness increases the maximum effect of novelty on impact. Second, our results further show that a higher level of uniqueness shifts the turning point of the inverted U-shaped relationship between novelty and

impact to lower values of novelty (H4b). Article's that are more unique have a lower optimal level of novelty to maximize impact. Taking these effects together (graphically displayed in Fig. 3) shows that on the one hand higher levels of uniqueness can increase the overall effect of novelty. Higher uniqueness levels allow novelty to have higher effects on impact, but only to a certain inflection point. It is this combined effect, the steepening of the inverted U-shaped relationship and the turning point shift to lower values of novelty that shows that the advantages of uniqueness, at some point transform to severe downturns. After reaching the inflection point, higher levels of uniqueness yield lower impact for novelty. Articles with high levels of uniqueness are punished for higher novelty compared to articles with less uniqueness. Being aware of this tradeoff is important to understand how much and thus how uniquely to align to a sub-research field before being punished for novelty. This finding further advances research into novelty (Boudreau et al. 2016; Uzzi et al. 2013), validating the important moderating role of uniqueness. It is especially above trade-off which has to be taken into consideration when analyzing the effects of novelty.

Thus, scientists should reflect on their level of uniqueness and adjust it to their article's novelty level. An article's higher level of uniqueness moves the optimum (the turning point) of novelty to lower levels. Thus, authors aligning uniquely to a certain community should use less novel recombinations of community specific jargon to maximize impact. It further seems that aligning to a certain community hinders higher levels of novelty to increase impact again slowing the advancement of novel scientific findings. Authors should therefore be aware when uniquely aligning to one community in their jargon, leads to less novel levels of language recombination to maximize impact.

Our results further enable insights for the evolution of scientific disciplines, here entrepreneurship science. Several authors argued that entrepreneurship is a mature research field by showing its convergence of different sub-research fields (Busenitz et al. 2014; Grégoire et al. 2006). However, it seems that equally combining language form different communities of the field of entrepreneurship (low values of uniqueness) is punished with less impact. Contrary uniqueness increases impact. Furthermore, the novel recombination of jargon can be understood as a measure for a scientific field's convergence (Busenitz et al. 2014; Grégoire et al. 2006). Convergence implies that when a research field matures, it is more and more characterized by a set of codified theories, models, and measures. Again, we want to point out that even referring to jargon, this language-culture expresses different theories, norms, measures and constructs (Kuhn 1996). Thus, more novel recombination of jargon/theory from different communities, combines and moves these communities' theories and norms closer together enabling convergence. Our results therefore show how citation behavior might affect convergence. In particular, we show that high levels of novelty and thus convergence of scientific communities in entrepreneurship research are punished with less impact resulting in a slower pace of convergence for the scientific domain of entrepreneurship. Furthermore, does the trade-off, authors face (the interplay between uniqueness and novelty) lead to higher uniqueness and less novel papers.

However, being aware of these effects might already affect the way reviewers and readers of articles might be rejected or drawn to certain articles. The danger lies in

the rejection of articles not uniquely focusing on the readers prioritized jargon—likely to affect reviewers and potential citing readers.

We further think that our findings are not only relevant for the research domain of entrepreneurship. Several scientific domains consist of very heterogeneous communities (e.g. strategic management) which are likely to underlie the here proposed rationales as well. It would therefore be of interest to analyze the effects shown here in different scientific disciplines not only related to management research but further natural sciences, sociology, etc.

Distinct potential limitations need to be highlighted: first, we use the widely accepted approach of document co-citation analysis to quantitatively define scientific communities. This approach is based on citation measurement and thus tends to focus on older articles, contrary to recent publications not having that many citations, yet. Furthermore, would it be of interest to see if our results hold in other research fields or even in the entirety of science. Another limitation is the restriction to abstracts of articles for the identification of linguistic characteristics. Even though being in line with previous research (Kaplan and Vakili 2015) using the full texts might give deeper insights.

Building up on Kuhn's (1996) argument, that different language-culture communities are like different theories, the used methodological approaches allow to measure the relatedness of theories and norms. Especially the combination of bibliometric approaches, to detect communities, and natural language processing, to match articles to these communities, enables several more research directories in different scientific domains. These potentials therefore do not only exist in the domain of scientific publications but further when analyzing knowledge spillovers for instance between organizations. Previous texts of organizations (publications, patents, external statements) allow to detect an organization's language or jargon. This further could potentially affect the ability to absorb foreign knowledge. Not only in the dimension of distant technologies (Jung and Lee 2016) but the dimensions of the similarity between the language is likely to have an effect on the ability to adopt foreign knowledge. This could therefore contribute to the scientific construct of absorptive capacity (Cohen and Levinthal 1990) analyzing not only the amount but further the content and the associated content distance of this knowledge.

Individuals might also be an interesting unit of analysis: for instance, detecting the language used by individuals in twitter tweets compared to e.g. mission statements of the organization they work in. Furthermore, comparison between knowledge of different contexts is possible using the methods used in this article: for example, the similarity between scientific knowledge of individuals (e.g. scientific paper publications) compared to hiring companies applicable knowledge base in form of patents. To conclude, several new research directories emerge via the methodological approaches presented here.

Future research could also dive more deeply into category spanning research when taking this study's findings into account. Research into category spanning has identified that spanning categories leads to less market acceptance (Hsu et al. 2009; Kovács and Johnson 2014). This could be an alternative explanation of our findings: regarding scientific communities in the field of entrepreneurship as different categories an article is aligning to (uniqueness) or combining (novelty) to gain market acceptance and thus

impact. Especially the interaction of these two effects could be of interest for research into category spanning, namely answering the research question if categories should be spanned when novel ideas or inventions are in consideration.

Another promising avenue for future research could be potential learning effects of authors. When being rejected several times, one is likely to learn that aligning to a certain community increases an article's legitimacy and impact and thus is likely to also affect arguments coming up in the review process. Thus, the question arises if authors are aware of these effects. One could assume that more experienced authors probably are aware of this and that less experienced authors will learn, e.g. via review processes, to align uniquely to communities in terms of jargon to increase legitimacy and thus impact (or getting accepted in the journal review process).

Finally, drawing attention towards the effects of language in scientific publications enables authors to reflect on their citation behavior, enabling them to identify potentially useful input to their research despite a lack of proximity. Thus, by showing that citing authors are influenced by language cultures, this implicit bias might better be understood and thus counteracted by author's when citing other research.

## 7 Conclusion

Our study aims to advance our understanding of how language complexity in scientific articles affects impact. In the context of entrepreneurship science, which represents an ideal context due to its heterogeneous, cross-disciplinary nature, we conduct a more fine-grained analysis of the effects of different community specific jargons. We disentangle language complexity in two forms: we detect an article's unique alignment to one community's jargon (article's uniqueness) and its novel recombination of community specific jargon (article's novelty). Our research suggests that articles' uniqueness affects their impact in a positive manner. We show that novelty of articles affects their impact in form of an inverted U-shape relationship. This inverted U-shape relationship between novelty and impact is moderated by the article's uniqueness in two ways. First, higher levels of uniqueness steepen the curve of the U-shaped relations between novelty and impact. The second effect shifts the optimal level of novelty for higher values of uniqueness. Combining these moderating relationships shows the trade-off that more uniqueness might increase the positive effect of novelty for lower levels of novelty, but after an inflection point higher novelty levels are punished in form of lower impact, contrary to holding lower levels of uniqueness. These findings not only have implications for authors and reviewer but also increase our understanding on the evolution of scientific fields and their convergence.

## Appendix

See Tables 7 and 8.

**Table 7** Test of a turning point shift

Value of mod- erator	Coef.	Std. Err.	z	P >  z	95% Conf. interval	
<b>Results based on Table 2/model 6: basic model</b>						
0.00	-764.39	111.57	-6.85	0.000	-983.07	-545.71
0.01	-776.63	112.47	-6.91	0.000	-997.06	-556.208
0.02	-779.44	112.49	-6.93	0.000	-999.91	-558.98
0.03	-780.51	112.41	-6.94	0.000	-1000.81	-560.20
0.04	-781.03	112.35	-6.95	0.000	-1001.23	-560.83
0.05	-781.32	112.31	-6.96	0.000	-1001.44	-561.20
0.06	-781.50	112.28	-6.96	0.000	-1001.56	-561.44
0.07	-781.62	112.26	-6.96	0.000	-1001.64	-561.59
0.08	-781.70	112.24	-6.96	0.000	-1001.69	-561.71
0.09	-781.76	112.23	-6.97	0.000	-1001.72	-561.79
0.10	-781.80	112.22	-6.97	0.000	-1001.75	-561.85
<b>Results based on Table 4/model 6: including journal dummies</b>						
0.00	-621.46	210.63	-2.95	0.003	-1034.25	-208.60
0.01	-622.63	212.15	-2.93	0.003	-1038.44	-206.83
0.02	-623.35	212.80	-2.93	0.003	-1040.43	-206.27
0.03	-623.80	213.11	-2.93	0.003	-1041.49	-206.12
0.04	-624.11	213.26	-2.93	0.003	-1042.09	-206.14
0.05	-624.33	213.33	-2.93	0.003	-1042.46	-206.21
0.06	-624.49	213.37	-2.93	0.003	-1042.69	-206.30
0.07	-624.62	213.38	-2.93	0.003	-1042.84	-206.39
0.08	-624.71	213.39	-2.93	0.003	-1042.94	-206.48
0.09	-624.79	213.39	-2.93	0.003	-1043.01	-206.56
0.10	-624.85	213.38	-2.93	0.003	-1043.06	-206.63
<b>Results based on Table 4/model 6: zero inflated negbin</b>						
0.00	-2121.61	814.04	-2.61	0.009	-3717.09	-526.13
0.01	-2122.46	815.56	-2.60	0.009	-3720.93	-523.99
0.02	-2122.94	816.21	-2.60	0.009	-3722.69	-523.19
0.03	-2123.23	816.52	-2.60	0.009	-3723.59	-522.88
0.04	NA	NA	NA	NA	NA	NA
0.05	-2123.57	816.77	-2.60	0.009	-3724.42	-522.78
0.06	-2123.67	816.82	-2.60	0.009	-3724.61	-522.78
0.07	-2123.74	816.85	-2.60	0.009	-3724.74	-522.74
0.08	-2123.80	816.87	-2.60	0.009	-3724.83	-522.76
0.09	-2123.84	816.88	-2.60	0.009	-3724.89	-522.79
0.10	-2123.88	816.88	-2.60	0.009	-3724.94	-522.81

**Table 8** Overview of clusters

Cluster title (total cites/# of articles/average cites per article) (year of first publication–year of latest publication)	Description
1. International entrepreneurship (6368/49/129,96) (1989–2008)	International entrepreneurship analyses the process of startups transforming into internationally active companies, it is regarded as a combination of the research fields of entrepreneurship and international business (McDougall and Oviatt 2000). Relevant aspects to the growth of such companies have been identified (Autio et al. 2000) as well as factors relating to the speed (Knight and Cavusgil 2004) and likelihood (Oviatt and McDougall 1994) of internationalization
2. University-industry relations and entrepreneurship (4240/51/83,14) (1987–2009)	This large cluster investigates the relation between university and industry. For example, the role of technology transfer offices in creating startups (Siegel et al. 2003) as well as the influence of national policies (Goldfarb and Henrekson 2003) or individual characteristics of entrepreneurs with scientific background (Murray 2004)
3. Venture capital policies and financing (3053/31/98,48) (1989–2006)	The match-making between venture capitalists and entrepreneurs is an important factor in entrepreneurship and hence subject to intense study. Contracts may need to take into account varying motivations by investors and entrepreneurs (Aghion and Bolton 1992) as well as moral hazard implicit to the process (Bergemann and Hege 1998) while entrepreneurs have to optimize the relation between additional funds and shares sold (Hsu 2004)
4. Macroeconomic/global and regional impact of entrepreneurship (2612/34/76,82) (1987–2008)	The effect of entrepreneurship on economic growth is subject of research in this cluster (Reynolds et al. 2005). Related topics addressed in this cluster are differences between countries which may lead to advantages for some entrepreneurs compared to those from areas less conducive to entrepreneurship (Busenitz et al. 2000) or shifts from managed to entrepreneurial economies in developed countries (Audretsch and Thurik 2000)
5. Entrepreneurship and liquidity (2202/8/275,25) (1989–2002)	Financial assets are an important precursor to entrepreneurial activity, whether they are the result of inheritance or investments (Blanchflower and Oswald 1998). A related aspect is the return on investment from entrepreneurship (Hamilton 2000)
6. Institutional entrepreneurship (2039/22/92,68) (1980–2009)	Institutional entrepreneurship is the focus of this cluster. Maguire et al. (2004) analyze institutional entrepreneurship in emerging fields, Lounsbury and Crumley (2007) develop a process model of new practice creation Beckert (1999) introduces the effect of strategic choice in this context. Furthermore, Fligstein (1997) introduce social skill in this context

**Table 8** (continued)

Cluster title (total cites/# of articles/average cites per article) (year of first publication–year of latest publication)	Description
7. Corporate entrepreneurship (1375/4/343,75) (1983–1995)	Corporate Entrepreneurship is regarded as means to improve a company's long term financial performance (Zahra and Covin 1995). Antecedents and effects are studied within this cluster
8. Social entrepreneurship (1196/15/79,73) (2000–2009)	Social entrepreneurship is compared to commercial entrepreneurship (Austin et al. 2006) to determine important differences such as special performance indicators (e.g. social needs) that are not covered by commercial entrepreneurship (Mair and Marti 2006)
9. Entrepreneurship and social network analysis (1102/6/183,67) (2001–2003)	In this cluster methods of social network analysis are applied to firm and entrepreneurial networks. E.g. theories on cohesive networks and networks with structural holes are related to firm success (Hite and Hesterly 2001) or the evolution of personal networks through different phases of entrepreneurship is described (Greve and Salaff 2003)
10. Entrepreneurship in family firms (967/12/80,58) (2003–2007)	Differences between family and non-family firms are the subject of research for this cluster (Zahra et al. 2004). Family structures influence entrepreneurship (Aldrich and Cliff 2003) and family firms exhibit special characteristics (Zahra 2003)
11. Entrepreneurial education and self-efficacy (877/8/109,63) (1997–2006)	Self-efficacy appears to play an important role in entrepreneurship. It influences venture growth (Baum and Locke 2004) and affects entrepreneurial learning (Zhao et al. 2005)
12. Entrepreneurial intention (794/3/264,67) (1988–2000)	This cluster distinguishes between different entrepreneurial intentions (Krueger et al. 2000). Further, entrepreneurs and managers are distinguished by their self-efficacy (Chen et al. 1998)
13. Immigration and entrepreneurship (708/9/78,67) (1985–1996)	This cluster, belonging to the field of sociology, analyses self-employment among immigrants. Relevant factors are, e.g., family structures (Sanders and Nee 1996) or performance of small business entrepreneurs (Portes and Zhou 1996). It appears to be related to the 'Entrepreneurship and family firms' cluster
14. Entrepreneurial opportunity detection and learning (666/10/66,6) (2005–2007)	This cluster explores how entrepreneurs detect business opportunities. E.g. pattern recognition, i.e. the ability to apply past experience to detect business opportunities before others, appears to be an important determinant (Baron and Ensley 2006)
15. Transition economies (518/5/103,6) (2001–2003)	This cluster concentrates on transition countries, especially Russia. Johnson et al. (2002) analyze the effect of property right on new firms reinvestment of profits. Peng (2001) studies how entrepreneurs create wealth in transition economies

**Table 8** (continued)

Cluster title (total cites/# of articles/average cites per article) (year of first publication–year of latest publication)	Description
16. Personal initiative (416/3/138,67) (1996–2000)	At the center of this cluster are studies that compare personal initiative in the former East and West Germany. The degree of control and complexity of work affects initiative, which can be regarded as similar to the concept of entrepreneurship (Frese et al. 1996)
17. Origin of entrepreneurs (412/5/82,4) (2003–2006)	An important antecedent for entrepreneurial behavior is the social context of the entrepreneur. Being embedded in a start-up friendly environment (Gompers et al. 2005) or in high performing academic environments with corporate links (Kenney and Goe 2004) tends to increase the likelihood of start-ups being created
18. Cultural entrepreneurship in the US (365/3/121,67) (1982–1991)	This cluster differs from the other clusters in that it encompasses papers on media and culture rather than economics. Papers in this cluster describe the role of individuals in cultural entrepreneurship, i.e. the creation of museums or operas (Dimaggio 1982)
19. Emerging economies (314/3/104,67) (2002–2008)	These articles analyze entrepreneurship in emerging economies (Bruton et al. 2008). Meyer and Peng (2005) concentrate on the context in Central and Eastern Europe. Specifically, three lines of theorizing have been advanced: organizational economics theories, resource-based theories and institutional theories
20. Narratives and presentation (311/4/77,75) (2007–2009)	This cluster analyses the relation between presenting a business model and investments. Signaling certain capabilities to potential investors is an important aspect observed in a study by Zott and Huy (2007). Short et al. (2009) points out that empirical evidence in this area is scarce which may negatively impact the application of theoretical concepts to managerial practice
21. Business incubators (307/5/61,4) (2002–2005)	Co-production of business assistance in business incubators (Rice 2002) and the effectiveness of business incubators (Colombo and Delmastro 2002)
22. Management buyouts (259/4/64,75) (1992–2001)	Papers in this cluster explore the effect of leveraged buyouts on companies and describes various types of LBOs (Wright et al. 2001). It is suggested the LBOs are beneficial to a company's performance and corporate entrepreneurship and does not impact negatively on a company's RandD efforts (Zahra 1995)

**Table 8** (continued)

Cluster title (total cites/# of articles/average cites per article) (year of first publication–year of latest publication)	Description
23. Entrepreneurship in the public sector (253/3/84,33) (1992–2000)	Noticeable is that all three articles are published in <i>Public Administration Review</i> . Articles in this cluster analyze that public entrepreneurs of the neo-managerialist persuasion pose a threat to democratic governance (Terry 1998) as well as implications of the reinvention movement for democratic governance and its glorification of entrepreneurial management (deLeon and Denhardt 2000)
24. Evans-Jovanovic entrepreneurial choice model (233/3/77,67) (1998–2003)	This cluster focusses on the Evans and Jovanovic (1989) entrepreneurial choice model, which states that the decision to become an entrepreneur can be modelled as optimization problem given equations for the income of wage workers and entrepreneurs
25. Habitual entrepreneurs (232/3/77,33) (1997–2003)	This cluster analyzes effects of entrepreneurs which were involved in more than one venture (Wright et al. 1997; Westhead and Wright 1998)
26. Entrepreneurship as a social construct (216/3/72) (2004–2006)	Steyaert and Katz (2004) consider Entrepreneurship as a societal rather than an economic phenomenon. Fletcher (2006) studies social constructionist thinking particularly with regard to opportunity formation processes
27. Cultural support for entrepreneurship (205/3/68,33) (2000–2007)	Related to the large cluster of international entrepreneurship this cluster focusses on cultural factors relevant to entrepreneurship. Individual factors shown to be important for entrepreneurial success vary by culture (Thomas and Mueller 2000). Social concepts such as shame related to failure or social status also explains entrepreneurial action (Begley and Tan 2001)
28. Venture capitalist investment decisions (198/3/66) (1992–1998)	The process underlying an investment decision is governed by variables attributable to the investor, such as his preference for national investments and variables describing the investment opportunity, such as the quality of the business idea. Papers in this cluster explore this relationship with methods to go beyond survey data, conjoint analyses (Muzyka et al. 1996) or policy capturing (Zacharakis and Meyer 1998)
29. Franchise I (185/3/61,67) (1988–1996)	This cluster concentrates on entrepreneurial franchise and starting reasons. Michael (1996) analyses decision rights and organizational form shares. Kaufmann and Dant (1996) show that capital acquisition is a relevant reason for engaging in franchising and not the assumption that franchisees manage the outlets better than company employees would if the unit were company owned

**Table 8** (continued)

Cluster title (total cites/# of articles/average cites per article) (year of first publication–year of latest publication)	Description
30. Women and entrepreneurship (143/4/35,75) (1991–2003)	Most cited articles in this cluster study differences between women and men entrepreneurs. DeMartino and Barbato (2003) explore family flexibility and wealth creation as career motivators and Caputo and Dolinsky (1998) analyze the role of financial and human capital of household members on women's choice to pursue self-employment
31. Gender and entrepreneurship (143/3/47,67) (2001–2005)	This cluster concentrates on gender diversities in Entrepreneurship. Not only differences and divisions between women business owners who are silent about gender issues and those who are not are explored (Lewis 2006) but also formal and informal sources of business funding to illustrate how this concept impacts upon women in self-employment (Marlow and Patton 2005)
32. Franchise II (118/4/29,5) (1996–1999)	This whole journal is published in the Journal of Business Venturing. It analyzes business-format franchising growth's in the U.S. (Lafontaine and Shaw 1998) as well as survival patterns among franchisee and nonfranchise small firms (Bates 1998)
33. Entrepreneurs vs. managers (112/3/37,33) (1987–1990)	This cluster analyzes differences between Entrepreneurs and Managers in small business firms (Begley and Boyd 1987) and in their motivational patterns (Miner 1990)
34. Alliances and Joint Ventures (111/3/37) (1994–1999)	Market valuation of joint ventures in terms of Joint venture characteristics and wealth gains (Park and Kim 1997) as well as opportunistic action within research alliances (Deeds and Hill 1999) are studies in this cluster
35. Entrepreneurs in organizations (100/3/33,33) (1986–1997)	Articles located in this cluster analyze implications for organizations' structures and their Human-Resource Management Practices to foster and facilitate entrepreneurship (Schuler 1986) and effects of managers' entrepreneurial behavior on subordinates (Pearce et al. 1997)

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aghion P, Bolton P (1992) An incomplete contracts approach to financial contracting. *Rev Econ Stud* 59(3):473–494
- Aldrich HE, Cliff JE (2003) The pervasive effects of family on entrepreneurship: toward a family embeddedness perspective. *J Bus Ventur* 18(5):573–596. [https://doi.org/10.1016/S0883-9026\(03\)00011-9](https://doi.org/10.1016/S0883-9026(03)00011-9)
- Aldrich HE, Fowler SW, Liou N, Marsh SJ (1994) Other people’s concepts: why and how we sustain historical continuity in our field. *Organ* 1:65–80
- Antons D, Piller R (2015) Opening the black box of “Not Invented Here”: attitudes, decision biases, and behavioral consequences. *Acad Manag Perspect* 29(2):193–217
- Antons D, Joshi AM, Salge TO (2019) Content, contribution, and knowledge consumption: uncovering hidden topic structure and rhetorical signals in scientific texts. *J Manag* 45(7):3035–3076
- Audretsch DB, Thurik AR (2000) Capitalism and democracy in the 21st century: from the managed to the entrepreneurial economy. *J Evol Econ* 10(1):17–34. <https://doi.org/10.1007/s001910050003>
- Austin J, Stevenson H, Wei-Skillern J (2006) Social and commercial entrepreneurship: same, different, or both? *Entrep Theory Pract* 30(1):1–22. <https://doi.org/10.1111/j.1540-6520.2006.00107.x>
- Autio E, Sapienza HJ, Almeida JG (2000) Effects of age at entry, knowledge intensity, and imitability on international growth. *Acad Manag J* 43(5):909–924. <https://doi.org/10.2307/1556419>
- Baron RA, Ensley MD (2006) Opportunity recognition as the detection of meaningful patterns: evidence from comparisons of novice and experienced entrepreneurs. *Manag Sci* 52(9):1331–1344. <https://doi.org/10.1287/mnsc.1060.0538>
- Bates T (1998) Survival patterns among newcomers to franchising. *J Bus Ventur* 13(2):113–130. [https://doi.org/10.1016/S0883-9026\(97\)00066-9](https://doi.org/10.1016/S0883-9026(97)00066-9)
- Baum JR, Locke EA (2004) The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *J Appl Psychol* 89(4):587–598. <https://doi.org/10.1037/0021-9010.89.4.587>
- Beckert J (1999) Agency, entrepreneurs, and institutional change. The role of strategic choice and institutionalized practices in organizations. *Organ Stud* 20(5):777–799. <https://doi.org/10.1177/0170840699205004>
- Begley TM, Boyd DP (1987) A comparison of entrepreneurs and managers of small business firms. *J Manag* 13(1):99–108. <https://doi.org/10.1177/014920638701300108>
- Begley TM, Tan WL (2001) The socio-cultural environment for entrepreneurship: a comparison between east Asian and Anglo-Saxon countries. *J Int Bus Stud* 32(3):537–553. <https://doi.org/10.1057/palgr ave.jibs.8490983>
- Bergemann D, Hege U (1998) Venture capital financing, moral hazard, and learning. *J Bank Finance* 22(3):703–735. [https://doi.org/10.1016/S0378-4266\(98\)00017-X](https://doi.org/10.1016/S0378-4266(98)00017-X)
- Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. O’Reilly Medi Inc., Sebastopol
- Blanchflower DG, Oswald AJ (1998) What makes an entrepreneur? *J Labor Econ* 16(1):26–60. <https://doi.org/10.1086/209881>
- Boudreau KJ, Guinan EC, Lakhani KR, Riedl C (2016) Looking across and looking beyond the knowledge frontier: intellectual distance, novelty, and resource allocation in science. *Manag Sci* 62:2765–2783
- Bruton GD, Ahlstrom D, Obloj K (2008) Entrepreneurship in emerging economies: where are we today and where should the research go in the future. *Entrep Theory Pract* 32(1):1–14. <https://doi.org/10.1111/j.1540-6520.2007.00213.x>
- Busenitz LW, Gomez C, Spencer JW (2000) Country institutional profiles: unlocking entrepreneurial phenomena. *Acad Manag J* 43(5):994–1003. <https://doi.org/10.2307/1556423>
- Busenitz LW, Plummer LA, Klotz AC, Shahzad A, Rhoads K (2014) Entrepreneurship research (1985–2009) and the emergence of opportunities. *Entrep Theory Pract* 38:981–1000
- Caputo RK, Dolinsky A (1998) Women’s choice to pursue self-employment: the role of financial and human capital of household members. *J Small Bus Manag* 36(3):8–17. <http://crawl.prod.proquest.com.s3.amazonaws.com/fpcache/6e8d087f7cf115612d32902f6ce5a1ac.pdf?AWSAc>

- cessKeyId=AKIAJF7V7KKNV2KKY2NUQ&Expires=1475069146&Signature=IkoxD05d2qDWT%2FCg4KkgOv3TJgY%3D
- Cawkell AE (1976) Understanding science by analysing its literature. *Essays Inf Sci* 2:543–549
- Chen CC, Greene PG, Crick A (1998) Does entrepreneurial self-efficacy distinguish entrepreneurs from managers? *J Bus Ventur* 13(4):295–316. [https://doi.org/10.1016/S0883-9026\(97\)00029-3](https://doi.org/10.1016/S0883-9026(97)00029-3)
- Cohen WM, Levinthal DA (1990) Absorptive capacity: a new perspective on learning and innovation. *Admin Sci Q* 35:128–152
- Cole S, Cole JR (1967) Scientific output and recognition: a study in the operation of the reward system in science. *Am Sociol Rev* 32:377–390
- Colombo MG, Delmastro M (2002) How effective are technology incubators? Evidence from Italy. *Res Policy* 31(7):1103–1122. [https://doi.org/10.1016/S0048-7333\(01\)00178-0](https://doi.org/10.1016/S0048-7333(01)00178-0)
- Cornelius B, Landström H, Persson O (2006) Entrepreneurship studies: the dynamic research front of a developing social science. *Entrep Theory Pract* 30(3):375–398
- De Bellis N (2009) *Bibliometrics and citation analysis: from the science citation index to cybermetrics*. Scarecrow Press, Lanham
- Deeds DL, Hill CWL (1999) An examination of opportunistic action within research alliances: evidence from the biotechnology industry. *J Bus Ventur* 14(2):141–163. [https://doi.org/10.1016/S0883-9026\(97\)00069-4](https://doi.org/10.1016/S0883-9026(97)00069-4)
- deLeon L, Denhardt RB (2000) The political theory of reinvention. *Public Admin Rev* 60(2):89–97. <https://doi.org/10.1111/0033-3352.00068>
- DeMartino R, Barbato R (2003) Differences between women and men MBA entrepreneurs: exploring family flexibility and wealth creation as career motivators. *J Bus Ventur* 18(6):815–832. [https://doi.org/10.1016/S0883-9026\(03\)00003-X](https://doi.org/10.1016/S0883-9026(03)00003-X)
- Dimaggio P (1982) Cultural entrepreneurship in 19th-century Boston, Part II: the classification and framing of American art. *Media Cult Soc* 4(4):303–322. <https://doi.org/10.1177/016344378200400402>
- Evans DS, Jovanovic B (1989) An estimated model of entrepreneurial choice under liquidity constraints. *J Polit Econ* 97(4):808–827. <https://doi.org/10.1086/261629>
- Fletcher DE (2006) Entrepreneurial processes and the social construction of opportunity. *Entrep Reg Dev* 18(5):421–440. <https://doi.org/10.1080/08985620600861105>
- Fligstein N (1997) Social skill and institutional theory. *Am Behav Sci* 40(4):397–405. <https://doi.org/10.1177/0002764297040004003>
- Foster JG, Rzhetsky A, Evans JA (2015) Tradition and innovation in scientists' research strategies. *Am Sociol Rev* 80:875–908
- Frese M, Kring W, Soose A, Zempel J (1996) Personal initiative at work: differences between east and west Germany. *Acad Manag J* 39(1):37–63. <https://doi.org/10.2307/256630>
- Garfield E (1973) Citation and distinction. *Nature* 242:485–485
- Garfield E, Malin MV, Small H (1978) Citation data as science indicators. In: Elkana Y, Lederberg J, Merton RK, Thackray A, Zuckerman H (eds) *Toward a metric of science: the advent of science indicators*. Wiley, New York, pp 179–208
- Gartner WB (1990) What are we talking about when we talk about entrepreneurship? *J Bus Ventur* 5(1):15–28
- Garud R, Gehman J, Giuliani AP (2014) Contextualizing entrepreneurial innovation: a narrative perspective. *Res Policy* 43(7):1177–1188
- Gephart RP (1988) *Ethnostatistics: qualitative foundations for quantitative research*. Sage, Newbury Park
- Gmür M (2003) Co-citation analysis and the search for invisible colleges: a methodological evaluation. *Science* 57:27–57
- Goldfarb B, Henrekson M (2003) Bottom-up versus top-down policies towards the commercialization of university intellectual property. *Res Policy* 32(4):639–658. [https://doi.org/10.1016/S0048-7333\(02\)00034-3](https://doi.org/10.1016/S0048-7333(02)00034-3)
- Gompers P, Lerner J, Scharfstein D (2005) Entrepreneurial spawning: public corporations and the genesis of new ventures, 1986 to 1999. *J Finance* 60(2):577–614. <https://doi.org/10.1111/j.1540-6261.2005.00740.x>
- Grégoire DS, Noël MX, Déry R, Bécharde J-P (2006) Is there a conceptual convergence in entrepreneurship research? A co-citation analysis of 'frontiers of entrepreneurship' research, 1981–2004. *Entrep Theory Pract* 30:333–373
- Greve A, Salaff JW (2003) Social networks and entrepreneurship. *Entrep Theory Pract* 28(1):1–22. <https://doi.org/10.1111/1540-8520.00029>
- Gross AG (1990) *The rhetoric of science*. Harvard University Press, Cambridge

- Gurses K, Ozcan P (2015) Entrepreneurship in regulated markets: framing contests and collective action to introduce pay TV in the US. *Acad Manag J* 58(6):1709–1739
- Haans RFJ, Pieters C, He Z-L (2016) Thinking about u: theorizing and testing U- and inverted U-shaped relationships in strategy research. *Strateg Manag J* 37:1177–1195
- Hall AR (1980) *Philosophers at war*. Cambridge University Press, Cambridge
- Hamilton BH (2000) Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *J Polit Econ* 108(3):604–631. <https://doi.org/10.1086/262131>
- Hausman J, Hall BH, Griliches Z (1984) Econometric models for count data with an application to the patents–R&D relationship. *Econometrica* 52:909–938
- Hite JM, Hesterly WS (2001) The evolution of firm networks: from emergence to early growth of the firm. *Strateg Manag J* 22(3):275–286. <https://doi.org/10.1002/smj.156>
- Honig B, Lampel J, Siegel D, Drnevich P (2014) Ethics in the production and dissemination of management research: institutional failure or individual fallibility? *J Manag Stud* 51(1):118–142
- Hsu DH (2004) What do entrepreneurs pay for venture capital affiliation? *J Finance* 59(4):1805–1844. <https://doi.org/10.1111/j.1540-6261.2004.00680.x>
- Hsu G, Kocak Ö, Hannan MT (2009) Multiple category memberships in markets: an integrative theory and two empirical tests. *Am Sociol Rev* 74(1):150–169
- Ireland RD, Reutzler CR, Webb JW (2005) Entrepreneurship research in AMJ: what has been published, and what might the future hold? *Acad Manag J* 48(4):556–564
- Jaccard P (1901) Distribution de la flore alpine dans le bassin des drouces et dans quelques regions voisines. *Bull Soc Vaud Sci Nat* 37:241–272
- Jaro MA (1989) Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *J Am Stat Assoc* 84:414–420
- Jaro MA (1995) Probabilistic linkage of large public health data files. *Stat Med* 14:491–498
- Johnson S, McMillan J, Woodruff C (2002) Property rights and finance. *Am Econ Rev* 92(5):1335–1356. <https://doi.org/10.1257/000282802762024539>
- Judge TA, Cable DM, Colbert AE, Rynes SL (2007) What causes a management article to be cited—article, author, or journal? *Acad Manag J* 50:491–506
- Jung HJ, Lee JJ (2016) The quest for originality: A new typology of knowledge search and breakthrough inventions. *Acad Manag J* 59:1725–1753
- Kaplan S, Vakili K (2015) The double-edged sword of recombination in breakthrough innovation. *Strateg Manag J* 36:1435–1457
- Katz R, Allen TJ (1982) Investigating the Not Invented Here (NIH) syndrome: a look at the performance, tenure, and communication patterns of 50 R&D project groups. *R&D Manag* 12(1):7–20
- Kaufmann PJ, Dant RP (1996) Multi-unit franchising: growth and management issues. *J Bus Ventur* 11(5):343–358. [https://doi.org/10.1016/0883-9026\(96\)00057-2](https://doi.org/10.1016/0883-9026(96)00057-2)
- Kenney M, Goe WR (2004) The role of social embeddedness in professorial entrepreneurship: a comparison of electrical engineering and computer science at uc Berkeley and Stanford. *Res Policy* 33(5):691–707. <https://doi.org/10.1016/j.respol.2003.11.001>
- Knight GA, Cavusgil ST (2004) Innovation, organizational capabilities, and the born-global firm. *J Int Bus Stud* 35(2):124–141. <https://doi.org/10.1057/palgrave.jibs.8400071>
- Knorr-Cetina K (1999) *Epistemic cultures: how the sciences make knowledge*. Harvard University Press, Cambridge
- Kovács B, Johnson R (2014) Contrasting alternative explanations for the consequences of category spanning: a study of restaurant reviews and menus in San Francisco. *Strateg Organ* 12:7–37
- Kramsch C (1998) *Language and culture*. Oxford University Press, Oxford
- Krueger NF, Reilly MD, Carsrud AL (2000) Competing models of entrepreneurial intentions. *J Bus Ventur* 15(5–6):411–432. [https://doi.org/10.1016/S0883-9026\(98\)00033-0](https://doi.org/10.1016/S0883-9026(98)00033-0)
- Kuhn TS (1996) *The structure of scientific revolutions*, 3rd edn. The University of Chicago Press, Chicago
- Lafontaine F, Shaw KL (1998) Franchising growth and franchisor entry and exit in the US market: myth and reality. *J Bus Ventur* 13(2):95–112. [https://doi.org/10.1016/S0883-9026\(97\)00065-7](https://doi.org/10.1016/S0883-9026(97)00065-7)
- Lampe HW, Hilgers D (2015) Trajectories of efficiency measurement: a bibliometric analysis of DEA and SFA. *Eur J Oper Res* 240:1–21
- Lee YN, Walsh JP, Wang J (2015) Creativity in scientific teams: unpacking novelty and impact. *Res Policy* 44:684–697
- Lewis P (2006) The quest for invisibility: female entrepreneurs and the masculine norm of entrepreneurship. *Gender Work Organ* 13(5):453–469. <https://doi.org/10.1111/j.1468-0432.2006.00317.x>

- Lind JT, Mehlum H (2010) With or without U? The appropriate test for a U-shaped relationship. *Oxf Bull Econ Stat* 72:109–118
- Locke K, Golden-Biddle K (1997) Constructing opportunities for contribution: structuring intertextual coherence and “problematizing” in organizational studies. *Acad Manag J* 40(5):1023–1062
- Long JS (1997) Regression models for categorical and limited dependent variables: advanced quantitative techniques in the social sciences. Sage Publications, Thousand Oaks
- Lounsbury M, Crumley ET (2007) New practice creation: an institutional perspective on innovation. *Organ Stud* 28(7):993–1012. <https://doi.org/10.1177/0170840607078111>
- Maguire S, Hardy C, Lawrence TB (2004) Institutional entrepreneurship in emerging fields: HIV/AIDS treatment advocacy in Canada. *Acad Manag J* 47(5):657–679. <https://doi.org/10.2307/20159610>
- Mair J, Marti I (2006) Social entrepreneurship research: a source of explanation, prediction, and delight. *J World Bus* 41(1):36–44. <https://doi.org/10.1016/j.jwb.2005.09.002>
- Marlow S, Patton D (2005) All credit to men? Entrepreneurship, finance, and gender. *Entrep Theory Pract* 29(6):717–735. <https://doi.org/10.1111/j.1540-6520.2005.00105.x>
- Martin BR, Irvine J (1983) Assessing basic research. *Res Policy* 12:61–90
- McCloskey D (1994) Knowledge and persuasion in economics. Cambridge University Press, Cambridge
- McDougall PP, Oviatt BM (2000) International entrepreneurship: the intersection of two research paths. *Acad Manag J* 43(5):902–906. <https://doi.org/10.2307/1556418>
- McKnight B, Zietsma C (2018) Finding the threshold: a configurational approach to optimal distinctiveness. *J Bus Ventur* 33(4):493–512
- Merton RK (1968) The Matthew effect in science: the reward and communication systems of science are considered. *Science* 159:56–63
- Merton RK (1973) The sociology of science: theoretical and empirical investigations. University of Chicago Press, Chicago
- Meyer KE, Peng MW (2005) Probing theoretically into Central and Eastern Europe: transactions, resources, and institutions. *J Int Bus Stud* 36(6):600–621. <https://doi.org/10.1057/palgrave.jibs.8400167>
- Michael SC (1996) To franchise or not to franchise: an analysis of decision rights and organizational form shares. *J Bus Ventur* 11(1):57–71. [https://doi.org/10.1016/0883-9026\(95\)00088-7](https://doi.org/10.1016/0883-9026(95)00088-7)
- Miller D (1986) Configurations of strategy and structure: towards a synthesis. *Strateg Manag J* 7:233–249
- Miner JB (1990) Entrepreneurs, high growth entrepreneurs, and managers: contrasting and overlapping motivational patterns. *J Bus Ventur* 5(4):221–234. [https://doi.org/10.1016/0883-9026\(90\)90018-O](https://doi.org/10.1016/0883-9026(90)90018-O)
- Mingers J, Xu F (2010) The drivers of citations in management science journals. *Eur J Oper Res* 205:422–430
- Moed HF (2005) Statistical relationships between downloads and citations at the level of individual documents within a single journal. *J Am Soc Inf Sci Technol* 56:1088–1097
- Mueller JS, Melwani S, Goncalo JA (2012) The bias against creativity. *Psychol Sci* 23:13–17
- Murray F (2004) The role of academic inventors in entrepreneurial firms: sharing the laboratory life. *Res Policy* 33(4):643–659. <https://doi.org/10.1016/j.respol.2004.01.013>
- Muzyka D, Birley S, Leleux B (1996) Trade-offs in the investment decisions of European venture capitalists. *J Bus Ventur* 11(4):273–287. [https://doi.org/10.1016/0883-9026\(95\)00126-3](https://doi.org/10.1016/0883-9026(95)00126-3)
- Navis C, Glynn MA (2011) Legitimate distinctiveness and the entrepreneurial identity: influence on investor judgements of new venture plausibility. *Acad Manag Rev* 36(3):479–499
- Nerkar A, Lahiri N (2014) The duality of knowledge networks: the impact of production and usage networks on academic citations. In: Aharonson BS, Stettner U, Amburgey TL, Ellis S, Drori I (eds) Understanding the relationship between networks and technology, creativity and innovation (Technology, innovation, entrepreneurship and competitive strategy, Vol. 13). Emerald Group Publishing Limited, Bradford, pp 165–197
- Newman JM, Cooper E (1993) Determinants of academic recognition: the case of the journal of applied psychology. *J Appl Psychol* 78:518–526
- Oviatt BM, McDougall PP (1994) Toward a theory of international new ventures. *J Int Bus Stud* 25(1):45–64. <https://doi.org/10.1057/palgrave.jibs.8490193>
- Park SH, Kim D (1997) Market valuation of joint ventures: joint venture characteristics and wealth gains. *J Bus Ventur* 12(2):83–108. [https://doi.org/10.1016/S0883-9026\(96\)00036-5](https://doi.org/10.1016/S0883-9026(96)00036-5)
- Pearce JA, Kramer TR, Robbins DK (1997) Effects of managers’ entrepreneurial behavior on subordinates. *J Bus Ventur* 12(2):147–160. [https://doi.org/10.1016/S0883-9026\(96\)00066-3](https://doi.org/10.1016/S0883-9026(96)00066-3)
- Pehlan SE, Ferreira M, Salvador R (2002) The first twenty years of the Strategic Management Journal. *Strateg Manag J* 23:1161–1168

- Peng MW (2001) How entrepreneurs create wealth in transition economies. *Acad Manag Exec* 15(1):95–108. <http://www.jstor.org/stable/4165713>
- Pennebaker JW, Booth RJ, Boyd RL, Francis ME (2015) Linguistic inquiry and word count: LIWC2015. Pennebaker Conglomerates, Austin
- Podsakoff PM, MacKenzie SB, Bachrach DG, Podsakoff NP (2005) The influence of management journals in the 1980s and 1990s. *Strateg Manag J* 26:473–488
- Podsakoff PM, MacKenzie SB, Podsakoff NP, Bachrach DG (2008) Scholarly influence in the field of management: a bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century. *J Manag* 34:641–720
- Portes A, Zhou M (1996) Self-employment and the earnings of immigrants. *Am Sociol Rev* 61(2):219–230. [https://www.sscnet.ucla.edu/soc/faculty/zhou/pubs/Portes\\_Zhou\\_SelfEmployment.pdf](https://www.sscnet.ucla.edu/soc/faculty/zhou/pubs/Portes_Zhou_SelfEmployment.pdf)
- Rafols I, Leydesdorff L, O'Hare A, Nightingale P, Stirling A (2012) How journal rankings can suppress interdisciplinary research: a comparison between innovation studies and business and management. *Res Policy* 41:1262–1282
- Reader D, Watkins D (2006) The social and collaborative nature of entrepreneurship scholarship: a co-citation and perceptual analysis. *Entrep Theory Pract* 30(3):417–441
- Rehurek R, Sojka P (2010) Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks, pp 45–50
- Reynolds P, Bosma N, Autio E, Hunt S, De Bono N, Servais I, Lopez-Garcia P, Chin N (2005) Global entrepreneurship monitor: data collection design and implementation 1998–2003. *Small Bus Econ* 24(3):205–231. <https://doi.org/10.1007/s11187-005-1980-1>
- Rice MP (2002) Co-production of business assistance in business incubators—an exploratory study. *J Bus Ventur* 17(2):163–187
- Robertson S (2004) Understanding inverse document frequency: on theoretical arguments for idf. *J Doc* 60:503–520
- Sanders JM, Nee V (1996) Immigrant self-employment: the family as social capital and the value of human capital. *Am Sociol Rev* 61(2):231–249. <http://www.jstor.org/stable/2096333>
- Sasabuchi S (1980) A test of a multivariate normal mean with composite hypotheses determined by linear inequalities. *Biometrika* 67:429–439
- Schildt HA, Zahra SA, Sillanpaa A (2006) Scholarly communities in entrepreneurship research: a co-citation analysis. *Entrep Theory Pract* 30:399–415
- Schoenmakers W, Duysters G (2010) The technological origins of radical inventions. *Res Policy* 39(8):1051–1059
- Schuler RS (1986) Fostering and facilitating entrepreneurship in organizations—implications for organization structure and human-resource management-practices. *Hum Resour Manag* 25(4):607–629. <https://doi.org/10.1002/hrm.3930250408>
- Schumpeter J (1939) *Business cycles: a theoretical historical and statistical analysis of the capitalist process*. McGraw-Hill, New York
- Shane S, Venkataraman S (2000) The promise of entrepreneurship as a field of research. *Acad Manag Rev* 25(1):217–226
- Short JC, Moss TW, Lumpkin G (2009) Research in social entrepreneurship: past contributions and future opportunities. *Strateg Entrep J* 3(2):161–194. <https://doi.org/10.1002/sej.69>
- Siegel DS, Waldman D, Link A (2003) Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: an exploratory study. *Res Policy* 32(1):27–48. [https://doi.org/10.1016/S0048-7333\(01\)00196-2](https://doi.org/10.1016/S0048-7333(01)00196-2)
- Simons HW (1990) *The rhetorical turn: invention and persuasion in the conduct of inquiry*. University of Chicago Press, Chicago
- Simonton DK (2004) *Creativity in science: chance, logic, genius, and zeitgeist*. Cambridge University Press, Cambridge
- Small H (1973) Co-citation in the scientific literature: a new measure of the relationship between two documents. *J Am Soc Inf Sci* 24:265–269
- Small H, Greenlee E (1980) Citation context analysis of a co-citation cluster: recombinant-DNA. *Scientometrics* 2:277–301
- Steyaert C, Katz J (2004) Reclaiming the space of entrepreneurship in society: geographical, discursive and social, dimensions. *Entrep Reg Dev* 16(3):179–196. <https://doi.org/10.1080/0898562042000197135>
- Stremersch S, Verniers I, Verhoef PC (2007) The quest for citations: drivers of article impact. *J Manag* 71:171–193

- Tauscher K, Bouncken RB, Pech R (2020) Gaining legitimacy by being different: Optimal distinctiveness in crowdfunding platforms. *Acad Manag J* (**forthcoming**)
- Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 29:24–54
- Terry LD (1998) Administrative leadership, neo-managerialism, and the public management movement. *Public Admin Rev* 58(3):194–200. <https://doi.org/10.2307/976559>
- Thomas AS, Mueller SL (2000) A case for comparative entrepreneurship: assessing the relevance of culture. *J Int Bus Stud* 31(2):287–301. <https://doi.org/10.1057/palgrave.jibs.8490906>
- Trapido D (2015) How novelty in knowledge earns recognition: the role of consistent identities. *Res Policy* 44:1488–1500
- Uzzi B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. *Science* 342:468–472
- Valentini G (2012) Measuring the effect of M&A on patenting quantity and quality. *Strateg Manag J* 33:336–346
- Van Noorden R (2015) Interdisciplinary research by the numbers. *Nature* 525:306–307
- Vilhena DA, Foster JG, Rosvall M, West JD, Evans J, Bergstrom CT (2014) Finding cultural holes: how structure and culture diverge in networks of scholarly communication. *Sociol Sci* 1:221–238
- Vuong Q (1989) Likelihood ratio test for model selection and non-nested hypotheses. *Econometrica* 57:307–333
- Wang J (2014) Unpacking the Matthew effect in citations. *J Info* 8:329–339
- Westhead P, Wright M (1998) Novice, portfolio, and serial founders: are they different? *J Bus Ventur* 13(3):173–204. [https://doi.org/10.1016/S0883-9026\(97\)90002-1](https://doi.org/10.1016/S0883-9026(97)90002-1)
- Winkler WE (1999) The state of record linkage and current research problems. In: *Stat Soc Can, Proc Surv Methods Sect*, pp 73–90
- Wright M, Robbie K, Ennew C (1997) Venture capitalists and serial entrepreneurs. *J Bus Ventur* 12(3):227–249. [https://doi.org/10.1016/S0883-9026\(96\)06115-0](https://doi.org/10.1016/S0883-9026(96)06115-0)
- Wright M, Hoskisson RE, Busenitz LW (2001) Firm rebirth: buyouts as facilitators of strategic growth and entrepreneurship. *Acad Manag Perspect* 15(1):111–125. <https://doi.org/10.5465/AME.2001.4251486>
- Zacharakis AL, Meyer GD (1998) A lack of insight: do venture capitalists really understand their own decision process? *J Bus Ventur* 13(1):57–76. [https://doi.org/10.1016/S0883-9026\(97\)00004-9](https://doi.org/10.1016/S0883-9026(97)00004-9)
- Zahra SA (1995) Corporate entrepreneurship and financial performance—the case of management leveraged buyouts. *J Bus Ventur* 10(3):225–247. [https://doi.org/10.1016/0883-9026\(94\)00024-0](https://doi.org/10.1016/0883-9026(94)00024-0)
- Zahra SA (2003) International expansion of us manufacturing family businesses: the effect of ownership and involvement. *J Bus Ventur* 18(4):495–512. [https://doi.org/10.1016/S0883-9026\(03\)00057-0](https://doi.org/10.1016/S0883-9026(03)00057-0)
- Zahra SA, Covin JG (1995) Contextual influences on the corporate entrepreneurship—a longitudinal analysis. *J Bus Ventur* 10(1):43–58. [https://doi.org/10.1016/0883-9026\(94\)00004-E](https://doi.org/10.1016/0883-9026(94)00004-E)
- Zahra SA, Hayton JC, Salvato C (2004) Entrepreneurship in family vs. non-family firms: a resource-based analysis of the effect of organizational culture. *Entrep Theory Pract* 28(4):363–381. <https://doi.org/10.1111/j.1540-6520.2004.00051.x>
- Zhao H, Seibert SE, Hills GE (2005) The mediating role of self-efficacy in the development of entrepreneurial intentions. *J Appl Psychol* 90(6):1265–1272. <https://doi.org/10.1037/0021-9010.90.6.1265>
- Zott C, Huy QN (2007) How entrepreneurs use symbolic management to acquire resources. *Adm Sci Q* 52(1):70–105

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.