Comparing Indicators of Scholarly Impact: A Journal Focused Analysis

Yi Chen, Ameya Karnad, Hui Soo Chae, & Gary Natriello

Teachers College Columbia University
Abstract

**Purpose:** Measuring the scholarly impact of publications in a corpus is an important and challenging problem in research evaluation. The scholarly impact has typically been equated with the citation count. But the information that can be obtained from the citation count is inherently limited. Therefore, new methodologies try to expand the scope of scholarly impact measurement through incorporating more research behaviors beyond citation count (e.g., readership analysis in Mendeley) or pay more attention to the context of publication (e.g., text mining). However, only a few studies have investigated the application of these methods and compared them with real data.

**Research Design:** In this paper, we provide an empirical analysis of three scholarly impact measures to assess 2570 feature articles from a representative journal in the field of education: *Teachers College Record*. The three measurements are citation count (drawn from Google Scholar and Scopus), altmetrics (using the Mendeley read count), and a text-mining indicator (Dynamic Influence Modeling).

**Conclusion:** We identify and explain the similarities and unique qualities among these measures for the same corpus. Furthermore, we also discuss the potential and the challenges of applying these methodologies in the field of Informatics more generally.

**Keywords:** Scholarly Impact, Altmetrics, Citation Analysis, Dynamic Influence Modeling, Informatics
Comparing Indicators of Scholarly Impact: A Journal Focused Analysis

INTRODUCTION

Identifying influential articles in a corpus is an essential and challenging task in informatics. Conventionally, citation count is most widely-used for measuring the impact of a publication (Eugene, 1998; Moed, 2005; Oppenheim, 1995), since citation is a recognition or validation of one’s research by other researchers (Wang, Song, & Barabási, 2013). Though citation count is powerful and succinct, the information that can be obtained from citation count is inherently limited (Waltman, 2016). With the expanding coverage of digital bibliographic data and continuous improvements in new methodologies, informaticians today have opportunities to measure the impact of a publication more comprehensively. During the past few decades, many new citation-based indicators and alternative measurement technologies have been introduced in the literature. However, the utility of these new measures and their comparative advantages and disadvantages remain to be assessed.

In this paper, we provide a comparative analysis of three different approaches to assessing scholarly impact by examining the data from one representative journal in the field of educational research: Teachers College Record. We compare the results from citation count (drawn from Google Scholar and Scopus), altmetrics (using the Mendeley read count), and text mining algorithms (Dynamic Influence Modeling). This case study provides empirical evidence of the benefits and drawbacks of data-intensive methodologies. The study aims to address two research questions: (1) what are the differences and similarities among these primary scholarly impact measures? (2) what are the possibilities and barriers to using these scholarly impact measurements in practice?
LITERATURE REVIEW

We begin by discussing bibliographic databases and the citation impact factors derived from them. We then proceed to consider new methodologies for developing measures of scholarly impact. This sets the stage for consideration of the three approaches we examine in this study.

Bibliographic Databases

Bibliographic databases contain descriptive records of published literature (e.g., journal articles, conference proceedings, reports, and books). They are organized collections of reference information, typically including author, title, and publisher information. These records generally contain rich subject descriptions provided in the subject classification terms, keywords, or abstracts (Feather & Sturges, 2003). Many digital bibliographic databases also include entries and metadata beyond the standard forms, including full text, numeric compilations (e.g., survey data files), images, video and audio files, or funding sources. These data provide more opportunities to apply advanced techniques (e.g., social network analysis, text mining, and deep learning) and to expand the tools available to the field of informatics.

Many informetric records are available in digital form, and sufficient computing power is readily available for users to deal with large-scale collections of scholarly resources. Three of the most popular and representative multidisciplinary digital bibliographic databases are Thomson Reuters’ Web of Science, Elsevier’s Scopus, and Google Scholar. Other accessible digital databases or academic search engines for scholarly document retrieval include Academic Search (EBSCO), CiNii (National Institute of Informatics), Mendeley, AMiner, DataBase systems and Logic Programming (DBLP), CiteSeer, and Microsoft Academic. Many other bibliographic
databases cover more specialized publications in specific scientific fields. For example, the Educational Resource Information Center (ERIC) provides access to over 1.5 million records dating back to 1966 in the field of education. Similarly, Lesson Planet contains over 400,000 teacher-reviewed classroom resources in K-12 education. A significant number of bibliographic databases are proprietary, available either by licensing agreements from vendors or directly from the indexing and abstracting services that create them (Reitz, 2004). For example, EBSCO, Web of Science, and Scopus are available through subscription. Other bibliographic databases are free and open, including Google Scholar, Mendeley, ERIC, and DBLP. The various databases are foundational for most efforts to assess scholarly impact.

Citation Impact Factors

Scholarly impact is one of the strongest currencies in academia. It has traditionally been equated with citation count—be it for individuals, articles, departments, universities, journals, or entire research fields (Aguinis, Suarez-Gonzalez, Lannelongue, & Joo, 2012). Moreover, citation counts have been used as the basis for many other scholarly impact metrics such as the $h$-index (Hirsch, 2005), $i$-10 index, and many other evaluation metrics for journals, conferences, researchers, and research institutes (Moed et al., 2012; Wildgaard, Schneider, & Larsen, 2014).

The modern era of examining the impact of academic journal content began over forty years ago with the founding of the Institute for Scientific Information (ISI) by Garfield. Over that time, the ISI has held a near-monopoly in citation analysis. Under ISI, the Thomson Scientific Journal Impact Factor (JIF; Garfield, 1999) has maintained a dominant position among all the evaluation metrics for journals. However, even ISI’s databases have long been
criticized by scholars for the limited number of journal titles (less than 4% of the world’s total journals) and enormous bias toward English publication (Kurmis, 2003).

The growing flood of scholarly literature has exposed the weakness of current citation-based methods for evaluating and filtering articles (Priem & Hemminger, 2010). Whether citation-based indicators are reliable measures for scholarly impact is an open question (Opthof, 1997; Seglen, 1997; Harter & Nisonger, 1997). Many informaticians reject the use of the citation count and other journal-level citation-based indicators for evaluating individual publications or academic institutions.

One reason for such rejection is that the frame of reference usually cannot easily be transferred across different fields, nations, and generations. For example, compared with natural sciences research, social sciences and humanities research has stronger national and regional orientations, relatively more publications and impact in book form as opposed to journal articles, a slower pace of theoretical development, less collaboration among scholars, and relatively more connections to the non-scholarly public (Waltman, 2016).

Additionally, it is a sobering fact that 90 percent of papers that have been published in academic journals are never cited. As many as 50 percent of publications are never read by anyone other than their authors, referees and journal editors (Meho, 2007). Citation count is a closed currency only used by scholars. Compared with the whole scholarly corpus, the citation network is sparse and limited. On the other hand, "article overload" besets evaluators and the informetric measures they employ during the decision-making process regarding grant funding and priority in assigning resources (Boyack & Jordan, 2011).

Finally, citation-based indicators are criticized for being slow (Brody, Harnad, & Carr, 2006), narrow (Priem & Hemminger, 2010), secretive as well as irreproducible (Rossner, Van
Epps, & Hill, 2008), open to gaming (Falagas & Alexiou, 2008), and biased (Meho, 2007). To be more specific, researchers are concerned with the issues of self-citation (i.e., scholars deliberately citing the article, journal or the academic institutions they are affiliated with), ceremonial-citation (i.e., scholars citing an authority in the field without ever having consulted the relevant work itself), cronyism (i.e., friends or colleagues reciprocally citing each other), and undifferentiated measurement (i.e., references or the authors from the same publication are measured equally).

As a result, many other citation-based metrics have been developed to address the issues mentioned above. For example, metrics use citation counts to generate various ratio-based indicators (e.g., Impact per Publication, CiteScore, Source-Normalized Impact per Paper, and Impact Factor), portfolio-based indicators (e.g., h-index), and network-based indicators (e.g., Relative Citation Ratio, SCImago Journal Rank, and Eigenfactor). These new indicators provide more standardized and comparable solutions and complement or compete with the conventional citation-based indicators such as the Journal Impact Factor (JIF).

**New Methodologies**

Informaticians have tried to improve the evaluation of scholarly publication impact in three ways: expanding the dimensions of the data source (so that more features of an article and more complex data structures can be used), extending the scope of the data source (so that more research fields and research units will be considered), and incorporating more advanced algorithms or metrics (so that more potential factors that influence the article’s impact can be included).
In terms of expanding the dimension and scope of the data source, informaticians have tried to incorporate more research behaviors beyond citation. Altmetrics (i.e., alternative metrics) pay attention to the broader population by including all possible audiences. Altmetrics usually calculate scholarly impact based on diverse online research output, such as social media (e.g., Twitter), online news media (e.g., technology blogs), and online reference managers. For example, researchers use Mendeley bookmarks (Mohammadi, Thelwall, & Kousha, 2014) and readership indicators (Haunschild & Bornmann, 2015; Mohammadi & Thelwall, 2014) to reflect the impact of a publication. Altmetrics use public APIs across platforms to gather data with open-source scripts and algorithms. Given that expressions of scholarship are becoming more diverse, these metrics seem suitable to determine the impact of research in a broader manner than with citation counts (Aguinis, Shapiro, Antonacopoulou, & Cummings, 2014; Dinsmore, Allen, & Dolby, 2014; Priem, Taraborelli, Groth, & Neylon, 2010).

In terms of improving the evaluation metrics, many scholars are starting to apply more advanced data science techniques into informatics. Google’s method of ranking web pages has inspired many measures that apply social network analysis in citation networks (Lazega, Wasserman, & Faust, 2006). For example, Aguinis et al. (2012) used the number of pages as indexed by Google to assess the scholarly impact on stakeholders outside and inside academia. Link prediction in the citation network can also be used to predict the impact and success of a publication (Bütün et al., 2017; Daud et al., 2017; Pobiedina & Ichise, 2016). Meanwhile, Article-Level Metrics (ALMs; Sparc & Tananbaum, 2013) seeks to incorporate data from social media mentions along with citations to present a more vibrant picture of how an individual article is being discussed, shared, and used. The development of algorithms and tools for
analyzing network/graphic data has also led to the popularity of Bibliographic Coupling Analysis (BCA), which provides other possibilities for tracing impact.

Though social network methodologies address some of the deficiencies in citation-analysis, the text content of articles is still ignored with these approaches. With the rapid development of data mining techniques and the massive improvement in computation and storage in the past two decades, scholars are now able to explore large numbers of publications in detail. Natural language processing (NLP) techniques (e.g., probabilistic topic modeling) have been used to evaluate research quality and productivity in informetrics (Li & Lei, 2019; Chen, Ding, Xu, & K, 2018; Li & Lei, 2019). Gerrish and Blei (2010) analyzed scholarly impact directly using the whole text context with text mining methods. There is also a growing interest in research applying data-intensive sentiment analysis using deep learning (Keramatfar & Amirkhani, 2018; Kalantari, Kamsin, & Kamaruddin, 2017). With the more advanced model design, other features of publication could also be incorporated into the algorithms.

In summary, informaticians always face a dilemma in maintaining the balance between simplicity and generalizability. What kind of publication should be selected? What document type (article, review, and/or editorial material) should be included? What language or languages should be included? How long should the citation window extend (two years, three years, or longer)? Which counting method (e.g., full counting or fractional counting) should be used? There is no generally applicable rule for these selection issues. And, it seems to be inevitable that selection always brings bias. For example, a short-term citation window will fail to capture “sleeping beauties” (i.e., those publications that only become influential over the long term). For a comprehensive analysis of scholarly impact, we cannot rely on any single measure, since there always exist some fundamental limitations. However, few research studies provide empirical
evidence regarding the similarities and differences among these measures in practice. Therefore, a comparative analysis is essential and necessary.

METHODS

For our comparative analysis of the different approaches, we examine the citation count approach, the altmetric approach, and the text mining approach by implementing them according to the standards and protocols established for each. In all cases, we used the complete records of all feature articles (2570 in total) published in the Teachers College Record from 1994 to 2018.

Citation Count Approach

In this study, citation counts were collected from Google Scholar and Scopus. Google Scholar is the world’s largest academic search engine and indexes most scholarly literature that is available online on the web. However, little is known about the coverage of Google Scholar since Google does not publish the size of their database. Michael (2018) estimated that the total number of publications indexed by Google Scholar, without any language restriction, is about 389 million. In contrast, Scopus is a subscription-based database. A detailed discussion of the data quality of Scopus is provided by Franceschini, Maisano, and Mastrogiacomo (2016).

Altmetrics Approach

For the altmetrics approach in this study, we selected one of the most widely-used altmetrics: Mendeley reader statistics. Mendeley is one of the most widely used reference management tools and social networks for scholars (Rodgers & Barbrow, 2013; Haunschild & Bornmann, 2016). Mohammadi, Thelwall, and Kousha (2016) show that “82% of Mendeley
users had read or intended to read at least half of the bookmarked publications in their personal libraries.” The Mendeley database has the advantage of counting unique users of scholarly materials through individual accounts. However, the statistics may be lost if a user removes an article from their library or closes their account.

**Text Mining Approach**

Text mining incorporates the text of the publication in the analysis and consequently is more data-intensive than other measurements. In this study, we take probabilistic topic modeling (in particular Dynamic Influence Modeling) as an example. A detailed explanation of probabilistic topic modeling is beyond the scope of this article. D. M. Blei (2011) provides an introduction to topic modeling from a statistical perspective.

Dynamic Influence Modeling (DIM) is a probabilistic topic modeling approach to measure the scholarly impact of documents over a year. It is based on Dynamic Topic Modeling (DTM, C. Wang, Blei, & Heckerman, 2008). It allows multiple threads of influence within a corpus. The influence of each article is encoded as a hidden variable, and posterior inference reveals the influential publications in the collection. The evolution of hidden variables is encoded in a time-series model of sequential document collections. The word distribution within each topic will change over time. The change in word distribution consequently represents the evolution of the topic. However, the length of the time period is a priori, which can be specified based on prior knowledge or systematic techniques, including cross-validation or non-parametric approaches. In this study, we selected a time period of one year since the typical journal volume covers a year, and we also incorporated the feature articles available in 2019 to measure the scholarly impact of 2018 articles in the following year.
DATA

*Teachers College Record* (TCR) is a journal of research, analysis, and commentary in the field of education, which has been continuously published since 1900. For TCR, the journal impact factor (JIF) is 0.91 (2018), Citescore is 1.34 (2018), source normalized impact per paper (SNIP) is 0.915 (2018), Scimago Journal Ranking (SJR) is 0.995 (2018), H-index (2018) is 78, and self-citation ratio is 6.81% (2018). *Figure 1* shows the number of articles published/included and the Journal Impact Factor for TCR from 1994 through 2018. It is in the first quintile in the field of education according to the JIF measure. TCR includes four types of publications: feature articles, commentaries, book reviews, and research notes. TCR feature articles include 17 categories (e.g., educational administration, adult education, assessment and evaluation, educational counseling, curriculum, teaching, and policy) and more than 80 sub-categories. In our analysis, we considered feature articles published in-print from 1994 to 2018 (2570 articles in total).

*Figure 1.* The number of articles and the JIF for TCR, 1994 through 2018
The citations count for both Google Scholar and Scopus was collected using Publish or Publish Software (Harzing, 2007), a tool designed for citation and research impact analysis. Through the Mendeley API, we can search published TCR articles by their title, author, and year of publication. Since not all feature articles in TCR have a DOI (Digital Object Identifier – a unique linking identifier for publication), we rely on the above-mentioned article features for data retrieval. Moreover, we need to use the fuzzy matching technique on title and authors to determine whether the search results we get from the Mendeley API indeed match with the article in TCR. The benchmark similarity ratio we selected for fuzzy matching is 85%, which means the article title and author that the Mendeley API provides should at least have 85% similarity with the real title and author information from TCR based on the Levenshtein distance. We collected the text context and publication year for each feature article. Then, we conducted text data processing operations like word segmentation, punctuation removal, deletion of numbers, word lowercase transformation, updating the stop words, stop word removal, and stemming on the article text. After building the corpus, we transformed the data into a document-term matrix and removed the sparse terms.

RESULTS

Citation Count Approach

There are apparent differences between the citation counts from Scopus and Google Scholar. In terms of coverage, the citation indicator for TCR on Scopus is available from 1996 onwards. Google Scholar covers 84.55% (2173) of the 2570 feature articles, and Scopus covers 63.85% (1641). Thus, neither source provides complete coverage of the journal content with nearly 15 percent omitted by Google Scholar and well over a third not covered by Scopus.
Meanwhile, *Google Scholar* generally collects more citations for the same article than *Scopus* (for those articles identified by both *Google Scholar* and *Scopus*, 95.62% of the articles have the same number or more citations in *Google Scholar* than in *Scopus*.

The most popular article in TCR (Mishra & Koehler, 2006) shows the highest number of citations in both *Google Scholar* and *Scopus*. Its popularity is partly because of its high searchability. When we searched this article online, we found at least three different web pages that provide free access to this article outside of TCR. However, this article is not openly accessible on TCR.

Although *Google Scholar* covers a broader base of citation sources than *Scopus*, the accuracy of the citation counts is questionable. For example, the second most cited article from *Scopus* (Feiman-Nemser, 2001), is not identified by *Google Scholar* as a TCR article. Instead, this article is recognized as an article from the Brandeis Institutional Repository, which provides open-source access. Given that articles can be available from many different sources (e.g., digital achieve and pre-print centers), these inaccuracies are likely to become even more prevalent moving forward.

The boxplots on the left side of Figure 2 show the distributions of citation counts from *Google Scholar* and *Scopus*. Both bibliographic databases present highly left-skewed patterns. The 10% most cited articles, according to *Google Scholar*, account for 62.31% of overall citations, while the 10% most cited articles, according to *Scopus* account for 66.14% of overall citations. This distribution pattern in citation counts is consistent with the findings in previous studies (e.g., Albarrán et al., 2011; Seglen, 1992). For example, Testa (2012) found that as few as 300 journals account for more than 50% of the total number of citations in *Scopus*. Consequently, the citation-based indicators for an academic institution or a journal
(publisher) are also strongly influenced by a limited number of popular publications. Additionally, this pattern will tend to be reinforced over time, as a few highly cited articles attract much more attention from the audience until the article naturally becomes less popular as the field progresses. Finally, the median citation count reported for TCR articles by *Google Scholar* was 13 (mean is 63) with a standard deviation of 224 and skewness of 27. For *Scopus*, the median citation count was 9 (mean is 24), with a standard deviation of 77 and skewness of 21. This indicates that citation count from *Scopus* is more stable but with more limited coverage.

![Graph showing citation comparison between Google Scholar, Scopus, and Mendeley](image)

*Figure 2. Citation Comparison of Google Scholar, Scopus, and Mendeley.*

Notes: 1. There are 2173 articles identified in Google Scholar, 1641 in Scopus, and 1576 in Mendeley; 2. Mishra & Koehler (2006) from Google Scholar (8663 citation count), Scopus (2486 citation count) and Mendeley (15932 read count), and Nemser (2001) from Scopus (985 citation count) are ignored since their citation counts are outliers, compared with other articles.

*Figure 3* shows the citation counts from *Google Scholar* and *Scopus* across different publication years. The general tendencies of data from *Google Scholar* and *Scopus* are aligned. The top plot of *Figure 3* shows the average citation count for all the articles published in the
same year. We can broadly divide the whole pattern into three stages. Before 2000, the annual number of articles fluctuated around 60, and the average citation count for each article is around 40. During the period from 2000 to 2008, this number climbs from approximately 100 towards over 200, while the average citation count for each article is stable around 90. The overall increase of average citation count can partially be explained by the launch of the TCR website, which allowed more a broader audience to read the articles digitally. After 2008, the number of articles is stable at 120, while the average citation count decreases gradually. The phenomenon can partially be explained by the sleeping beauty effect. Another possible explanation is the increasing competition from other journals in the same field (in particular the journals from large publishing companies) who obtain more favorable positions on the digital search engines. The outliers in the patterns can also be explained by the instability of the number of publications over time. For example, 2008 witnessed a noticeable decrease in the average citation count while the number of articles captured in this year is much higher than all the other years.

The bottom plot of Figure 3 shows the average annual citation count increase since release. To control somewhat for the effect of time since publication, we divide the average citation count by the number of years since the articles were published. Compared with the top plot of Figure 3, the general trend is similar except for the recently published articles (i.e., articles from 2010 to 2018). For example, the articles published in 2015, on average, will obtain 6.11 citations per year, which is higher than for articles published before 2000. There is a slight increase in citation count for the articles published in 2015 and 2018. However, there is an obvious decrease in citation count in 2016 for both plots. The field of education has a pace of theoretical development that is slower than some other fields. Thus, the citation count, in general,
is fairly stable over the years. As time goes by, we would expect articles appearing more recently to garner more citations.

*Figure 3. Citation Count across Publication Years.*

Notes: Plot is based on the 2173 publications identified in Google Scholar and 1641 in Scopus.

**Altmetrics Approach**

Out of 2570 articles, we identified 1576 articles (61.32%) in *Mendeley* using the API. 250 (15.86%) of these identified articles contained a DOI. Similar to the pattern in the citation count, the *Mendeley* reader count is also highly left-skewed (see bottom panel of *Figure 2*). The article with the highest read count (15932) is Mishra & Koehler (2006); this is 3 time more than the second-highest read count. The median read count is 31.5 (mean is 107.21), with a standard deviation of 497.25. The skewness is 22.09, which is similar to *Scopus* and smaller than *Google Scholar*. In the *Mendeley* data, only 53.61% of the articles have larger read counts than citations according to *Google Scholar*; 88% have larger read counts than citations, according to *Scopus*. This is consistent with the known audience coverage limitations of *Mendeley* data.
The top panel of Figure 4 shows the average read count across different publication years. Similar to the patterns in citation count, there are approximately three stages: before 2000, from 2000 to 2008, and after 2008. However, the average read count is more unstable, and the articles published earlier accumulate higher counts. Again, the peak of 2006 is also partially because of the high traffic generated by Mishra & Koehler (2006).

The bottom plot of Figure 4 is the average read count increase after publication. Compared with citation count, this statistic is stable before 2014. However, recent articles (i.e., article from 2014 to 2018) do show a significant increase. This may be due to the act of reading articles being more immediately responsive to their publication than the act of citing them in subsequent work. Or, it may be an artifact of the evolution of Mendeley. As Mendeley has become increasingly popular in recent years, the new users are more likely to read the recently published articles and consequently reinforce the time effect. In conclusion, the overall sleeping beauty effect disappears in altmetrics. Instead, much more attention is paid to the new research.

Figure 4. Mendeley Read Count across Years of Publication
Notes: Plot is based on the 1576 identified publications in Mendeley.
Mendeley reader statistics also include information on the academic status, discipline, and country for readers of each article. However, the accuracy and representativeness of these data are questionable. Many Mendeley users do not provide their personal information, and they may not frequently update these statistics in their profile. With these caveats, aggregating across all of the articles, we can construct a Mendeley reader profile for TCR in general. Among all identified readers, 39.5% are Ph.D./doctoral students, 17.7% are master’s students, and 11.4% are professors or researchers. Most readers belong to social science disciplines (49.6%). Other disciplines include Psychology (9.1%), Art and Humanities (6%), Business, Management and Accounting (5.1%), and philosophy (1.9%). Additionally, most readers come from the United States (41.1%), the United Kingdom (9%), and Canada (4.6%).

Text Mining Approach

In applications of probabilistic topic modeling, the number of topics is a modeling choice, which we need to specify a priori. A more complex model has a more significant number of parameters and requires more computational resources. In addition, the added complexity often leads to difficulty in interpreting the results. A common practice for choosing the number of topics is cross-validation (Browne, 2000). Based on the evidence of deviance statistics in cross-validation, we selected the six topics for our analysis. DIM captures the dynamic evolution of the distribution of topics in the TCR article corpus over time. Based on the word distribution in general, we can summarize the primary meaning of each topic. The summary of each topic is subjectively using the evidence from the posterior word distribution within each topic. However, the topics of TCR articles could be harder to distinguish compared to a journal in natural science. Because the theories and concepts in education research are more general and less specialized
than in some more technical fields, the terms scholars use to discuss different research topics tend to be more similar. However, the widespread topic distribution and its evolution can still provide insight into the changing nature of scholarship in the field.

We found that the six major topic areas in the TCR corpus are: children & learning (children, parent, learn, develop, and new), school & teaching (school, student, teacher, education, instruct, curriculum, and classroom), adult education (college, institution, university, adult, research, and program), research & assessment (test, assess, effect, research, and standard), social (social, public, cultural, identity, and racial), others (practice, role, policy, media, information and social). The words in parentheses are the representative words with the highest posterior probability for each topic.

We also found the most representative (most frequently appearing) words for the articles in each category. The six significant topics we defined based on DIM are somewhat consistent with the evidence from the categories defined by the TCR editors and used to categorize articles on the TCR website. For example, the DIM derived topic “social” shares similar representative words as the TCR defined category “Social Context” (social, public, and American), “Diversity” (racial, American, black, and cultural), and “International Education” (social, global, public, nation). Similarly, the DIM derived topic “school & teaching” share similar representative words with the TCR defined categories “Curriculum” (school, teacher, curriculum, and classroom), “Teaching” (teacher, student, classroom, and instruct), and “Teacher Education” (teacher, student, school, develop, and profession). The TCR Categories “Higher Education” and “Adult Education” also have a similar topic word distribution as the DIM topic “adult education”. Since probabilistic topic modeling applies a mixed-membership framework, the same word can belong to multiple topics with different posterior probabilities. This phenomenon can also be identified
in the empirical findings in TCR defined categories. A word like “teacher”, “education”, “learn”, and “student” are common for every topic. Consequently, this presents more challenges for the task of distinguishing different topics and their evolution.

*Figure 5* shows the proportion of each of the six major DIM derived topics in each year of publication. While the topic distribution, in general, is stable over time, the topics “research & assessment” and “adult education” are becoming more prevalent. Generally, topics “social” and “school & teaching” are the most popular. We also compare this evolution of the DIM topic distribution with the TCR editorial category distribution. We did not provide the same visualization here since 17 categories are hard to portray visually. The most significant difference with the TCR expert-defined category distribution is that it does *not* evolve as smoothly as the DIM topic distribution. But, the increasing popularity of the categories “Higher Education”, “Adult Education”, “Assessment & Evaluation”, and “Research Methods” is evident. Moreover, categories like “Teaching”, “Social”, “Policy”, and “Curriculum” are also popular in general over time.

*Figure 5. Topic Distribution by Year*
To illustrate the application of the text-mining approach, we selected the DIM derived “social” topic as an example. In Figure 6 we show the seven words most often associated over the years with our selected topic “social” to indicate the evolution of the topic over time.

Research frameworks, concepts, and methodologies in the field of education, similar to other social science fields, do not develop as rapidly as in the natural sciences. However, we can identify the general tendencies in this topic. For example, we can see the increasing popularity of “evidence” and “interview” around the year 2000.

*Figure 6. Topic Top Words Update and Tendency*
For each topic, we can identify the most influential articles based on the estimated influence factor from the DIM model. The influence factor is a measure of how much one article could affect a specific topic distribution in the following years. In other words, impactful publication have stronger effects in predicting the topic distribution of future articles since the future articles will adopt their ideas and consequently share a similar topic. We also aggregate the influence factor across different topics for each article by calculating the weighted average influence factor. The weights are each topic’s proportion in the article. Consequently, the aggregated influence factor indicates the overall influence level of the article. *Figure 7* shows that TCR articles had the most influence on the social topic in 2013. For 2018, the aggregated influence factor is close to zero since these papers are published recently.

*Figure 6. Aggregated Influence Factor from DIM*

**Comparative Analysis**

In this section, we compare the three scholarly impact indicators together to identify their similarities and differences. First, we examine the overall Kendall rank correlations among the
different scholarly impact measures. Rank correlation measures an ordinal association (ranking similarity) between two variables. We used only the feature articles that can be identified by Google Scholar, Scopus, and Mendeley. Table 1 shows the overall correlation among the different scholarly impact measures. In general, the correlation between Scopus and Google Scholar is relatively high, since they are both citation counts. The Mendeley read count correlated better with Google Scholar than Scopus. The aggregated influence indicator in DIM has a low correlation with all the other indicators.

Table 1: Correlation Analysis of different measurements

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Google Scholar</th>
<th>Scopus</th>
<th>Mendeley</th>
<th>DIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Scholar</td>
<td>1</td>
<td>0.636</td>
<td>0.273</td>
<td>0.117</td>
</tr>
<tr>
<td>Scopus</td>
<td>0.636</td>
<td>1</td>
<td>0.286</td>
<td>0.153</td>
</tr>
<tr>
<td>Mendeley</td>
<td>0.273</td>
<td>0.286</td>
<td>1</td>
<td>0.118</td>
</tr>
<tr>
<td>Topic Modeling</td>
<td>0.117</td>
<td>0.153</td>
<td>0.118</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: 1. the correlation metrics are Kendall rank correlation; 2. The correlation is based on the publications identified by all four measurements.

In Table 2, we summarize the most impactful three articles from varying measures in each publication year. We highlight the top article shared with different metrics in each year. Though the overall rank correlations between topic modeling and other impact indicators are relatively low, there are still many overlaps among the most impactful articles. For some specific publication years (e.g., 1994, 1995, and 2004), topic modeling even has more overlaps with citation counts than the Mendeley read count. Generally, older articles have more shared items. This indicates that there is an increasing convergence among these metrics as their impact becomes more stable over time.
### Table 2: Comparative Analysis of most impactful 3 articles according to different measures

<table>
<thead>
<tr>
<th>Year</th>
<th>Google Scholar</th>
<th>Scopus</th>
<th>Mendeley</th>
<th>Topic Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>B. Tatum (585); J. Ogbu (564); J. Darby (134);</td>
<td>S. Hollingsworth (1572); W. Rohwer Jr. (890); A. Lewy (769);</td>
<td>V. Gadsden (1.84); J. Ogbu (1.52); J. Darby (1.47);</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>J. Oakes (388); J. Ross (299); J. Anyon (218);</td>
<td>G. Orfield (1463); G. Ladson-Billings (907); R. Feinberg (392);</td>
<td>R. Elmore (1.37); J. Ross (1.18); J. Anyon (1.16);</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>R. Connell (977); A. Lieberman (461); W. Parker (283);</td>
<td>R. Connell (266); A. Lieberman (127); W. Parker (86);</td>
<td>NCTAF (1.89); A. Lieberman (1.29); A. Gitlin (1.25);</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>J. Oakes (475); D. Grissmer (372); J. Spillane (304);</td>
<td>J. Oakes (170); D. Grissmer (64); S. Lawrence (55);</td>
<td>L. Smith (2.57); M. Katz (1.83); H. Becker (1.82);</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>L. Cuban (489); J. Xu (276); C. Chinn (269);</td>
<td>L. Cuban (137); J. Xu (120); C. Chinn (106);</td>
<td>G. Salomon (80); L. Cuban (58); K. Sheldon (33);</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>S. Katz (356); B. Smerdon (252); J. Kailin (206);</td>
<td>S. Katz (103); J. Kailin (62); B. Smerdon (47);</td>
<td>A. Steinberg (198); M. Cochran-Smith (76); S. Katz (39);</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>D. Silva (520); E. Moje (491); D. Ballou (350);</td>
<td>D. Cohen (275); E. Moje (166); D. Silva (94);</td>
<td>A. Wells (2427); A. Wells (1178); P. Slezak (979);</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>R. Croninger (1238); K. Alexander (1033); A. Hargreaves (944);</td>
<td>S. Feiman-Nemser (985); P. Grossman (505); R. Croninger (317);</td>
<td>S. Feiman-Nemser (597); P. Grossman (397); A. Hargreaves (203);</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>C. Rodgers (1858); Y. Zhao (1411); B. Achinstein (820);</td>
<td>C. Rodgers (446); Y. Zhao (330); B. Rowan (259);</td>
<td>P. Grossman (3.11); N. Griffin (1.59); K. Alexander (1.56);</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>V. Richardson (1041); J. Little (997); K. Zeichner (601);</td>
<td>J. Little (304); V. Richardson (230); K. Zeichner (165);</td>
<td>N. Saito (4085); V. Richardson (581); E. Gordon (340); K. Zeichner (2.66); G. Farkas (2.07); G. Stanic (1.83);</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Google Scholar</td>
<td>Scopus</td>
<td>Mendeley</td>
<td>Topic Modeling</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
<td>--------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>2004</td>
<td>L. Darling-Hammond (975); D. Mitra (632); J. Diamond (586);</td>
<td>D. Mitra (214); L. Darling-Hammond (204); J. Diamond (195);</td>
<td>M. Futrell (1708); M. Smylie (1701); D. Mayrowetz (685);</td>
<td>D. Ingram (1.82); L. Darling-Hammond (2.04); K. Dorgan (1.53);</td>
</tr>
<tr>
<td>2005</td>
<td>R. Rumberger (873); G. Fenstermacher (704); Y. Zhao (698);</td>
<td>R. Rumberger (297); Y. Zhao (216); G. Fenstermacher (116);</td>
<td>J. McBrien (2571); J. Frechette (2571); P. Brown (1171);</td>
<td>G. Hu (1.83); E. Graue (1.63); G. Nuthall (1.33);</td>
</tr>
<tr>
<td>2006</td>
<td>P. Mishra (8663); K. Dougherty (411); J. Jackson (319);</td>
<td>P. Mishra (2486); D. Berliner (202); K. Dougherty (113);</td>
<td>P. Mishra (15932); B. Endres (938); P. Kelleher (683);</td>
<td>L. Flowers (1.32); D. Berliner (0.97); R. Farmer-Hinton (0.77);</td>
</tr>
<tr>
<td>2007</td>
<td>Y. Goddard (938); F. Hess (590); S. Kardos (368);</td>
<td>Y. Goddard (202); B. Rubin (133); G. Biesta (102);</td>
<td>C. Thorn (2287); P. Moss (2287); N. Burbules (838);</td>
<td>B. Long (1.73); S. Goldrick-Rab (1.37); V. Louie (1.18);</td>
</tr>
<tr>
<td>2008</td>
<td>N. Carr (1390); J. Boaler (693); T. Howard (565);</td>
<td>J. Boaler (198); T. Howard (130); R. Bomer (76);</td>
<td>M. Huckabee (1708); J. Mcnelly (1411); G. Fenstermacher (1049);</td>
<td>S. Zirkel (1); C. North (0.91); P. Iatarola (0.9);</td>
</tr>
<tr>
<td>2009</td>
<td>J. Cohen (1610); P. Grossman (1072); W. Penuel (458);</td>
<td>J. Cohen (464); P. Grossman (390); W. Penuel (163);</td>
<td>K. Ohkura (600); J. Cohen (536); P. Grossman (446);</td>
<td>J. Cohen (1.39); C. Coburn (0.78); D. Garcia (0.75);</td>
</tr>
<tr>
<td>2010</td>
<td>G. Camilli (802); G. Borman (351); L. Perry (343);</td>
<td>G. Camilli (287); L. Perry (123); G. Borman (97);</td>
<td>M. Packer (1819); B. Merino (382); G. Camilli (363);</td>
<td>M. Lynn (1.06); T. Timar (0.98); P. Carter (0.94);</td>
</tr>
<tr>
<td>2011</td>
<td>S. Shapiro (273); S. Järvelä (259); M. Windschitl (174);</td>
<td>A. Hadwin (137); S. Järvelä (103); M. Windschitl (88);</td>
<td>J. Neapolitan (605); A. Hadwin (270); S. Järvelä (224);</td>
<td>K. Brown (1.61); M. Militello (1); N. Nasir (0.99);</td>
</tr>
<tr>
<td>2012</td>
<td>S. Moore Johnson (691); J. Marsh (238); S. Rojstaczer (166);</td>
<td>M. Bleeker (173); J. Marsh (132); A. Daly (64);</td>
<td>E. Britton (1914); P. Jones (1270); P. Jackson (1131);</td>
<td>J. Marsh (1.85); M. Van Houtte (1.81); D. Larkin (1.73);</td>
</tr>
<tr>
<td>2013</td>
<td>B. Means (596); D. Berliner (295); B. Fishman (222);</td>
<td>B. Means (194); A. Claessens (84); D. Berliner (71);</td>
<td>S. Burdick-Shepherd (1009); D. Kennedy (1009); B. Means (666);</td>
<td>S. Green (3.56); J. Mehta (3.41); T. Good (3.41);</td>
</tr>
<tr>
<td>2014</td>
<td>C. Robertson-Kraft (367); D. Berliner (123); S. Turkan (98);</td>
<td>C. Robertson-Kraft (90); W. Au (44); S. Turkan (43);</td>
<td>J. Thelin (598); C. Robertson-Kraft (351); J. Bempechat (202);</td>
<td>T. McCormack (2.46); M. Cochran-Smith (2.33); E. McGee (2.3);</td>
</tr>
</tbody>
</table>
### Measurement of Scholarly Impact

<table>
<thead>
<tr>
<th>Year</th>
<th>Google Scholar</th>
<th>Scopus</th>
<th>Mendeley</th>
<th>Topic Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>N. Simon (403); K. Zeichner (117); A. Datnow (113);</td>
<td>N. Simon (70); K. Zeichner (44); K. Schildkamp (42);</td>
<td>V. Jones (2373); N. Simon (137); J. Marsh (120);</td>
<td>J. Marsh (1.67); N. Gerzon (1.63); L. Perez-Felkner (1.59);</td>
</tr>
<tr>
<td></td>
<td>JP. Winne (32); J. Marsh (30); N. Shah (30);</td>
<td>J. Thompson (24); H. Nienhusser (14); J. Nelson (12);</td>
<td>J. Murphy (4567); S. Nichols (2431); J. Joyce (2042);</td>
<td>K. Zeichner (1.56); J. Neumann (1.54); S. Athamnases (1.38);</td>
</tr>
<tr>
<td>2016</td>
<td>N. Shah (30); J. Marsh (30); T. Howard (29);</td>
<td>A. Barton (31); P. Winne (15);</td>
<td>J. Freedman (683); P. Winne (435);</td>
<td>V. Collet (0.8); K. Zumwalt (0.73);</td>
</tr>
<tr>
<td></td>
<td>R. Toutkoushian (55); F. Mensah (21); Y. Kanno (16);</td>
<td>F. Mensah (14); R. Toutkoushian (9);</td>
<td>W. Tierney (1708); L. Zenk (982); Z. Xu (457);</td>
<td>H. Shaked (0.08); E. Hutt (0.08);</td>
</tr>
<tr>
<td>2017</td>
<td>N. Shah (30); J. Marsh (30); T. Howard (29);</td>
<td>A. Barton (180);</td>
<td>A. Barton (180);</td>
<td>T. Howard (0.69);</td>
</tr>
<tr>
<td></td>
<td>R. Toutkoushian (55); F. Mensah (21); Y. Kanno (16);</td>
<td>G. Mensah (14); R. Toutkoushian (9);</td>
<td>W. Tierney (1708); L. Zenk (982); Z. Xu (457);</td>
<td>H. Shaked (0.08); E. Hutt (0.08);</td>
</tr>
<tr>
<td>2018</td>
<td>N. Shah (30); J. Marsh (30); T. Howard (29);</td>
<td>A. Barton (180);</td>
<td>A. Barton (180);</td>
<td>T. Howard (0.69);</td>
</tr>
<tr>
<td></td>
<td>R. Toutkoushian (55); F. Mensah (21); Y. Kanno (16);</td>
<td>G. Mensah (14); R. Toutkoushian (9);</td>
<td>W. Tierney (1708); L. Zenk (982); Z. Xu (457);</td>
<td>H. Shaked (0.08); E. Hutt (0.08);</td>
</tr>
</tbody>
</table>

Notes: (1) Scopus does not contain the citation records for the publication before 1996; (2) For the simplicity of expression, we only use the first author and the corresponding measurement indicator for each article; (3) Topic modeling used the aggregated influence factor, and we multiply the original influence factor by 100; (4) There are 2173 articles identified in Google Scholar, 1641 in Scopus, and 1576 in Mendeley.

### Discussion

Overall, the indicators of scholarly impact capture the effect of the articles from different perspectives. Citation indicators describe the impact of publications through reference within academia. The underlying assumption is that an impactful publication will be recognized since other scholars will cite its ideas. Meanwhile, altmetrics assume that user behaviors (e.g., searching, reading, and adding publications into personal libraries) can be considered as indicators of quality. For text mining, the underlying assumption is that an influential article will affect how future articles are written. Consequently, this effect can be detected by examining the evolution of topic distribution in the corpus over time.

Of the citation count indicators, Google Scholar defines the academic community more broadly than Scopus since it captures more online publications. However, it incorporates pre-
prints, conference papers, technical reports, and abstracts. It treats these sources no differently than peer-reviewed journal articles. Consequently, informaticians criticize the utility and quality of the Google Scholar index. Other reasons also include inadequateness and frequency of updates (Falagas, Pitsouni, Malietzis, & Pappas, 2007), coverage differences across different disciplines (in particular, ambiguous coverage in the arts and humanities; Kousha & Thewall, 2007; Jody, 2017), and being open to gaming (e.g., predatory pseudo-science journals are included; Jeffrey, 2014). As a representative subscription-based database, Scopus focuses on a smaller set of specific journals and reviews these journals every year to ensure that high standards are maintained. Thus, the difference between Google Scholar and Scopus is mainly in the balance between coverage and quality of the index. If we want to measure the impact of one publication within the narrow and higher standards of academia, Scopus tends to be a better choice than Google Scholar. Otherwise, Google Scholar provides a broader but unstable estimation of the impact on a more loosely defined academy community.

Altmetrics plays a supplementary role in measuring scholarly impact. These new technologies become even more important as more students and scholars today obtain their information in digital ways. Although altmetrics cannot describe the entire picture of how knowledge and information are searched, obtained, and shared, they provide additional insight into the patterns of usage for scholarly articles beyond citation. After all, not all readers of academic publications are researchers themselves. Altmetrics that extend beyond citations can capture the learner community that extends beyond the limited circle of authors. Readership, bookmarking, pageviews, or duration time on an article is calculated based on the online audience. The limitations of this approach are also clear: (1) it cannot measure the physical learning and reading behaviors; (2) it is highly affected by uncontrollable factors like interface
(UI) design, advertising, and random visiting, as these factors are difficult to normalize; (3) it is only focusing on a limited number of users who are actively using the digital platform like Mendeley. However, altmetrics do indeed provide potential solutions for problems that are hard to solve with citation indicators. For example, some questions that arise are: how to distinguish the impact of vast numbers of articles without citations? How to help journal publishers quickly capture the research tendencies and the interests of their audience? Once the traffic of a publication becomes stable and large, altmetrics tends to provide measurements closer to reality and so are useful.

Text mining is the most computationally complex and data-intensive approach. Topic modeling provides more detailed information through the analysis of the context. However, the estimation time would increase exponentially if an increasing number of publications are included in the corpus. Consequently, this method may be useful to identify the research trends in a specific field or for some specific journals, but it is not generally applicable across different domains. This problem is common when informaticians try to apply new technologies (e.g., neural network and complex social network analysis) with more information. To alleviate this problem, two main approaches have been discussed. First, we need to work with statisticians and computer scientists to find more advanced estimation techniques to reduce the computational complexity of algorithms. For example, the stochastic variational inference has been designed for analyzing very large data sets (D. M. Blei, Kucukelbir, & McAuliffe, 2017). Second, identifying the most influential information from a publication rather than incorporating the whole context. For example, we could input only the abstract and keywords to reduce the computational complexity and improve the generalizability of text mining technologies.
CONCLUSION

In this study, we are not aiming at identifying the best scholarly impact measure. Instead, we emphasize the uniqueness of different types of measurements through comparisons. Our data is limited to a single journal. However, focusing on one specific journal allows us to provide a more comprehensive and detailed analysis and concerning editorial curation work. Meanwhile, all of the citation and readership data we collected represents a unique look. It has limitations in detecting gradual change and tendencies. Consequently, we recommend more research to replicate and extend our data analysis in the future.

In this study, we compare three different types of scholarly impact measurements: citation count, altmetrics, and text mining indicators. The selection of a measure is based on different measurement goals and the information available. The contribution of this research is three-fold: (1) for researchers, this study shows more options beyond citation-based indicators for measuring the impact of scholarly publications, and these alternative measures provide more possibilities for searching useful learning resources; (2) for informatics, this study provides empirical evidence of the similarities and differences of different scholarly impact measurements and their potential and challenges in the application; (3) for journal publishers like TCR, this study shows how an academic publisher can analyze their readership and determine reader interests.

For informaticians, we recommend additional studies that might: (1) provide more empirical evidence about how different scholarly impact measures are applied across various disciplines in practice; (2) pay more attention to the theoretical foundations of different measures and clarify their strengths and limitations under different application situations; (3) motivate
improving the coverage and quality of data sources and suggest principles and standards for data collection and cleaning; and (4) compare different measurement approaches and combine the results from different approaches to provide a more comprehensive analysis. Additionally, we also need to face the new challenges of informatics in the digital revolution. There is increasing use of preprint, digital archive, and academic social networking and collaboration platforms. Linking and identification of academic content across these new distribution channels become more difficult. Finally, we need to rethink the fundamental definition of scholarly impact. Because no matter what measures we apply, it may not just be about the quality of a publication in terms of innovation, integrity, and recognition from the academic community. It is also about its searchability and open-source status.

For journal publishers, improving the visibility of publications is important. Using linking indicators (e.g., DOI for publications and ORCiD for authors) will increase the searchability of the publication across different platforms and improve the quality of bibliometrics databases. Additionally, journal publishers should set up frameworks to provide more detailed descriptions of scholarly contributions and impact. For example, distinguishing the contribution of each author in terms of design and analysis, data collection, and writing might be reported along with the currently used indicators. Meanwhile, using altmetrics and new data-intensive technologies (e.g., text mining, neural network, and social network analysis) will provide useful and detailed information for decision-makers.
References


Hirsch, J. E. (2010). An index to quantify an individual’s scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics.* https://doi.org/10.1007/s11192-010-0193-9


