Using field and quasi experiments and text-based analysis to advance international business theory

Ravi S. Ramani a,1,*, Herman Aguinis b,1

a Earl G. Graves School of Business and Management, Morgan State University, 4200 Hillen Road, Baltimore, MD 21218, USA
b School of Business, The George Washington University, 2201 G St. NW, Washington, DC 20052, USA

ARTICLE INFO

Keywords:
• Research methodology
• Research design
• Analytical methods
• Sentiment models
• Experiments

ABSTRACT

Methodological developments are critical for driving theoretical advancements in international business (IB) due to the field’s diversity regarding disciplinary, theoretical, and conceptual bases. We provide an accessible introduction to two methodological approaches—one related to design and one to analysis—that are currently underutilized in IB despite their great potential: (a) field and quasi experiments, and (b) text-based analysis. We describe each method and provide examples of how they can be used to make advancements in several IB domains and theories including internationalization process theory; ownership, location, internalization (OLI) paradigm; knowledge-based view of multinational enterprises; dynamic capability theory; and international entrepreneurship.

1. Introduction

Advancing knowledge in international business (IB) and all scholarly fields requires ongoing theory improvements, which are usually produced by the use of innovative methodologies (Aguinis et al., 2023; Boyd et al., 2017; Hennart & Sutherland, 2022; Kuhn, 1967; Nielsen et al., 2020). However, most researchers only become familiar with and use a limited set of methodological approaches (Cornelissen, 2017; Welch et al., 2022). While deep knowledge and expertise in a handful of familiar methods is useful, and perhaps even necessary, methodological specialization can also constrain awareness, understanding, and adoption of alternative methodologies that can be helpful in advancing theory (Bramen, 2022; Knight et al., 2022). Stated differently, researchers often fail to examine “phenomena and problems to which” their “familiar toolkits cannot be as easily applied” (Delbridge & Fiss, 2013, p. 328). In turn, using just a handful of familiar methods precludes researchers from pursuing the “theoretical and methodological pluralism” needed “to promote complementary ways to address new and difficult research questions and enhance the overall development of the field” (Cuervo-Cazurra et al., 2016, p. 883).

Accordingly, following in the tradition of methodological articles such as Decker (2022), Gaur and Kumar (2018), Lindner et al. (2021), Lukoianove et al. (2022), Nguyen and Tull (2022), Nielsen and Raswant (2018), Richter and Hauff (2022), Schotter et al. (2018), and Vashchilko et al. (2022), our goal is to increase awareness about and spur the dissemination of methodological approaches that can serve as catalysts for theory advancements in IB. Specifically, our article addresses how the following two methodological approaches can be used by IB researchers to advance theory: (a) field and quasi experiments, and (b) text-based analysis. While each of these methodologies are known and regularly employed in other fields, a review by Nielsen et al. (2020) revealed that they are underutilized in IB research.

Our article provides an accessible introduction to each of these methodological approaches, with examples of how they can advance IB theory. As a preview of their potential, Table 1 outlines how each of these approaches may be applied to questions and phenomena in widely-studied IB domains (cf. Nambisan et al., 2019), and the context in which these issues may be studied. We also discuss examples from some of these domains and theories throughout our article as we describe their application.

---

We thank Journal of World Business (JWB) Editor-in-Chief Ajai Gaur and three JWB anonymous reviewers for offering highly constructive feedback that allowed us to improve our manuscript substantially. A previous version of this manuscript was presented at the meetings of the Academy of International Business US Northeast, Washington DC, October 2022.

* Corresponding author.
E-mail address: ravisramaniphd@gmail.com (R.S. Ramani).
1 Both authors contributed equally to this research.

Received 29 August 2021; Received in revised form 30 April 2023; Accepted 18 May 2023
1090-9516/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
2. Field and quasi experiments

The search for causal relations is at the heart of all scientific research (Aguinis et al., 2021; Eden, 2017; Zellmer-Bruhn et al., 2016). Establishing evidence regarding causality requires three criteria: a) covariation between the independent (i.e., predictor) and dependent (i.e., criterion or outcome) variables; b) temporal precedence (i.e., variation in the independent variable precedes variation in the dependent variable); and c) ruling out alternative explanations for the observed relation (Heath & Luff, 2018; Podsakoff & Podsakoff, 2019). Thus, regardless of the sophistication of the analytical method, commonly used research designs in IB, such as observational studies based on surveys or secondary data from archival databases, are inadequate when seeking evidence about causal relationships (Shadish, 2010; Zellmer-Bruhn et al., 2016). We therefore suggest implementing experimental designs to better test and understand causal relationships in IB.

It is unsurprising that IB researchers may hesitate to conduct traditional experiments because of two reasons. First, experimental research is often associated with highly-controlled laboratory settings, and IB researchers usually cannot manipulate variables such as buy-or-build decisions, risk assessments, or internationalization decisions for large multinational enterprises (MNEs) (Aguinis et al., 2023; Reeb et al., 2012). Additionally, even if managers are willing to participate in experiments (e.g., using simulations or vignette studies), there may be concerns about whether behaviors or choices made in these hypothetical situations mirror how they would act in real ones (Highhouse, 2009). Second, IB researchers may believe that creating and conducting experiments is difficult, time-consuming, and expensive (Aguinis & Lawal, 2012). Together, these challenges can seem insurmountable, leading the vast majority of IB researchers to abandon the greater internal validity and causal advantages of experiments in favor of the seemingly greater external validity and yet non-causal results of non-experiments (e.g., collecting data using surveys in a passive observation design, analyzing archival data, qualitative case studies).

We posit that there is an underutilized way to balance these seemingly inescapable tradeoffs: Field and quasi experiments. Although the terms are often mistakenly used interchangeably, they represent two distinct varieties of experimental designs, as summarized in Table 2. At the heart of field experiments are the researcher’s ability to (a) manipulate the independent variable (i.e., treatment) that is hypothesized to affect the dependent variable (i.e., outcome), and (b) randomly assign participants to treatment (i.e., a group that experiences the manipulation) and control (i.e., a group that does not receive the treatment) conditions (Baldassarri & Abascal, 2017; King et al., 2013). Because field experiments examine phenomena in natural settings while also employing research design features that support causal inference, they are considered the “gold standard” of experimentation (Eden, 2017, 2021; Shadish & Cook, 2009).

There are three critical questions that IB researchers need to consider in deciding whether a field experiment may be feasible and appropriate for a specific research situation:

1. Can we manipulate the independent variable? For example, while we may be able to manipulate certain aspects related to expatriate staffing (e.g., training related to deployment, opportunities to build social networks in host country after deployment), we are unlikely to be able to decide when or how to make changes to an organization’s executive leadership. Therefore, certain phenomena, such as the composition of the board of directors or the degree to which an organization engages in corrupt business practices, are not suited for field experiments.

2. Can we randomly assign participants to conditions? Random assignment is perhaps the most important feature of field experiments, as it allows for the control of multiple potential confounders, including those that we may not even be aware of (Dunning, 2008; Gerber & Green, 2012). In random assignment, each participant has an

<table>
<thead>
<tr>
<th>IB Domains and Theories</th>
<th>Illustrative Questions</th>
<th>Methodological Approach and Application Context</th>
<th>Text-based Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internationalization Process Theory</td>
<td>• How do multinational enterprises (MNEs) use incremental or springboard internationalization strategies? • How do board structure and corporate governance practices of business groups change during and after internationalization?</td>
<td>Adoption and enforcement of global trade agreements or international treaties (e.g., intellectual property) Collecting data on roughly equivalent business groups within an industry as some internationalize while others do not</td>
<td>Applying structural topic modeling (STM) to examine a firm’s press releases, 10-Ks, SEC filings, and external trade magazine and news outlet articles Analyzing documents related to environmental, social, and corporate governance (ESG) before and after internationalization using latent Dirichlet allocation (LDA) and STM</td>
</tr>
<tr>
<td>Ownership, Location, Internalization (OLI) or Eclectic Paradigm</td>
<td>• How do MNEs make choices regarding location and timing of foreign direct investment (FDI)?</td>
<td>Impact of governmental policy changes on carbon emission policies in emerging market economies (EMEs)</td>
<td>Analysis of company’s annual reports “Management discussion and analysis” sections using LDA</td>
</tr>
<tr>
<td>Knowledge-Based View of MNEs</td>
<td>• What are different post-entry expansion strategies of MNEs in foreign markets? • What role do human resource management (HRM) practices play in knowledge-transfer?</td>
<td>Implementation of business analytic capabilities of different firms in managing supply chains and post-entry expansion strategies Field experiment where the choice of who attends training is randomized to create a treatment- and control-group</td>
<td>Using LDA to analyze communications (e.g., emails, memos) between MNE headquarters and local subsidiaries to uncover knowledge-sharing practices and their relation to expansion strategies Using LDA and natural language processing (NLP) to analyze documents related to job design, performance management and rewards, and training</td>
</tr>
<tr>
<td>Dynamic Capability Theory</td>
<td>• What is the best way to reconfigure a firm’s dynamic capabilities to better launch new products and brands?</td>
<td>Usefulness of intelligent fast failure (IFF) frameworks for product launches across different subsidiaries/ branches of a firm</td>
<td>Analyzing CEO focus by examining: 1) Letters to shareholders, 2) published interviews, 3) speeches, 4) press releases, and 5) quarterly earnings calls using LDA, linguistic inquiry word count (LIWC), and support vector machines (SVM)</td>
</tr>
<tr>
<td>International Entrepreneurship Perspective</td>
<td>• What factors influence international entrepreneurship efforts of small and medium enterprises (SMEs)?</td>
<td>Privatization of state-owned firms/ sectors or political turbulence as drivers of international entrepreneurship efforts</td>
<td>Analysis of firm social media (e.g., Twitter, Instagram, Facebook) content and documents prepared for venture/angel investors using bing and national research council Canada (NRC) lexicons and LDA</td>
</tr>
</tbody>
</table>

equivalent chance of being assigned to each condition. Referring back to the example of expatriate staffing, candidate selection for these positions is likely based on pertinent factors (e.g., previous performance, company tenure, candidate knowledge, skills, and abilities). By randomly assigning candidates to different conditions, we can control for these and many other potential alternative explanations of the effects of the manipulation (Eden, 2017).

3. What level of “fieldness” is required for our study? Field experiments can vary in terms of, for example, the setting for the study, the authenticity of the treatment, the influence of observation on behavior (i.e., Hawthorne effect), and the salience of the outcome measures (Baldassarri & Abascal, 2017; Gerber & Green, 2012; Hansen & Tummers, 2020).

Each of the three aforementioned questions is relevant in our illustrative study about expatriate staffing. For example, considering training, should it be conducted before the expatriate is deployed, or after? Should training be confined to a single session, making it more manageable, or should researchers strive for authenticity by holding multiple sessions? Should researchers inform those in the treatment condition that they are being or will be observed? Should outcomes be subjective (e.g., attitude, feelings) or objective (e.g., relevant individual or organizational performance metrics)?

Although conducting a field experiment for the first time may seem daunting, it is a type of research design that is more tractable than many assume. As an example, consider the study by Liu and Meyer (2020), which used a case-study approach and found that training and workshops played an important role in the ability of boundary spanners to engage in reverse knowledge transfer. Because many employees will likely need to attend these trainings and workshops, and it is improbable that all of them do so at the same time, this situation is ripe for a field experiment. Specifically, as managers are often indifferent as to the order in which employees attend a training or a workshop, researchers can partner with an organization to implement a randomized-rollout or waiting-list design where the choice of who attends training first is randomized, thus creating a treatment group (i.e., those who attend training first) and a control group (i.e., those who attend training later), and conducting a natural field experiment (Eden, 2017; Harrison & List, 2004).

There are many other IB phenomena that lend themselves well to the use of field experiments. For example, researchers adopting a knowledge-based view of the firm might manipulate the presence and degree to which employees have access to different tools (e.g., customer interactions, expert feedback, ability to share knowledge across divisions) to examine the drivers of innovation in MNEs. Similarly, researchers relying on transaction cost economics and international corporate governance theories can manipulate the type (i.e., coercive versus enabling) and extent of formalization in the procedures used to conduct or report on sustainability programs and examine its impact on attitudes, compliance, and engagement with sustainability efforts. As yet another example, researchers studying the liability of foreignness and competitive strategy can manipulate the timing and use of marketing and community-building strategies to understand how MNEs and their subsidiaries develop locally-relevant and authentic presence. We refer readers to Table 1 for additional examples of how we can use field experiments to address questions pertaining to several IB domains and theories.

In contrast to field experiments, in quasi experiments researchers do not control the manipulation, but rather take advantage of circumstances that divide participants into treatment and control conditions (Aussens et al., 2011; Zellmer-Bruhn et al., 2016). Participants in quasi experiments may self-select into treatment or control conditions, or there may be naturally occurring events (e.g., new governmental regulations, geopolitical events) that allow researchers to treat participants as if they were randomly assigned to conditions (Campbell & Stanley, 1971; Dunning, 2008; Podsakoff & Podsakoff, 2019). Returning to our running example of the role of training and workshops in allowing boundary spanners to engage in reverse knowledge transfer, the design would be quasi-experimental if the order in which employees received training was not random but instead based on particular factors (e.g., seniority, job function) (Chatterji et al., 2016). Alternatively, the study may have started as a field experiment, but become quasi-experimental because external circumstances (e.g., promotions, strategic changes) altered the randomized process used to select participants for training (Eden, 2017). By accommodating situations like these, quasi experiments allow researchers to conduct studies with greater internal validity compared to survey or archival data studies, while still maintaining high external validity (Aguinis et al., 2020; Grant & Wall, 2009; Podsakoff & Podsakoff, 2019).

Quasi experiments offer four key benefits. First, this approach is well suited to examine phenomena that occur in large organizations with multiple sites, perhaps in different states or countries or even continents, where changes are likely to be more long-lasting (Grant & Wall, 2009). These features make quasi experiments especially conducive to the types of phenomenon typically studied by IB researchers. Quasi experiments are also useful for exploring the microfoundations view of IB, which examines how the actions of individuals and groups affect the strategic decisions and choices of MNEs (Boilinger et al., 2022; Foss & Pedersen, 2019). Accordingly, quasi experiments complement approaches such as participant interviews (e.g., Ferraris et al., 2022) and the analysis of archival data (e.g., Castellani et al., 2022). In addition, because quasi experiments address questions relevant to IB domains and theories.

### Table 2
Summary of Differences among Field Experiments, Quasi Experiments, and Non-experimental Designs.

<table>
<thead>
<tr>
<th>Research Design</th>
<th>Manipulation of Independent Variable(s)</th>
<th>Assignment to Condition (Treatment versus Control)</th>
<th>Control Group or Comparison Condition</th>
<th>Illustrative Studies Addressing International Business Phenomena</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field experiment</td>
<td>By researcher</td>
<td>Random assignment by researcher</td>
<td>Identified a-priori by researcher</td>
<td>Fu et al. (2022)</td>
</tr>
<tr>
<td>Quasi experiment</td>
<td>Occurs naturally (i.e., not by researcher)</td>
<td>Assignment based on particular criteria (i.e., non-random); possible to self-select into treatment or control condition</td>
<td>Identified after assignment to condition</td>
<td>Buckley et al. (2018)</td>
</tr>
<tr>
<td>Non-experimental designs (e.g., surveys, archival data, qualitative case studies)</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>Fan and Harring (2017)</td>
</tr>
</tbody>
</table>

2 While some researchers distinguish between experiments where participants self-select into conditions versus those where they cannot, both experimental designs are similar in that manipulation and assignment are not controlled by the researcher. Accordingly, we use the term quasi experiments to refer to both types of designs.
experiment results allow for causal inferences, researchers can not only examine what has happened, but also make predictions about the future. Third, quasi experiments allow IB researchers to examine questions that involve sensitive or proprietary information that firms may not want to disclose, hypothetical or yet unobserved situations, and untested theoretical models such as, for example, those proposed in conceptual articles. Finally, because quasi experiments identify causal relationships and not just covariation, they are useful for helping IB researchers offer practical advice to leaders and policy makers.

An example of an area ripe for quasi experimental research is the relationship between internationalization and business groups, where most (i.e., 89%) studies have relied on secondary sources (Aguilera et al., 2020). For example, one of the causal hypotheses of interest proposed by Aguilera et al. (2020) focuses on how business-group board structures and their corporate governance practices change with internationalization. Researchers using quasi experiments may start by focusing on a particular industry within a country, and identifying business groups that are roughly equivalent to each other in terms of current board structure and corporate governance practices, but which have not yet ventured into international operations. Following these firms as some undergo internationalization (i.e., the intervention), while others do not (i.e., control group), and collecting data on the changes (or lack thereof) in parameter estimates would provide data to test hypotheses. Researchers could also collect data on other factors (e.g., political structures or laws in different countries, degree of foreign ownership of international venture) to account for alternative explanations, and collect data at different points in time (e.g., same business group entering different countries, different levels of foreign ownership when internationalizing) to build support for or to counter these alternative explanations.

There are several other IB phenomena that would benefit from using quasi experiments. For example, consider researchers investigating how executive’s heuristics, psychic distance, and agency concerns may affect resource allocation to subsidiaries and subsequent subsidiary performance. We are not able to manipulate executive’s perceptions of psychic distance or who is appointed to lead subsidiary operations, precluding a field experiment. However, we can study these issues via quasi experiments by partnering with an MNE and examining how natural variance between different executives interacting with subsidiary managers whose names are stereotypically representative of the country of the subsidiary (e.g., “Patel” for India, “Pierre” for France, “Isabel” for Spain) versus those whose names are representative of the headquarters (e.g., “Suzuki” or “Tanaka” for Japan, “Park” or “Kim” for South Korea). As another example, researchers using internationalization theory can take advantage of naturally occurring phenomena (e.g., global financial crises, the 2020 withdrawal of the United Kingdom from the European Union) to examine changes in risk taking propensity among MNEs. Similarly, based on the ownership, location, internalization (OLI) paradigm, we can examine how the implementation of carbon emissions trading in different markets (e.g., European Union, China, India) affected subsequent foreign direct investment (FDI) in those markets. Table 1 includes additional illustrations of questions across IB domains and theories that can be answered by using quasi experiments.

Finally, data from a quasi experiment can be analyzed using one of three fairly novel approaches in IB: (a) Difference-in-differences design (DDD), (b) regression discontinuity design (RDD), and (c) synthetic control method (SCM). In DDD, firms that are equivalent to each other are subject to similar time-related effects (e.g., market fluctuations) are examined before and after the intervention. Assuming certain conditions are met (e.g., no spillover between firms, equivalent effect of time), then observed differences in the outcome can be attributed to the intervention (Lonati et al., 2018). In RDD, the assignment of firms to intervention or control group is based on certain variable(s), but the firms do not have control over whether or not they receive the intervention (Lonati et al., 2018). For example, some of the industries in which the business groups operate may be the subject of government policy changes, which affects the decision to internationalize, but over which business groups do not have power. Finally, SCM is useful in small-n studies or when data unavailability does not allow us to establish equivalence prior to intervention (as in DID), or there is no variable that determines the assignment of business groups to different conditions (as in RDD) (Lonati et al., 2018). To overcome this limitation, we can match the outcomes of a focal business group with that of a simulated business group which is created as a weighted average of all other business groups. Furthermore, we can conduct “placebo tests” by conducting simulations where the non-intervention business groups receive the intervention at varying times. If there is no statistically significant impact of the intervention in these simulations, we can infer that those differences in outcomes are due to the intervention in the treatment unit (Abadie, 2021).

In sum, field and quasi experiments provide IB researchers with a useful approach to ascertain causal relations. In addition to the sources cited in the preceding section, Appendix 1 lists additional resources useful for those interested in exploring and potentially applying field and quasi experiments in their own work.

3. Text-based analysis

Text-based analysis (TBA) is a particularly relevant methodology given the rapid increase in the amount of structured and unstructured text available to IB researchers (e.g., annual reports, firm websites, press releases, governmental filings, social media sites). Many IB scholars are probably already familiar with some versions of TBA, such as the use of notes from participant observations or interviews to create codes and categories using software such as NVivo (e.g., Jasovska et al., 2023; Peltokorpi, 2022). Another likely familiar form of TBA is bibliometric analysis, which focuses on documents (e.g., academic articles), to which researchers apply existing or custom-built codes to create databases that can be summarized in tables and graphs, such as when conducting systematic literature reviews (e.g., Buckley et al., 2023; Debellis et al., 2021; Zhao et al., 2022).

In addition to these perhaps familiar forms, there are newer TBA applications that can help generate theory advancements by providing fresh perspectives on existing areas of inquiry, or by helping answer questions for which it is not possible to find answers using more traditional approaches. For example, consider a study examining the nature and configuration of dynamic capabilities, and their relation to a firm’s organizational adaptation and success in international markets. Traditional methodological approaches for answering this question might involve, for example, a multiple case-study approach (e.g., Zeng, 2022), archival data (e.g., Fu et al., 2022), or perhaps a mixed-methods approach (e.g., Ferraris et al., 2022). However, a particular challenge when studying dynamic capabilities is that the fast-changing nature of adaptation processes exacerbates concerns regarding the scope and timing of archival data collected, interview recall error, survey demand effects, common method variance, and even endogeneity (Bartholomew & Smith 2006; Eisenhardt & Martin 2000; Laaksonen & Peltoniemi, 2018; McKenny et al., 2018). These are precisely the challenges for which newer and more advanced forms of TBA are particularly useful because they allow researchers to quickly analyze large volumes of near real-time data (Gaur & Kumar, 2018; McKenny et al., 2018; Piepenbrink & Gaur, 2017).

Specifically, researchers can use TBA by implementing topic modeling using supervised or unsupervised machine learning algorithms such as, for example, those employed in bag-of-words (BOW), structural topic modeling (STM), and natural language processing (NLP) approaches (Banks et al., 2018; Gaur & Kumar, 2018). For example, companies regularly update their websites in response to changing crises, the 2020 withdrawal of the United Kingdom from the European Union, archival data (e.g., Fu et al., 2022), or perhaps a mixed-methods approach (e.g., Ferraris et al., 2022). However, a particular challenge when studying dynamic capabilities is that the fast-changing nature of adaptation processes exacerbates concerns regarding the scope and timing of archival data collected. However, it is not possible to find answers using more traditional approaches. For example, consider a study examining the nature and configuration of dynamic capabilities, and their relation to a firm’s organizational adaptation and success in international markets. Traditional methodological approaches for answering this question might involve, for example, a multiple case-study approach (e.g., Zeng, 2022), archival data (e.g., Fu et al., 2022), or perhaps a mixed-methods approach (e.g., Ferraris et al., 2022). However, a particular challenge when studying dynamic capabilities is that the fast-changing nature of adaptation processes exacerbates concerns regarding the scope and timing of archival data collected, interview recall error, survey demand effects, common method variance, and even endogeneity (Bartholomew & Smith 2006; Eisenhardt & Martin 2000; Laaksonen & Peltoniemi, 2018; McKenny et al., 2018). These are precisely the challenges for which newer and more advanced forms of TBA are particularly useful because they allow researchers to quickly analyze large volumes of near real-time data (Gaur & Kumar, 2018; McKenny et al., 2018; Piepenbrink & Gaur, 2017).

Specifically, researchers can use TBA by implementing topic modeling using supervised or unsupervised machine learning algorithms such as, for example, those employed in bag-of-words (BOW), structural topic modeling (STM), and natural language processing (NLP) approaches (Banks et al., 2018; Gaur & Kumar, 2018). For example, companies regularly update their websites in response to changing crises, the 2020 withdrawal of the United Kingdom from the European Union, archival data (e.g., Fu et al., 2022), or perhaps a mixed-methods approach (e.g., Ferraris et al., 2022). However, a particular challenge when studying dynamic capabilities is that the fast-changing nature of adaptation processes exacerbates concerns regarding the scope and timing of archival data collected, interview recall error, survey demand effects, common method variance, and even endogeneity (Bartholomew & Smith 2006; Eisenhardt & Martin 2000; Laaksonen & Peltoniemi, 2018; McKenny et al., 2018). These are precisely the challenges for which newer and more advanced forms of TBA are particularly useful because they allow researchers to quickly analyze large volumes of near real-time data (Gaur & Kumar, 2018; McKenny et al., 2018; Piepenbrink & Gaur, 2017).
scraping content displayed on company websites over time and by examining CEO letters. This content can be analyzed using topic modeling approaches such as latent Dirichlet allocation (LDA), an unsupervised BOW algorithm that analyzes text to form latent categories reflecting unobserved structures or topics contained within a document (Blei et al., 2003; Hartmann et al., 2019). Analyzing this content would reveal how capabilities within the firm changed over time, as reflected in text on the company website or in CEO letters. Furthermore, researchers can use more intricate methods such as STM, which allows for the inclusion of covariates, and employ regression techniques (e.g., two stage least squares approach; 2SLS) to examine how covariates affect the presence and ubiquity of topics and make predictions about future changes in capabilities (Banks et al., 2018; Puranam et al., 2017). Finally, to triangulate findings, researchers can utilize more computationally sophisticated approaches such as NLP, which employs a fully automated process that aims to mimic how people use language by encoding rules about grammar and sentence structure and utilizing the context in which words are embedded. Therefore, NLP can be used to produce additional insights by analyzing related data such as text from CEO press releases or interviews in trade magazines (Kobayashi et al., 2018; Piepenbrink & Gaur, 2017).

Another TBA approach involves making inferences about underlying sentiment as expressed through text (Hartmann et al., 2019). Specifically, sentiment analysis aims to infer “people’s opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes” (Liu, 2012, p. 1). A critical component in sentiment analysis is the “dictionary,” which contains category names, definitions, or assignment rules to classify text into categories, and lists of exemplars or words that belong to each category (Weber, 1990). Researchers can rely on preexisting validated dictionaries, such as the Linguistic Inquiry and Word Count (LIWC; Tausczik & Pennebaker, 2010), the bing lexicon (Hu & Liu, 2004), or the national research council Canada (NRC) word-emotion lexicon (Mohammad & Turney, 2010), which assign words to psychologically meaningful affect categories. For example, Antioco et al. (2023) used LIWC to analyze 2223 press releases from companies in four countries to understand how the use of English as the lingua franca for business affected MNE communication. Similarly, Tian (2022) used LIWC to analyze more than 480,000 social media posts to examine the relationship between a host country’s political hostility and the communication rhetoric used by foreign subsidiaries. Finally, Järvinen et al. (2020) used the bing lexicon to analyze sentiments expressed in 20,153 tweets to show how pre-internationalization firms can use TBA to improve their knowledge about foreign markets.

Alternatively, IB researchers can create custom dictionaries for specific contexts using machine learning methods such as support vector machines (SVM). When using SVM, researchers first manually analyze a number of documents regarding sentiment (i.e., the training dataset), then re-examine the same documents using the SVM algorithm to train the classifier, with corrections made to improve performance. Finally, the SVM algorithm is applied to a new, larger set of similar documents to infer sentiment. For example, Sheng et al. (2019) used SVM to analyze more than 40,000 hotel reviews and 23,000 responses over a 15-year period. After manually coding 350 examples (i.e., the training dataset), Sheng et al. (2019) were able to use SVM to analyze the rest of the documents and examine how managerial actions in response to reviews influenced firm dynamic capabilities, and subsequently, customer satisfaction. As another example, Itikhar and Khan (2020) used SVM to analyze over 1.2 million Twitter tags and Facebook comments to recommend how a retail organization operating across multiple continents should reconfigure its dynamic capabilities to meet demand in response to changing consumer sentiments.

As an illustrative example of using TBA to examine changes in company strategy and dynamic capabilities, we analyzed Amazon CEO’s annual letters to shareholders between 1997 (when the company went public) and 2022 (most recent letter as of the writing of our article). Our illustration focused on three analytical areas: word frequencies, sentiment analysis, and topic models using LDA. For brevity, we do not detail all considerations in each step of the process, nor do we discuss all potential analytical options available. However, the online supplement provides the complete annotated R code (Appendix A) and example output and figures from our analysis (Appendices B and C). Furthermore, to enhance transparency and replicability (Aguinis et al., 2017, 2018), the Amazon dataset as well as an additional dataset (MetLife CEO letters to shareholders) is available from the authors upon request. Researchers can use these datasets to practice TBA by replicating our analysis and using them as instructional materials in classroom and workshop settings.

As a summary and preview, we employed the eight-step process outlined in Fig. 1 to explore changes in Amazon’s strategy or dynamic capabilities over the last 26 years. While these steps are presented sequentially, TBA is frequently iterative, with researchers refining the analytical process based on the idiosyncrasies of their data and the results of previous operations. The first step, after formulating pertinent hypotheses or research questions, is identifying the types of documents (e.g., website text, social media posts, press releases, emails) that will be used. In our example, we used letters to shareholders (1997–2022) from Amazon’s CEOs. Because TBA relies on probabilistic methods, ensuring there is a sufficient volume of data available for analysis is critical for generating accurate and reliable models (Marshall et al., 2022). Furthermore, researchers should consider both the number of documents and the amount of content in each document. For example, Twitter posts are relatively short (280 characters), while annual reports and SEC filings can run into hundreds of pages.

Next, data need to be acquired, perhaps via downloading or scraping from relevant websites. While a discussion of web scraping is beyond the scope of this article, there are several excellent resources (e.g., Bradley & James, 2019; Krotov & Tennyson, 2021; Landers et al., 2016) that explain the process using both R and Python, as well as paid services (e.g., import.io). In our example, we downloaded all letters directly from Amazon’s website. Third, we imported these letters into R. These letters constituted the “corpus” for our analysis.

The fourth step is cleaning the data. This comprised of several tasks, including: a) tokenization (converting documents to units of text such as individual words which represent the simplest and most meaningful component of the document), b) removal of punctuation marks and numbers, c) converting all letters to lowercase (analytical methods are case-sensitive; “Amazon” and “amazon” count as separate words), d) stemming (reducing words to their base forms; transforming “accommodate” and “accommodating” into the base form “accommod”), and e) removing “stop words” (non-text-specific common words that do not add value to the analysis, such as “the,” “and,” “of,” and “is”). A good practice is to inspect the data as each of these tasks is executed. For example, as shown in Appendix B of the online supplement (Output B1 and B2 respectively), for our data, the 2022 and 2021 CEO letters were the most verbose (with 5245 and 4562 tokens respectively) before removing stop words. After stop word removal, the two most verbose CEO letters were 2022 and 2013 (with 2286 and 1956 tokens respectively). Similarly, the three most common tokens in our data before stop word removal were “and” (n = 2135), “the” (n = 2008), and “to” (n = 1925). After stop word removal, the top three were “custom” (stem of customer/customers; n = 625), “amazon” (n = 394), and “busi” (stem of business/businesses; n = 232). We also visually inspected the most common tokens using a word cloud (Appendix C of online supplement, Figure C1).

3 https://ir.aboutamazon.com/annual-reports-proxies-and-shareholder-letters/default.aspx
4 Alternatives to individual words (unigrams) include sequences of words such as bigrams (two words) and trigrams (three words). A good practice is to examine if conclusions change based on the type of token used.
The fifth step is to create the database for analysis, or “document term matrix” (DTM), in which each row represents a document, columns represent tokens, and cells indicate how often each token was used in each document. Depending on the package used in R, many of the preprocessing tasks from step four and the creation of the DTM can be accomplished using a single line solution. That is, steps four and five are often combined in practice. In our example, we used the “tidytext” package, but R provides a number of alternatives (e.g., quanteda, tm) that accomplish the same function.

In the sixth step, we computed some basic text descriptives. These included examining how often particular words were used in the texts (term frequency or tf), assigning weights such that more common words were assigned a lower weight (inverse document frequency or idf), and multiplying tf and idf to measure how important a word was to the corpus (tf-idf). While simple, examining the data closely at this stage is critical. Just as data need to satisfy the assumptions underlying linear regression for researchers to obtain valid results, checking for errors at this step is crucial to ensure that more sophisticated analyses are not being biased by inaccurate data. In our data, “custom” was the most frequent token across all 26 letters (n = 625), while the term “employee” (stem of employer/employee) was in 19th position (n = 103), possibly indicating Amazon’s strategic priorities over the years. Furthermore, the token “covid” (from Covid-19) was the 21st most important term in the corpus (tf-idf = 0.023), despite only appearing in three letters (2019, 2020, and 2021), which is understandable given the oversized impact of Covid-19 on Amazon’s operations and strategy. A visual representation of the five most important tokens (i.e., highest tf-ids) for the last 10 letters is shown in Figure C2 in Appendix C of the online supplement. Focusing on the 2022 letter, while “custom” (n = 60; tf = 0.026), “Amazon” (n = 55; tf = 0.024), and “busi” (n = 53; tf = 0.023) were the most-common terms (i.e., highest tf), idf indicates that the most important words included “infer” (stem of inference/inferential; n = 6; idf = 3.26), “ilm” (abbreviation for Large Language Models; n = 6; idf = 3.26), and “macroeconom” (stem of macroeconomic; n = 5; idf = 3.26). Reading the 2022 letter we see that these results align with the strategic areas of emphasis highlighted by the CEO.

Next, we conducted sentiment analysis using the “bing” lexicon to assign positive and negative sentiments to each token. Results show that CEO letters from 2013 to 2014 were most positive, while 2003 and 2016 were the only two letters with an overall negative score (Appendix B of online supplement, Output B3). Examining the letters, we see that the 2013 and 2014 letters introduced several new Amazon offerings (e.g., Kindle, Prime Video, Amazon Web Services), events associated with positive emotions. In contrast, the negative tone of the 2003 and 2016 letters is likely reflective of the stock market downturns in those years (Dougherty, 2016; Jones, 2003), and can be seen in how the CEO defends changes to Amazon’s business practices that sacrifice short-term revenue for long-term profitability. We also computed the most positive and negative sentiment words in the DTM, and visually examined the data using a word cloud (Appendix C of online supplement, Figure C3). This check showed that the word “cloud” features prominently as an indicator of negative sentiment. However, “cloud” in our dataset refers to Amazon’s cloud-based web services, not negative sentiment (e.g., “future is cloudy”), again reflecting the need to use an iterative approach and conduct periodic data checks when employing TBA.

Finally, in the eighth step, we conducted topic modeling using LDA. As a reminder, LDA is an unsupervised BOW approach wherein researchers only specify the number of models, and the algorithm generates these models based on patterns in the data (Welbers et al., 2017). Because these models are data-driven, it is critical to carefully examine the results to ensure they have substantive meaning and are not simply statistical artifacts (Banks et al., 2018; Blei et al., 2003). The recommendation is to initially consider between one and 100 topics, depending on the size of the corpus (Banks et al., 2018). Given that we only had 26 letters in our corpus, we began by running seven models with varying numbers of topics (i.e., 2, 4, 7, 9, 11, 13, & 15 topics) using the “textmimR” package. After running our models, we needed to determine the optimal number of topics. As there is no “one best way,” choosing how many topics requires consideration of several evaluative criteria, as well as the researcher’s informed judgment (Kobayashi et al., 2018). We began by examining “phi,” which represents the distribution of words over topics such that words with high phi have greater frequency within a topic (Jones et al., 2019). In our example, models with nine or 11 topics seemed to be a good fit. Next, we inspected model R². Similar to regression, the goal is to balance explanatory power (i.e., higher R²)

---

5. The tidytext package does not technically create a DTM, but instead data in a long format. However, converting tidy data to DTMs is relatively straightforward.

---

**Fig. 1. Summary of Steps for Conducting Text-based Analysis: Illustration Using Amazon CEO’s Letters to Shareholders.**
with parsimony (i.e., number of topics) (Banks et al., 2018; Welbers et al., 2017). So, while the model with 15 topics had the highest $R^2$ (0.64), models with nine (model $R^2=0.53$) or 11 ($R^2=0.57$) topics also seemed acceptable. We then examined coherence, which is the degree of semantic similarity between high scoring words in a topic (ranging from $-1$ to 1), and helps distinguish between topics that are semantically interpretable and those that are simply statistical inference artifacts (Jones et al., 2019). Results showed that the model with 11 topics had the highest mean (0.09) and median (0.09) coherence. We also examined prevalence, which is the probability of topic distribution in whole document (Jones et al., 2019). This helps investigate if words within a particular topic are unlikely to occur together in the text, even if they are statistically related to each other. Again, the model with nine or 11 topics seemed to be a good fit for our data. We also created a summary of all these measures (i.e., phi, $R^2$, coherence, and prevalence) for all models (Appendix B of online supplement, Output B4), and compared features of the models most likely to be useful (Appendix B of online supplement, Output B5), before making a decision on which model to choose for our data.

Our analysis illustrates how TBA can be used to advance IB theory. Specifically, CEOs play an important role in shaping an organization’s dynamic capabilities, and in turn, the firm’s ability to build or maintain a competitive advantage under changing conditions (Kor & Mesko, 2013; Schilke et al., 2018). Therefore, in addition to letters to shareholders, researchers can use TBA to examine other documents (e.g., CEO interviews, transcripts of earnings calls with investors) before and after a CEO succession event to better understand how this change affects the firm’s strategic priorities and dynamic capabilities. For example, the two letters from Amazon’s current CEO (i.e., 2021 and 2022) suggest that artificial intelligence, predictive analytics, employee relations, and environmental impacts are more salient priorities than under the previous CEO. Using the letters as a training dataset, researchers can also employ supervised machine learning (e.g., naïve Bayes, random forests) to make predictions about what they expect to find in press releases issued since the CEO succession event. Alternatively, researchers can implement STM by including covariates and a regression framework to examine how particular factors are linked to the topics discussed in the letters and press releases. Furthermore, theoretical advancements can be made by examining the phenomenon of CEO succession in different contexts (Makadok et al., 2018). For example, how do changes in dynamic capabilities due to CEO successions in technology firms (e.g., Amazon, Microsoft, Google, WeWork) differ from those in the automotive (e.g., Ford, Subaru, Toyota) or finance (e.g., Mastercard, Visa) industries? Similarly, does the nature of the succession, such as the change from a founder-CEO (e.g., Jeff Bezos to Andy Jassy at Amazon) versus a change from a non-founder-CEO (e.g., from Steve Ballmer to Satya Nardella at Microsoft), have differential effects on this phenomenon?

In sum, TBA allows IB researchers to take advantage of increasing access to large volumes of structured and unstructured text to examine previously unanswered research questions. TBA is also able to generate new insights by providing a fresh perspective on existing knowledge about a phenomenon. In addition to the sources cited in the preceding section of our article, Appendix 2 lists additional resources for those interested in exploring and potentially applying TBA in their own work.

4. Conclusions

Given the diversity in disciplinary, theoretical, and conceptual bases in IB, research methodologies are critical in driving theoretical advancements (Joh, 2015; Hennart & Sutherland, 2022; Teagarden et al., 2018). Our article described two underutilized approaches that can help IB scholars expand their methodological toolkits. While both of these approaches can, by themselves, be sufficient to answer interesting and impactful questions, researchers can derive even more value by using these methods simultaneously—for example, by using a field experimental design to gather data in the form of text (e.g., reports, memos, emails), and subsequently analyzing them using TBA. Moreover, although we believe that these approaches have great potential, they are certainly not “methodological silver bullets” that will help solve all methodological challenges in IB research. However, we hope our article will result in the greater use of these methodologies to further advance IB theory.

Data availability

The Amazon dataset as well as an additional dataset (MetLife CEO letters to shareholders) is available from the authors upon request.

Supplementary materials

Supplementary material including the complete annotated R code (Appendix A) and example output and figures from our analysis (Appendices B and C) can be found in the online version, at doi:10.1016/j.jwb.2023.101463.

Appendix 1. Additional resources on field and quasi experiments


Appendix 2. Additional resources on text-based analysis


References
