COLIN I.S.G. LEE

Big Data in Management Research

Exploring New Avenues

Big Data in Management Research:

Exploring new avenues

Big Data in Management Research

Exploring new avenues

Big Data in management onderzoek: Verkenning van nieuwe perspectieven

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on
Friday, March 11th, 2016 at 13:30 hours

by

Colin Idzert Sarkies Gregory Lee
born in Utrecht

Erasmus University Rotterdam

Doctoral Committee

Promotor: Prof.dr. S.J. Magala

Other members: Prof.dr. Y. Baruch
Prof.dr. D. Dierendonck
Prof.dr. P.G.W. Jansen

Copromotor: Dr. W.A. Felps

Erasmus Research Institute of Management – ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ERIM Electronic Series Portal: http://hdl.handle.net/1765/1

ERIM PhD Series in Research in Management, 365
ERIM reference number: EPS-2016-365-ORG

Cover art: Tommy Kleyn www.blueboost.nl
Design: B&T Ontwerp en advies www.b-en-t.nl

This publication (cover and interior) is printed by haveka.nl on recycled paper, Revive®.
The ink used is produced from renewable resources and alcohol free fountain solution.
Certifications for the paper and the printing production process: Recycle, EU Flower, FSC, ISO14001.
More info: http://www.haveka.nl/greening

Acknowledgements

A huge number of people have helped make this dissertation possible. First of all, I would like to thank my mentors. Sławek Magala, thank you for helping me recognize and pursue my interests in research and in teaching; for making me relax every once in a while and putting things in perspective. Will Felps, thank you for the many hours that you have invested in teaching me the ropes and the patience and flexibility you showed in letting me find my own way. When discussing your input with fellow PhD candidates I often downplayed the extent of our collaboration, but I would still make most of them jealous. In extension, I would also like to thank Eliza Byington. Eliza, thank you for the massive amount of feedback you have provided over the years, for cheering me on, and for encouraging me to keep going and improving. I also want to thank Pretty Liem, once again. Your belief in me has been immensely empowering and I often wonder where I would have been without your support.

I would also like to thank my co-authors. Yehuda Baruch, it has been an honor to work with you and your work ethic has been a true inspiration. Flavius Frasincar, thank you for your enthusiasm and immensely valuable counseling. I would also like to thank the co-authors with whom I have conducted research that has not become part of this dissertation, but from whom I have learned a great deal. These include Maria Markoulli, Torsten Biemann, and Robbie Finch. Lastly, Piers Steel, thank you for giving me the room to develop my own research agenda, hopefully I will soon be able to join you in Calgary.

I was fortunate in finding several outstanding companies and institutions willing to provide their support. Many thanks to those who gave me access to ATS data, including the people at Connexys/NuWerk, TextKernel, and AIAS. Many of you deserve special mention, but I am afraid I can only mention some. To start, I would like to thank Ivo Tanis. None of this would have been possible if I hadn't bumped into you on the subway. Hopefully I can finally take you up on that dinner now that this dissertation has been finalized. Special thanks also go to Guus Meijer, who gave me the freedom to engage in this project from my own angle and on an academic timeline. Hopefully the results have made it worth the wait. Many thanks to Lia Dreef and Sander van Lingen for the time they spent on compiling the data. At TextKernel I would like to extend my gratitude to Jakub Zavrel, for helping me leverage the immense ATS database using the powerful TextKernel tools. I also want to give special mention to Doris Hoogeveen who has been a tremendous help in the process, even after moving to Melbourne.

Many thanks to the people at CWTS in Leiden. Thank you for supporting me in integrating tools and ideas from distant disciplines in my research. And special thanks to Ludo Waltman and Nees Jan van Eck, for the tons of advice that you have provided along the way, plus for pointing out to my supervisor that it would be wasteful to use me as coding stock. Much appreciated!

Thank you to the people at the University of New South Wales, in Sydney, where I spent my research visit, including Anton, David, Esha, Jeff, Jane, Lee, Maria, Mirjam, Tobias, Xuefeng. You immediately made me feel at home. Also many thanks to the people at the VU and UVA for letting me drop-in every so often and giving me advice on my work including Brian, Carolin, Chen, Evgenia, Paul Janssen, Svetlana Khapova, and the entire Eduworks Group; I hope to keep in touch. And Brian, just so that you know, I hope to visit you in Iceland sometime.

From the Erasmus University I want to thank Jasper op de Coul for his selfless guidance, advice, and support while I learned how to program and Cécile Reulen for her council as I learned about careers and HR. I would like to thank ERIM and the Erasmus Trust Fund for their kind support and I thank Patrick, Finn, Tineke, Olga, Natalija, Miho, Marisa, Elaine, and Kim for their help. I would also like to thank Babs, Dicea, and Kelly for their support in administration and their general advice.

I would also like to thank my former colleagues from the Rotterdam School of Management. I count myself lucky to have shared an officice with two brilliant minds, Julija and Giorgio. I will look back at our time in the office with great fondness. Julija we shared an office for a long time and I wouldn't have missed it for the world. You made me look forward to coming into the office. You would always have a solution ready, from simply clearing the mind by enjoying a beer in the sun to perhaps the most elaborate relocation schedule ever created. Giorgio, your concern for others is incomparable and your erudition immeasurable. Like our supervisor you are a true intellectual, a rare breed that should be cherished. Visit me in Calgary if you want another workshop in how to play Super Mario Cart (64 not SNES of course).

Thank you to the many more friends and colleagues from Rotterdam. Basak and Tim, thank you for your support throughout the last couple of years. Ning, Irina, Misagh, Pouriyha, Jorien, Merlijn, Pierre and others I look back at our “MPhil period” with great fondness. I am also thankful for the many coffees, cakes, pie’s, lunches, and drinks I got to share with Anne, Burcu, Dirk, Gabriela, Gijs, Inga, Janine, Jasmien, Jasmien, Juup, Hendra (see you in Canada?), Kate, Lisanne, Maartje, Marja, Meir, Murat, Nathan, Nüfer, Philip, Saskia, Shira, Steffen, Teodor, Tobias, Yingjie and many more.

Thanks to my old friends Andre, Martijn, Tamar and Thom and thank you to my new friends from the “vriendengroep”. Thank you to my family from the UK, for letting me stay over when I needed to and for giving me advice and support. Thank you to the Van der Schansjes and Breederlands whom so warmly welcomed me into their family. And especially to Tommy Kleyn for designing the amazing front cover.

The last few lines I have kept for those dearest to me and those who’ve had most to endure as I finished my doctorate. This is true for my brothers and especially for my parents. You were always there for me, helping me in whatever way you could. I will never be able to express my appreciation to you in full. From now on, I’ll at least make sure to pick up my phone whenever you call.

Finally, Astrid. My pride and joy. I like to believe I have done some memorable things in my life, such as writing this doctorate, but I know that convincing you to stick with me has been my greatest achievement by far. You’re the most cheerful, kind, and fun person that I have ever met. Thank you for putting up with my grumpiness, for always being there for me, for knowing how to make me smile, and never complaining. I’m most grateful that we met and know that we’ll be good, whatever happens. Thank you.

Colin I.S.G. Lee,

January 2016, Rotterdam

Table of Contents

List of Tables

List of Figures

Chapter 1: General Introduction

"Big data" has become a hot topic. The first documented use of the term Big Data, with clear reference to its current meaning, dates back to 1997 (c.f., Diebold, 2012; Press, 2013). Two NASA scientists, Michael Cox and David Ellsworth, described the challenge that some visualizations posed for computer systems at the time: "[…] data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk" and added "We call this the problem of *Big Data*" (M. Cox & Ellsworth, 1997, p. 1, emphasis in original). In 2001, Doug Laney published a redefinition of the term, describing Big Data as data that is of high volume, variety, and velocity (Laney, 2001) where 1) *volume* refers to the size of the data, 2) *variety* refers to the number of different data types, variables, and levels the data has, and 3) *velocity* refers to the rate at which the data is entered into the database or storage system (with *veracity* later added by Zikopoulos et al., 2013). In subsequent years, interest in Big Data rose; until, in 2011, there was an explosion in the number of news headlines (e.g., Google, 2015) and academic publications on the topic (e.g., Thomson Reuters, 2015), plus a surge in interest from industry (e.g., Bloomberg Businessweek, 2011; IBM, 2011; Manyika et al., 2011).

The rising interest in Big Data can be attributed to developments in technologies that capture data digitally (Castells, 2011; Wenzel & van Quaquebeke, in press). Enhancements in network speed and network connectivity (OECD, 2013) plus intensified research and development on small, energy-efficient computer processors (e.g., Hoque, Siekkinen, & Nurminen, 2014) have led to the invention of a range of products that have made digitalization more pervasive. Prototypical inventions, in this respect, are web 2.0 applications (i.e., websites that emphasize user-generated content and interoperability, such as social media websites; DiNucci, 1999; O'Reilly, 2005), smart mobile devices (e.g., smartphone, tablet, and smart watches), cloud computing systems (i.e., network access to a shared pool of configurable computing resources […] that can be rapidly provisioned and released; Mell & Grance, 2011, p. 2), and products that are part of the Internet of things (e.g., TVs, lamps, refrigerators, weighing scales, or baby monitors, with embedded systems connected to a network or to the internet; Yan, Zhang, Yang, & Ning, 2008). These products have greatly extended the variety of actions and processes that are computer-based or computer-mediated. The ubiquity of these products, the emphasis on user- and devicegenerated data, and advancements in our ability to capture and store that data centrally, provide a formula for high-magnitude datasets (Wenzel & van Quaquebeke, in press). Furthermore, these developments have accelerated the rate at which individuals leave digital traces as a by-product of their actions (Dutta & Mia, 2009; Giles, 2012).

For academia, Big Data could provide many new opportunities. Big data promises more generalizable analyses (e.g., Fan, Han, & Liu, 2014), research on questions that could not be addressed before (e.g., Sarcevic et al., 2012), and conclusions or results delivered at a high rate or even continuously (e.g., Antenucci, Cafarella, Levenstein, Ré, & Shapiro, 2013; Bruns, 2013; Hogenboom, Hogenboom, Frasincar, Schouten, & Meer, 2012). However, scholars can experience obstacles in utilizing Big Data. First, working with Big Data can provide challenges in terms of the skills required. For instance, scholars might need to have knowledge of database software. Conventional statistical software (e.g., SPSS) might not be able to process the data (e.g., Madden, 2012). Alternatively, since Big Data is often unstructured (Troester, 2012) and frequently contains inconsistent semantics (Bizer, Boncz, Brodie, & Erling, 2012; Kaisler, Armour, Espinosa, & Money, 2013), scholars might need to be able to engage in semantic analyses or computer programming in order to restructure or clean data (e.g., Dasu & Johnson, 2003). Second, Big Data also tests the limits of conventional resources used for data collection, storage, processing, and analysis. A personal computer does not always suffice when working with Big Data and more powerful computers or servers may be required (e.g., De Witte et al., 2013; Jacobs, 2009; D. Singh & Reddy, 2014). In fact, according to some, if the data can be processed on the average personal computer, it should not be considered Big Data (e.g., Dumbill, 2013; Manyika et al., 2011; Turnbull et al., 2010). Lastly, Big Data also challenges some of the entrenched norms in research. For instance, conventional standards with respect to data cleaning, such as a reliance on manual coding, might be replaced with a combination of rule-based data-cleaning (Dasu & Johnson, 2003; Tang, 2014), requirements regarding the extensive or supplementary reporting of methods or programming code (c.f., Bruns, 2013, p. 4)., and possibly a higher tolerance for noise or uncertainty (Mayer-Schönberger & Cukier, 2013).

Taken together, the novel skills, extensive resources, and the tension between Big Data research and conventional research practices can provide a barrier to Big Data research and can thus stand in the way of the optimal use of Big Data opportunities. Embracing Big

Data might be especially challenging for management research. Management research does not have the data-oriented tradition that is found in other business disciplines such as logistics, marketing, and finance (Manyika et al., 2011; Wenzel & van Quaquebeke, in press). While managers and companies seem intent on jumping on the Big Data band-wagon (e.g., FreshMinds, CapGemini, & EMC, 2015), few management scholars have the capabilities and resources required to respond (Wenzel & van Quaquebeke, in press). Indeed, courses on research methods in doctoral programs of management schools rarely touch on Big Data skills (Putka & Oswald, 2015). This suggests that the management scholarship community needs to learn how to leverage Big Data opportunities or risk being superseded by practitioners or other business disciplines (George, Haas, & Pentland, 2014). Up to now, however, relatively little has been done to explore the various ways in which Big Data can be used in management scholarship. As argued by Wenzel and Quackebecke (in press), discussion of Big Data in the management field has hardly surpassed deliberations on commercial value (e.g., H. Chen, Chiang, & Storey, 2012) or privacy (e.g., Lyon, 2014).

In this dissertation, I describe several projects in which different Big Data opportunities are explored. Specifically, I demonstrate some of the tools that can be used when working with Big Data. I not only focus on the use of Big Data in research, but also describe how some of the tools and principles behind Big Data can be used in order to change the way in which we disseminate research so as to make science more efficient and effective.

Dissertation Overview

This dissertation provides illustrations of how Big Data can be productively used in management research. Together with a number of colleagues, I engaged in a range of projects that leverage Big Data opportunities. I present a collection of the three most crucial studies in this endeavor. These projects, presented in chapter two through four, have been described in separate papers and can also be read as such. Collectively, however, they show the range of different opportunities that the Big Data phenomenon can provide. Although I was the primary author on each of these projects, I will hereafter use the first person plural "we" instead of the singular form "I", so as to acknowledge my collaborators' valuable contributions. As per the doctoral regulations, the contributions from my coauthors are described after the outline of the dissertation.

Dissertation Outline

In chapter two, we use semantic tools and visualization techniques to provide an overview of an academic field, focusing on the field of career studies in particular. People's careers are determined by multiple factors at various levels. At the individual level, apart from the obvious demographics such as age, gender, race, and nationality, careers are influenced by personality differences, values, attitudes, education, and training, to name just a few. At the organizational level, careers are affected by strategies and structures. At the societal level, careers are affected by industry characteristics, labor market, economy, and national culture. Reflecting this reality, careers research is highly diverse. Since much can be learned by studying careers from a variety of perspectives (Arthur, 2008; Khapova & Arthur, 2011), this diversity has its benefits. However, to the extent that scholars fail to recognize relevant research from other disciplines or subfields, this diversity can also lead to fragmentation. Using advanced bibliometric mapping techniques (van Eck & Waltman, 2010; Waltman, van Eck, & Noyons, 2010), we provide a systematic review of the 3,141 articles on careers published in the management literature between 1990 and 2012. We (1) map key terms to create a systematic taxonomy of career studies within the field of management studies, (2) provide a synthetic overview of each topic cluster which extends prior reviews of more limited scope, and (3) identify the most highly influential studies on careers within each cluster. To classify a broad range of research opportunities for career scholars, we also create a "global" map of 16,146 career articles from across the social sciences and compare and contrast the local career studies research with the global research on careers.

In chapter three, we describe an endeavor that could allow continuous, detailed insights to be attained from the literature, with low levels of supervision. We present a software platform called "ReNotate", which captures the highlights and annotations of people reading academic publications in order to allow that data to be searched and queried. The chapter first presents the basic tenets of the platform. We then go into detail on the algorithms required to aggregate the data from readers. Afterward, we present three possible future extensions of the platform, including (1) the use of ReNotate for attaining evolving taxonomies of the concepts in the literature, (2) the use of ReNotate in classroom exercises, and (3) the use of ReNotate in making data coding for meta-analytic research easier.

In chapter four, we address the fundamental, yet understudied question: On what basis do hiring professionals make their decisions about whom to invite for a job interview? In order to answer this question, we apply Big Data analytical techniques to Applicant Tracking System (ATS) data from 48 companies (N = 441,769 applicants). ATSs are software and database infrastructures that allow posting, tracking, and storing of vacancy and applicant data (including resumes or CVs and contact information), monitoring applicants during the applications process (viewing assessment results and interview memos), and communicating decisions to applicants and other recruiting staff. We use synthetic validity (Lawshe, 1952; Steel, Huffcutt, & Kammeyer-Mueller, 2006) and relative weight analysis (Tonidandel & LeBreton, 2010; Tonidandel, LeBreton, & Johnson, 2009) to identify the relative importance across occupations and industries of 18 factors relating to the applicant (i.e., demographic factors, work and education experience), the vacancy (i.e., relevance of applicant education, skills, and prior experience), the company (i.e., external or internal applicant), the applicant pool (i.e., number of applications received), and the labor market (i.e., occupation vacancy rate). Building synthetic validity equations on a training sample of the first 90% of the applicants and testing them on the remaining 10% of the sample, we were able to correctly predict invitation outcomes for 30,415 out of 44,187 applicants (68.8%).

The final chapter provides a brief summary of the contributions and main findings of the projects described. I close with a general discussion of the lessons learnt, focusing on what the key Big Data opportunities and pitfalls might be for management scholars and management scholarship.

Contributions

Here, I will list the contributions I provided to each of the chapters in this dissertation and acknowledge the input provided by co-authors and others. I conducted the majority of the work for all chapters and wrote the introduction (i.e., chapter one) and conclusion (i.e., chapter five) independently.

For the second chapter (i.e., “Toward a Taxonomy of Career Studies”), I collected and processed the data. Data coding was conducted in collaboration with Dr. Will Felps. I completed the majority of the writing, but received substantial and valuable input from both Dr. Will Felps and Prof.dr. Yehuda Baruch. Eliza Byington and Sherry Sullivan provided extensive friendly reviews. Ludo Waltman, Nees Jan van Eck, and Cathelijne Waaijer provided advice on the creation of science maps. A version of the second chapter has been published in the Journal of Vocational Behavior (C. I. S. G. Lee, Felps, & Baruch, 2014b) and a condensed version featured in the Academy of Management Proceedings (C. I. S. G. Lee, Felps, & Baruch, 2014a). The research was supported in part by a grant from Trustfund Erasmus University Rotterdam.

I conducted most of the work for the third chapter (i.e., “Collaborative Tagging of Academic Articles”) independently. This work included software and web programming and the creation of an ontology. Dr. Will Felps gave valuable input on the writing and Dr. Flavius Fransincar provided excellent comments and edits. Additional advice was given by Jasper op de Coul from the Erasmus University. A version of this chapter has been submitted to the Academy of Management conference for 2016.

For the fourth chapter I received support from several companies in data collection and coding. The data, owned and released by 48 different companies, was attained the servers of Connexys (www.connexys.nl), a distributor and manager of Applicant Tracking Systems. TextKernel (www.textkernel.nl) conducted the extraction and parsing of résumé data. I conducted all other data coding and analysis independently. I completed the majority of the writing with some valuable input provided by Dr. Will Felps. Eliza Byington gave an extensive friendly review. Prof. dr. Piers Steel gave additional advice on the use of synthetic validity. A version of this chapter has been submitted to the Academy of Management conference for 2016.

Chapter 2: Toward a Taxonomy of Career Studies through Bibliometric Visualization[1]

Abstract

One of the greatest strengths and liabilities of the career field is its diversity. This diversity allows for wide coverage of relevant career dynamics across the lifespan and across levels of analysis. However, this diversity also reflects fragmentation, with career scholars failing to appreciate how the insights from other thought worlds can advance their own work. Using advanced bibliometric mapping techniques, we provide a systematic review of the 3,141 articles on careers published in the management literature between 1990 and 2012. In doing so, we (1) map key terms to create a systematic taxonomy of career studies within the field of management studies, (2) provide a synthetic overview of each topic cluster which extends prior reviews of more limited scope, and (3) identify the most highly influential studies on careers within each cluster. Specifically, six local clusters emerged – i.e., international careers, career management, career choice, career adaptation, individual and relational career success, and life opportunities. To classify a broad range of research opportunities for career scholars, we also create a "global" map of 16,146 career articles from across the social sciences. Specifically, six global clusters emerged – i.e., organizational, individual, education, doctorate careers, high-profile careers, and social policy. We describe and compare the clusters in the map with an emphasis on those avenues career scholars in management have yet to explore.

---

[1] A version of this chapter has been published in the Journal of Vocational Behavior and a condensed version was published in the Academy of Management Proceedings:

Lee, C. I. S. G., Felps, W., & Baruch, Y. (2014a). Mapping career studies: A bibliometric analysis. *Academy of Management Proceedings*, *2014*. http://doi.org/10.5465/AMBPP.2014.284

Lee, C. I. S. G., Felps, W., & Baruch, Y. (2014b). Toward a taxonomy of career studies through bibliometric visualization. *Journal of Vocational Behavior*, *85*(3), 339–351. http://doi.org/10.1016/j.jvb.2014.08.008

Introduction

Career studies is an active area of inquiry which cuts across a variety of disciplines, domains of work, and levels of analysis. CEOs, surgeons, politicians, actors, scientists, and line workers all have careers, where a "career" is defined as an "evolving sequence of work experiences" (Arthur, Hall, & Lawrence, 1989a, p. 8). The career literature draws from and contributes to a variety of disciplines, with clear links to management, psychology, sociology, economics, and education. This diversity of research reflects the reality that peoples' careers are determined by multiple factors at various levels, with various stakeholders involved.

Despite a plethora of strong scholarship, the very virtue of conceptual and methodological diversity within career studies is also an obstacle to accretive progress. Critics have pointed out that the career field includes various disjointed research streams, which do not systematically connect to each other even when they concern the same phenomena (Arnold & Cohen, 2008; Arthur, 2008). This suggests that serious attention should be paid to evaluating the extent to which career findings from across research communities might be integrated (Arthur, 2008; Arthur et al., 1989a; Gunz & Peiperl, 2007). Importantly, we do *not* advocate reducing the variety of approaches or presenting a single over-arching theory on careers. Instead, we posit that a systematic taxonomy of career studies, which groups together related concepts, would go a long way toward providing the intellectual architecture for bridging thought worlds. In other words, we believe that it is possible to make accretive intellectual progress and retain diversity, but only within a framework that facilitates the exchange of concepts, methods, and findings across streams.

The purpose of this study is to present a systematic overview of the career literature using the bibliometric technique known as *science mapping*. Science mapping offers a visualization of the relationships between scientific objects, such as research topics (Callon, Courtial, Turner, & Bauin, 1983). We employ science mapping to create local (i.e., management) and global (i.e., all of social science) maps of the research topic clusters of peer-reviewed journal articles on careers. This allows us to provide a structured overview and synthetic taxonomy of the state of the career literature in management and in the social sciences more broadly.

Reviewing the Literature through Science Mapping

Several past reviews provide a high quality overview of the streams of career studies within management (e.g., Baruch & Bozionelos, 2010; Feldman, 1989; Sullivan, 1999). In addition, a small number of reviews go beyond the field of management (e.g., Maranda & Comeau, 2000; Özbilgin & Tatli, 2011). In both cases, however, the reviews rely on the authors' subjective view of the field. Other notable contributions to understanding career studies come from edited volumes (Arthur, Hall, & Lawrence, 1989b; Gunz & Peiperl, 2007; Inkson & Savickas, 2013a, 2013b, 2013c, 2013d). These collections let the structure of the field emerge from the input of several selected scholars. Still, such contributions are dependent upon how the scholars are selected and on the idiosyncrasies of these scholars' sensemaking. In an effort to achieve standards of rigor comparable to primary researchers (Cooper, 1989; Fitzgerald & Rounds, 1989, p. 106) we use science mapping to conduct a comprehensive analysis of thousands of studies on careers and develop a *taxonomy* of the career literature. A taxonomy is a classification based on *empirical* evidence of correspondence between characteristics (Bailey, 1994). Taxonomies can be contrasted with typologies, which are based on the comparison of *conceptual* categories across more than one conceptual dimension (c.f., Baruch & Bozionelos, 2010). As such, our scientific mapping approach reveals topics and relationships between topics that have been absent from others' frameworks.

Science mapping employs innovative bibliometric techniques to create visual representations of academic research. Much like an architectural drawing, maps of academic literatures can help create shared understanding and bridge diverse knowledge domains (Carlile, 2002). Science maps can help scholars with highly specialized knowledge overcome barriers to discussion and collaboration across disconnected research communities, which can advance theoretical and conceptual progress (Börner, Boyack, Milojevic, & Morris, 2012; Fitzgerald & Rounds, 1989, p. 107; Rafols, Leydesdorff, O'Hare, Nightingale, & Stirling, 2012). Finally, science maps provide a tool for creating synthetic reviews and complement meta-analyses. Indeed, whereas meta-analyses focus on a particular research topic, science maps have the capability to zoom out further, and empirically capture the relationships between multiple topic areas.

Science mapping has been around for several decades. However, early science mapping generally relied on manual coding of articles (e.g., Fitzgerald & Rounds, 1989). This introduced evaluative criteria into the process and was laborious; limiting the number of articles that could be incorporated into a map. Recent advances in information technology enabled the introduction of text extraction and normalization techniques for science mapping, which has removed these limitations. Within the field of management, novel mapping techniques have been used to visualize research in areas like international management (Acedo & Casillas, 2005), strategy (Ramos-Rodríguez & Ruíz-Navarro, 2004), and business ethics (Özmen Uysal, 2010). While these science maps are still relatively rare, those published have been well received within the academic community.

Methods

As stated above, we created maps of "local" (management) and "global" (social science) career topics. Our sample for both maps includes all articles published between 1990 and 2012 containing the term "career" (plus any suffix) in the title or abstract. Studies published before 1990 were excluded because these rarely have an abstract in the Web of Science database. Studies after 2012 were left out in order to prevent preprint bias. The local map is based on all journals listed under the management category in the Web of Science (hereafter WoS) or under the Harzing (2013) category "Organization Science/Organization Behavior, Human Resource Management/Industrial Relations". The search resulted in 3,141 publication abstracts and titles related to careers within management journals. For the global map, we obtained the articles published in any social science journal from the WoS containing the term "career" in the title or abstract. Articles that used the term "career" in ways that were unrelated to work (c.f., Arthur et al., 1989a, p. 8) were excluded by discarding those containing terms that signify non-work careers – i.e., the terms "treatment careers", "drug", "abuse", "diagnosis", "illness", "cancer", or "AIDS". This article identification process resulted in 16,146 relevant journal articles in social science journals. The interested reader can find both search phrases in Appendix 2-A.

We created the science maps using the *VOSviewer* software developed by van Eck and Waltman (van Eck & Waltman, 2011; Waltman et al., 2010). Various studies have successfully utilized the VOSviewer software for science mapping, including visualization of the research on renewable energies (Rizzi, van Eck, & Frey, 2014), the relations between journals in the business field (Rafols et al., 2012), and the topics covered in the editorials of Nature and Science (Waaijer, van Bochove, & van Eck, 2010, 2011). Indeed, VOSviewer combines the most advanced and valid techniques for every step in the science mapping process, including: (1) term extraction and selection, (2) visual mapping of relatedness, and (3) clustering of science objects. Across our maps, we use the default settings in the software, which generally represent the best practice in the science mapping literature.

In the first step of the process, *noun phrases* (i.e., groups of nouns and preceding adjectives) that occur in the abstract or title of at least 10 different documents are extracted. This method is effective for the extraction of technical terminology, regardless of the domain of the text (Justeson & Katz, 1995). The next step is to remove generic noun phrases like "research" or "method" using the Kullback-Leibler distance (van Eck & Waltman, 2011, p. 2). Such phrases co-occur indiscriminately across the corpus and thus are not helpful for distinguishing specific topic areas. In addition, drawing from the conceptual scheme developed by Brinberg and McGrath (1982) we manually coded terms into six categories: concept, data collection, data analysis, substantive actor, substantive industry, and substantive geography. Terms that do not denote a concept, a method, or a sample are removed (e.g., *Elsevier*, *Academy*, and *Administrative Science Quarterly*). This term-identification process produced 400 terms for the local map and 1,780 terms for the global map.

After computing the relevance and coding the terms, the relatedness of the terms was determined using the *association strength measure* (Rip & Courtial, 1984; van Eck & Waltman, 2009). The association strength between two terms is the ratio between the observed number of co-occurrences of two terms and the expected number of co-occurrences of the two terms. A key technical advantage of the association strength measure over alternative measures of (dis-)similarity (e.g., *Jaccard index, cosine*) is in the way it corrects for the number of times objects occur (Luukkonen, Tijssen, Persson, & Sivertsen, 1993; van Eck & Waltman, 2009; Zitt, Bassecoulard, & Okubo, 2000). The association strength values were used as input for the "Visualization Of Similarities" (VOS) mapping technique, which creates a two-dimensional representation of term relatedness (van Eck, Waltman, Dekker, & Van den Berg, 2010). The VOS mapping technique has been shown to be especially effective in a) dealing with large numbers of null values (which is typical of term co-occurrence data) and b) big differences in frequency of term occurrence (van Eck et al., 2010). The distance between any two terms reflects their relatedness, such that terms located close together tend to occur in the same article abstracts and titles relatively frequently. This also means that terms at the center of the map co-occur with a wider range of terms than terms at the periphery of the map.

The last step is to assign terms to *clusters*. The VOS clustering technique is based on the same assumptions as the VOS mapping technique (Waltman et al., 2010). Terms that are strongly associated are placed in the same cluster, and colored accordingly. As such, the VOS clustering technique provides the basis for an emergent taxonomy of the literature.

Results

The results of our analysis are presented visually in Figure 2-1 and 2-2. In each map, terms that occur more frequently are presented as larger than terms occurring less frequently. The local map (Figure 2-1) shows the terms divided into six clusters, identifiable by the different colors. To gain a more refined view and dynamically explore the figure in more detail, the reader can access the map via the following link: http://bit.ly/CareerLocal. The global map (Figure 2-2) shows a visual representation of the career literature across the social sciences and also contains six clusters. This map, including the terms with smaller nodes, can be explored at: http://bit.ly/CareerGlobal.

In analyzing these maps, we will discuss the most prominent areas of research, selectively review developments in each area, and compare our maps to prior classifications, noting points of overlap and departure.

Local Map: The Career Field within Management

The map in Figure 2-1 provides a descriptive overview of the term clusters in career studies in management. In order to provide insight into the clusters, we carefully reviewed the top 50 most highly cited publications of which (a) more than 50% of the terms are located in one cluster (see Table 2-1 for the top five per cluster), and (b) there were at least two terms from the article's title/abstract. Using the terms from the map as a guide, we identified recurring themes.

International careers

The international careers cluster (cyan) is the smallest cluster on the map. It focuses on the career effects of international assignment (both organizational- and self- initiated). This includes questions regarding motivations to engage in international assignment (Dickmann, Doherty, Mills, & Brewster, 2008), benefits of international assignment (Jokinen, Brewster, & Suutari, 2008), and repatriation (Suutari & Brewster, 2003).

In the recent literature in this cluster, many studies have looked at the aims and expectations of both parties in the international assignment. Evidence suggests that career benefits of expatriate assignment are conditional on the alignment between the motivations of the organization, the characteristics of the assignment, and the expatriate's career orientation (Dickmann et al., 2008; Jassawalla & Sashittal, 2009; Richardson, McBey, & McKenna, 2008). How precisely to align these three factors has been considered in a number of recent studies (Bolino, 2007; Cerdin & Pargneux, 2009; Haslberger & Brewster, 2009).

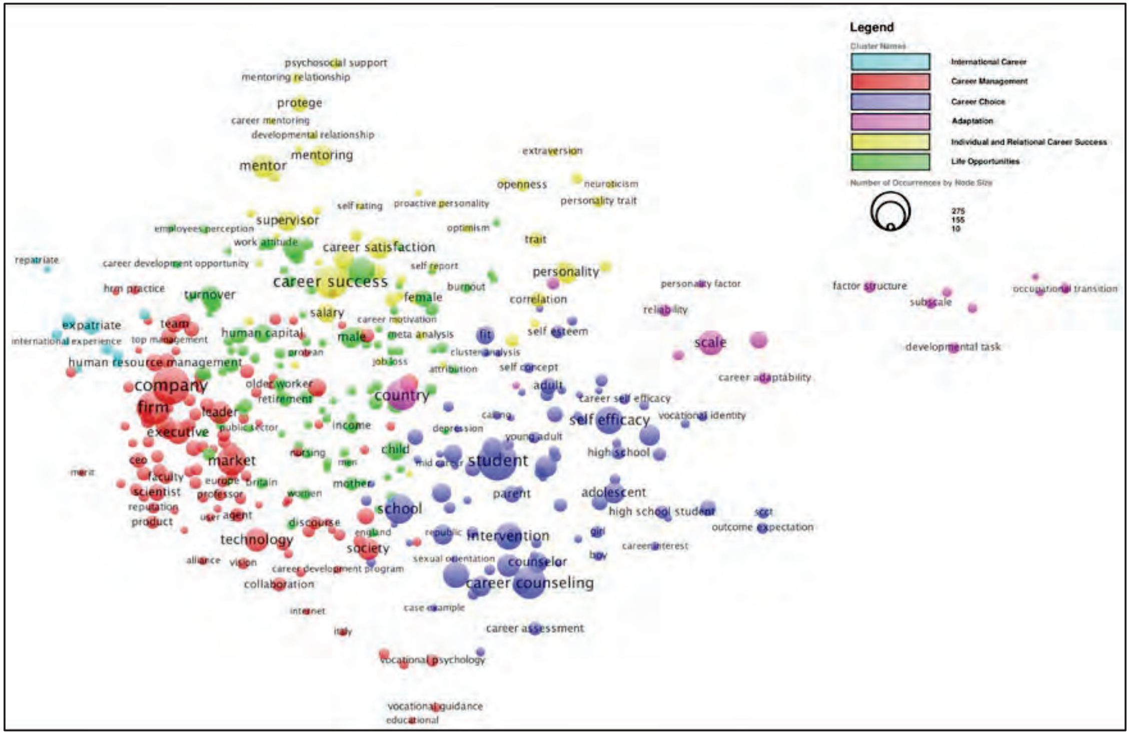


Figure 2-1. Map of the Management Literature on Careers

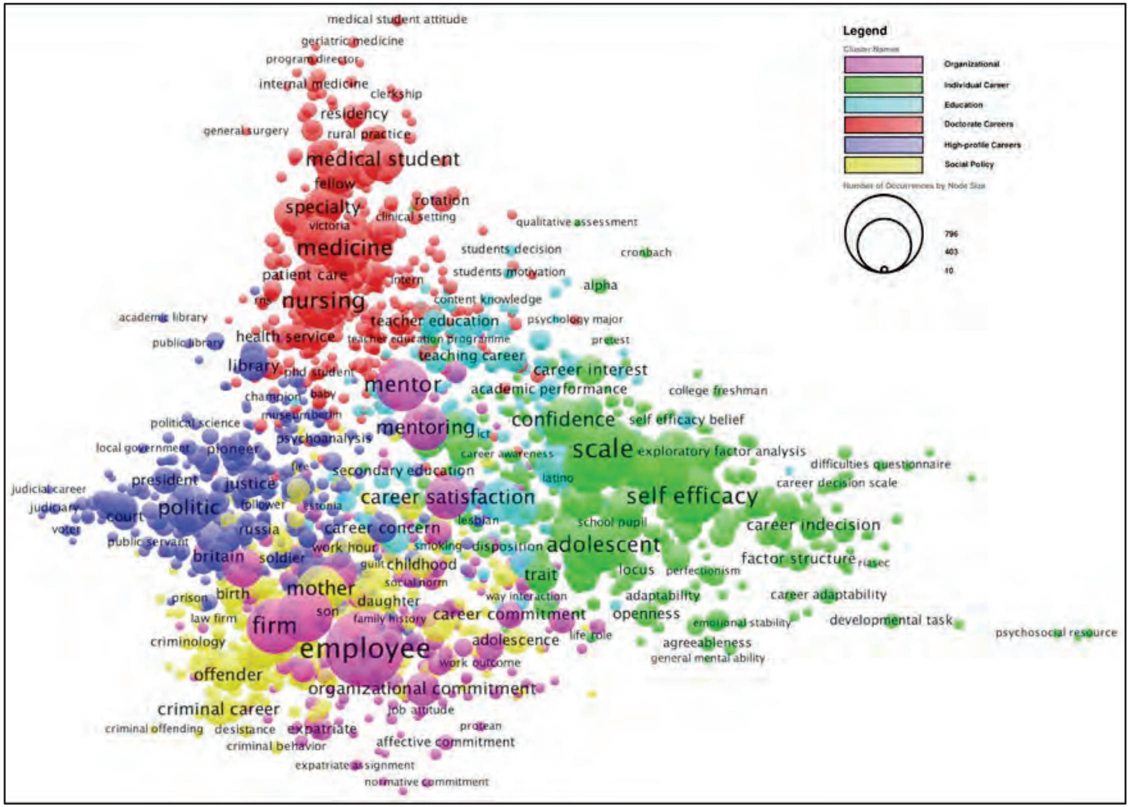


Figure 2-2. Global Map of the Career Literature Across the Social Sciences

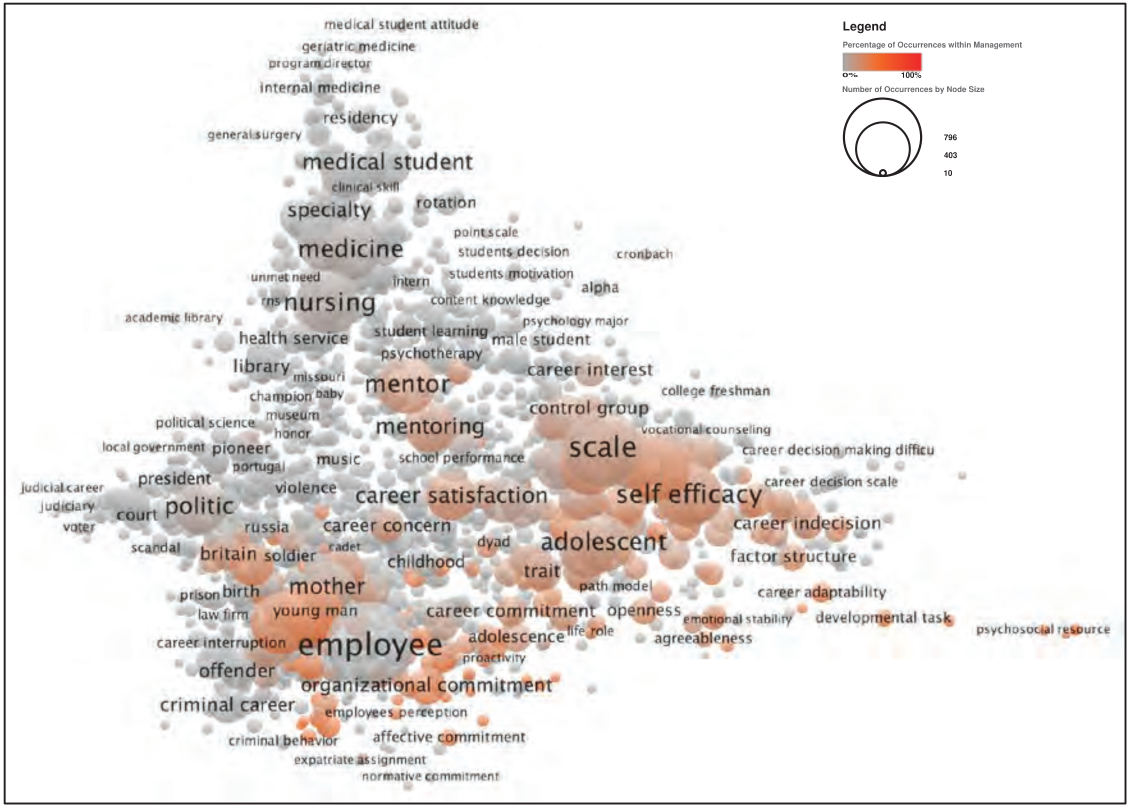


Figure 2-3. Overlay Map Indicating the Position of the Management Literature According to the WoS Classification

Table 2-1. Overview of the Clusters in the Management Literature on Careers

Cluster	Definition	Prominent Papers[a] [b]
International Careers	Research focusing on the career effects of international assignment.	Bolino, 2007 Dickmann, Doherty, Mills, & Brewster, 2008 Jokinen, Brewster, & Suutari, 2008 Lazarova & Cerdin, 2007 Suutari & Brewster, 2003
Career Management	Research on the management of careers.	Bechky, 2006 Briscoe & Hall, 2006 Hedlund, 1994 Lee, Trauth, & Farwell, 1995 Sullivan & Arthur, 2006
Career Choice	Research on how (young) people choose their careers.	Duffy & Sedlacek, 2007 Hall & Chandler, 2005 Lent, Brown, & Hackett, 1994 Savickas, 1997 Schoon & Parsons, 2002
Career Adaptation	Research focusing on individuals' adaptation to change.	Brown, Bimrose, Barnes, & Hughes, 2012 Frese, Fay, Hilburger, Leng, & Tag, 1997 Koen, Klehe, van Vianen, Zikic, & Nauta, 2010 Savickas & Porfeli, 2012 van Vianen, Klehe, Koen, & Dries, 2012
Individual & Relational Career Success	Research on individual and relational determinants of career success	Allen, Eby, Poteet, Lentz, & Lima, 2004 Greenhaus, Parasuraman, & Wormley, 1990 Judge, Higgins, Thoresen, & Barrick, 1999 Ng, Eby, Sorensen, & Feldman, 2005 Seibert, Kraimer, & Liden, 2001
Life opportunities	Research on exogenous factors that constrain and enable people to achieve their career related goals.	Anderson, Coffey, & Byerly, 2002 Parasuraman, Purohit, Godshalk, & Beutell, 1996 Tharenou, Latimer, & Conroy, 1994 Thompson, Beauvais, & Lyness, 1999 Wang, Zhan, Liu, & Shultz, 2008

[a] Five selected papers from the 50 most highly cited per cluster with more than 50% of the extracted terms in the cluster. [b] References can be found in Appendix 2-B.

Career management

The literature on career management is wide, and covers several issues and perspectives. The main focus of the career management cluster (red) is on how individuals plan and manage their own career. A complementary perspective deals with the way organizations plan and manage the careers of their employees. Very few discuss the wider system at a national or global level (M. C. Higgins, 2005).

Within the individually oriented literature, a major focus is Arthur and Rousseau's (1996) boundaryless career theory. This theory was developed to address how people navigate their careers given changes in the nature of the economy and employee-organization relationship. While some wholeheartedly subscribe to this theory, others question its validity (Inkson, Gunz, Ganesh, & Roper, 2012; Rodrigues & Guest, 2010), and still others call for a balanced view (Baruch, 2006; Lips-Wiersma & Hall, 2007). Two other leading concepts are employability (Kanter, 1990) and the protean career (Hall, 2004). Linking theories of boundaryless careers, employability, and protean careers is the assertion that individuals must take proactive ownership of their careers given decreased job security in many sectors and countries at the same time that job opportunities have become more dynamic, diverse, and global (Arthur, 2008; Sullivan & Baruch, 2009).

Within the career management cluster, a complementary organizational-level perspective considers how HRM practices create internal labor markets that enable and constrain careers (Birdi, Allan, & Warr, 1997). These different HRM practices vary in their level of sophistication and level of involvement (Baruch & Peiperl, 2000).

Career choice

The career choice cluster (dark blue) pertains to how people choose their careers. Seminal in this cluster is Holland's occupational themes theory (1973). The theory distinguishes six personality types, each of which is suited for a different kind of work and occupational environment.

In the contemporary literature, the Social Cognitive Career Theory (SCCT) is studied particularly often. SCCT argues that people's career goals and career choices are a function of the interaction between their confidence in their success in a profession and the estimated benefits and costs associated with each occupation (Lent, Brown, & Hackett, 1994, 2000). Importantly, several scholars in this area contend that women, ethnic minorities, and those from lower socioeconomic status may be culturally conditioned to have artificially low expectations of success in several "advanced" occupations, which can be self-limiting (Correll, 2004; Lent et al., 2005; Spelke, 2005; Whiston & Keller, 2004).

Another major career choice theme is the role of career counselors. A striking feature of the recent research on counseling is that it explicitly considers the "whole person" in helping people to find a career they will enjoy, find meaningful, and have success within (R. A. Young et al., 2011). Relatedly, career counselors are urged to help individuals find their "calling" (Duffy & Sedlacek, 2007). Calling can be defined as "work that a person perceives as his [or her] purpose in life" (Hall & Chandler, 2005, p. 160). Recent research has considered the antecedents (Hirschi, 2011; I. Hunter, Dik, & Banning, 2010) and consequences of having a calling or searching for a calling (Duffy, Allan, & Dik, 2011; Duffy & Sedlacek, 2007; Elangovan, Pinder, & McLean, 2010; Hagmaier & Abele, 2012).

Career adaptation

The career adaptation cluster (purple), close to the career choice cluster, only recently developed into a succinct area within the local career literature. The literature in this area focuses on the adaptation of individuals to change. The core literature in this cluster originated from career construction theory (Savickas, 2002, 2005). This elaborate theory considers the causes and consequences of individuals' vocational self-concepts, which are constructed over the life-course as individuals attempt to adapt themselves and their environments in order to evolve toward occupational success and career satisfaction.

Prominent in this area is the research on the Career-Adapt Abilities Scale (CAAS). In a cross-national collaborative endeavor, a group of 29 scholars operationalized the "individual's ability to adapt" (i.e., career adaptability resources) as a hierarchical construct with four reflective components: *concern*, *control*, *curiosity*, and *confidence* (Savickas & Porfeli, 2012). The scale was tested in 13 different countries. The results provide considerable support for its validity and reliability.

Individual and relational career success

The literature on the individual-level and relational determinants of career success (yellow) include factors like personality (Seibert & Kraimer, 2001) and proactivity (Fuller, Jr. & Marler, 2009; Kim, Hon, & Crant, 2009). Relational determinants of career success include network structure (e.g., Seibert, Kraimer, & Liden, 2001) and relationship quality (e.g., Dutton & Ragins, 2007).

As an extension to the literature on the relational determinants of career success, the yellow cluster also contains literature on social support and mentoring (e.g., Eby, Allen, Evans, Ng, & DuBois, 2008). This body of work focuses on workplace mentoring in particular (rather than youth and academic mentoring) and the different types of mentoring relationships that one can have in and around the workplace (Carraher, Sullivan, & Crocitto, 2008; Parker, Hall, & Kram, 2008). In recent years, an increasing number of the mentoring studies have focused on the mentors themselves (O'Brien, Biga, Kessler, & Allen, 2010; Parker et al., 2008) and the effect mentoring has on the mentor's career (Lentz & Allen, 2009).

Another key focus of this cluster is the distinction between extrinsic/objective career success (such as salary and promotion) and intrinsic/subjective career success (such as career satisfaction and meaningful life)(Abele & Spurk, 2009; Heslin, 2005). An open question is the relative importance of each type of success (e.g., King & Napa, 1998).

Life opportunities

The life opportunities cluster (green) focuses on the relationship between one's career and larger life ambitions and trajectories. A number of studies focus on the way careers unfold as life progresses (Duffy, Dik, & Steger, 2011), and in particular, the relationship between work and family life (Valcour, Ollier-Malaterre, Matz-Costa, Pitt-Catsouphes, & Brown, 2011).

An important theme in this cluster is that one's demographic background and family characteristics influence career perceptions, progress, and power. There are unique career constraints associated with being divorced, a dual career couple, a parent, or aging (Hammer, Allen, & Grigsby, 1997; Parasuraman, Purohit, Godshalk, & Beutell, 1996;

Wang, Zhan, Liu, & Shultz, 2008). These constraints affect the individual's ability to accumulate human capital and can limit career success (Probert, 2005).

Global Map: The Study of Careers across the Social Sciences

In order to facilitate links with wider areas of career research we also provide a taxonomy of the field of career studies across the social sciences. To show the relationship between the local and the global map, we created an *overlay map* (Rafols, Porter, & Leydesdorff, 2010) which indicates the extent to which the management literature is related to other clusters in the global map. In this map, depicted in Figure **2-3**, the node of each term is colored according to the relative frequency of occurrence of the term in the management literature. The overlay map reveals that the management literature is most concentrated at the bottom and right of the global map, focusing on organizational and individual psychological factors associated with careers.

A striking finding from our analysis is that overall, less than one fifth of the career literature was published in management journals – i.e., 3,141 articles out of 16,146 articles. This implies that research by management scholars should benefit from the exchange of insights and findings with the broader social sciences (e.g., psychology, sociology, labor economics, political science, education, scientometrics, etc.) in order to fully understand career dynamics inside and outside organizations (Arthur, 2008). We develop the global map to systematically identify such opportunities.

To discover themes within clusters, we a) consider the top terms in each cluster (Table 2-2) and b) carefully review the 50 most highly cited publications of which more than 50% of the terms are located in one cluster (Table 2-3). In our discussion of the global clusters, we focus on scholarship that goes beyond topics and approaches usually considered by the management literature on careers. Similarly, we do not review the organizational cluster (in purple at the right lower half of Figure 2-2), since the associated literature was already covered in the local map discussion.

Individual

The individual cluster (green) has a number of characteristic themes. First, virtually all of the top 50 papers belonging to this cluster are resolutely individual. Their methodological designs use individual level surveys and variables. Indeed, this uniformly individual-level approach is the rationale behind labeling it the *individual* cluster. Three research issues dominate the articles belonging to this cluster: (a) the determinants of career choices, (b) the individual-level determinants of career success (e.g., personality), and (c) career counseling.

The majority of the articles associated with this cluster are published in psychology journals, although a substantial minority are in career journals (e.g., Lent et al., 1994; Savickas & Porfeli, 2012; Seibert & Kraimer, 2001; Tokar, Fischer, & Subich, 1998). The overlay map (Figure **2-3**) reveals that management career journals publish on these topics as well, to some extent. Moreover, this cluster has substantial overlap with research found in the local map's clusters on career choice (blue) and the personality component of the mentoring cluster (yellow).

Education

The education cluster (cyan) concerns the relationship between education and careers. There are two key streams within this cluster. The first stream is premised on the notion that educational factors serve as key mediators between a host of factors and ultimate career success. Some studies in this stream focus on psychological factors, like goals and self-efficacy (Bandura, Barbaranelli, Caprara, & Pastorelli, 2001; Harackiewicz, Barron, Tauer, & Elliot, 2002). Other studies adopt a sociological approach, explaining differences in academic achievement in terms of opportunity structures and barriers (Aschaffenburg & Maas, 1997; Else-Quest, Hyde, & Linn, 2010; Hillmert & Jacob, 2010; Hindman, Skibbe, Miller, & Zimmerman, 2010; Tyson, Castellino, & Darity, 2005). Of particular interest is the question of why certain demographic groups are less inclined to engage in science, technology, engineering, and mathematics (e.g., Archer et al., 2010; Else-Quest et al., 2010; Stout, Dasgupta, Hunsinger, & McManus, 2011), since these domains are often associated with rewarding career opportunities.

Table 2-2. Terms Associated with the Six Global Cluster of the Career Literature

Cluster	Concepts	Data Collection	Data Analysis	Substantive Actors	Substantive Industry/Work Environment	Substantive Geography
Organizational	Firm, Mentoring, career satisfaction, Turnover, organizational commitment	longitudinal design, self rating, longitudinal investigation, cross sectional design, archival data	structural equation modeling, structural equation, hierarchical regression analysis, LISREL, hierarchical multiple regression analysis	employee, mentor, executive, CEO, expatriate	company, multinational corporation, manufacturing, hotel, large organization	home country, Chinese context
Individual	career counseling, self efficacy, career decision, personality, confidence	scale, correlation, subscale, interest inventory, predictive validity	reliability, factor analysis, confirmatory factor analysis, factor structure, coefficient	client, college student, adolescent, counselor, high school student	career counselor, sport, career center, football, elite sport	Mexican American
Education	educational career, socioeconomic status, academic achievement, school career, teacher education	mixed methods study, control condition	structural equation model, structural model, chi square analysis, multilevel analysis, constant comparative method	girl, boy, female student, pupil, male student	high school, secondary education, teaching profession, chemistry, teacher education program	Asian American, urban school, Asian, Massachusetts, mainland China
Doctorate Careers	nursing, medicine, specialty, clinical practice, research career	citation, cross sectional survey, postal questionnaire, h index, action research	chi square test, beta, chi square, univariate analysis, thematic analysis	physician, medical student, resident, trainee, psychiatrist	medical school, psychiatry, health service, surgery, primary care	rural area, rural community, American college, Victoria, New South Wales
High-profile Careers	politic, career concern, justice, political career, election	ethnography, ethnographic research, life history interview, natural experiment	discourse analysis, exploratory analysis, post hoc analysis	entrepreneur, elite, politician, librarian, president	library, court, film, army, military	Britain, Russia, Poland, German, Brazil
Social Policy	marriage, wage, human capital, earning, criminal career	longitudinal data, panel study, national longitudinal survey, panel data, longitudinal sample	event history analysis, odds ratio, confidence interval, hierarchical level, dynamic model	mother, spouse, father, offender, wife	prison, law school, baseball, vocational school	household, Spain, Belgium, Denmark, Great Britain

Table 2-3. Overview of the Clusters in the Career Literature across the Social Sciences

Cluster	Definition	Prominent Papers[a b]
Organizational	Research on the interplay between the careers, management, and organization.	Allen, Eby, Poteet, Lentz, & Lima, 2004 Ashforth, Harrison, & Corley, 2008 Crant, 2000 Greenhaus, Parasuraman, & Wormley, 1990 Higgins & Kram, 2001
Individual	Research on the individual-level determinants that influence career progression.	Judge, Higgins, Thoresen, & Barrick, 1999 Lent, Brown, & Hackett, 1994 Judge, Thoresen, Pucik, & Welbourne, 1999 Nosek, Greenwald, & Banaji, 2005 Savickas et al., 2009
Education	Research on the relationship between education and careers.	Else-Quest, Hyde, & Linn, 2010 Harackiewicz, Barron, Tauer, & Elliot, 2002 Lankford, Loeb, & Wyckoff, 2002 Tschannen-Moran & Hoy, 2007 Tyson, Castellino, & Darity, 2005
Doctorate Careers	Research on the careers of highly educated individuals.	Buerhaus, Staiger, & Auerbach, 2000 Hojat et al., 2002 Landon, Reschovsky, Pham, & Blumenthal, 2006 Sambunjak, Straus, & Marusic, 2006 Willis-Shattuck et al., 2008
High-profile Careers	Research on careers in professions that are prominent or conspicuous.	Gibbons & Murphy, 1992 Hong & Kubik, 2003 Lerner & Tirole, 2002 Simonton, 2003 Walder, 1995
Social Policy	Research on careers with implications for social policy.	Heilman, Wallen, Fuchs, & Tamkins, 2004 Kalev, Dobbin, & Kelly, 2006 Laub, Nagin, & Sampson, 1998 Nagin & Land, 1993 Neal & Johnson, 1996

[a] Five selected papers from the 50 most highly cited per cluster with more than 50% of the extracted terms in the cluster. [b] References can be found in Appendix 2-B.

The second stream within this cluster considers teaching careers. A key focus is how processes of identity development shape teachers' career paths (Alger, 2009; Beijaard, Verloop, & Vermunt, 2000; Day, Elliot, & Kington, 2005). For example, Kelchtermans (2009) suggests that successful teachers take teaching "personally", in that they use their personal commitments and vulnerabilities to perform their jobs. This stream of research is reminiscent of the scholarship on work as a calling, which has received a surge of interest within the management discipline (Elangovan et al., 2010).

Doctorate careers

The doctorate careers cluster (red) focuses on the careers of highly educated individuals, exemplified by those with doctorates – e.g., physicians and academic researchers. In the main, the journals publishing research on this topic are either medical (e.g., JAMA – Journal of the American Medical Association) or devoted to understanding academics (e.g., Journal of Higher Education, Scientometrics). The main research questions in this cluster include: a) developing sound conceptualizations and measures of career satisfaction and extrinsic success in doctoral careers (Egghe, 2006; Williams et al., 1999), b) identifying the determinants of career satisfaction and success in doctorate careers (e.g., Kaplan et al., 1996; Linzer et al., 2000), and c) using labor force projections to predict whether a societally beneficial number of people will choose careers in under-served occupational specialties (e.g., Bodenheimer & Pham, 2010; Rabinowitz, Diamond, Markham, & Paynter, 2001).

The doctorate careers cluster may have several implications for management career scholars. Articles in this cluster tend to focus on a subset of professional occupations. There are several benefits to this approach. First, it allows for the development of very precise measures of internal and external career success (e.g., Egghe, 2006; Williams et al., 1999). This is in contrast to management career studies that tends to develop omnibus measures of career satisfaction and success, which imperfectly cover a wide variety of occupations (e.g., Spurk, Abele, & Volmer, 2011).

Second, focusing on a specific range of occupations encourages descriptive accounts of the level of career success associated with different specialties (e.g., primary care versus surgeon; social scientist versus physical scientist). These descriptive differences between specialties are documented in large-scale studies, which allow for population-level inferences (Landon, Reschovsky, & Blumenthal, 2003; Leigh, Kravitz, Schembri, Samuels, & Mobley, 2002). This sort of data should be incredibly helpful for people engaged in making career choices.

A third benefit of focusing on a narrow range of occupations is that it facilitates developing labor force projections. Thus, there are a number of papers that carefully estimate the number of workers in different occupations and the competencies demanded of these future professionals (D. M. S. Lee, Trauth, & Farwell, 1995; Rabinowitz et al., 2001). Several of these papers then suggest concrete reforms to educational and public policy in order to effectively manage the supply and demand of professionals across specialties. Management career scholars might consider engaging in similar estimating exercises, as the implications for education and policy could be profound.

High-profile careers

The high-profile careers cluster (dark blue) represents a collection of work from political science, law, sociology, and labor economics. By simply looking at the terms on the map, the underlying nature of this cluster is not immediately clear. However, upon reading the articles associated with this cluster, it becomes apparent that this cluster focuses on high-profile careers – e.g., securities analysts (Hong, Kubik, & Solomon, 2000), elite executives (Maclean, Harvey, & Chia, 2012), politicians (Desposato, 2006; Fiorina, 1994), judges (Gennaioli & Rossi, 2010), artists (Menger, 1999), and entrepreneurs (Carter, Gartner, Shaver, & Gatewood, 2003). By "high-profile", we mean careers that are intentionally conspicuous, where positive attention is key to career success.

A preoccupation of individuals in high-profile careers is what economists and political scientists call "*career concerns*". Career concerns involve career behaviors that are not currently valued but that are predicted to be valued in the future (Gibbons & Murphy, 1992). A practice, for example, that can be explained by career concerns is open source software development (Lerner & Tirole, 2002). This type of software development is not remunerated directly, but it can lead to future job offers and shares in open source-based companies. In academia, volunteering to be a journal editor or on an editorial board might be rationalized in terms of career concerns – i.e., such service raises one's profile in the academic community, leading to future opportunities (Baruch, Konrad, Aguinis, & Starbuck, 2008). Career concerns have been incorporated into elegant mathematical models of career behavior (Dewatripont, Jewitt, & Tirole, 1999; Tirole, 1994). This sort of formal theory is rarely seen in the study of careers within management. As such, career concerns is a concept that management career scholars could benefit from attending to.

Social policy

The social policy cluster (yellow) draws from sociology, and the related specialty of criminology. The cluster studies careers in relation to three broad subthemes: a) work-family policy, b) inequality, and c) deviant behavior.

Research on work and family life from the social policy cluster is closely related to studies on work-family issues in the management literature (Eby, Casper, Lockwood, Bordeaux, & Brinley, 2005). Areas of overlap include the effects of family life factors such as dual earning and family formation on career success (Parasuraman et al., 1996). However, the scope of the social policy research on work and family life is broader. For example, the social policy cluster considers how the timing of marriage and children hinges upon career factors (Blossfeld & Huinink, 1991; Griskevicius, Delton, Robertson, & Tybur, 2011; Sweeney, 2002). A particularly interesting finding within this cluster is the phenomenon of "scaling back", where both individuals in a dual career couple attempt to moderate their work commitments in order to buffer the family (P. E. Becker & Moen, 1999).

A second theme in this cluster is inequality. Processes that produce career inequalities include (implicit) biases (Nosek, Banaji, & Greenwald, 2002; Schmader, Johns, & Barquissau, 2004), ineffective organizational diversity policies (Kalev, Dobbin, & Kelly, 2006), underprivileged childhoods (Neal & Johnson, 1996), and cumulative disadvantage (DiPrete & Eirich, 2006).

Deviant career behavior is the third theme. Studies on deviant behavior include research on the effects of low socioeconomic status on careers, the effects of substance abuse, and criminal careers. The criminal career is by far the most studied subtopic within this theme (DeLisi & Piquero, 2011; Piquero, Farrington, & Blumstein, 2003). Topics covered within this research stream include crime cessation and recidivism (Laub, Nagin, & Sampson, 1998; Warr, 1998) and negative effects of criminal behavior on salary (Western, 2002). The research in this area has also put considerable effort into methodological issues related to the estimation of career effects. For instance, studies within the cluster use panel data, event history analysis, semi-parametric group-based models, corrections for endogeneity, and within-individual Hierarchical Linear Modeling (Blokland & Nieuwbeerta, 2005; DiPrete & Eirich, 2006; Piquero et al., 2003; Western, 2002). These analytic techniques allow for stronger empirical claims. Each of these techniques could be put to use by management career scholars.

Revealing Gaps

The maps depict the research areas covered by the local and global literature on careers. In order to provide insight into what areas remain to be explored, Table 2-4 organizes the research clusters in terms of three research focuses (i.e., pre-work career, career development, and career outcomes) by three levels of analysis (i.e., individual, organizational, and societal/economic).

The table reveals that career development at the individual level is well studied, both within and outside of the management literature. This is not a surprise, since a career is an individual level phenomenon. Striking, however, is the lack of any major body of work on career outcomes at an organizational level, while it is accepted that the organization is a major stakeholder in the career of its employees. With the exception of the work on the benefits of expatriation for the organization (e.g., Dickmann & Doherty, 2008; Jassawalla & Sashittal, 2009) and a handful of publications looking at the benefits of career related HR policies for the organization (e.g., Baruch, 1999), there is little research that looks at how career systems affect organizational performance or survival. This is a major opportunity for future management research.

Table 2-4. Overview of the Career Research Focuses and Levels of Analysis Covered by the Topic Areas in the Careers Literature

<table>
<tr><th></th><th>Pre-work Career</th><th>Career Development</th><th>Career Outcomes</th></tr>
<tr><td rowspan="2">Societal/Economic</td><td colspan="2">Social Policy (e.g., work-family policy and criminal careers)[a]</td><td>Work Force Planning in Skilled Occupations[a]</td></tr>
<tr><td></td><td>Life Opportunities</td><td rowspan="4">[Missing as a major area of inquiry]</td></tr>
<tr><td rowspan="3">Organizational</td><td rowspan="3">Career Counseling[b]</td><td>Relationships and Mentoring[c]</td></tr>
<tr><td>HRM Policies[d e]</td></tr>
<tr><td rowspan="2">International Careers</td></tr>
<tr><td rowspan="9">Individual</td><td rowspan="4">Degree Choice and Academic Achievement[f]</td><td rowspan="7">Career Success[h]</td></tr>
<tr><td>Career Choice</td></tr>
<tr><td>Life Opportunities</td></tr>
<tr><td>Career Self Management[g]</td></tr>
<tr><td colspan="2">Career Adaptation</td></tr>
<tr><td colspan="2">Individual Differences (e.g., personality)[c e]</td></tr>
<tr><td rowspan="2"></td><td>High Profile Careers</td></tr>
<tr><td>Teacher Careers[f]</td><td rowspan="2">Performance Quality[f i]</td></tr>
<tr><td colspan="2">Doctorate Careers</td></tr>
</table>

[a] Social policy cluster. [b] Career choice cluster. [c] Individual and relational career success cluster. [d] Career management cluster. [e] Organizational cluster. [f] Education cluster. [g] Individual career cluster. [h] Studied in all clusters, but most prominent in the individual and relational career success cluster, the organizational, and the individual cluster. [i] Doctorate careers cluster.

Discussion

Using a novel science mapping approach, we created a map of the career literature within the field of management, as well as a map of career studies across the social sciences. On the basis of term co-occurrence in the abstracts and titles of scientific publications, distinct topic areas have emerged.

To the best of our knowledge, this article provides the first empirically grounded taxonomy of career studies. We identified six career topics within management journals: international careers, career management, career choice, career adaptation, individual and relational career success, and life opportunities. Importantly, less than one fifth of the studies on careers are found in management journals, which implies that career scholars may also benefit from looking outside of management for insights, concepts, and ideas. To aid in that effort, we also identified six broader clusters of career topics from across the social sciences, namely: organizational, individual, education, doctorate careers, high-profile careers, and social policy.

Our taxonomy provides a framework that can be a helpful tool to outsiders trying to learn what career studies is about as well as clarifying the matter for those already engaged in career studies. Accordingly, our schema can guide interested practitioners in their search for actionable knowledge on careers and help aspiring career scholars navigate the literature and find ways in which to contribute. Moreover, the maps can help established career scholars in the design of new endeavors. The current fragmentation hinders the exchange of valuable insights between career researchers in different sub-fields. Familiarity with the concepts, methods, and objects of study in near and distant career literatures allows career scholars to leverage the insights of fellow career researchers.

Of particular value for future research, our review of the global map highlights a number of concepts, methods, and approaches that could be productively imported and adopted by management career scholars. To recap, the education cluster has an under-appreciated body on work as an identity-based calling. The doctorate careers cluster highlights the value of studying specific occupations, in particular the ability to create tailored measures, provide guidance about the most rewarding careers, and develop labor force projections. The high-profile careers cluster not only provides rich descriptive detail about a number of interesting careers, it also develops rigorous formal theories around the insightful concept of “career concerns”. Finally, the social policy cluster not only identifies intriguing work-family dynamics (e.g., “scaling back” amongst dual-career couples), but also employs sophisticated statistical techniques, which allow for more precise estimates of causal relationships.

We do wish to call the reader’s attention to several of the limitations of our study. While the clusters are empirically grounded, some subjective judgment is involved in their naming and interpretation. In addition, the precise number of clusters (but not term placement) depends on the clustering parameter. We chose the default VOS settings (i.e., with a clustering coefficient of 1), but there is no guarantee that other insights might come from setting a resolution parameter that leads to more or fewer clusters. However, in contrast to other forms of reviewing and categorization, science mapping provides transparency regarding the basis on which these inferences are drawn and the maps hopefully pave the way to informed discussion regarding the “true” nature of the field. Interested readers can alter the resolution parameters themselves in the online version of the maps.

These science maps are likely to be particularly helpful for making sense of the career literature, because research on careers is diverse and studied across many fields (Arthur et al., 1989a, pp. 9–10). More broadly, we believe that term mapping is a helpful tool for reviewing a field’s major themes and underlying structure. It is more systematic and disciplined than traditional narrative reviews, and may represent the future of literature reviews.

Appendix 2-A: Search Phrases

For the maps on career studies within management and career studies across the social sciences, the following two search phrases were run on the Web of Science Core Collection. Afterwards we removed articles where the string "career" only occurred in the keywords, thereby reducing the corpus to articles where the string "career", including any inflection of the term, only occurs in the title or abstract of the article.

Search Phrase for Local Map: The Career Field within Management

(TS=(career) AND IS=(0002-7642 OR 0003-6870 OR 0311-6336 OR 1030-7222 OR 0007-1080 OR 1052-150X OR 0962-8770 OR 0889-4019 OR 1363-3589 OR 0142-5455 OR 0959-6801 OR 0894-4865 OR 0018-7259 OR 0895-9285 OR 0199-8986 OR 0019-7939 OR 0019-8676 OR 0019-8692 OR 1470-5958 OR 1039-6993 OR 0167-7187 OR 0147-1767 OR 1360-3736 OR 0020-7780 OR 1382-340X OR 0892-7626 OR 0889-3268 OR 0167-4544 OR 0022-0027 OR 0309-0590 OR 0022-166X OR 0022-1856 OR 0195-3613 OR 0262-1711 OR 0001-8791 OR 1363-6820 OR 0023-656X OR 0160-449X OR 0098-1818 OR 1358-6297 OR 0361-6843 OR 0742-7301 OR 1532-3005 OR 0730-8884 OR 0267-8373 OR 0950-0170 OR 1038-4111 OR 1362-0436 OR 1352-7606 OR 1359-432X OR 0968-6673 OR 1059-6011 OR 0018-7267 OR 1044-8004 OR 0090-4848 OR 0954-5395 OR 1053-4822 OR 0958-5192 OR 0143-7720 OR 0965-075X OR 0268-3946 OR 0963-1798 OR 0894-3796 OR 0160-8061 OR 1742-7150 OR 1048-9843 OR 0268-1072 OR 1350-5084 OR 1047-7039 OR 0170-8406 OR 0749-5978 OR 0090-2616 OR 0031-5826 OR 0048-3486 OR 0191-3085 OR 1046-4964 OR 1941-6520 OR 1537-260X OR 0001-4273 OR 1558-9080 OR 0363-7425 OR 1012-8255 OR 1476-7503 OR 0001-8392 OR 0742-3322 OR 1360-2381 OR 0217-4561 OR 1472-4782 OR 0312-8962 OR 1746-5265 OR 0340-5370 OR 1045-3172 OR 0964-4733 OR 0008-1256 OR 0825-0383 OR 1750-614X OR 0927-7099 OR 1938-9655 OR 0964-8410 OR 1535-3958 OR 1475-9551 OR 1545-8490 OR 0011-7315 OR 0965-3562 OR 1212-3609 OR 1389-5753 OR 1019-6781 OR 1042-9247 OR 0013-791X OR 1751-6757 OR 0263-2373 OR 0926-2644 OR 0017-8012 OR 0018-9391 OR 1471-678X OR 0960-6491 OR 1366-2716 OR 0019-8501 OR 1617-9846 OR 0378-7206 OR 1471-7727 OR 1047-7047 OR 1385-951X OR 0121-5051 OR 1447-9338 OR 0959-6119 OR 0169-

2070 OR 0957-4093 OR 1367-5567 OR 1460-8545 OR 0144-3577 OR 0960-0035 OR 0263-7863 OR 1756-6517 OR 1648-715X OR 0267-5730 OR 0266-2426 OR 0969-6016 OR 0092-2102 OR 0021-8863 OR 0021-9010 OR 0735-3766 OR 1058-6407 OR 0923-4748 OR 0277-6693 OR 0268-3962 OR 0047-2506 OR 1075-4253 OR 1367-3270 OR 0149-2063 OR 0742-1222 OR 1056-4926 OR 1833-3672 OR 0022-2380 OR 0966-0429 OR 0272-6963 OR 0160-5682 OR 0953-4814 OR 0737-6782 OR 1478-4092 OR 1757-5818 OR 0047-2778 OR 0888-4773 OR 1523-2409 OR 0892-9912 OR 1477-8238 OR 0024-6301 OR 1523-4614 OR 0960-4529 OR 1044-5005 OR 0893-3189 OR 0025-1747 OR 0938-8249 OR 1350-5076 OR 1740-8776 OR 0025-1909 OR 1540-1960 OR 0276-7783 OR 1532-9194 OR 0748-4526 OR 1048-6682 OR 0305-0483 OR 0030-364X OR 1094-4281 OR 8756-9728 OR 1471-9037 OR 0033-6807 OR 0034-7590 OR 1806-4892 OR 0048-7333 OR 0895-6308 OR 0889-938X OR 1863-6683 OR 0378-9098 OR 0956-5221 OR 0971-7218 OR 1862-8516 OR 0264-2069 OR 0921-898X OR 1932-4391 OR 1476-1270 OR 0143-2095 OR 1359-8546 OR 0883-7066 OR 1094-429X OR 1092-7026 OR 0953-7325 OR 0166-4972 OR 1478-3363 OR 0261-5177 OR 0963-1690 OR 1740-4754 OR 1554-7191 OR 1480-8986 OR 1546-2234 OR 1936-9735 OR 2222-3436 OR 0302-3427)) AND Language=(English) AND Document Types=(Article OR Review)

Timespan=1990-2012. Databases=SSCI.

Search Phrase for Global Map: The Study of Careers across the Social Sciences

(TS=((career) NOT ("treatment careers" OR drug OR abuse OR addict OR diagnosis OR illness OR cancer OR aids))) AND Language=(English) AND Document Types=(Article OR Review)

Timespan=1990-01-01 - 2012-12-31. Databases=SSCI.

Appendix 2-B: References from Tables

Allen, T. D., Eby, L. T., Poteet, M. L., Lentz, E., & Lima, L. (2004). Career benefits associated with mentoring for protégés: A meta-analysis. *Journal of Applied Psychology*, *89*, 127–136.

Anderson, S. E., Coffey, B. S., & Byerly, R. T. (2002). Formal organizational initiatives and informal workplace practices: Links to work-family conflict and job-related outcomes. *Journal of Management*, *28*, 787–810.

Ashforth, B. E., Harrison, S. H., & Corley, K. G. (2008). Identification in organizations: An examination of four fundamental questions. *Journal of Management*, *34*, 325–374.

Bechky, B. A. (2006). Gaffers, gofers, and grips: Role-based coordination in temporary organizations. *Organization Science*, *17*, 3–21.

Bolino, M. C. (2007). Expatriate assignments and intra-organizational career success: Implications for individuals and organizations. *Journal of International Business Studies*, *38*, 819–835.

Briscoe, J. P., & Hall, D. T. (2006). The interplay of boundaryless and protean careers: Combinations and implications. *Journal of Vocational Behavior*, *69*, 4–18.

Brown, A., Bimrose, J., Barnes, S.-A., & Hughes, D. (2012). The role of career adaptabilities for mid-career changers. *Journal of Vocational Behavior*, *80*, 754–761.

Buerhaus, P. I., Staiger, D., & Auerbach, D. I. (2000). Implications of an aging registered nurse workforce. *Journal of the American Medical Association*, *283*, 2948–2954.

Crant, J. M. (2000). Proactive behavior in organizations. *Journal of Management*, *26*, 435–462.

Dickmann, M., Doherty, N., Mills, T., & Brewster, C. (2008). Why do they go? Individual and corporate perspectives on the factors influencing the decision to accept an international assignment. *International Journal of Human Resource Management*, *19*, 731–751.

Duffy, R. D., & Sedlacek, W. E. (2007). The presence of and search for a calling: Connections to career development. *Journal of Vocational Behavior*, *70*, 590–601.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127.

Frese, M., Fay, D., Hilburger, T., Leng, K., & Tag, A. (1997). The concept of personal initiative: Operationalization, reliability and validity in two German samples. *Journal of Occupational and Organizational Psychology*, *70*, 139–161.

Gibbons, R., & Murphy, K. J. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *The Journal of Political Economy*, *100*, 468–505.

Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experiences, job-performance evaluations, and career outcomes. *Academy of Management Journal*, *33*, 64–86.

Hall, D. T., & Chandler, D. E. (2005). Psychological success: When the career is a calling. *Journal of Organizational Behavior*, *26*, 155–176.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology*, *94*, 562–575.

Hedlund, G. (1994). A model of knowledge management and the N-form corporation. *Strategic Management Journal*, *15*, 73–90.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, *89*, 416–427.

Higgins, M. C., & Kram, K. E. (2001). Reconceptualizing mentoring at work: A developmental network perspective. *Academy of Management Review*, *26*, 264–288.

Hojat, M., Gonnella, J. S., Nasca, T. J., Mangione, S., Vergare, M., & Magee, M. (2002). Physician empathy: Definition, components, measurement, and relationship to gender and specialty. *American Journal of Psychiatry*, *159*, 1563–1569.

Hong, H., & Kubik, J. D. (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *Journal of Finance*, *58*, 313–351.

Jokinen, T., Brewster, C., & Suutari, V. (2008). Career capital during international work experiences: Contrasting self-initiated expatriate experiences and assigned expatriation. *International Journal of Human Resource Management*, *19*, 979–998.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, *52*, 621–652.

Judge, T. A., Thoresen, C. J., Pucik, V., & Welbourne, T. M. (1999). Managerial coping with organizational change: A dispositional perspective. *Journal of Applied Psychology, 84*, 107–122.

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, *71*, 589–617.

Koen, J., Klehe, U.-C., van Vianen, A. E. M., Zikic, J., & Nauta, A. (2010). Job-search strategies and reemployment quality the impact of career adaptability. *Journal of Vocational Behavior*, *77*, 126–139.

Landon, B. E., Reschovsky, J. D., Pham, H. H., & Blumenthal, D. (2006). Leaving medicine: The consequences of physician dissatisfaction. *Medical Care*, *44*, 234–242.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*, 37–62.

Laub, J. H., Nagin, D. S., & Sampson, R. J. (1998). Trajectories of change in criminal offending: Good marriages and the desistance process. *American Sociological Review*, *63*, 225–238.

Lazarova, M. B., & Cerdin, J.-L. (2007). Revisiting repatriation concerns: Organizational support versus career and contextual influences. *Journal of International Business Studies*, *38*, 404–429.

Lee, D. M. S., Trauth, E. M., & Farwell, D. (1995). Critical skills and knowledge requirements of IS professionals: A joint academic-industry investigation. *MIS Quarterly*, *19*, 313–340.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*, 79–122.

Lent, R. W., Brown, S. D., Sheu, H. B., Schmidt, J., Brenner, B. R., & Gloster, C. S. (2005). Social cognitive predictors of academic interests and goals in engineering: Utility for women and students at historically black universities. *Journal of Counseling Psychology*, *52*, 84–92.

Lerner, J., & Tirole, J. (2002). Some simple economics of open source. *Journal of industrial Economics*, *50*, 197–234.

Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity - specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, *31*, 327–362.

Neal, D. A., & Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *The Journal of Political Economy*, *104*, 869–895.

Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, *58*, 367–408.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*, 166–180.

Parasuraman, S., Purohit, Y. S., Godshalk, V. M., & Beutell, N. J. (1996). Work and family variables, entrepreneurial career success, and psychological well-being. *Journal of Vocational Behavior*, *48*, 275–300.

Sambunjak, D., Straus, S. E., & Marusic, A. (2006). Mentoring in academic medicine: A systematic review. *Journal of the American Medical Association*, *296*, 1103–1115.

Savickas, M. L. (1997). Career adaptability: An integrative construct for life-span, life-space theory. *Career Development Quarterly*, *45*, 247–259.

Savickas, M. L., Nota, L., Rossier, J., Dauwalder, J.-P., Duarte, M. E., Guichard, J., … van Vianen, A. E. M. (2009). Life designing: A paradigm for career construction in the 21st century. *Journal of Vocational Behavior*, *75*, 239–250.

Savickas, M. L., & Porfeli, E. J. (2012). Career Adapt-Abilities Scale: Construction, reliability, and measurement equivalence across 13 countries. *Journal of Vocational Behavior*, *80*, 661–673.

Schoon, I., & Parsons, S. (2002). Teenage aspirations for future careers and occupational outcomes. *Journal of Vocational Behavior*, *60*, 262–288.

Seibert, S. E., Kraimer, M. L., & Liden, R. C. (2001). A social capital theory of career success. *Academy of Management Journal*, *44*, 219–237.

Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior the integration of product, person, and process perspectives. *Psychological Bulletin*, *129*, 475–494.

Sullivan, S. E. E., & Arthur, M. B. (2006). The evolution of the boundaryless career concept: Examining physical and psychological mobility. *Journal of Vocational Behavior*, *69*, 19–29.

Suutari, V., & Brewster, C. (2003). Repatriation: Empirical evidence from a longitudinal study of careers and expectations among Finnish expatriates. *International Journal of Human Resource Management*, *14*, 1132–1151.

Tharenou, P., Latimer, S., & Conroy, D. (1994). How do you make it to the top: An examination of influences on women and men managerial advancement. *Academy of Management Journal*, *37*, 899–931.

Thompson, C. A., Beauvais, L. L., & Lyness, K. S. (1999). When work-family benefits are not enough: The influence of work-family culture on benefit utilization, organizational attachment, and work-family conflict. *Journal of Vocational Behavior*, *54*, 392–415.

Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, *23*, 944–956.

Tyson, K., Castellino, D. R., & Darity, W. (2005). It's not "a black thing": Understanding the burden of acting white and other dilemmas of high achievement. *American Sociological Review*, *70*, 582–605.

Van Vianen, A. E. M., Klehe, U.-C., Koen, J., & Dries, N. (2012). Career Adapt-Abilities Scale - Netherlands form: Psychometric properties and relationships to ability, personality, and regulatory focus. *Journal of Vocational Behavior*, *80*, 716–724.

Walder, A. G. (1995). Career mobility and the communist political order. *American Sociological Review*, *60*, 309–328.

Wang, M., Zhan, Y., Liu, S., & Shultz, K. S. (2008). Antecedents of bridge employment: A longitudinal investigation. *Journal of Applied Psychology*, *93*, 818–830.

Willis-Shattuck, M., Bidwell, P., Thomas, S., Wyness, L., Blaauw, D., & Ditlopo, P. (2008). Motivation and retention of health workers in developing countries: A systematic review. *BMC Health Services Research*, *8*: 247-254.

Chapter 3: Collaborative Tagging of Academic Articles - Making Management Publications Machine-Readable

Abstract

This paper introduces the "ReNotate" software platform. ReNotate uses the highlights and annotations of people reading academic publications in order to make those publications machine-readable. For individuals this platform consists of a software application that helps in reading and reviewing academic articles. It enables highlighting and annotating (i.e., *tagging*) and individuals can subsequently query their tags relatively easily. For the collective the platform consists of a large database of validated and aggregated data from the tags of all individuals combined. The platform could thus allow publications to be found based on study characteristics, concepts and their definitions to be traversed to their respective publications, and the relationships between concepts to be queried. Thereby the platform has the potential to improve the accessibility and efficiency of the sciences in general and the field of management in particular. We provide an overview of the functionality of the ReNotate platform, describe the algorithms used to aggregate data from readers, and put forward an ontology specifically designed for concept tagging in the management literature.

Introduction

How do we know what we collectively know? With at least 20,800 peer-reviewed journals and 55 million publications (Scopus, 2014), the academic literature is overwhelming and its rate of growth is only increasing (de Solla Price, 1970; P. O. Larsen & von Ins, 2010). The size of the academic literature reflects the enormity of our collective knowledge. Peer-reviewed publications chronicle the findings of academic research and thinking, thereby providing insight into a vast array of topics within numerous domains. At the same time, however, the sheer magnitude of the academic corpus poses a challenge to those who aim to navigate through the scientific literature. As the size of the academic literature increases, it becomes more difficult to find what one is looking for or even to know what one should be looking for (e.g., Thomson, 1984).

This also holds true for the management literature. There are nearly 130,000 journal articles in the management literature from over 185 journals, and currently at least 7,500 additional articles are published each year. [2] This vastness is exacerbated by a general disorganization of the academic literature. One contributor to this disorganization is that different academic communities frequently use varying terms to refer to the same phenomena, or use the same terms to refer to different phenomena (Block, 1995). Indeed, concept proliferation seems highly prevalent within management scholarship (Bosco, Uggerslev, & Steel, 2014; Hallberg & Schaufeli, 2006; Le, Schmidt, Harter, & Lauver, 2010; Li & Larsen, 2011; Schwab, 1980; J. Singh, 1991). And the more difficult it becomes to know what we collectively know, the harder it is to establish what is unknown or to become aware of prior knowledge that we can build upon. Both the magnitude and disorganization of the academic literature can thus impede the cumulative development of science (Attwood et al., 2009; Blalock, 1968; Le et al., 2010; Tesser & Krauss, 1976; Thomson, 1984).

In order to cope with this bewildering array of scholarship, we suggest using the collaborative efforts of readers to add semantic tags and annotations to academic publications in order to extract the key data from articles. Many students

[2] This is the count of articles and reviews in the 185 journals classified as "Management Journal" by the Web of Science (WoS), retrieved on September 7th, 2015. Scopus provided 130,458 articles from these journals and WoS listed 128,240 articles. This is a conservative estimate, as the WoS Management journal list does not contain all journals that could be specified as Management journals (Harzing, 2013).

and scholars already highlight and annotate academic publications to help make sense of and remember what they have read (Porter-O'Donnell, 2004). This note-taking can be, and increasingly is, done digitally due to the digitalization of journal publications alongside the increasing access to laptops and tablets (e.g., Connell, Bayliss, & Farmer, 2012; Glover, Xu, & Hardaker, 2007) and the multitude of software tools that support annotations (e.g., Adobe Reader, Foxit Reader, Mac OSX Preview, Mantano Reader, etc.). These digital highlights and annotations can serve as user-generated "tags". As we will describe later in the paper, the extraction of these digital tags enables the aggregation of the sensemaking efforts of individuals to the collective level. The use of this tagging data holds the promise of "bottom-up" syntheses of the literature and swift access to the key content. As such, the advocated approach constitutes an elaborate form of *crowdsourced* or *collaborative* tagging (Golder & Huberman, 2006) applied to academic documents.

To structure and attain readers' tags and tagged content, we have developed a software platform called "ReNotate". [3] In this chapter, we describe both the basic tenets of the platform and the specific technologies that it employs. The platform has a dual purpose. First of all, it helps students and scholars annotate the articles they read in such a way that their *own* annotations become (machine-) searchable – e.g., a person can search through all their own tags across documents that are related to a concept they are interested in (such as "empathy" or "employability"). Secondly, it allows anybody to draw from a validated version of the *collective* highlights and annotations of a large research community. Accomplishing these two aims involves simultaneously extending prior technological capabilities, mostly from the field of information science, and accommodating to existing social routines.

To preview the structure of this chapter, we start by providing a brief summary of the diverse literature on document highlighting and annotation, from the pen and highlighter, to current annotation software, to the ReNotate document reading and tagging software. Afterward, we present the ReNotate platform for

[3] The features of the ReNotate platform, as described in this chapter, have been scheduled for a public Alpha release mid-2016. Some of the features will already be available in the pre-Alpha release in February 2015, but this is mainly intended for the training of the platform's algorithms and will not be public. The software will become available via http://renotate.com , where we will also be posting regular updates regarding its development.

crowdsourcing document coding, describing a categorization scheme (i.e., "ontology") that focuses on extracting concepts from management publications and providing ways to link data from different readers and articles. The tool allows for a large crowdsourced database to be developed that extracts key information – e.g., concept names, definitions, operationalizations, and relationships to other concepts – from a large number of publications. The resulting database allows those interested to find which publications feature which concepts and which relationships. Such information is helpful for literature reviews, for example to search for studies with particular characteristics, concepts and their respective definitions, or even the relationships between concepts. In addition, we describe how the ReNotate platform, with the default settings, is particularly helpful for identifying opportunities for meta-analysis. In our discussion section, we go into the broader benefits of the platform and provide an outline for several future uses. These future uses include (a) improvements that would allow more complex taxonomies to be attained, (b) capability enhancements that would allow the ReNotate platform as a pedagogical tool, and (c) making ReNotate useful for meta-analytic data collection beyond the identification of relevant studies.

First Purpose of the ReNotate Platform: A New Approach to Note-Taking

Since at least the 3rd Century BC, people have been underlining and annotating their readings (Dickey, 2007). Such note-taking involves two basic routines: (1) in some way identifying which text is especially important (traditionally through underlining text, but currently often done through "highlighting"), and (2) making comments on the text (which is referred to as "annotating"). These basic routines can be made more sophisticated. For example, some people use different kinds of highlighting (e.g., different colors) for different kinds of important concepts. Another helpful routine involves linking these highlights to annotations – i.e., in order to explain why the highlighted text is important. These traditional note-taking practices serve two key functions. First, they provoke readers to actively reflect on the text they read (e.g., Wolfe, 2000). Second, they serve as a memory aid, such that the text can be quickly reviewed and comprehended upon revisiting (e.g., Blanchard & Mikkelson, 1987; Hartley, Bartlett, & Branthwaite, 1980; McAndrew, 1983).

Within the last 20 years, a number of new technologies have led to substantial changes in reading practices and, subsequently, in note-taking. In these years, academic journals have started making their publications available in digital format, and books can also increasingly be read electronically. This has extended to many older publications too, which have been scanned and made available online (Coyle, 2006; Lopatin, 2006). At the same time, the advent and availability of light laptop computers, tablets, and e-readers has made it much easier to read these digital formats (Zickuhr & Rainie, 2014). One advantage of this digitalization is that it allows notes (i.e., highlights and annotations) to be electronically saved and shared across computers. Some of the newer document readers have even evolved to the extent that they actively use the highlighted material. Readers using an Amazon Kindle device, for instance, can see what parts of the document they are reading are commonly highlighted (i.e., "Popular Highlights") and can thereby easily see which parts of the document might be the most interesting or important.

However, there is a key limitation to (even the newer versions of) existing document readers: they do not allow for annotations or highlights to be easily grouped by type or topic such that individuals can search for all the annotations and highlights in a given category (Lin, Davis, & Zhou, 2009; Shepitsen, Gemmell, Mobasher, & Burke, 2008; Specia & Motta, 2007). For example, let us imagine a scholar has highlighted all conceptual definitions of the concept "Social Capital" as she came across them in texts. With traditional document readers, there would be no way for her to easily search for all definitions she has identified across documents, much less a specific one.

Partly to address the previously mentioned limitation, we envision a software platform that allows individuals to put annotations and highlights into categories. These categories and the content within them are searchable across all documents "tagged" in the software. In order to allow for hierarchical categories to develop, the appropriate tags to be used for different fields, and the tagged material to be more easily searched through, the software uses a "code-tree" or "coding scheme" similar to those used in qualitative textual analyses (e.g., Corbin & Strauss, 1990; Glaser & Strauss, 1967). Thereby, structure is imposed on readers' highlights and annotations.

In Figure 3-1 we give an impression of how highlights and annotations are structured in the default ReNotate code-tree. The reader can make use of three different highlight colors ("Concepts", "Methods", and "Personal"). In turn, these highlights can be linked to certain annotations that define the properties of the highlighted text. In this case, the text "This syndrome can be defined as the periodic itch to move from a job in one place to some other job in some other place" is annotated so as to indicate that it is a definition of the concept "Ghiselli's hobo syndrome" (Ghiselli, 1974, p. 81; Woo, 2011, p. 461).

After having tagged the data, the user can peruse what he or she has tagged in a separate table. This allows errors in the text to be corrected and enables the reader to search through the highlights and annotations, make changes, export the tags to a spreadsheet application, or even create summaries using the data. Figure 3-2 shows an example of how one would search for the definition of the concept "Social Capital". In this example, the user looks for the definitions by searching for instances under the parent node "Social Capital". She specifies her search further by restricting the search to "concepts".

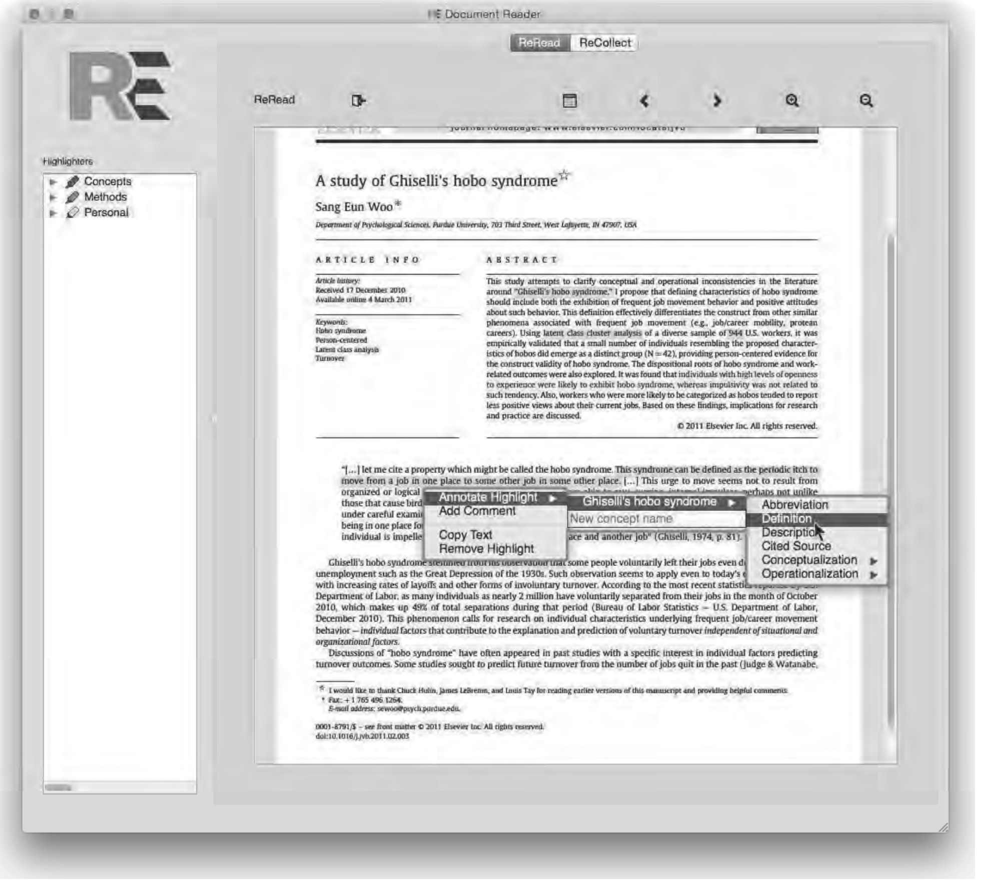


Figure 3-1. Highlighting and Annotation in the ReNotate Document Reader

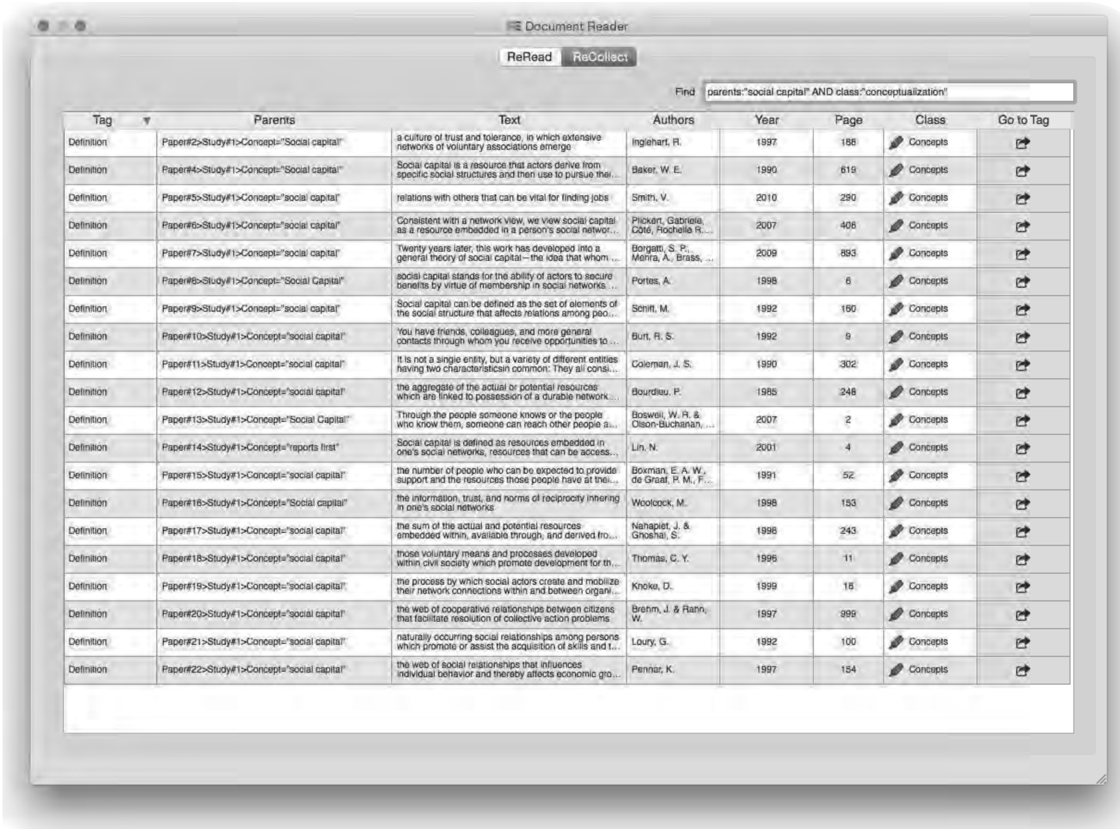


Figure 3-2. Searching for Definitions of "Social Capital" in the ReNotate Document Reader Software

Most prior developments in electronic document readers have been geared toward making the experience more like reading from paper. The digital document reader we describe diverges from this tradition and provides unique advantages over the reading of documents in paper format. Specifically, it structures highlights and makes searching through tagged data easier and faster. While these advantages alone would no doubt provide an important contribution for scholars, students, and even librarians in the fields of management and the sciences in general, ReNotate has further capabilities and ambitions.

Second Purpose of the ReNotate Platform: Aggregating Notes from Multiple Readers

Although searchable categorization is attractive for the various types of readers that exist, a true breakthrough could be achieved by aggregating tags across individuals. This allows for all the ReNotate tags – i.e., properties linked to annotations linked to highlights linked to categories – on a particular publication to be combined. Such aggregation allows individuals to access the collective notes of the scholarly community. To accomplish this, we employ several tools from information science. In particular, we build on semantic web technologies. Let us first provide a brief description of what the semantic web entails and what it has and has not been able to achieve in making science machine-readable, before explaining how semantic web technologies can be used to leverage the collective wisdom of readers.

Making Science Machine-Readable Using Only the Semantic Web

The semantic web can be thought of as an extension of the existing web. It provides structure to the content of a website so as to facilitate the machine-reading of data (Berners-Lee, Hendler, & Lassila, 2001). Applied to academic articles, this method is referred to as *semantic publishing* (Peroni & Shotton, 2012; Shotton, 2009). Semantic publishing refers to the use of semantic web technologies to:

> "*enhance the meaning of a published journal article, to facilitate its automated discovery, to enable its linking to semantically related articles, to provide access to data within the article in actionable form,*

> *and to facilitate integration of data between articles*" (Peroni & Shotton, 2012, p. 33).

Concretely, semantic publishing entails adding tags (generally hidden from direct view of the reader) to publications such that elements of the publication can be recognized for what they are – e.g., title, authors, abstract, definitions, hypothesis, findings – and then machine-read, extracted, and searched through.

Some argue that semantic publishing could revolutionize the ability to aggregate and interrogate the scholarly literature when applied to the content of academic publications (Ciancarini, Di Iorio, Nuzzolese, Peroni, & Vitali, 2013; Shotton, 2009). This revolution would involve the development of powerful tools geared toward making it easier to browse and extract information from academic articles and books. Several such tools have been developed already, mostly in the natural sciences. In the life sciences, for instance, semantic publishing is used to link data on biological species through their taxonomic names and persistent identifiers (e.g., Penev, Agosti, et al., 2010; Penev, Kress, et al., 2010). In neuroimaging, the capabilities have been taken even further through the combination of semantic publishing and meta-analysis (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011), allowing for automated meta-analyses of neuroimaging results. As the adoption of semantic publishing increases, the power and availability of such tools is likely to increase as well. Conceivably, tools could ultimately be developed that allow instant overviews of what we collectively know about any research question one is interested in, at a level of rigor similar to what is currently accomplished only through labor-intensive meta-analytic studies.

Sadly, however, publishers have little incentive to embrace semantic publishing technologies. Indeed, some applications of semantic publishing might go against the interests of the publishers, as these could enable readers to attain access to information from articles that would otherwise have required a paid journal subscription. In addition, publishers would have to come to an agreement on how articles should be tagged. Therefore, some have argued that a move to open-access publishing is necessary for semantic publishing to truly become a success (Peroni, Dutton, Gray, & Shotton, 2015; Peroni & Shotton, 2012). These conclusions are discouraging given the structural impediments to the widespread adoption of open-access publishing (Harnad et al., 2004).

Hopefully, the ReNotate platform that we describe can provide an alternative route to making research in the field of management (and perhaps beyond) machine-readable. Instead of waiting for publishers to further implement semantic web practices or waiting for open access to become more widespread, we suggest using the collaborative efforts of readers to add semantic tags and annotations to academic publications. Moreover, since many semantic web technologies are useful in aggregating and disseminating the collective knowledge of readers, several semantic web technologies are integrated into the platform.

Making Science Machine-Readable using Semantic Web Tools in the ReNotate Platform

For the aggregation of tagged data, ReNotate uses three core semantic web technologies: (1) the "Resource Description Framework" (RDF) model, (2) an ontology language, and (3) the SPARQL query language. Below, we describe each of these technologies and explain how and why they have been integrated into the ReNotate platform. We also describe how "fuzzy-matching" (Gravano et al., 2001) and "record-linkage" algorithms (Fellegi & Sunter, 1969; Newcombe, Kennedy, Axford, & James, 1959) are used to validate data from different readers of the same article and link tag content from different articles.

Using (RDF) Triplets

The semantic web is built on a common technological standard for the tagging of webpages – that of the World Wide Web Consortium (W3C). The tags are generally used to classify web content using predefined vocabularies. This allows web shops, for instance, to indicate what texts denote the prices and what parts of the website represent the related products. The fact that these tags are standardized enables third parties to extract the raw data from web pages. For example, price comparison websites and search engines can use semantic web tags to extract price and product information from the websites. These tags also enable structured queries on the crowdsourced online encyclopedia, Wikipedia, as much of its content has also been tagged (e.g., Bizer et al., 2009). For example, one can use the tags on Wikipedia to create a query that retrieves all people who have won the Nobel

Peace Prize and this can even be specified further, for instance to attain only the names of those born in a certain country.

In order to allow the tagged data from readers of academic publications to be aggregated and released on the semantic web, the structure of the code tree in the ReNotate document reader corresponds with the W3C Resource Description Framework (RDF) model (Brickley & Guha, 1999; Knublauch, Oberle, Tetlow, & Wallace, 2006). The RDF model helps structure knowledge such that it can more easily be extracted and used by computer software. This is achieved by using subject-predicate-object expressions to encode knowledge regarding resources. The "subject" is the resource in question – i.e., the entity regarding to which some knowledge is encoded. The "predicate" refers to a characteristic of that resource and can also be referred to as the "property" (and we will do so for the remainder of the chapter). Lastly, the "object" is the value of the resource on the characteristic and is a resource in itself.

Traditionally, the resources encoded with RDF referred to (representations of) resources on the World Wide Web. But in theory, nearly everything can be described in this format. Indeed, RDF is a generic way of structuring knowledge, which can be applied beyond the online world (Bergman, 2009). For example, RDF can be used to structure information about an academic article that has never been posted online. To illustrate, the RDF model can capture that a certain study (e.g., "Study of Manager Charisma") uses the method of analysis "Structural Equation Modeling". The *subject* would be "Study of Manager Charisma", the *property* would be "uses Data Analysis", and the *object* would be "Structural Equation Modeling". Thus, the subject-predicate-object triplet would be "Study of Manager Charisma" "uses Data Analysis" "Structural Equation Modeling".

RDF triplets make use of Uniform Resource Identifiers (URI) (Berners-Lee et al., 2001). These are strings that provide (1) unique identifiers for resources and their properties, and (2) adhere to the URI syntax (for a description of the required syntax see Masinter, Berners-Lee, & Fielding, 2005). A URI can be used to identify both on- or off-line resources. This stands in contrast to the more commonly known Uniform Resource Locator (URL), which is a type of URI that always refers to a web resource. To continue the prior illustration, if the RDF model were used to classify the study on manager charisma, it would provide separate URIs for the study "Study of Manager Charisma", for the property "uses Data

Analysis” and for the method “Structural Equation Modeling”, thus allowing the URIs to be used to look for characteristics uniquely related to this particular “Study of Manager Charisma”, the general property “uses Data Analysis”, or to the method of analysis “Structural Equation Modeling”.

In essence, RDF is a model for capturing *graph data*. In the field of management, graph data are most commonly used in network analysis. Like RDF triplets, graph data consist of edges (i.e., properties) connected to nodes (i.e., resources). Unlike graph data, however, RDF models work with URIs, which serve as unique identifiers of the content that is tagged. In addition, RDF models are the standard on which many other semantic web tools have been built, including the ones that are used in the platform: the “ontology language” (discussed in the next section) and the database query language “SPARQL”.

Using Ontologies

The RDF model is made specific in conjunction with an ontology. These ontologies consist of classes, attributes, and relationships from a domain (Gruber, 2007, 2009) and specify what these entities are meant to denote in a machine-readable fashion (Gruber, 1993). Different ontologies have been developed by scholars and computer scientists (e.g., Ding, Kolari, Ding, & Avancha, 2007; Ruiz-Iniesta & Corcho, 2014). As such, existing ontologies can be used to embed tagged content within a larger semantic web. Ontologies can add meaning to tagged content by giving it certain properties and putting it in context by specifying its relationship to other content.

Ontologies add structure and, thereby, computational logic to RDF triplets. The key uses of ontology languages are to allow the creator of the ontology to (1) add explicit rules of inference (e.g., any study that contains the property “uses Data analysis” is an empirical study), (2) restrict the range of certain properties (e.g., “uses Data Analysis” can refer to a study, but not to a concept in the study), and (3) specify the types for objects (e.g., value tagged with “sample size/n” should be a numerical value). Once specified, these rules enable computer-based logical inferences to derive consequences for the tagging data (Bry & Schaffert, 2002; Grosof, Horrocks, Volz, & Decker, 2003; Gruber, 2009). For instance, if someone were to conduct a query looking for all empirical studies on manager charisma (e.g., in order to see whether there is sufficient data to conduct a meta-analysis on the topic) and the ontology were to specify that a study that "uses Data analysis" is an empirical study, the query would automatically show the aforementioned study (i.e., "Study on Manager Charisma") without requiring that study to be tagged explicitly as an empirical study.

The use of an ontology for the ReNotate platform ensures that the data can be aggregated more easily and more reliably. Therefore, the code tree in the ReNotate platform is built on an ontology. It is possible to let an ontology grow organically by allowing individuals to define new categories and rules whenever they are needed or use already created categories or rules whenever possible. For ReNotate, this would mean that readers could determine the categories and annotations relatively freely. However, in ontology languages it can be helpful to have an authority or domain expert oversee their development (Gruber, 2007). Like a language-regulating body for a conventional language (e.g., "Nederlandse Taalunie" for the Dutch language, "Asociación de Academias de la Lengua Española" for Spanish, the "Académie Française" for French in France, and "Office Québécois de la langue Française" for French in Quebec), the domain expert publishes the prescribed structure and dictionary for the language, responding to developments in his or her domain and in the use of the language. Having a domain expert oversee the development of an ontology is helpful in ensuring that there is a relatively stable and common language with which resources are classified, and safeguards the consistency in the underlying logic (e.g., Baclawski, Kokar, Waldinger, & Kogut, 2002; Simperl, 2009).

Considering the benefits of letting ontologies be developed by domain experts instead of allowing users of the ReNotate document reader freedom in determining which elements of ontologies to (re-)use, the ReNotate platform allows the user only a limited amount of control over the code tree and ontology from within the ReNotate document reader. In order to provide a reasonable amount of room to accommodate changes in requirements and also to allow the platform to extend to other fields (or even uses) in the future, we encourage ontologies and corresponding code trees to be created by domain experts in negotiation with the broader community (Gruber, 2007). As with ontologies in general, different code trees can be interconnected by sharing common fields when appropriate. In the ReNotate document reader, users can then select the ontology they wish to use. We provide a default ontology ourselves, to be used for annotating academic documents.

A ReNotate Ontology

As a default for the document reader and example ontology for this chapter, we have created a basic ontology helpful in tagging the key data from academic publications, with a secondary aim of allowing the identification of articles for meta-analytic studies within the field of management. This includes the identification within publications of concepts and corresponding conceptualizations, operationalizations, types of methods used and results reported, and of the sample sizes. Figure 3-3 provides a schematic overview of the key components of the ontology. Understanding the ontology requires understanding their (a) classes, (b) subclass relationships, (c) object properties, and (d) data properties. Each is considered in turn and data properties are described in detail. For a concrete example on how the default ReNotate ontology stores tagging data, we refer to Appendix 3-A. The full ontology, including its documentation, can be found at http://purl.org/net/re.[4]

Classes. The bigger nodes in Figure 3-3, in dark and light blue, are classes. Classes represent groups of resources with similar characteristics (Bechhofer et al., 2004; Bock et al., 2009). The base class "Work" represents "works that are published or potentially publishable, and that contain or are referred to by bibliographic references, or entities used to define bibliographic references" (Shotton, 2015). Once a paper is added to the ReNotate document reader it is defined to be of "type" "Work" and thereby becomes an instance of this class. It is subsequently specified further to, for example, "Research Paper" or "Working Paper". This dark blue class "Work" is external, meaning that it comes from another ontology[5]. The light blue classes are internal – i.e., they are from the ReNotate ontology.

[4] Dependent upon the type of request the server receives, the URL either leads to the documentation or to the source code of the actual ontology. More concretely, when accessed through a browser (and thus supposedly by a human being), the URL leads to the documentation of the ontology, including descriptions of the characteristics of its elements and definitions of the classes and properties. When accessed by a machine or software application, the source code of the ontology is provided. Such publication of the ontology is common practice for semantic web ontologies as it makes reuse of parts of the ontology easier – i.e., people can easily find out how the ontology can be used, and computers can be programmed to automatically retrieve and use the latest version of the ontology.

[5] The "Work" class is connected to a range of other classes and properties, corresponding with Peroni and Shotton's suggested approach to the classification of academic articles (Peroni & Shotton, 2012). The "Work" class also provides a container for the article metadata, such as "Title", "DOI", and "Creators". Due to space constraints, these are not discussed here. The interested reader is referred to Peroni and Shotton's excellent description of this ontology for the classification of academic articles.

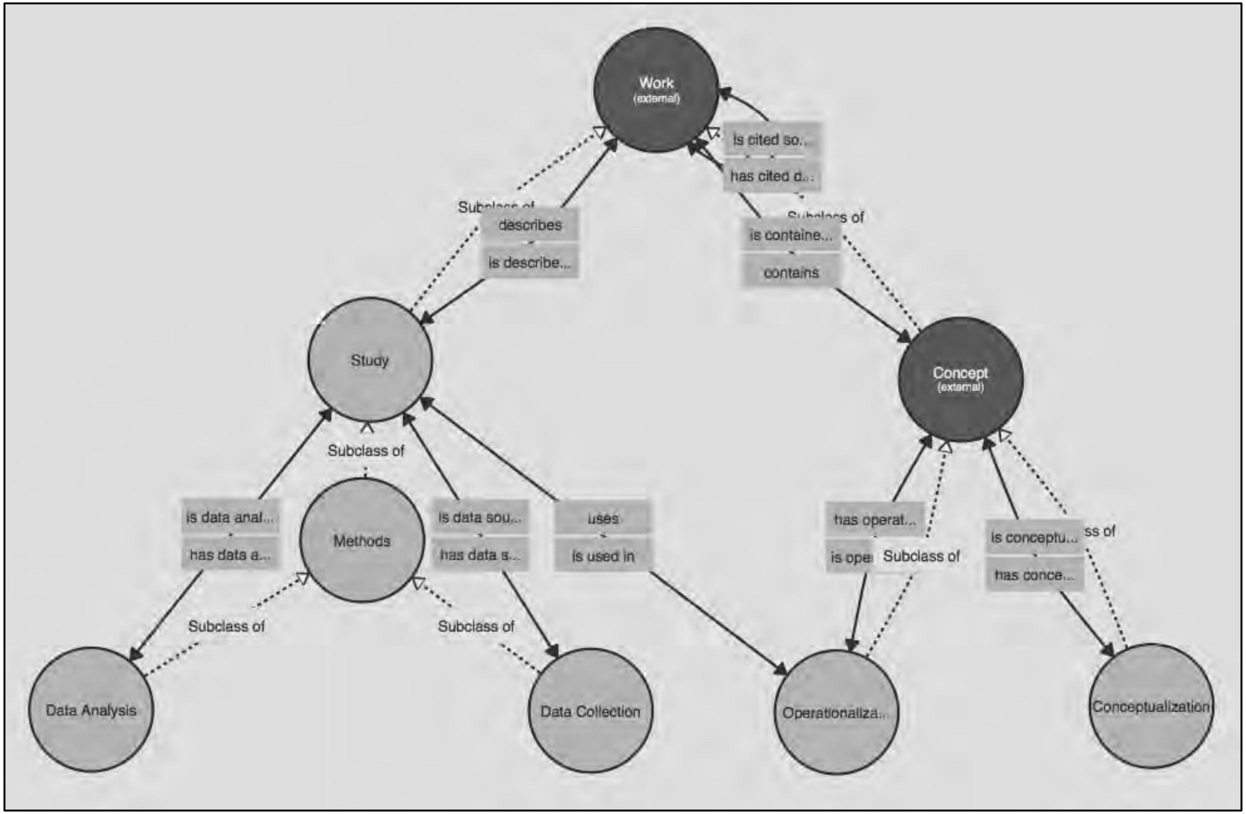


Figure 3-3. A Schematic Overview of the Default Ontology of the ReNotate Platform

Object properties. The range of possible *object properties*, or relationships between classes, is provided in blue on top of the arrows drawn between classes. They indicate what relationships (i.e., properties or predicates) there can be between instances of classes. They show, for example, that a "Work" can "describe" instances of the type "Study" (and inversely, a "Study" "is described in" a "Work"), where a "Study" is defined as "a detailed inquiry and analysis of a topic or context". This structure allows people to distinguish works that have no studies (e.g., a conceptual paper) from those that have one or more studies. In addition, for those works that have multiple studies, it becomes possible to specify what study each different underlying element relates to. Besides "describing" a "Study", a "Work" can also "contain" instances of the type "Concept". A "Concept" is defined as "an idea or notion; a unit of thought" (Miles & Bechhofer, 2009). In turn, the "Study" class and the "Concept" class have object properties that relate them to other classes – e.g., a "Study" "has" "data analysis" "Data analysis" and a "Concept" "has operationalization" "Operationalization".

Subclass properties. Some classes have black dotted lines drawn between them, and these lines carry the label "Subclass of". A "Subclass" relationship indicates that one class "is part of" another class. There is such a line, for instance, drawn from the "Data Analysis" class, pointing to the "Methods" class. This means that "Data Analysis" is a "Subclass" of "Methods". In other words, "Data Analysis" "is part of" "Methods". A subclass relationship allows for the "generalization" of properties from the parent class to the subclasses as the relationship between a class and its subclasses is transitive – i.e., if something applies to the parent class, it also applies to the subclass. In the presented ontology, for instance, when a certain "Methods" section "has real sample size" (i.e., "sample size after data cleaning or [an approximation of] the average sample size across all parts of the analysis") 364, the "real sample size" for the "Data Analysis" and "Data Collection" is also 364. When there is a discrepancy between the "real sample size" reported in the "Data collection" and "Data analysis" of a study, the "real sample size" can be set separately for each child class instead.

Data properties. Classes are linked to raw data or "literal values" through "data properties" (Bock et al., 2009). For the ReNotate platform these literal values come from the highlighted text in academic publications. The data properties are not shown in Figure 3-3, in order not to clutter the figure. Instead, we will describe the key data properties here.

Some data properties are shared by all classes in the ontology. Every class instance can have a label (i.e., a term that describes the instance), an abbreviation, a cited source, and a page number (or starting and ending page, if the class' text is spread over multiple pages). The other data properties pertain to only several or just one of the classes.

The "Methods" class and its subclasses are linked to the data property "has real sample size" (i.e., "number of cases in the sample before the data have been cleaned or a subsample has been selected"). The "Data collection" class is uniquely linked to the data properties "has real sample size (i.e., "sample size after data cleaning or [an approximation of] the average sample size across all parts of the analysis") and "has data collection type" (with ten preset values that categorize the method of data collection used, such as "interview", "survey", and "focus group"). The "Data Analysis" property has the unique property "has analysis sample size", which allows the sample size to be set for a specific analysis. If the "real sample size" has not been set explicitly for the "Data Collection" class and one or multiple values of the "analysis sample size" are provided, the average of the "analysis sample size" is used as the "real sample size".

Every instance of a "Concept", including its underlying components "Conceptualization" and "Operationalization", allows a separate "definition" and "description" to be provided. As such, a general concept definition and description can be provided, but it is also possible to provide a separate conceptual and operational definition or more detailed descriptions of a concept's conceptualization and operationalization. The links between "Conceptualization", "Operationalization" and their respective definitions are "functional", meaning that each instance of these classes can have only one definition (Bock et al., 2009). Should a new definition be provided, it is considered to be a separate conceptualization or operationalization of the parent concept. Lastly, the "Operationalization" has the unique data property "has measure", which allows the specification of measurement items (such as the questions in a survey) or the tagging of other means to gauge the value of a concept. Moreover, if the document contains a single study, it is assumed that the "Operationalization" was "used in" that study. If there are multiple studies, the "Operationalization" can be explicitly linked to a study. This helps to distinguish concepts that were merely mentioned in the study (Liakata & Soldatova, 2008) from those which were actually used (and which the study thus provides data on).

To summarize, the use of ontologies has several advantages for a collaborative data tagging platform such as ReNotate. Most importantly, by pre-specifying an ontology that is kept and developed separately from the tagged data, we ensure that the categorization to which data are connected is well developed and logically consistent. Furthermore, ontologies provide support for the correct reuse of common elements in different coding schemes, so that these data can be collectively aggregated and connected. Ontologies also allow logically consistent rules to be specified, which enables elements of the data to be inferred instead of being explicitly specified. These are especially helpful when used in combination with the SPARQL query language discussed on page **Error! Bookmark not defined.**. Before going into SPARQL, however, we explain how the data from different users reading and tagging documents can be validated and aggregated.

Data Validation and Aggregation

After a user has finished reading a document, their notes are synchronized between the ReNotate document reader and the server. This includes only the tag data (i.e., tagged text, highlight, and annotation) and the article metadata. The tags correspond to the ontology and therefore provide structure to the tagged data. However, even when readers use the same ontology, they might not always provide the same values. Therefore, before the data from the reader can be added to the searchable database, the data first needs to be compared with the data that is already available. This process is described below and includes (1) validating and merging or rejecting data from different coders of the same document, and (2) establishing common elements between documents. Before data are sent to the server, however, some steps are taken to make the aggregation more reliable.

In order to ensure that highlights and annotations of users are linked to the correct documents on the server, the platform uses strict matching criteria for the classification of the source document. When users import a document into ReNotate, the software uses *regular expression* to scan the document in search of a Digital Object Identifier (DOI; mainly for articles) or International Standard Book Number (ISBN or ISBN-A; mainly for books)[6] and, when found, article metadata are retrieved from the Crossref database (http://www.crossref.org). If a DOI is not available (e.g., for older articles or for books not registered in the DOI system), metadata are sought based on the data provided by the user (at minimum this requires the title and author to be specified). Secondly, common issues with PDF text extraction are corrected (i.e., words that are split across lines are merged to a single string and so is text that continues on a next page or column). In addition, tags with a numerical range are cleared from any non-numeric characters.

Comparing Notes Linked to the Same Publication

Ideally, the coordinates of the highlights, annotations, and text could be used to establish agreement between readers of the same article. Regrettably, however, we cannot assume that readers have used the same source document. Even when reading the same article, readers' copies may differ for several reasons. There are often various scanned versions of articles; there are versions distributed by the author(s); and certain databases add front or back pages to the document, making comparison difficult. Direct comparison between the source documents would require copies of the documents to be uploaded to the server, which would not only require huge amounts of bandwidth but also infringe upon copyrights.

Still, by using matching algorithms, it is possible to establish agreement between readers' notes without the source document available. When the text and tag from different readers of the same document (i.e., same DOI, not necessarily identically digitalized versions) are identical, it can reasonably be assumed that the tag-data is a true match. When one user has tagged an item while the other has not (e.g., one specifies the items of the measurement scale and the other does not), the text can generally still be accepted under normal circumstances. Since people tag with different purposes and different levels of diligence (Noll & Meinel, 2008; Yeung, Noll, Gibbins, Meinel, & Shadbolt, 2011), we can assume that neither user is at fault. However, when the text linked to the same tag is not identical it becomes more complicated, as this does not necessarily mean that the two taggers

[6] The ISBN-A, or actionable ISBN, standard incorporates the unique book identifier (ISBN) into the DOI identifier, such that it provides a unique and persistent link to a book's main webpage. Not every ISBN has an ISBN-A. It only exists for ISBN numbers that have been registered in the DOI system (https://www.doi.org/factsheets/ISBN-A.html).

disagree. One person might have highlighted more of the same text than the other. For instance, one person might highlight just the definition of the concept, while another also adds the qualifications that are provided in the next line. It is also possible that users have had different PDF source documents, where one of the documents converted the source text slightly differently.

In order to establish whether there is true agreement – and distinguish differences in tag length from actual disagreement – a q-gram metric is used (Gravano et al., 2001). This is a fuzzy string-matching algorithm that measures similarity by counting the pairs of q-character substrings that two strings have in common. We use substrings of character length 3 (i.e., 3-grams), thus each string of N characters is cut into $N - 3$ substrings. For example, imagine that two people reading the same article tagged the concept "self-efficacy", but one of the two only tagged the last part of the concept "efficacy" and left off the first part of the string. When determining whether the string pair j, comprising of string A (i.e., "self-efficacy") and string B (i.e., "efficacy"), are similar we first split them into 3-grams and thereby end up with the following two lists of lengths $n_a = 11$ and $n_b = 6$:

['sel', 'elf', 'lf-', 'f-e', '-ef', 'eff', 'ffi', 'fic', 'ica', 'cac', 'acy']

['eff', 'ffi', 'fic', 'ica', 'cac', 'acy']

Next, we remove common elements pairwise. Note that the pairwise removal is intentional; after a substring has found its equal, it cannot be matched a second time:

['sel', 'elf', 'lf-', 'f-e', '-ef', ~~'eff'~~, ~~'ffi'~~, ~~'fic'~~, ~~'ica'~~, ~~'cac'~~, ~~'acy'~~]

[~~'eff'~~, ~~'ffi'~~, ~~'fic'~~, ~~'ica'~~, ~~'cac'~~, ~~'acy'~~]

Finally, we take the sum total of the number of remaining substrings in list one (i.e., 5) and list two (i.e., 0) and this provides the difference score $d = 5$.

Two issues remain, however. Firstly, all characters occur in three-character substrings except for the first and last two characters. The first and last characters occur only once and the second and second-to-last characters occur twice. Due to this, a difference in the first and last two characters would not affect the difference score to the extent that a difference in any of the other characters would. To correct for this, each string gets a two-character prefix ("##") and suffix ("%%") (Gravano et al., 2001), extending the two lists to lengths $n_a = 15$ and $n_b = 10$:

['##s', '#se', 'sel', 'elf', 'lf-', 'f-e', '-ef', ~~'eff'~~, ~~'ffi'~~, ~~'fic'~~, ~~'ica'~~, ~~'cac'~~, ~~'acy'~~, ~~'cy%'~~, ~~'y%%'~~]

['##e', '#ef', ~~'eff'~~, ~~'ffi'~~, ~~'fic'~~, ~~'ica'~~, ~~'cac'~~, ~~'acy'~~, ~~'cy%'~~, ~~'y%%'~~]

which changes the difference score d to 9.

Second, the difference score is affected by the length of the strings. A comparison of two long strings is far more likely to lead to a high difference score than a comparison of two short strings. Therefore, we divide the difference score by the maximum difference score possible $n_a + n_b$ (i.e., the sum total of elements in the lists).

$$\gamma_j = 1 - \frac{d}{(n_a + n_b)}$$

Finally, by inverting difference score d we attain the probability score γ for the string pair j:

$$\gamma_j = 1 - \frac{9}{(15 + 10)} = .64$$

which gives us agreement score $\gamma_j = .64$ for the string pair j (i.e., "self-efficacy" versus "efficacy").

The text that is assigned to the tag and eventually provided in the database is the part of the text most commonly tagged, where the tag frequency is weighed by the proficiency scores of the taggers. For determining the proficiency of the taggers we employ Noll and colleagues' (2009) and Yeung and colleagues' (2011) heuristic, differentiating users, based on (1) the extent to which their tags match those of others, and (2) their ability to contribute relevant tags before others do. In order to prevent "spamming", a common nuisance in tagging systems (Gruber, 2007), negative tagging (i.e., "down voting") of aggregated data is possible, and these tags also affect the proficiency score of the taggers whose tags received a negative tag. The exact weighing of proficiency scores is to be determined empirically with data attained with the pre-Alpha release of the software, starting February 2016.

Establishing Links between Different Publications

Once agreement between readers of the same article has been established and the values for the different tags in the article have been selected, common elements between articles are sought. This is necessary in order to make it easier to query the collective data. Specifically, the database is much more easily navigable if the relatedness between instances is determined and synonyms are identified. Plus, the advocated approach allows polysems (i.e., different words with the same name, such as "book" referring to a written or printed work or "to book" as in to reserve or buy a ticket in advance) to be identified, which is especially relevant for concept tagging, but also relevant for, for instance, author names.

Let us elaborate using an example. Imagine that two articles have been tagged that both contain the concept "work-life conflict" – i.e., "a form of inter-role conflict in which the role pressures from the work and family domains are mutually incompatible in some respect" (Greenhaus & Beutell, 1985: 77). However, one of the articles uses the name "work/non-work conflict" instead of "work-life conflict" and the two have a relatively low agreement score $\gamma_j \approx .636$. Had one of the articles mentioned both names the concepts would have been linked (as the key classes can be given multiple names in the ontology we provided), but this is not the case. Thus a different approach is required to allow such concepts to be matched.

To enable true matching fields from different articles to be matched, even when the values on those fields do not match, we also take the "context" of the tags into consideration. For each class, this context includes all subclasses (i.e., classes that "are part of" the class) and properties. We compare the context using the q-gram score within a probabilistic record-linkage model based on the Fellegi and Sunters (SFS) metric (Fellegi & Sunter, 1969; Newcombe et al., 1959). When, for a given class, the context matches with the context of the alter(s), the likelihood of that tag being similar to the alter(s) is also higher. This metric was initially invented for database merging, as it allows, for instance, street names to be merged when the city and zip code are alike. It is most commonly used to match people's medical records across separate databases (Clark, 2004; Herzog, Scheuren, & Winkler, 2007; Winkler, 1999). In web-tag merging, a similar context-matching approach has been found to be highly effective (e.g., Abel, Henze, & Krause, 2008; Cattuto, Benz, Hotho, & Stumme, 2008). In the interest of brevity, we limit the context variables used in the example to "Concept Name", "Concept Definition", and the "Cited Source" of the definition in the following values for two articles that are to be compared:

Table 3-1. Comparison of Two Coded Elements

	Article 1	Article 2	q-metric (γ_j)
Concept Name	"work-life conflict"	"work/non-work conflict"	.59
Concept Definition	"a form of inter-role conflict in which the role pressures from the work and family domains are mutually incompatible in some respect."	"occurs when there are incompatible demands between the work and family roles of an individual that makes participation in both roles more difficult."	.37
Cited Source	Greenhaus & Beuttel, 1985	Greenhaus & Beuttel, 1985	1

The probabilistic record-linkage model that we use is based on two types of error with respect to each value (Fellegi & Sunter, 1969): (1) the probability m that the value agrees, given that the pair of records j truly are a match, and (2) the probability u that the value agrees, given that the pair is not a true match. For a concept, m is a function of the amount of disagreement due to spelling differences or mistakes plus the inverse of the incidence of synonymy in the dataset. This can be the incidence for the specific concept, but if this is not available (or the amount of data is too limited), the incidence in the dataset as a whole can be used. The probability u, in turn, is equal to the incidence of polysemy (i.e., two identical terms having the same meaning). Currently, we do not yet have the data required to derive these probabilities for the three fields, so let us assume the following for the purpose of demonstration:

Table 3-2. Match (m) and Unmatch (u) Probabilities for Each Field

	m	u
Concept Name	.75	.03
Concept Definition	.5	.0001
Cited Source	.4	.001

Thus the probability of the concept names agreeing, when the data truly concerns the same concept, is estimated at .75 (i.e., 75%) and the probability of the concept names being the same, while in truth relating to different concepts, is estimated to be .03 (i.e., 3%).

Next, the records from the different articles are compared. When the two records agree on a field, an agreement weight is computed by dividing the probability m by the probability u and taking the $\log_2$ of the quotient. For the above estimates of Concept Name the weight would thus be $\log_2(\frac{.75}{.03}) \approx 4.64$. When the fields are not in agreement, a disagreement weight is computed by dividing 1 minus the probability m by 1 minus the probability u. For the Concept Name field this would be $\log_2\left(\frac{1-.75}{1-.03}\right) \approx -1.96$.

The probabilistic weight is the sum of all weights for the fields of the compared records. We do not take agreement as a binary variable, however, but instead use the degree of agreement and disagreement (i.e., the inverse of agreement) from the q-gram metric and multiply this by the SFS agreement and disagreement weight: [7]

$$w_j = \sum_{k=1}^{n} \gamma_k^j \log_2\left(\frac{m_k}{u_k}\right) + \left(1 - \gamma_k^j\right) \log_2\left(\frac{1 - m_k}{1 - u_k}\right)$$

where w_j is the weight for record pair j and is the sum of the weights on the n fields, the value y_j is the q-gram metric score for the value pair (1 = agree, 0 = disagree), m the estimated agreement rate for true matches, and u the estimated agreement rate for non-matching value pairs.

Applying this to the elements from the example, the overall agreement weight w_j increases by ~2.96 due to the relatedness of the concepts, ~4.24 due to the relatedness of the definitions, and ~8.64 because of the agreement on the cited source of the definition. This brings the final weight (i.e., the sum of the three) to $w_j \approx 15.13$.

In the final step the weight w_j is compared to the "true-link" threshold (where it can safely be assumed that the classes are similar). The pair is considered "linked" and the classes are merged if the value is above the threshold and dissimilar if it is below the threshold (Grannis, Overhage, Hui, & McDonald, 2003). It is also possible to have a certain range where the pairs are kept separate for human review (Fellegi & Sunter, 1969). This threshold will be determined using training data.

---

[7] This diverges from the more commonly used approach to weighting (Grannis, Overhage, Hui, & McDonald, 2003; Wajda & Roos, 1987; Winkler, 1990, 1999), which is based on a string comparison metric by Jaro (1989) and generally has a positive agreement weight for a field even when the degree of agreement between the strings is low:

$$w_j = \sum_{k=1}^{n} \log_2\left(\frac{m_k}{u_k}\right)^{\gamma_k^j} \log_2\left(\frac{1 - m_k}{1 - u_k}\right)^{1-\gamma_k^j}$$

In contrast, the q-gram metric is continuous and some agreement can be expected even between unrelated strings. Thus, the approach described above is used, as it still provides a negative agreement weight when the degree of agreement between the strings is low.

The described algorithm is applied to every newly coded document, comparing its coded instances against the existing instances in the database. In the interest of computational efficiency, matched value pairs are combined (Christen & Goiser, 2007) for the comparison to consecutive value pairs using the most common values for each field in the comparison. This also ensures transitivity between articles – e.g., if concept x_1 in article A equals concept x_2 in article B, and concept x_2 in article b equals concept x_3 in article c, concept x_1 equals concept x_3 – which can otherwise be a concern in record linkage (Cohen, Ravikumar, & Fienberg, 2003).

Releasing the Data

The aggregated RDF data are stored online and can most easily be accessed using the SPARQL query language. SPARQL stands for "Simple/SPARQL Protocol and RDF Query Language". SPARQL was created specifically for queries on RDF graph data. It uses pattern matching (Pérez, Arenas, & Gutierrez, 2006). In pattern matching (Gimpel, 1973), the user provides a sequence of tokens representing a range of possible patterns and an algorithm is used to find instances that match this pattern.[8] Likewise, in SPARQL, the user provides a search pattern representing graph data and SPARQL tries to match this pattern to the available data.

There are several advantages to the use of SPARQL over the use of other query languages (e.g., the commonly used relational database query language "Structured Query Language", SQL) for the tagged data. Firstly, SPARQL queries are relatively intuitive. Data are queried using properties, which are explicit descriptions of the relationships. In contrast, in the queries for relational databases, the nature of these relationships is not always clear. Secondly, SPARQL leverages the underlying ontologies to derive properties not made explicit (e.g., making it possible to use the inverse of properties) and include inherited properties (e.g., the "Conceptualization" and "Operationalization" also get the name of their parent class "Concept") or enable references to parent properties (e.g., select "Concept" that is "is contained in" a scholarly "Work" that "has Creator" with "family Name" "Granovetter"). Third, it can make use of the metadata of resources. For example, users

8 Regular expression, used to find the DOIs in documents as described earlier, uses pattern matching to match a search pattern to character strings (Mitkov, 2005).

could search for data entered after a certain date or articles coded by X number of readers. Finally, it also makes use of the URIs, making it possible to distinguish resources by their unique identification numbers. The use of SPARQL would not only make it easier to distinguish polysems; it would also make it possible to distinguish between articles with the same title or between authors with the same name.

A possible disadvantage of letting the data be queried through SPARQL is that the queries can be relatively complex and therefore difficult to create correctly. Therefore, in order to make the database accessible to those who do not feel comfortable with SPARQL, we will also be providing a visual query interface, based on the Resource Descriptive Framework Graphical Language "RDF-GL" (Hogenboom et al., 2010).

Discussion

This chapter presents the basic functionality of the ReNotate software platform, including a description of the semantic web technologies used, the algorithms employed to validate and aggregate the data from different users, and an explanation of how this aggregated data can be queried. The platform serves two primary purposes. Firstly, it provides a document reader that (a) helps scholars and students structure the highlights and annotations of their readings, and (b) enables those highlights and annotations to be quickly searched and reviewed in a separate personal database. Secondly, it aggregates the data from the individual readers in order to provide a searchable database across all articles highlighted and annotated. As such, it promises an overview of the literature that is emergent, evolving, and crowdsourced.

As such, the platform caters to a broad range of users. We put forward an ontology that aims to be helpful for the individual, the collective, and especially for those who seek to conduct meta-analyses. Furthermore, the ReNotate platform allows multiple ontologies to be used in conjunction, thereby greatly extending its potential range. For instance, ontologies could relatively easily be extended to capture (a) hypotheses and their results, (b) theories that are drawn upon, (c) formulas, or (d) even the reasoning used (e.g., Chesñevar et al., 2006; Rahwan, Zablith, & Reed, 2007; Schneider, Groza, & Passant, 2013).

The approach described does, however, have its limitations. First, although the framework provides flexibility, it is not optimally suited for several specific-use cases for which it could be a highly effective tool. For example, while the example ontology we put forward promises to help attain an overview of the concepts in the literature, it does not capture relationships between concepts. In order to capture such information, extensions would need to be created. Below, we discuss several options as to how this could be addressed, such that the ReNotate platform can be used to attain a hierarchical taxonomy of academic concepts.

Second, the reliance on readers to attain data is both a strength and a weakness. When relying on "ordinary" readers only, there is no guarantee that the notes will systematically cover either (a) the relevant literature, or (b) that all the elements in the default ontology will be entered. Indeed, it is highly likely that certain "core" readings will be covered extensively, while peripheral readings may not be coded by anyone. Moreover, there is no guarantee that all the relevant text will be coded. In order to counter this and ensure better coverage of the literature, we discuss two possible future extensions to the platform. These include an extension that makes ReNotate more useful as a pedagogical tool, plus an extension that would make the ReNotate document reader more useful for individuals coding articles for their own meta-analytic studies. Like the ReNotate platform as a whole, these extensions build on the notion that the platform should independently serve the interest of the individual and the collective in order to attain a strong user base and maximize its benefits.

Potential ReNotate Extension #1: An Evolving Taxonomy of Academic Concepts

The ReNotate platform that we describe allows concept tagging and enables people to look up definitions of concepts, the related measures, and the source articles and pages of those concepts. However, it does not currently allow relationships between concepts to be specified. Fortunately, ReNotate could be extended in this direction. It can even be argued that the semantic tools that are used, RDF in combination with an ontology, are optimally suited (or even built) for the creation of a hierarchical classification of concepts (e.g., Berners-Lee et al., 2001; Gruber, 2007). The semantic web allows relationships to be specified, permits synonyms to be captured, and can help in keeping polysems separate or even separating the polysems through rule-based inference. It can also let concepts be located in different places in a hierarchy simultaneously (e.g., an ethnography is both a method of analysis and a method of data collection), which could be considered more realistic than a rigid mutually exclusive structure (Gruber, 2007).

However, properly linking ontologies requires considerable deliberation (e.g., Floridi, 2009; Shirky, 2005). Inconsistencies are to be expected even when combining two planned and well-structured ontologies (e.g., Baclawski et al., 2002; Simperl, 2009). When letting ontologies emerge without supervision, such inconsistencies can easily become unmanageable. So, although it is possible to use the ReNotate concept ontology to create a hierarchical classification of the literature – i.e., by letting readers specify the relationships between concepts – it might not lead to a coherent image. Within the academic literature, even though concepts and their relationships are particularly well thought-through, discrepancies can be expected just the same. Different classifications of the same concepts are not always in agreement and thus, when combined, might not be logically consistent. Indeed, a large part of the scholarly discussion revolves around finding the most appropriate definitions of and most significant relationships between concepts (e.g., Li & Larsen, 2011). From this we can, at best, capture the consensus at certain points in time. And considering its rule-based nature, a semantic web ontology might not be the best tool to use – at least not by itself.

Reasonably, one could expect the best hierarchical classification of concepts when combining the strengths of the ReNotate platform as a way of crowdsourcing data from academic publications, with other (semantic) techniques (Hitzler & Harmelen, 2010). In this respect, one of the most promising approaches at present is arguably the one described and used by Li and Larsen (2011), which is based around the "Inter-Nomological-Network" (INN) analysis (Cook, Larsen, Sakraida, & Pedro, 2012; K. R. Larsen, Lee, Li, & Bong, 2010; K. R. Larsen & Monarchi, 2004). Larsen and colleagues have collected concepts and the relationships between concepts (amongst other things) using mostly Natural Language Processing (NLP) techniques. In their approach, hypotheses are extracted using regular expression. From these hypotheses, concepts and the hierarchical relationships between the concepts are identified using NLP algorithms for the detection of grammatical structures in sentences and to find the dimensions underlying a range of concepts.[9]

Using components of the INN analysis within the ReNotate platform could provide a powerful approach to the creation of a concept classification from the academic literature. Specifically, a "hypotheses" tag could be incorporated into an ontology and code tree. Next, the INN approach to the extraction of relationships between concepts could be applied to the tagged and aggregated hypotheses. Not only would this provide a way to continuously model the consensus in the literature as readers' tagging data are attained, it would also relieve the reader from having to specify the relationships between concepts – thereby decreasing the reliance on their interpretation and, most likely, more closely reflecting the evolving consensus in the literature instead. In comparison to the regular INN analysis, which uses regular expression to find hypotheses following the common structure, the ReNotate platform would likely provide a much higher recall of hypotheses from the literature.

However, before this approach can be put into practice, further research is required is in order to know the precision and recall that can be expected within the ReNotate platform versus the INN analysis. This enquiry is not only technical, as the best outcome would also need to make optimal use of the capabilities of individual readers while ensuring that the data that are tagged are also of interest to them, even before they are aggregated. More concretely, how likely is a user of the ReNotate document reader to tag a hypothesis and the concepts within it? And how accurate are the aggregated data?

Potential ReNotate Extension #2: Developing a Platform for Classroom Exercises

The success of the ReNotate platform depends on the number of people using the ReNotate document reader to highlight and annotate academic publications, the quality of their tags, and the degree of coverage of the tags over the literature. One way to increase the likelihood that there will be a large user base, one for which there is control over the quality

[9] Due to space restrictions we were only able to provide a highly condensed description of a selection of relevant techniques used in the INN endeavor. We acknowledge that the INN approach encompasses considerably more than what is described here.

of the tags and whose tags cover a relatively broad range of publications, is by creating an extension for the platform such that the ReNotate document reader becomes a helpful pedagogical tool for classroom exercises.

Highlighting and annotation are generally considered to be the most frequently used learning strategies (e.g., Feito & Donahue, 2008; Simpson & Nist, 1990). On the one hand, they enhance reading comprehension. On the other, they can improve writing skills; by understanding the structure of academic texts, one can get a better sense of how to structure one's own academic writing (Marshall, 1997, 1998). However, the benefits to highlighting and annotation are minimal if the reader does not review the notes afterward (e.g., Hartley et al., 1980; L. L. Johnson, 1988; Moreland, Dansereau, & Chmielewski, 1997). In order to help students, some have advocated that courses should be provided that teach annotation and highlighting skills, along with periodic review (e.g., T. E. Johnson, Archibald, & Tenenbaum, 2010; Porter-O'Donnell, 2004). The ReNotate document reader can be used for such classroom exercises. Not only does it help the reader structure highlights and annotations, but it also helps in reviewing the annotated data. Thereby the text extraction capabilities can be leveraged to give the teacher more detailed insight into the highlights and annotations of the learner. Moreover, it could be used for student evaluation and for the allocation of course credit.

In order to make sure the platform is used and is useful as a pedagogical tool (beyond what can currently be done using simple printed materials), a web interface would need to be built that provides several basic capabilities for instructors. The web interface would allow instructors to select the preferred ontology, add or hide tags, and send that ontology to the students. This is because, for some classroom exercises, only parts of an ontology might be relevant. For instance, if the course focuses on research methods, only the methods part of the default ontology would need to be visible for students. For other courses, it might be necessary to add classes and properties to the ontology. For example, a course on Communication or on Logic could have students code for Toulmin's argumentative structures (e.g., Chesñevar et al., 2006; Rahwan et al., 2007; Schneider et al., 2013).

The web interface should also allow instructors to peruse the highlights and annotations of students. For this, we can also provide some basic supporting metrics. These metrics could include the comparison between a student's tags and the "correct" or "preferred" tags. The correct tags could come from the instructor or could consist of a comparison between students who were assigned the same article using a similar metric as is used for the coding proficiency. In order to establish the validity of these metrics as a teaching and evaluation tool, however, further research would be required.

The instructions for teachers would include the following steps: (1) setting the ontology, (2) selecting the readings for students or setting the domain, (3) setting the "correct" or "preferred" tags for each paper (optional), (4) providing instructions on how to tag, (5) evaluating students' notes, and (6) sending feedback and/or assigning credit. For students the instructions would build on the following points: (1) read an instruction manual for the ReNotate document reader, (2) use library access to find articles, (3) code the assigned articles, checking the ontology definitions whenever it is unclear what to highlight, (4) compare personal notes with notes from others and analyze discrepancies, and (5) recode the articles based on the teacher's feedback.

These sets of routines turn coding publications into a pedagogically beneficial experience for university students and instructors. Provided the default ontology is used frequently in these classroom exercises, this set of routines holds the potential to ensure comprehensive coverage of the relevant literature, such that individuals can do things like search for all the definitions of a concept found in the literature. With this comprehensive coverage, it also helps facilitate the identification of studies for meta-analyses.

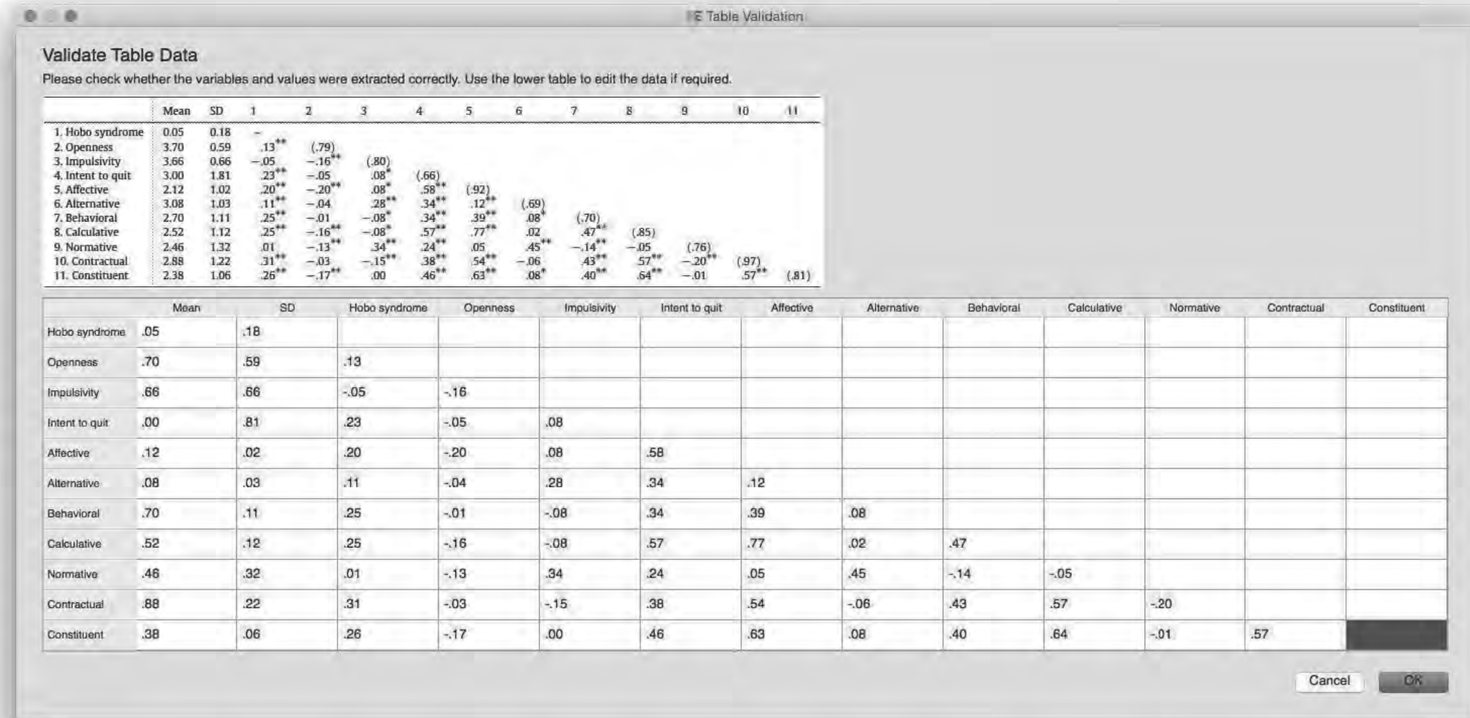
Potential ReNotate Extension #3: Coding Meta-analytic Studies

The ReNotate platform, as presented in this paper, holds the potential of making meta-analytic studies easier to conduct. It can provide the information necessary to know (1) whether there are enough data available to conduct a meta-analysis on a certain topic and (2) where those data can be found. Especially if the coverage of the field is high, this should save a considerable amount of time. The contribution of the ReNotate platform could extend further, however, if it were to help meta-analysts code studies. As a corollary, this would provide an additional source of expert coding of the data for the platform. Developing these capabilities would require several ReNotate software extensions.

As a first step, data extraction from academic articles could be made easier. Specifically, table data are important (Cheung, Samwald, Auerbach, & Gerstein, 2010), as tables often contain the effect sizes, sample sizes, and standard deviations on which meta-analyses are based. Currently these data are often collected using simple copy and paste or non-specialized text extraction software.

In order to show how the document reader can help in collecting meta-analytic data, we will be including table data extraction functionality in the Alpha and Beta releases of the ReNotate document reader (due in February 2015 and mid 2016 respectively). When tagging table data, the user is led through several steps in order to ensure that the data are correctly extracted. First, the user needs to draw a box around the table and identify its main axis. Next, the software extracts the content of the table and looks for links between the variables coded in the text and the terms in the table. In the final step (shown for a correlation table in Figure 3-4), the user can validate the data and change the variables and the values when necessary. This is not just helpful for the meta-analyst. Should this functionality be used frequently, it would provide the data required for near-instant (rudimentary) meta-analyses.

Still, data extraction alone may not be sufficient for making the ReNotate document reader an attractive tool for the coding of meta-analytic studies. A second potential modification is the ability to interact with the aggregated data – i.e., to create new "compound variables" (consisting of a combination of different variables) or re-categorize some of the operationalizations to fit a meta-analyst's conceptual schema. Lastly, the ontology should contain items for common codes in meta-analyses, such as a code for the "quality" of the study and classes that allow study-level moderators.



	Mean	SD	1	2	3	4	5	6	7	8	9	10	11
1. Hobo syndrome	0.05	0.18	–										
2. Openness	3.70	0.59	.13**	(.79)									
3. Impulsivity	3.66	0.66	−.05	−.16**	(.80)								
4. Intent to quit	3.00	1.81	.23**	−.05	.08*	(.66)							
5. Affective	2.12	1.02	.20**	−.20**	.08*	.58**	(.92)						
6. Alternative	3.08	1.03	.11**	−.04	.28**	.34**	.12**	(.69)					
7. Behavioral	2.70	1.11	.25**	−.01	−.08*	.34**	.39**	.08*	(.70)				
8. Calculative	2.52	1.12	.25**	−.16**	−.08*	.57**	.77**	.02	.47**	(.85)			
9. Normative	2.46	1.32	.01	−.13**	.34**	.24**	.05	.45**	−.14**	−.05	(.76)		
10. Contractual	2.88	1.22	.31**	−.03	−.15**	.38**	.54**	−.06	.43**	.57**	−.20**	(.97)	
11. Constituent	2.38	1.06	.26**	−.17**	.00	.46**	.63**	.08*	.40**	.64**	−.01	.57**	(.81)

	Mean	SD	Hobo syndrome	Openness	Impulsivity	Intent to quit	Affective	Alternative	Behavioral	Calculative	Normative	Contractual	Constituent
Hobo syndrome	.05	.18											
Openness	.70	.59	.13										
Impulsivity	.66	.66	-.05	-.16									
Intent to quit	.00	.81	.23	-.05	.08								
Affective	.12	.02	.20	-.20	.08	.58							
Alternative	.08	.03	.11	-.04	.28	.34	.12						
Behavioral	.70	.11	.25	-.01	-.08	.34	.39	.08					
Calculative	.52	.12	.25	-.16	-.08	.57	.77	.02	.47				
Normative	.46	.32	.01	-.13	.34	.24	.05	.45	-.14	-.05			
Contractual	.88	.22	.31	-.03	-.15	.38	.54	-.06	.43	.57	-.20		
Constituent	.38	.06	.26	-.17	.00	.46	.63	.08	.40	.64	-.01	.57	

Figure 3-4. Validating Data Extracted from a Correlation Table

Besides individual meta-analysts, this extension could possibly also benefit a much grander project called MetaBUS (Bosco, Steel, Oswald, Uggerslev, & Field, 2015; Bosco et al., 2014). With this project, Bosco, Ugerslev, Steel, and other collaborators aim to collect the data required for meta-analyses and release it for public use, such that instant meta-analyses can be conducted. Within the last two years, all articles from 15 core management journals have been coded (Bosco, Aguinis, et al., 2015). Moreover, the aim is to ultimately extend this to other journals and fields (Bosco et al., 2014).

A dedicated coding tool could be valuable for this endeavor. Although highly effective, the project requires a considerable amount of resources, as data entry involves a dedicated team of highly trained coders. Although progress in coding Management articles has been impressive already, the use of the ReNotate platform by MetaBUS could accelerate the coverage of the scientific literature.

Conclusion

In conclusion, the platform we describe has the potential to make science more effective and efficient, within the field of management and beyond. The ReNotate platform should make it easier to recognize what has been studied and what is still unknown. Especially as the discussed extensions are developed, ReNotate holds the promise of providing an overview of a broad range of literature and would have the ability to continuously leverage the sensemaking efforts of numerous readers, thereby providing an evolving overview of the field. In essence, ReNotate could allow individuals to more readily access the collective wisdom of the scholarly community.

Appendix 3-A: Capturing Tagging Data with the Default ReNotate Ontology

Since the notion of an ontology and the description of the ReNotate ontology are relatively abstract, we provide a concrete demonstration of the ReNotate ontology in practice.[10] Consistent with norms from the computer science field, we have used the following abbreviations in order to make the syntax more succinct and readable:

- "re" refers to "http://purl.org/net/re#",
- "fabio" to "http://purl.org/spar/fabio/",
- "skos" to "http://www.w3.org/2004/02/skos/core#", and
- "xsd" to "http://www.w3.org/2001/XMLSchema#".

FaBiO is an ontology intended for the encoding of "bibliographic records of scholarly endeavors" (Peroni & Shotton, 2012), SKOS is a more general ontology for the classification of knowledge (Miles & Bechhofer, 2009), and XSD is used for defining data types (e.g., "integer", "string", "date").

The tagging data that we show here come from an article entitled "A Preliminary Exploration on the Measurement of Expertise: An Initial Development of a Psychometric Scale" by Germain and Tejeda (2012). As indicated in the second line of the syntax, the article describes two studies, which are referred to as "<study1>" and "<study2>". It also contains a concept labeled "expertise" which has a conceptualization coming from a certain book "<book>" and an operationalization with its own a) label, b) abbreviation, and c) measurement items:

```
<journalarticle> a fabio:JournalArticle;
      re:describes <study1>,
         <study2>;
      re:mentions [ a skos:Concept;
         re:has_concept_name "expertise";
         re:has_conceptualization [ re:has_cited_source <book>;
                 re:has_page_number 205
                 re:has_defintion "Displayed behavior within a specialized
                 domain and/or related domain in the form of consistently
```

[10] This example has been written in "Turtle", a syntax for the expression of RDF data. For further documentation on Turtle we refer to Becket and colleagues (2011).

```
                demonstrated actions of an individual that are both
                optimally efficient in their execution and effective in
                their results" ];
        re:has_operationalization [ re:has_abbreviation "GEM";
            re:has_concept_name"generalized expertise measure";
            re:has_starting_page 204
            re:has_ending_page 220
            re:has_definition "The GEM is based on employee expertise as
            perceived and reported by another person.";
            re:has_measure "This person conducts research related to their
            field.", "This person has knowledge about their field.", "This
            person has knowledge that is specific to his or her field of
            work.", "This person has the qualifications required to be an
            expert in their field.", "This person shows that they have the
            education necessary to be an expert in their field." […]
```

Below we show the tagging data attained for the article's second study – i.e., "study2". It describes how a "snowball method" was used to collect the sample for "study2" and that this data source was analyzed using "method of analysis" "<analysis3>" and "<analysis4>":

```
    <study2> a re:Study;
            re:has_data_source [ a re:DataSource;
                    re:had_method_name "snowball method";
                    re:has_page_numer 215;
                    re:has_n_initial 307;
                    re:has_method_description "A one-time e-mail solicitation
                    to participate in an anonymous study was sent to
                    participants.[…] The e-mail invited those 10 participants
                    to forward the online survey to coworkers, hence using a
                    snowball method of data collection." ];
            re:has_method_of_analysis <analysis3>,
                    <analysis4> .
```

Finally, to indicate how an analysis is classified using the ReNotate ontology in this chapter, we provide the below syntax, which shows that a subsample was selected from the data source described earlier in order to conduct an Exploratory Factor Analysis:

```
    <analysis3> re:has_method_name "Exploratory Factor Analysis";
            re:has_method_abbreviation "EFA";
            re:has_analysis_sample_size 142;
```

```
re:has_starting_page 216;
re:has_ending_page 217;
re:has_method_description "The purpose of the exploratory analysis
was twofold: (1) to determine whether the factor structure
developed by the panelists was manifest in the data collected and
(2) to examine the statistical patterns of loadings in the data
reduction that could guide the elimination poor items toward and
thereby create a final scale […]
```

As such, an ontology allows complex data to be structured and connected with existing data or even existing ontologies (e.g., Fabio skoes, xsd). This structure can then be used to query the data.

Chapter 4: Who gets Invited for a Job Interview?

Abstract

On what basis do hiring professionals decide whom to invite for a job interview? In order to address this fundamental but understudied question we draw on applicant tracking data from 48 companies (*N* = 441,769 applicants). We use synthetic validity and relative weight analysis to identify the relative importance across occupations and industries of 18 factors including demographics (i.e., age, gender, nationality, distance from hiring company), biodata (i.e., experience, level of education), the application (i.e., internal or external candidate, applied before deadline) and the applicant pool (i.e., applicant count, average percentage invited, occupation vacancy rate). We correctly predict invitation outcomes on a holdout sample for 2,539 of the 3,656 applicants invited (sensitivity of 69.45%) and 27,876 of the 40,531 that were not invited (specificity of 68.78%). For those vacancies that did not seem to require a cover letter the accuracy was even higher with 151 out of 183 invited (82.51%) and 2,832 out of 3,528 not invited (80.27%) correctly classified. Age and experience emerged as the most relevant in applicant prescreening. Results further suggest that hiring professionals' are relatively inconsistent in how they adjust their evaluations with respect to demographics and biodata to the occupation and industry.

Introduction

As the employee selection process has moved online, the number of applications companies receive per vacancy has grown substantially (Dunleavy, Mueller, Buonasera, Kuang, & Dunleavy, 2008; Gilster, Davison, & Dickmeyer, 2001). While studies suggest that most applicants were already rejected in the first stage of the selection process before (Carlson, 1972; Cole, Rubin, Feild, & Giles, 2007; Levine & Flory, 1975), the rise in the application counts will likely have further increased hiring professional's reliance on prescreening, thus rejecting applicants before conducting interviews or assessments. Prescreening is where most applicants get rejected and yet the employee selection literature currently provides limited insight into the factors companies rely on when prescreening applicants, the relative importance of screening factors, and as such, offers relatively little guidance on what companies can do to make the prescreening stage more effective.

Most of what we know about applicant screening comes from scenario studies in which a number of key attributes are manipulated in the application in order to find how these attributes affect the artificial selection decision (e.g., Derous, Ryan, & Nguyen, 2012; Knouse, 1994; Schröder, Rogers, Ike, Mell, & Scholl, 2013) and from policy capturing studies (e.g., McKinney, Carlson, Mecham, D'Angelo, & Connerley, 2003; Thoms, McMasters, Roberts, & Dombkowski, 1999; Young, Young, & Oto, 2011), generally using résumés of students' or job incumbents in order to see how résumé attributes affect the evaluation of applicants (Karren & Barringer, 2002). While often insightful, the reliance on artificially constructed studies and non-representative samples raises some questions about the generalizability of the current knowledge on applicant selection and applicant prescreening. Results from student samples or job incumbents might not hold with respect to job applicants in general (Breaugh, 2009). In many ways students' résumés are not representative; for instance, they are much less likely to have prior work experience and possibly provide restricted variance on levels of education in comparison to job applicants in general. Moreover, comparisons of job incumbents with job applicants have shown that measures that help predict the best job incumbents can have much lower validity (Stokes, Hogan, & Snell, 1993) or no-validity (Bliesener, 1996; Harold, McFarland, & Weekley, 2006) in finding the best job applicants. Finally, some have raised concerns about the ecological validity of policy capturing studies, because these studies may be influenced in complex ways by social desirability, as this varies with age, experience, and design (Mazen, 1990).

To help advance research on applicant prescreening, this study takes up the challenge of testing the relative importance of the various screening predictors in a field setting across occupations and industries. We examine variables that prior research has shown to be important and test their predictive value using state-of-the-art tools for data collection and analysis. In doing so we develop a framework that provides support to companies in making pre-screening decisions. And most importantly, through this study, we are able to describe (and replicate) how applicant prescreening is currently conducted as a benchmark for future scholarship.

In this paper we first provide an overview of the key variables used in applicant prescreening. Next, we describe how we can predict the selection decisions of candidates on specific occupations in different industries using a data-analytic technique called *synthetic validity* (Balma, 1959; Lawshe, 1952; Steel, Huffcutt, & Kammeyer-Mueller, 2006) and applying this technique on data from Applicant Tracking Systems (ATSs). Using *relative weight analysis* (Tonidandel & LeBreton, 2010, 2011) we show the relative importance of the various predictors across occupations and industries. Finally, we discuss potential topics for further research on prescreening specifically and on applicant selection more generally.

Applicant Prescreening

We will first discuss the factors that prior research show to be most important in the evaluation of applicants in applicant screening. Considering the low number of studies that looked specifically at prescreening, we also draw from the general literature on employee selection.

Demographics

Most demographics should not play a big role (or any role) in applicant selection. The consideration of demographics such as gender, ethnicity, age, and marital status within the selection procedure goes hand-in-hand with certain occupation-specific stereotypes. Selection on the basis of such stereotypes implies distinguishing people based on a group or category instead of individual merit, which is discrimination. Still, research has shown that many of these factors affect how applicants are evaluated. Especially when the stereotypical traits of the social category that the applicant is perceived to belong is considered undesirable for the occupation, discrimination on the basis of that social category is likely to occur (Arvey, 1979; Glick, Zion, & Nelson, 1988).

Demographic factors that have been found to play a role in applicant selection include gender (Cole, Feild, & Giles, 2004; Curseu & Boros, 2008; Devlin, 1997; Glick et al., 1988), ethnicity (Derous et al., 2012; Werbel & Others, 1989), (perceived) age (Zebrowitz, Tenenbaum, & Goldstein, 1991), marital status (Oliphant & Alexander, 1982), and, when a photograph is attached, physical appearance of the candidate (e.g., Krueger, Stone, & Stone-Romero, 2014; Watkins & Johnston, 2000; Zebrowitz, Tenenbaum, & Goldstein, 1991). The address of the applicant has also been found to be considered relatively important by professionals in charge of the selection decision (e.g., Ross & Young, 2005). Conceivably, this is because the distance between the applicant's home address and the hiring company can be considered relevant for different reasons. For some positions, for instance, employees should be able to arrive at their work location swiftly after being called-in (e.g., when on stand-by for emergency services). Alternately, potential employees who will have to commute long-distances may be (perceived to be) more likely to quit their jobs.

Moving on to age, there appears to be U-shaped relationship between age and performance, such that performance increases with age initially but then declines (e.g., Ng & Feldman, 2008; Sturman, 2003). It is plausible that this dynamic will be anticipated by those responsible for the hiring decision. Indeed, job prospects are generally poorer for the oldest and youngest job seekers (e.g., Ministerie van Social Zaken en Werkgelegenheid, 2015; U.S. Bureau of Labor Statistics, 2010, 2012).

In sum, prior research using scenario and policy-capturing studies invariably indicate that demographic factors – i.e., age, gender, nationality, marital status, and distance from the hiring companies – affect selection decisions.[11]

[11] Only a small number of résumés in the sample contained photographs of the applicant. Therefore, neither the inclusion nor the evaluation of the photographs on the résumés was included in the model for this study.

Biodata

In the context of the application process, "the core attribute of biodata items is that the items pertain to historical events that may have shaped the person's behavior and identity" (Mael, 1991, p. 763). Biodata, for example, refers to past jobs, courses, or education. Defined as such, the advantage of biodata is that it is relatively verifiable (Asher, 1972) and evaluations based on biodata are seen as justifiable (Mael, 1991). In addition, given that biodata is historical and verifiable, biodata is less susceptible to faking than most criteria used in the selection process (Asher, 1972; T. E. Becker & Colquitt, 1992; Cascio, 1975; Harvey-Cook & Taffler, 2000; Shaffer, Saunders, & Owens, 1986).

The main biodata available on the typical résumé are academic qualifications and work experiences. These two factors have often been found to be the key criteria in the decision of whether to invite or hire an applicant (e.g., Cole et al., 2007; Hutchinson, 1984; Hutchinson & Brefka, 1997; Piotrowski & Armstrong, 2006), despite research suggesting that these factors are weakly related to job performance (McDaniel, Schmidt, & Hunter, 1988; Quińones, Ford, & Teachout, 1995). However, the relative importance of each part of the résumé can vary by occupation (Ross & Young, 2005).

Like age, experience has been found to have a non-linear relationship with performance (Sturman, 2003). There can be considerable initial returns to experience, but these returns diminish over time. Further experience can even have a negative effect on performance as one becomes more likely to acquire or be influenced by non-relevant knowledge and skills (Dokko, Wilk, & Rothbard, 2009).

Biodata is often used to draw inferences about underlying attributes (Ash & Levine, 1985; Levine & Flory, 1975). For work-experience and education this includes inferences regarding personality (Burns, Christiansen, Morris, Periard, & Coaster, 2014; Cole, Feild, Giles, & Harris, 2008; Cole, Feild, & Stafford, 2005; Cole, Feild, Giles, & Harris, 2004) ability, motivation (Brown & Campion, 1994) and professional knowledge (Chen, Huang, & Lee, 2011). For extracurricular activities different activities have been found to correlate with communication, initiative, decision-making, and teamwork (Nemanick & Clark, 2002; Rubin, Bommer, & Baldwin, 2002) – interpersonal skills that are highly valued by recruiters

(Graves & Karren, 1992).[12] Although the inferences tend to be incorrect, especially for attributes such as agreeableness and neuroticism, these perceptions do, ultimately, affect how the résumé is evaluated (Burns et al., 2014; Cole et al., 2008, 2005).

In line with prior research on biodata, measures for the amount and relatedness of experience, as well as the level of education and the relatedness of skills and education to the vacancy are taken into consideration.

Analysis of Applicant Data

The analysis of applicant data is not a straightforward task. In particular, the relative weight of the various prescreening factors can vary between occupations and industries. This poses a challenge given that there are hundreds or even thousands of occupations.

Fortunately, there is an analytic approach - called synthetic validity – that can help cope with this complexity. This approach to applicant selection was introduced in the 1950's (Lawshe, 1952). The chief purpose of this approach is the prediction of performance of candidates in specific occupations. With synthetic validity, the model is *inferred* for a specific situation based on the key characteristics of that situation. Thus, instead of either using a job matching model for a specific occupation or organization (i.e., *situational validity*) or having a generic model that is used across occupations and organizations (i.e., *generalized validity*), the characteristics of the occupation are used to infer how to value different predictors in the selection of applicants (Lawshe, 1952; Scherbaum, 2005). If the relationship between the characteristics of the occupation and the weight of the predictors is know (e.g., through prior research, a training sample, or the estimates of subject matter experts), only the job characteristics need to be determined before synthetic validity can be applied. While generic validity requires a minimum amount of data from an organization before predictions can be made with respect to vacancies in that organization, synthetic validity only requires sufficient estimates, data, or research to establish the relationships between the predictors and the outcome variable at different levels of the job characteristics (Ghiselli, 1959). Synthetic validity is therefore uniquely suited for small businesses.

[12] Extracurricular activities have been deemed relevant too. However, as will be explained in the methods section, we were unable to take the extracurricular activities into consideration due to data constraints.

Recently two suggested improvements to the method synthetic validity have been put forward, which increase the potential value of the technique for both small and large organizations. Instead of just allowing the parameters to vary by job characteristics, they allow the parameters to vary by context too. First, McCloy (1994, 2001) introduced techniques which incorporate notions from multilevel-analyses into synthetic validity. Later, Steel, Huffcutt, and Kammeyer-Mueller (2006) further extended the approach, showing how synthetic validity could benefit from meta-analytic techniques as well. Most importantly, Steel and colleagues demonstrate that the estimates of the validity coefficients (i.e., the magnitude of the correlation) relating predictors and performance are much more accurate when Weighted Least Squares is used (Steel & Kammeyer-Mueller, 2009).

For this study, synthetic validity provides an advantage over alternative methods (e.g., OLS regression or HLM) for several reasons. Firstly, synthetic validity allows predictions to be made for occupations that were not included in the sample on which the model was trained. This is helpful in showing the predictive validity of the model and making full use of the available data. Secondly, the use of meta-analytic techniques (as suggested by Steel et al., 2006; Steel & Kammeyer-Mueller, 2009) and especially Weighted Least Squares, allows the appropriate weighting of data in building a predictive model. Lastly, for the aforementioned reasons, the technique could also be valuable to practitioners and we see value in demonstrating this, plus there is considerable evidence that the use of synthetic validity in practice is legally defensible (for a discussion on the legal defensibility we refer the reader to Steel et al., 2006).

Methods

In order to be able to use synthetic validity on existing and future occupations, one can either (a) use subject matter experts to estimate the validity coefficients or (b) derive validity coefficients from a database of occupations, their respective components, a variety of predictors, and a job-related outcome variable. We employ the latter approach.

The ATS Sample Used

In order to attain the data required for full-scale synthetic validity, this study uses data from Applicant Tracking Systems (ATSs). ATSs contain the data required to unveil generalizable insights on a broad range of topics with regard to recruitment and selection due to the range of variables stored in the systems, the amount of data, and the accessibility of the data. Although several have described these databases when they were first adopted by companies (Hendrickson, 2003; McCrory & Mueller, 2000), and they have been studied more generally under the umbrella of Human Resource Information Systems and e-HRM since then (Stone, Deadrick, Lukaszewski, & Johnson, 2015; Strohmeier, 2007), their use in HRM research is still surprisingly limited (Stone et al., 2015).

In this study, we used data from ATSs deployed and managed by a Dutch organization called Connexys (www.connexys.com). We attained the data from 48 different companies using a Connexys ATS. These companies gave permission for the use of the ATS data in return for detailed, tailored analyses of their data. In order to preserve anonymity, company specific data fields were not accessed and identifying information such as applicants' names and their street addresses were not collected. The names and street addresses were considered identifying information of no special value to the study. For an approximation of the home address of the applicant the postal code was sufficient as Dutch postal codes specify the street or the part of the street where the address is located. Names were not required since for each organization every applicant had a unique ID number.

The combined ATSs of the 48 companies contained data from 467,320 applicants across 1,563 different occupations and 21,743 vacancies. Data for internships and open applications (i.e., unsolicited applications) were not taken into consideration, since the selection process for these are different from the selection process for regular vacancies. We also excluded application processes that were still on going.

The raw résumé and vacancy data needed to be *parsed* and *normalized* and this was accomplished with the help of TextKernel TextKernel – a company specialized in information extraction, web mining, and semantic searches for Human Resource practices (www.textkernel.com). Data parsing is the syntactic analysis of text so as to derive the meaning of the text (e.g., Date, 2003). Normalization is the transformation of text into a canonical form (Codd, 1974). In other words, data was extracted, categorized, and standardized to correspond to structured sets of titles and categories including the O*NET job-classification (e.g., Peterson et al., 2001) and TextKernel's personal job classification, "Jobfeed".

TextKernel parsed and normalized the résumé (or CV) and vacancy data using an innovative approach developed by Zavrel and colleagues (Spitters, Bonnema, Rotaru, & Zavrel, 2010; Zavrel & Daelemans, 2003). The backbone of the approach is a granular taxonomy. First the different sections of the résumé are distinguished – e.g., job experience, academic qualifications, skills, and extracurricular activates. Next, data from each section is linked to the taxonomy by taking into consideration the relatedness of the terms to terms from the taxonomy as well as the likelihood of consecutive co-occurrence of the terms (where terms denote entities like job-titles or educational degrees). The likelihood of co-occurrence is established by using a Markov-Chain Model. The optimal classification is attained through bootstrapping (Spitters et al., 2010). In other words, applied to job titles, there are several iterations; with each iteration, a different job title is selected as a starting point. The job title that is selected first is simply matched to an element from the taxonomy (e.g., to a specific job title). Afterwards each consecutive job titles is matched to the taxonomy on the basis of both its similarity to an existing title in the taxonomy and the likelihood of that title co-occurring with the terms that were matched before it. Finally, the classification is selected that has the highest overall sum likelihood of the job titles having been accurately matched to the normalized job titles from the taxonomy. Attesting to the accuracy of the classification, the reliability scores (i.e., the correspondence of a job title and its description to an existing job from a classification) for the TextKernel classification and O*NET classification had no significant correlation with the error term for the holdout sample. In other words, the reliability of the classification did not have a significant effect on the accuracy of the prediction.

In order to ensure the quality of the different indicators that were used, we manually reviewed random samples of data at each company. In addition, we examined the generic ATS system interface to find out what the standard validation parameters are (i.e., identifying optional fields from required fields, noting restrictions on input values allowed, and finding out to what extent ATS system interfaces vary between companies). For fields on which we attained data from both the résumé and ATS we further validated the data by comparing the data across the two sources.

Whenever dealing with massive databases compiled by people, there are likely to be some errors. To minimize these errors, we first identified cases with values that fell outside the theoretical boundaries for one of the variables. This lead to the exclusion of applicants who were (a) under 16 years old when applying (i.e., for which education was still compulsory and the job thus must have constituted a side-job) or (b) would have been younger than 12 years old (legal age at which one can start working) when starting their first job according to their résumé. This incorrect data could have been due to data-entry errors or mistakes in the optical character recognition for scanned résumé files). These errors appeared for 5,569 of the 467,320 applicants (i.e., 1.19% of the original dataset). As is standard practice with large datasets with some entry errors, we also identified statistically improbable outliers using *centered leverage values* (Aguinis, Gottfredson, & Joo, 2013). The centered leverage values indicate to what extent a case is an outlier based on its likeness to other cases on all predictors in the model combined. We used a high cut-off value so as to only remove extreme outliers (i.e., 3k/n, where k denotes the number of predictors and n is the sample size; Cohen, Cohen, West, & Aiken, 2003). This led to the exclusion of an additional 19,982 applicants (4.28%). Thus, the final sample contained 441,769 applicants and 21,694 distinct vacancies.

The data was split into a training sample and a holdout sample. In order to emulate applicant selection, we used the first 90% of the applicants from each company, sorted by date applied, in order to "train" the model and attain the cutoffs for the predictions. Next, we applied the model and the cutoffs on the remaining 10% of the sample.

Missing data was not replaced for the creation of the synthetic equations from the holdout sample. One advantage of Synthetic Validity is that its estimates are robust to missing data. There simply needs to be sufficient data to allow the relationships between the variables to be inferred. However, we did impute missing data in order to (1) allow the centered leverage values to be computed for all variables (centered leverage requires full data) and (2) attain estimates for all cases in the holdout sample. A robustness check ascertained that the use of imputed data had a minimal effect on the outcomes (for these outcomes we refer the reader to Appendix 4-A).

Predicting Applicants' Invitation Success

In our analysis we used all 42 Generalized Work Activities from the Occupational Network Online (O*NET) database version 18. The synthetic equations were further moderated by the industry of the hiring organization. In other words, the beta weights were allowed to differ for the same position in a different industry. Industry classification was based on the Dutch Standard Company Classification 2008 (SBI) codes (Centraal Bureau voor de Statistiek, 2013), as provided by the Dutch chamber of commerce. This lead to 9 industries to which 38 companies were allocated: "Utilities", "Healthcare", "IT", "Wholesale", "Financial", "Applicant Placement", "Engineering", "Public Administration", and "Facility Management". The contrast group contained 8 companies from different industries.[13]

Demographic Characteristics

Demographic variables in the sample include (1) age, (2) gender, (3) whether the applicant reported a Dutch nationality, (4) whether the applicant had formally recognized personal relationship (e.g., registered partner or marriage) or not (which is often included in a Dutch résumé), and (5) distance between home address and company.

For the distance between the applicant's home address and the vacancy's primary location, the postcodes were converted to geolocations (i.e., latitudes and longitudes) using the address key of the Dutch land register (BAG). The distances between the geolocations were computed using the Haversine formula (Robusto, 1957).

Biodata

The biodata variables that were included are (1) years of experience (i.e., difference between the start of an applicant's first job and the date of application, minus the number of years unemployed in the interim), (2) level of education, (3) relatedness of experience, (4) relatedness of education, and (5) relatedness of skills.

The level of education of each applicant was recoded using the 11 levels of the Dutch education system, as recognized by the Dutch ministry of education (Rijksoverheid, 2005). This was recoded in order to make the level relative to the minimum required level of education according to the vacancy profile set for the position. We split this measure into a measure of *over-education* (i.e., number of levels above the minimum required level of

[13] Industries for which only one organization was in the sample are not named to assure that the companies are not identifiable

education) and *under-education* (i.e., number of levels below the minimum required level of education).

We considered a candidate's experience, education, or skills to be more related to the vacancy if individuals who listed the respective occupation on their CV had respectively held similar jobs, degrees, or skills. The relatedness scores were attained using the *association strength measure* (van Eck & Waltman, 2007, 2009; van Eck, Waltman, van den Berg, & Kaymak, 2006), also known as the *proximity index* (e.g., Peters & van Raan, 1993; Rip & Courtial, 1984) or the *probabilistic affinity index* (e.g., Zitt, Bassecoulard, & Okubo, 2000). The formula for the association strength measure is relatively simple:

$$\text{association strength} = \frac{c_{ij}}{s_i s_j}$$

Where c_{ij} denotes the co-occurrence of node i with node j and s_i and s_j denote the sum of co-occurrences of node i and node j respectively (van Eck & Waltman, 2009: 1637 equation 6). Thus, the association strength constitutes the total number of observed co-occurrences between the item and the job title, divided by the product of the number of occurrences of the item and the job title. This measure has been chosen over other more commonly used co-occurrence measures, such as the Euclidean distance or cosine measure (Robusto, 1957), because it is relatively insensitive to "size effects" (van Eck & Waltman, 2009). Where other measures tend to show a relationship between the overall frequency of occurrence of items and the items' relatedness scores, the association strength most effectively distills similarities between items from differences in overall occurrence counts.

Since the association strength measure already corrects for size effects, no weighting was used for the degrees and skills. Jobs were weighted by the job tenure for the position, such that jobs that were held for a longer period of time were weighed more heavily than jobs held for a shorter period of time.

Context Variables

Context variables included are (1) whether the person applied after or before the company attained the initial target number of hires, (2) whether the person was already working in the company or not, (3) the number of other applicants competing divided by the number of vacancies available, and (4) the demand and supply in the labor market.

The measure of demand and supply in the labor market was attained by using data from the Dutch Central Bureau of Statistics (CBS) on the number of vacancies per 1000 jobs per quarter for each of the 19 Dutch Standard Company Classification 2008 (SBI) sectors (Centraal Bureau voor de Statistiek, 2013). We linked the 19 SBI sectors to the 25 related job classes in our job classification as a measure of the *vacancy rate* – i.e., a measure of supply and demand for the occupation's labor market.

Transformations

Squared terms were added for the variables age and experience years. The different measures of relatedness (i.e., experience relatedness, skills relatedness, and education relatedness), the distance from the company, and the number of other applicants for the vacancy were skewed (with skew statistics ranging from 2.38 to 49.07). Since a normal distribution is assumed for the OLS estimations in synthetic validity we transformed these variables to reduce skewness using the Box-Cox transformation (Box & Cox, 1964; Osborne, 2010).[14]

Using Synthetic Validity

We applied Synthetic Validity using the approach described by Steel and colleagues (2006) and put into practice by Steel and Kammeyer-Mueller (2009). The synthetic equations were attained from the training sample (i.e., the first 90% of the applicants from each company). First, we attained the *validity coefficients* of the predictors for each occupation – i.e., the correlations between the predictors and the preselection decision. Following Steel and Kammeyer-Mueller (2009), only occupations with at least 10 applicants in the specific industry were taken into consideration. Since the outcome variable is dichotomous (i.e., invited/not invited) and could be considered to have an underlying continuity (i.e., some are rejected with more certainty than others), we provide results in Appendix 4-A where the correlations between the outcome variable and the predictors were

[14] Since a dichotomous outcome variable requires the error term to be distributed evenly over the predictor, a Box-Tidwell transformation was also tested (Box & Tidwell, 1962). This provided comparable, but inferior results.

corrected using a Point-Biserial correction (Kemery, Dunlap, & Griffeth, 1988), but this did not have any meaningful effect on the results.

Second, in order to improve the accuracy of the validity coefficients (especially for occupations with few observations) and allow predictions over occupations-industry combinations that were not in the training sample but were present in the holdout sample, we used Weighted Least Squares to predict or adjust the coefficients from the work characteristics and the nine industries. The observed validity coefficients were weighted by their inverse sampling error variance (Hedges & Olkin, 1985; Steel et al., 2006). Next, for each occupation in every industry, a performance-by-predictor correlation matrix was created and an Ordinary Least Squares regression was run. This provides beta-weights for the job characteristic and industry variables, with which the validity coefficients of the predictors can be estimated.

Third, for each occupation in each of the industry groups, the correlation matrix (as shown in Table 4-1) between the predictors (which is the same for each) and the decision to invite or reject the applicant (the grey area in Table 4-1, different for each equation) was used (similar to Table 1 in Steel et al., 2006) to attain the synthetic equation with which the decision to invite an applicant can be predicted.

Finally, using the synthetic equations developed from the training data (i.e., the first 90% of the applicants per company) we predicted the odds of being invited for every applicant in the holdout sample (i.e., the last 10% of the applicants per company).

Table 4-1. Correlations Matrix from Training Sample ($\tilde{N}$=366,364) [a]

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Invited (1=invited)	.11	.31	-																		
2. Age	36.66	11.91	-.064	-																	
3. Age^2	1485.54	931.78	-.07	.99	-																
4. Gender (1=female)	.5	.5	-.024	-.082	-.079	-															
5. Dutch nationality	.66	.47	-.01	-.145	-.138	.052	-														
6. Registered relationship	.26	.44	-.002	.245	.225	-.054	.11	-													
7. Distance from company (in km)[b]	27.05	30.69	-.01	.072	.067	-.16	-.047	.034	-												
8. Experience years	15.1	10.51	-.058	.85	.85	-.091	-.098	.217	.058	-											
9. Experience years^2	338.47	411	-.064	.792	.817	-.087	-.09	.181	.054	.959	-										
10. Undereducated	.63	1.36	-.034	.138	.146	-.04	-.021	.03	.003	.138	.141	-									
11. Overeducated	1.1	1.83	.006	-.071	-.069	.062	.007	-.052	-.066	-.069	-.062	-.276	-								
12. Experience relatedness	6.55	2.07	.109	.094	.083	-.033	-.033	.043	.061	.083	.059	-.029	-.053	-							
13. Skill relatedness	7.07	1.09	.075	.069	.058	-.077	-.082	.035	.118	.038	.025	-.024	-.088	.362	-						
14. Education relatedness	6.43	.85	.045	.036	.027	-.071	-.044	.043	.123	.014	.005	-.025	-.106	.274	.404	-					
15. Applied after target reached	.19	.39	.029	-.143	-.132	-.028	.148	-.076	-.008	-.11	-.095	.037	-.001	-.045	-.089	-.092	-				
16. External applicant	.89	.31	-.042	-.001	.005	-.022	.017	-.003	.005	.022	.028	.044	-.087	.005	-.037	.003	.037	-			
17. Number of other applicants[b]	237.59	759.66	-.086	-.089	-.074	.113	.131	-.058	-.035	-.05	-.034	.033	-.033	-.073	-.148	-.113	.322	.194	-		
18. Avg. % of applicants invited by company	.15	.09	.137	.056	.045	-.022	-.222	.024	.115	.036	.027	-.077	-.007	.08	.185	.186	-.239	-.025	-.26	-	
19. Occup. vacancy rate (per 1000)	18.99	10.34	.02	.018	.011	-.06	.008	.026	.06	.011	.003	.012	-.092	.039	.062	.077	.065	-.076	-.013	.031	-

[a] Point-biserial correction not applied [b] Correlations attained from transformed variables, mean and standard deviation from original variable

Results

The underlying odds of being invited lie on a continuous measurement scale. Different cutoff values can be used to make the outcome dichotomous and the cutoff has consequences for the relative accuracy of the prediction for “invited” versus not “invited candidates”. Therefore, we first show how the selected cutoff affects relative accuracy for each category (i.e., invited and not invited) for both the training sample and the holdout sample. Next, we select a specific optimum for the classification (but by no means the only optimum), in order to demonstrate how these estimates would be used in practice. Afterward we review the relative importance of the various predictors including the moderating effects of the industry and job characteristics.

General Accuracy of the Predictions

In Figure 4-1, a general overview is given of the accuracy across the possible cutoff values for the predicted odds (based on Sing, Sander, Beerenwinkel, & Lengauer, 2005). The accuracy measure that is used provides the accuracy relative to chance and is commonly referred to as ESS (Yarnold & Soltysik, 2005). ESS consists of the sum of the specificity values (i.e., percentage invited correctly predicted) and sensitivity values (i.e., percentage not invited correctly predicted). ESS is normed such that -100 is perfect inaccuracy, 0 is the level of accuracy expected when classification is based on chance (i.e., random), and 100 equals perfect accuracy.

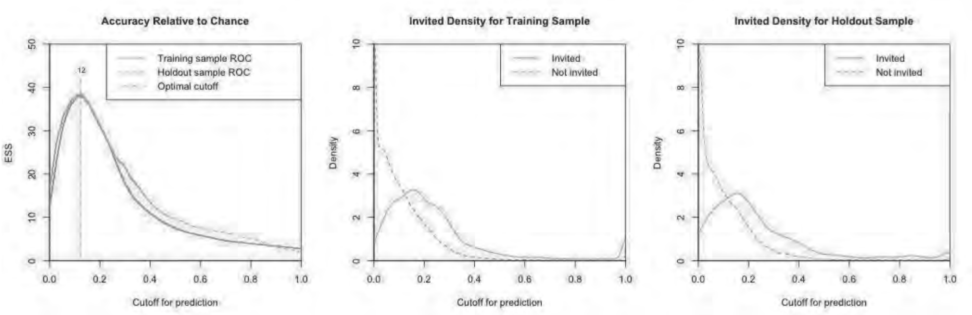


Figure 4-1. Classification Performance at Different Cutoff Values

The two graphs on the right in Figure 4-1 depict the density of invited and not invited applicants over all cutoff values for the training sample and holdout sample. These graphs serve to show how well the predictions help distinguish those invited from those not invited. They reveal that a large number of those not invited have predicted odds close to zero, but that the predicted odds of those who were in reality invited are relatively low too, leading to some overlap between the two groups.

Comparing the accuracy and density in Figure 4-1 for the training sample with the holdout sample, it is evident that the accuracy was similar across the two samples. Across most cutoff values, the accuracy is even slightly higher in the holdout sample than in the training sample, but the difference is arguably negligible. These comparable levels of accuracy provide convincing evidence that, despite the large number of different equations and parameters, the use of synthetic validity did not lead to over fitting of the data. Moreover, it testifies to the predictive validity of the approach as future cases could be classified just as proficiently as the cases used to form the synthetic equations.

We selected a cutoff that maximizes the ESS accuracy value (i.e., highest sum of sensitivity and specificity) such that concrete binary predictions (i.e., invited and not invited) could be made and could be compared to the observed outcome. The highest ESS accuracy for the training sample was attained at a cutoff value of .12 (as shown in Figure 4-1), with a corresponding ESS value of 38.07. Following the general rule of thumb suggested by Yarnold and Soltysik (2005), this can be considered a moderately accurate prediction (as it falls between 25 and 50). When this cutoff is used for the training sample, out of the 44,187 candidates that were invited to the second round of the application process, 31,141 were correctly identified. Thus the sensitivity of the analysis when making predictions from the sample on which it was trained is 71.74%. Out of the 354,081 applicants that were not invited, 234,849 were correctly identified, which equals a specificity of 66.33%.

At the cutoff attained from the training sample (.12) the ESS accuracy value for the holdout sample (i.e., last 10% of the applicants) is 38.22, which is also a moderate level of accuracy and very similar to the accuracy for the training sample (i.e., + .16). As can be seen in Table 4-2, the holdout sample does have a slightly lower percentage of applicants correctly identified as invited in comparison to the training sample (i.e., sensitivity of 69.45% compared to 71.74%). Meanwhile, the percentage invited correctly identified was slightly higher in the holdout sample than in the training sample (i.e., specificity of 68.78% compared to 66.33%).

Table 4-2. Contingency Table Predicting Invitation to Next Round for Holdout Sample [a]

	Predicted		Percentage Correct
Observed	Invited	Not Invited	
Invited	2,539	1,117	69.45%[b]
Not Invited	12,655	27,876	68.78%[c]

[a] Cutoff .12, attained from training sample [b] Percentage of invited correctly predicted, also known as sensitivity [c] Percentage not invited correctly predicted, also known as specificity

One of the key advantages of synthetic validity is that it allows predictions for jobs that were not in the training sample. The holdout sample contained vacancies 135 novel occupations, which attracted 3,200 applicants in total. From these applicants, 253 were invited to the second round of the selection process out of which 158 were correctly classified. This equals a sensitivity rate of 62.45%. Of the 2,947 applicants that did not get invited, 2,078 were correctly classified, which equals a specificity rate of 70.51%. The ESS accuracy of the classification for these occupations was 32.96, only slightly lower than the accuracy for the jobs that were in the training sample.

Additional analyses revealed that the accuracy of the predictions is high for vacancies for which none of the applicants submitted a cover letter, thus suggesting that a cover letter was not required. For those vacancies for which none of the applicants submitted a cover letter, 151 out of 183 invited (82.51%) and 2,832 out of 3,528 not invited (80.27%) were correctly classified in the holdout sample. This suggests that the accuracy of the predictions could be improved by taking the cover letters into consideration.

Predictors of the Prescreening Decision

The synthetic equations provide different beta weights for the predictors of the prescreening decision in different industries and occupations. This means that there are thousands of different beta weights and it is not feasible to discuss these individually. However, we can discuss the general direction of the effects and the relative importance of each variable. In order to determine the relative importance of the variables, we conducted *Relative Weight Analysis* (RWA) for models with a dichotomous outcome variable (Tonidandel & LeBreton, 2010, 2011). RWA decomposes the variance explained by the model into the relative contribution each variable makes to the total predicted variance for the model (Johnson & LeBreton, 2004).

Table 4-3 shows the relative importance of each predictor across occupations and industries. Besides the predictors' mean standardized Beta weights (β), the unstandardized weights (B), and the standard deviation of the unstandardized weights (B_{SD}) across occupations, we provide the "relative weight" (Var) in terms of variance explained per variable for the holdout sample. These percentages add up to the total variance explained, which is approximately 16.74%. We also provide the confidence intervals for the relative weights (95% CI). These confidence intervals were obtained through bootstrapping.[15] Since explained variance is always positive, the confidence interval cannot include zero and significance cannot be derived from the confidence interval. Therefore, significance for the relative weights was established at $\alpha \leq .05$ by a comparison with a random variable attained through Monte Carlo simulation (Tonidandel, LeBreton, & Johnson, 2009).

Out of all variables included in this study, age has the highest relative weight. It explains approximately 8.2% of the total variance (age + age^2). The effect is curvilinear, with the youngest and oldest applicants having the lowest odds of getting invited. Other demographic variables have considerably lower relative weights. The applicants' gender (.14%), holding Dutch citizenship (.08%), having formally recognized relationship (i.e., married or registered partnership)(.09%), and distance from the hiring company (.39%) explain less than one percent of the variance in total.

Out of all biodata variables, the years of work experience of the applicant has the highest relative weight. This explains approximately 3.15% of the total variance (years of experience + years of experience^2). In agreement with research on the relationship between experience and performance (Sturman, 2003), experience is beneficial for the odds of being invited only up to a certain point, after which the odds of being invited decrease. The relatedness of experience is relatively important for the selection decision too. Its mean beta weight is relatively high and the variable accounts for .99% of the variance. The more related

[15] We split the sample into 6 subsamples to avoid reaching the maximum vector size limit in R and ran each subsample on a separate Amazon Web Service cluster computer instance. Each instance had 16 virtual Core Processing Units (vCPUs) 30GB RAM and a High Frequency Intel Xeon E5-2680 v2 (Ivy Bridge) processor and every subsample took approximately seven hours to process.

the applicant's experience is to the vacancy in question, the higher his or her odds of getting invited.

In comparison, the level of education, education relatedness, and skills relatedness, are of much lesser importance in applicant prescreening. As can be expected, being undereducated has a negative effect on the odds of being invited, but the effect is not very strong and explains only .06% of the variance. The effect of being overeducated is even weaker, with a similar relative weight. With explained variance scores of respectively .22% and .12%, skill relatedness and education relatedness carry slightly more weight. Nevertheless, the results suggest that experience is generally valued over education in applicant prescreening. Surprisingly, education relatedness has a negative effect on the odds of being invited for many occupations, suggesting that for some occupations and in some industries, educational backgrounds that differ from the ordinary are valued most. However, these effects are extremely small.

Besides demographics and biodata, we also looked at context variables, specifically with relation to the application (i.e., whether the applicant applied after the target number of hires for the vacancy was reached, whether it concerned an external applicant) and the applicant pool (i.e., the number of other applicants applying to the vacancy, the average percentage of applicants invited by the company, and the number of vacancies per 1000 jobs for the occupation's job category for the quarter of the year in which the applicant applied). Although the effect sizes and variance explained by these variables is relatively low (with relative weights ranging from .45% for applied after target reached to .96% for number of other applicants), there are two notable effects. Firstly, applicants applying after the target number of hires was reached sometimes have a higher chance of being invited. There are several possible reasons why this may be the case, including self-selection (e.g., only the more suitable candidates apply after the vacancy deadline has closed) or it may concern applicants that were asked to apply by (someone in) the hiring company. Secondly, in general, it is not beneficial to be an external applicant, but there are some occupations in which coming from outside the company has a small positive effect on the odds of making it through prescreening.

Table 4-3. Mean Predicted Beta Weights and Relative Variance Explained for Holdout Sample (DV=Invited)

		Unstandardized		Relative Weight	
Variable	$\boldsymbol{\beta}$	$\boldsymbol{B}$	$\boldsymbol{B_{SD}}$	**Var**[a]	**95% CI**[b]
Age	.096	.003	.007	4.05%	(4.01%, 4.09%)
Age^2	-.171	.000	.000	4.15%	(4.11%, 4.19%)
Gender (1=female)	.017	.013	.026	.14%	(.14%, .15%)
Dutch nationality	.012	.008	.021	.08%	(.08%, .09%)
Registered relationship	.015	.011	.020	.09%	(.09%, .1%)
Distance from company (in km)[c]	-.034	-.009	.009	.39%	(.38%, .4%)
Experience years	.014	.000	.005	1.78%	(1.76%, 1.8%)
Experience $years^2$	-.023	.000	.000	1.45%	(1.43%, 1.47%)
Undereducated	-.012	-.003	.006	.06%	(.06%, .07%)
Overeducated	.002	.000	.006	.06%	(.06%, .06%)
Experience relatedness	.105	.018	.011	.99%	(.98%, 1.0%)
Skill relatedness	.042	.014	.014	.22%	(.22%, .23%)
Education relatedness	-.023	-.010	.015	.12%	(.12%, .12%)
Applied after target reached	.020	.017	.063	.45%	(.44%, .45%)
External applicant	-.052	-.059	.114	.71%	(.7%, .72%)
Number of other applicants[c]	-.088	-.035	.033	.96%	(.95%, .98%)
Avg. % of applicants invited by company	.009	.037	.317	.54%	(.53%, .55%)
Occupation vacancy rate (per 1000)	.032	.001	.003	.47%	(.46%, .49%)

[a] Total variance explained is 16.74% [b] All weights are significant at $\boldsymbol{\alpha \leq .05}$ (established by comparison with a random variable) [c] Transformed variable

Relevance of Industry and Occupation

Table 4-4 provides an overview of the moderators and their relative importance. k is the sum total number of occupations for which data was used to compute the predictor's validity coefficient (where occupations were counted separately per industry). $\widetilde{N}$ is the average number of applicants across the k occupations. ρ is the corrected correlation across all occupations (also known as the *validity coefficient*), with ρ_{SD} showing the standard deviation in these correlations. The adjacent column provides the 95% confidence interval for the correlations (95% CI), showing the lower and the upper bound respectively. In the column next to the confidence interval is the 90% credibility interval (90% CV), which is the range within which 90% of the ρ values lie. The next two columns show the extent to which differences in the relative weight of each predictor in the selection process are explained by differences between industries ($Var_{industry}$) or in job characteristics (Var_{job}). Combined, these form the total variance explained. These relative weights have also been attained using relative weight analysis (Tonidandel & LeBreton, 2011) and were computed using the predictor validity coefficient by industry and job characteristic correlation matrix weighted by the inverse sampling error. Finally, I^2 indicates the proportion of total variation in validity coefficients that is due to heterogeneity between the k occupations (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003).

With the exception of the variables relating to the context of the vacancy, the weighting of different selection criteria across occupations and industries is relatively inconsistent. The distance from the hiring company has the highest rate of consistency with an I^2 score of 49.21%. The remaining 50.79% of the variance in how the distance from the hiring company is evaluated can be attributed to chance or inconsistency – i.e., a moderately low rate of consistency (Higgins et al., 2003). For other demographic variables the I^2 consistency scores are lower. Gender is lowest, with 37.18% of the differences in how it's weighted relating to true variation underlying the results and the remainder being compatible with chance.

For the biodata variables education, experience relatedness, and education relatedness, the consistency is 56.39%, 55.45%, and 44.94%. For experience years the consistency is 31.57%, meaning that the hiring party's valuation of this attribute is mostly random. For under education the I^2 is 17.56% and for education relatedness it is 11.44%

and thus nearly all of the variance in how these predictors are weighted can be attributed to chance.

Discussion

Using advanced data sources and analytic techniques we were able to study prescreening in a real-world field setting, across occupations and companies. We used data from Applicant Tracking Systems (ATSs) of 48 companies. From these ATS databases we acquired the résumés, background information, and selection decisions related to 441,769 applicants. We show the predictive validity of the key variables in applicant prescreening by using an advanced type of synthetic validity analysis based on meta-analytic techniques in cross-validation (Lawshe, 1952; Steel et al., 2006; Steel & Kammeyer-Mueller, 2009).

In support of the predictive validity of the approach, the difference between the accuracy in the training sample and the accuracy for the holdout sample was marginal. In addition, although the accuracy of the predictions were moderate for the holdout sample as a whole (i.e., correctly classified 69.45% of the invited applicants and 68.78% of the not invited applicants), the accuracy was relatively high when focusing on vacancies for which none of the applicants had submitted a cover letter (i.e., correctly classified 82.52% of the invited applicants and 80.17% of the applicants not invited).

This study also demonstrates the power of the use of ATS data in combination with synthetic validity for research on applicant selection. The ATSs provide access to large data samples that enable a person-centered approach in which parameters vary for different sub-populations (Wang & Hanges, 2011). This allows applicant selection to be studies in a field setting, with a level of ecological and predictive validity – and a potential for generalizability – that was not possible before (Breaugh, 2009).

The findings reiterate the potential value of the use of synthetic validity for decision support systems in applicant selection. It allows the selection decision to be predicted at various levels of confidence for different applicants. When synthetic validity is used on ATS data, without non-essential demographic predictors, the predicted level of performance and rate of confidence of the classification of each individual applicant could be used as an indication of which applicants deserve further evaluation (e.g., Grannis, Overhage, Hui, & McDonald, 2003; Winkler, 1999).

Table 4-4. Moderator Analysis: Variance in Predictor Weight and Heterogeneity Across Occupations and Industries

	k	$\tilde{N}$	ρ	ρ_{SD}	95% CI	90% CV	$\text{Var}_{\text{industry}}$	$\text{Var}_{\text{occupation}}$	I^2
Age	1915	195	.010	.099	(-.062, -.054)	(-.17, .054)	4.32%	7.88%	47.48%
Age^2	1915	195	.009	.095	(-.067, -.059)	(-.165, .039)	4.21%	8.42%	43.03%
Gender (1=female)	1858	204	.008	.088	(.008, .016)	(-.076, .101)	1.78%	9.78%	37.18%
Dutch nationality	1911	200	.008	.087	(.003, .01)	(-.077, .09)	2.05%	4.39%	44.29%
Registered relationship	1915	198	.007	.082	(-.002, .005)	(-.064, .068)	2.88%	3.53%	39.74%
Distance from company (in km)[c]	1645	194	.010	.101	(-.046, -.037)	(-.157, .074)	1.94%	24.73%	49.21%
Experience years	1854	170	.009	.093	(-.056, -.048)	(-.139, .035)	4.76%	8.81%	31.57%
Experience years^2	1854	170	.007	.086	(-.061, -.053)	(-.121, .008)	4.26%	9.6%	21.71%
Undereducated	1817	193	.005	.072	(-.036, -.029)	(-.033, -.033)	2.93%	3.63%	17.56%
Overeducated	1813	193	.008	.090	(.012, .02)	(-.072, .104)	.63%	5.25%	56.39%
Experience relatedness	1749	158	.013	.116	(.079, .09)	(-.055, .224)	3.94%	19.18%	55.45%
Skill relatedness	1844	172	.010	.102	(.049, .058)	(-.058, .164)	2.91%	14.67%	44.94%
Education relatedness	1718	148	.008	.088	(.017, .026)	(-.031, .074)	1.02%	4.78%	11.44%
Applied after target reached	676	422	.008	.088	(.008, .021)	(-.107, .136)	5.97%	19.51%	75.55%
External applicant	1438	245	.018	.135	(-.055, -.041)	(-.244, .148)	17.64%	16.83%	85.2%
Number of other applicants[c]	1833	208	.020	.140	(-.08, -.067)	(-.274, .127)	11.19%	16.8%	78.99%
Avg. % of applicants invited by company	1748	216	.014	.120	(.037, .049)	(-.12, .206)	10.29%	15.71%	76.86%
Occupation vacancy rate (per 1000)	1432	251	.018	.135	(.034, .048)	(-.156, .238)	8.88%	9.98%	82.99%

Note: k is the sum total number of occupations for which data was used to compute the predictor's validity coefficient. $\tilde{N}$ is the average number of applicants across the k occupations. ρ is the corrected correlation across all occupations, ρ_{SD} shows the standard deviation in these correlations. 95% CI provides the 95% confidence interval for the correlations. 90% CV provides the 90% credibility interval. $Var_{industry}$ shows the variance explained by differences in industries. Var_{job} shows the variance explained by differences in occupational characteristics. I^2 indicates the proportion of total variation in validity coefficients that is due to heterogeneity between the k occupations.

Besides providing insight into the relative accuracy of the predictions with the selected approach, we were also able to provide insight into the relative importance of the different variables in prescreening by looking at the weighted Beta coefficients and using *Relative Weight Analysis* (Tonidandel & LeBreton, 2010, 2011). The results indicate that age and experience are highly valued, each with a curvilinear effect. Thus, for both age and years of work experience, odds of being selected increase initially, but then decline. Other demographic variables and biodata carried surprisingly little weight.

The study also shows that a lot of the variation in the weighting of the criteria across occupations and industries comes from inconsistency or chance. The high levels of inconsistency with respect to the weighting of biodata suggest that hiring parties' criteria for selection across occupations and industries is random to a very large extent. This is especially disconcerting in light of prior research that has established that there is a high level of consistency in how performance varies with occupations and occupational characteristics (Steel & Kammeyer-Mueller, 2009).

Alongside the aforementioned contributions, we also wish to point out several caveats and questions for future research. First, the present study did not include all data available to the hiring party. The application form for some of the vacancies in our sample included job- or organization-specific fields and these were not available to us. Considering that organizations find these fields important enough to be included in the application form, the data might be relevant in the prescreening decision and should be taken into consideration in future studies. Also, it would be advisable to include the applicants' cover letter as these likely explain a considerable portion of the remaining variance. Cover letters were not included in this study because of privacy concerns.

Second, the job characteristics for the occupations come from O*NET General Work Activities (GWAs). The GWAs have been compiled based on occupations in the United States while this study was conducted using data from the Netherlands. Although the matching of job titles and descriptions to a Dutch job classification justifies the use of the O*NET job characteristics for Dutch occupations, further validation of the job characteristics could help affirm the findings from the present study. This would likely help improve the variance explained by the job characteristics (Mossholder & Arvey, 1984).

Future research would also do well to look into the application of synthetic validity on a sample with a dichotomous outcome variable. This is the first study applying synthetic validity to a binary or discrete outcome variable. We apply the corrections used in meta-analytic studies in order to adjust for the violation of the non-linearity assumption with respect to the binary outcome variable and the predictors – i.e., using the Point-biserial correction (Kemery et al., 1988). This approach was taken because the approach we used (i.e., Steel et al., 2006; Steel & Kammeyer-Mueller, 2009) relies on the Least Squares Method to infer the synthetic equations, while the conventional method for the inference of models with a binary outcome variable, logistic regression, requires the maximization of a log likelihood function over the observed cases (Cox, 1958). However, the applied corrections do not seem to improve the accuracy of the equations (i.e., the difference is negligible).

Lastly, this study looks at prescreening decisions and therein it captures the decision-making behavior of professionals hiring new employees for their employer. As such, we provide insight into what these professionals currently do. However, this data provides little room for further prescriptive claims about the efficacy of the decision-making. In order to know what professionals should be doing, further research is required that compares hiring decisions and the related variables with the performance of those who were selected. In other words, we now have a much better understanding of who gets invited for a job interview. The next step is to see whether the people invited are the ones who should be invited.

Appendix 4-A: Robustness of the Outcome

Table 4-5. Contingency Table Predicting Invitation to Next Round for Holdout Sample with Point Biserial Correction Applied[a b]

	Predicted		Percentage Correct
Observed	Invited	Not Invited	
Invited	2,334	1,050	68.97%[c]
Not Invited	11,330	26,177	69.79%[d]

[a] ESS accuracy = 38.76 [b] Cutoff (.13) attained from training sample (also using Point Biserial correction)
[c] Percentage of invited correctly predicted, also known as sensitivity [d] Percentage not invited correctly predicted, also known as specificity

Table 4-6. Contingency Table Predicting Invitation to Next Round for Holdout Sample without the use of Imputed Values[a b]

	Predicted		Percentage Correct
Observed	Invited	Not Invited	
Invited	2,347	1,037	69.36%[c]
Not Invited	11,597	25,91	69.08%[d]

[a] ESS accuracy = 38.44 [b] Cutoff (.12) attained from training sample (training sample same as before) [c] Percentage of invited correctly predicted, also known as sensitivity [d] Percentage not invited correctly predicted, also known as specificity

Chapter 5: General Discussion

This dissertation reveals a variety of avenues through which Big Data opportunities can be leveraged in management research. It also highlights some of the distinctive practices one faces when dealing with Big Data with respect to data cleaning, data restructuring, measure creation, analysis, and knowledge dissemination. In this chapter I provide a brief summary of the main contributions and key findings of the projects discussed in chapter two through four. Afterward I discuss several key lessons that can be drawn from these studies on what the rise of the Big Data phenomenon means to management research.

Summary of the Contributions and Key Findings

In chapter two we use several bibliometric tools (van Eck & Waltman, 2010; Waltman et al., 2010) in order to provide the first empirically grounded taxonomy of the field of career studies. A science map was created, consisting of key terms extracted from article abstracts and titles from peer-reviewed publications in the field. Using the science map we provide a framework that can help outsiders in making sense of what career studies is about, as well as open up discussion on the current and desired state of the field for those already engaged in career studies research. We also compare and contrast career studies research in the field of management with non-management research on careers and identify some distinct opportunities for novel research. The comparison between the "local" (i.e., management) and "global" (i.e., all of social science) literature on careers reiterates that the field of career studies is fragmented (Arnold & Cohen, 2008; Arthur, 2008) and indicates that patches of understudied careers topics exist, such as research on the impact of careers at the organizational level (c.f., Baruch, 1999; Gunz & Jalland, 1996). At the same time, the study shows how Natural Language Processing (NLP) and science mapping tools can help create more systematic, disciplined, and transparent reviews of academic fields in comparison to traditional, narrative reviews.

Chapter three presents a new tool that goes beyond a static review of the literature, promising emergent, evolving, and searchable insights into detailed aspects of the literature in the near future. In order to deal with concept proliferation in the field of management (e.g., Bosco et al., 2014; Hallberg & Schaufeli, 2006; Le et al., 2010; Li & Larsen, 2011; Schwab, 1980; J. Singh, 1991), we propose a software platform called "ReNotate" that helps individuals highlight parts of the text, annotate text, keep track of notes, and review their own highlights and annotations. This platform should already prove useful for the individual, but the collective knowledge that is created by the sensemaking activities of individuals could play an even more valuable role in making the management literature more navigable. By aggregating data highlighted and annotated by individuals, it becomes possible to attain validated, crowdsourced insights into the literature. A searchable database can be created, allowing quick access to, for example, study characteristics, concepts and their operationalizations, and relationships between concepts. Thereby, navigating the literature could become far easier and less time-intensive. Thus, with sufficient interest in the use of ReNotate as a reading and sensemaking tool for individuals, the platform has the potential to make science in general, and the field of management in particular, more effective and efficient.

Where chapters two and three focus on how Big Data could help improve management scholarship, chapter four focuses on how Big Data can be used in management research. Challenges in data collection have seemingly hindered prior studies from examining prescreening decisions (i.e., whether the applicant is allowed to proceed to the second round of the selection process) in a field setting. We are able to show the relative importance of 18 factors relating to the applicant (i.e., demographic factors, work and educational experience), the vacancy (i.e., relevance of applicant education, skills, and prior experience), the company (i.e., external or internal applicant), applicant pool (i.e., number of applications received), and the labor market (i.e., occupation vacancy rate) by using data from Applicant Tracking Systems (ATSs) – i.e., database infrastructures used to capture, store, and communicate applicant data such as CVs, application decisions, and correspondence with the applicant. We draw on ATS data from 48 companies and over 441,769 applicants and predict whether applicants will be accepted into the second round of the selection process. Predictive validity is shown through cross validation, with a training sample consisting of the first 90% of the applicants and a testing sample of the remaining 10%. We use an analytical approach called synthetic validity (Lawshe, 1952; Steel et al., 2006), which infers the importance of the predictors with respect to the prescreening decision from the relationship between the predictors and a range of job characteristics. This minimizes overfitting, allows parameters to vary across subpopulations, and makes it possible to attain valid estimates for occupations on which there was little or no data in the training sample. Predictions were moderately accurate according to the rule of thumb for ESS accuracy scores (Yarnold & Soltysik, 2005), with 69.45% of the 3,656 invited correctly classified and 68.78% of the 40,531 candidates not invited correctly identified. For vacancies for which no cover letter was required, the prediction was powerful, with over 80% in the holdout sample accurately classified for both the invited and not-invited candidates. We use *relative weight analysis* to provide insight into the relative weight of the various predictors (Tonidandel & LeBreton, 2010, 2011). This shows that age and experience are the most relevant in applicant prescreening by far. Consistency scores further suggest that hiring professionals are relatively inconsistent in how they adapt their evaluations of demographic predictors and biodata to the occupation or industry.

New Avenues for Management Research

In this dissertation Big Data is defined according to the commonly used definition from Doug Laney, which postulates that Big Data is more than just data of high volume, as it is also delineated by high variety, and velocity (Laney, 2001). This resonates with the view of Big Data as a relatively recent "phenomenon" (Diebold, 2012), driven by advances in technology and a subsequent acceleration in the variety of data traces that people leave behind as a by-product of their interactions with computers (Dutta & Mia, 2009; Giles, 2012). These digital traces are generally of high volume as they come from individual users' interactions that have purpose at the individual level (i.e., adding data to the database is not the main intent), but they are also characterized by high variety as the data can come from many different types of devices, sensors, or interactions (e.g., smartphones, tracking devices, social media) and high velocity as these devices and their interconnectivity allow a constant stream of data to be provided. This definition of Big Data emphasizes the aspects of Big Data that are relevant to social scientific research. In particular, Big Data could empower the social sciences by providing "the ability to understand the patterns of human life by analyzing the digital traces that we leave behind" (Dutta & Mia, 2009, p. 75; Giles, 2012). This view of Big Data (i.e., as the result of the ubiquity of digital devices) further highlights several distinct potential gains for management research.

For management research, the technologies linked to Big Data not only allow behaviors to be studied in situ (Wenzel & van Quaquebeke, in press), but they could also make it easier to attain data from within organizations and across organizations. This can be achieved by using unobtrusive designs. Such designs include the analysis of email messages (Diesner, Frantz, & Carley, 2006), for instance, or the use of data from business intelligence systems such as the ATS system used in chapter four (Putka & Oswald, 2015). Alternatively, this data can also be obtained by purposefully creating an incentive for the individual to provide the data required. An example of such an approach is provided in chapter three – i.e., making the management literature machine-readable by creating a platform on which individuals can structure their reading and aggregating the data. Other examples include, for instance, the use of a website on which individuals can see how their salary compares to the salary of others to attain a continuous overview in salary trends for a range of occupations across many different countries (Tijdens et al., 2002), or the use of CAPTCHAs (Completely Automated Public Turing tests to tell Computers and Humans Apart) to prevent automated programs from abusing web services, while at the same time deciphering and digitalizing old printed material (Ahn, Maurer, McMillen, Abraham, & Blum, 2008).

Big data can also lead to a more data-driven stream of research in the field of management. While the contemporary management literature is heavily flavored toward deductive reasoning (Putka & Oswald, 2015; Wenzel & van Quaquebeke, in press) and there is a strong norm of the use of theory as a starting point (Hambrick, 2007), large representative samples of behavioral data from organizations could arguably allow a shift in the balance between induction and deduction, for instance by fueling research that originates from trends observed in society (Putka & Oswald, 2015; Wenzel & van Quaquebeke, in press). Indeed, the recently established journal Academy of Management Discoveries, which focuses on the publication of empirical studies on poorly understood phenomena, suggests that this shift might already be taking place (Wenzel & van Quaquebeke, in press).

An increase in the prevalence of the use of Big Data could also lessen the field of management's reliance on experimental- and survey-data (Podsakoff, MacKenzie, & Podsakoff, 2012). This is highly desirable when it comes to self-report measures of variables for which more objective equivalents have become available. Previously, for example, prescreening studies relied on hiring professionals' ratings of the employability of fabricated candidate data, while those ratings might not be representative of the actual prescreening decision, especially when the sample is not representative of an applicant pool (Breaugh, 2009). As shown in chapter four, ATS data allows the actual hiring decision to be used in a research design. However, other commonly used measures in the field of management might not be easily found in Big Data databases. This is likely to be true for psychometric scales. Although it would take time alternatives could be developed. For some of the most well known measures, such as personality, there are indications that reliable measures can be attained in creative ways, for instance by using Natural Language Processing tools on text written by an individual (Oberlander & Nowson, 2006) or even from the individual's social media webpage (Back et al., 2010).

In conclusion, Big Data could provide a range of opportunities for management research, but there are some barriers to overcome. Big data research is at odds with the type of research traditionally conducted in the field of management, presenting hurdles in terms of skill and resource requirements and, perhaps most of all, the prevailing norms in the field. These norms concern the current focus on deductive research and the prevailing use of psychometric measures. Undeniably, Big Data challenges the current status quo. Whether this will damage the field or make it stronger remains to be seen.

References

Abele, A. E., & Spurk, D. (2009). How do objective and subjective career success interrelate over time? *Journal of Occupational and Organizational Psychology*, *82*, 803–824.

Abel, F., Henze, N., & Krause, D. (2008). Ranking in folksonomy systems: Can context help? In *Proceedings of the 17th ACM Conference on Information and Knowledge Management* (pp. 1429–1430). New York, NY: ACM Press.

Acedo, F. J., & Casillas, J. C. (2005). Current paradigms in the international management field: An author co-citation analysis. *International Business Review*, *14*, 619–639.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*, 270–301.

Ahn, L. von, Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: Human-based character recognition via web security measures. *Science*, *321*, 1465–1468.

Alger, C. L. (2009). Secondary teachers' conceptual metaphors of teaching and learning: Changes over the career span. *Teaching and Teacher Education*, *25*, 743–751.

Antenucci, D., Cafarella, M. J., Levenstein, M., Ré, C., & Shapiro, M. (2013). Ringtail: Feature selection for easier nowcasting. Presented at the Sixteenth International Workshop on the Web and Databases, New York, NY. Retrieved from http://www-cs.stanford.edu/people/chrismre/papers/webdb_ringtail.pdf

Archer, L., Dewitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). "Doing" science versus "being" a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity. *Science Education*, *94*, 617–639.

Arnold, J., & Cohen, L. (2008). The psychology of careers in industrial and organizational settings: A critical but appreciative analysis. In G. P. Hodgkinson & J. K. Ford (Eds.), *International Review of Industrial and Organizational Psychology* (pp. 1–44). London: Wiley.

Arthur, M. B. (2008). Examining contemporary careers: A call for interdisciplinary inquiry. *Human Relations*, *61*, 163–186.

Arthur, M. B., Hall, D. T., & Lawrence, B. S. (1989a). Generating new directions in career theory: The case for a transdisciplinary approach. In *Handbook of career theory* (pp. 7–25). Cambridge, United Kingdom: Cambridge University Press.

Arthur, M. B., Hall, D. T., & Lawrence, B. S. (Eds.). (1989b). *Handbook of career theory*. Cambridge, United Kingdom: Cambridge University Press.

Arthur, M. B., & Rousseau, D. M. (Eds.). (1996). *The boundaryless career: A new employment principle for a new organizational era*. New York, NY: Oxford University Press.

Arvey, R. D. (1979). Unfair discrimination in the employment interview: Legal and psychological aspects. *Psychological Bulletin*, *86*, 736–765.

Aschaffenburg, K., & Maas, I. (1997). Cultural and educational careers: The dynamics of social reproduction. *American Sociological Review*, *62*, 573–587.

Asher, J. J. (1972). The biographical item: Can it be improved? *Personnel Psychology*, *25*, 251–269.

Ash, R. A., & Levine, E. L. (1985). Job applicant training and work experience evaluation: An empirical comparison of four methods. *Journal of Applied Psychology*, *70*, 572–576.

Attwood, T. K., Kell, D. B., McDermott, P., Marsh, J., Pettifer, S. R., & Thorne, D. (2009). Calling international rescue: Knowledge lost in literature and data landslide! *Biochemical Journal*, *424*, 317–333.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, *21*, 372–374.

Baclawski, K., Kokar, M. M., Waldinger, R., & Kogut, P. A. (2002). Consistency checking of semantic web ontologies. In *The Semantic Web-ISWC 2002* (pp. 454–459). Berlin, Germany: Springer-Verlag.

Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques*. Thousand Oaks, CA: SAGE.

Balma, M. J. (1959). The concept of synthetic validity. *Personnel Psychology*, *12*, 395–396.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, *72*, 187–206.

Baruch, Y. (1999). Integrated career systems for the 2000s. *International Journal of Manpower*, *20*, 432–457.

Baruch, Y. (2006). Career development in organizations and beyond: Balancing traditional and contemporary viewpoints. *Human Resource Management Review*, *16*, 125–138.

Baruch, Y., & Bozionelos, N. (2010). Career issues. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology* (pp. 67–113). Washington, DC: American Psychological Association.

Baruch, Y., Konrad, A. M., Aguinis, H., & Starbuck, W. H. (Eds.). (2008). *Opening the black box of editorship*. Basingstoke, United Kingdom: Palgrave Macmillan.

Baruch, Y., & Peiperl, M. (2000). Career management practices: An empirical survey and implications. *Human Resource Management*, *39*, 347–366.

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., & Stein, L. A. (2004). OWL web ontology language reference. (M. Dean & G. Schreiber, Eds.). W3C. Retrieved from http://www.w3.org/TR/owl-ref/

Becker, P. E., & Moen, P. (1999). Scaling back: Dual-earner couples' work-family strategies. *Journal of Marriage and Family*, *61*, 995–1007.

Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology*, *45*, 389.

Beckett, D., & Berners-Lee, T. (2011). Turtle-terse RDF triple language. W3C. Retrieved from http://www.w3.org/TeamSubmission/turtle/

Beijaard, D., Verloop, N., & Vermunt, J. D. (2000). Teachers' perceptions of professional identity: An exploratory study from a personal knowledge perspective. *Teaching and Teacher Education*, *16*, 749–764.

Bergman, M. (2009). Advantages and myths of RDF. Retrieved from http://23.21.221.45/wp-content/themes/ai3/files/2009Posts/Advantages_Myths_RDF_090422.pdf

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, *284*, 28–37.

Birdi, K., Allan, C., & Warr, P. (1997). Correlates and perceived outcomes of four types of employee development activity. *Journal of Applied Psychology*, *82*, 845–857.

Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of Big Data: Four perspectives – four challenges. *ACM SIGMOD Record*, *40*, 56–60.

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia - A crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, *7*, 154–165.

Blalock, H. M. (1968). The measurement problem: A gap between the languages of theory and research. In H. M. Blalock & A. B. Blalock, *Methodology in Social Research* (pp. 5–27). New York, NY: McGraw-Hill.

Blanchard, J., & Mikkelson, V. (1987). Underlining performance outcomes in expository text. *The Journal of Educational Research*, *80*, 197–201.

Bliesener, T. (1996). Methodological moderators in validating biographical data in personnel selection. *Journal of Occupational and Organizational Psychology*, *69*, 107–120.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*, 187–215.

Blokland, A., & Nieuwbeerta, P. (2005). The effects of life circumstances on longitudinal trajectories of offending. *Criminology*, *43*, 1203–1240.

Bloomberg Businessweek. (2011). *The current state of business analytics: Where do we go from here?* Bloomberg Businessweek Research Services. Retrieved from http://www.sas.com/resources/asset/busanalyticsstudy_wp_08232011.pdf

Blossfeld, H.-P. P., & Huinink, J. (1991). Human-capital investments or norms of role transition: How women's schooling and career affect the process of family formation. *American Journal of Sociology*, *97*, 143–168.

Bock, C., Fokoue, A., Haase, P., Hoekstra, R., Horrocks, I., Ruttenberg, A., … Smith, M. (2009). OWL 2 web ontology language: Structural specification and functional-style syntax. *W3C Recommendation*, *27*, 1–133.

Bodenheimer, T., & Pham, H. H. (2010). Primary care: Current problems and proposed solutions. *Health Affairs*, *29*, 799–805.

Bolino, M. C. (2007). Expatriate assignments and intra-organizational career success: Implications for individuals and organizations. *Journal of International Business Studies*, *38*, 819–835.

Borman, W. C. (1996). The Occupational Information Network: An updated Dictionary of Occupational Titles. *Military Psychology*, *8*, 263–265.

Börner, K., Boyack, K. W., Milojevic, S., & Morris, S. A. (2012). An introduction to modeling science: Basic model types, key definitions, and a general framework for the comparison of process models. In *Models of science dynamics: Understanding complex systems* (pp. 3–22). Heidelberg: Models of Science Dynamics.

Bosco, F. A., Aguinis, H., Uggerslev, K. L., Steel, P. D. G., Singh, K., & Pierce, C. A. (2015, August). *The one-minute meta-analyst: A metaBUS tutorial for synthesizing management research*. Presented at the Academy of Management Annual Meeting, Vancouver, Canada.

Bosco, F. A., Steel, P. D. G., Oswald, F., Uggerslev, K., & Field, J. (2015). Cloud-based meta-analysis to bridge science and practice: Welcome to metaBUS. *Personnel Assessment and Decisions*, *1*, 3–17.

Bosco, F. A., Uggerslev, K., & Steel, P. D. G. (2014). Scientific findings as Big Data for research synthesis: The metaBUS project. In *2014 IEEE International Conference on Big Data* (pp. 18–22). Washington, DC.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, B*, *26*, 211–252.

Box, G. E., & Tidwell, P. W. (1962). Transformation of the Independent Variables. *Technometrics*, *4*, 531–550.

Bradford, R. B. (2008). An empirical study of required dimensionality for large-scale latent semantic indexing applications. In *Proceedings of the 17th ACM*

Conference on Information and Knowledge Management (pp. 153–162). New York, NY: ACM Press.

Breaugh, J. A. (2009). The use of biodata for employee selection: Past research and future directions. *Human Resource Management Review*, *19*, 219–231.

Brickley, D., & Guha, R. V. (1999). Resource Description Framework (RDF) schema specification. W3C. Retrieved from http://www.w3.org/TR/PR-rdf-schema/

Brinberg, D., & McGrath, J. E. (1982). A network of validity concepts within the research process. *New Directions for Methodology of Social & Behavioral Science*, *12*, 5–21.

Brown, B. K., & Campion, M. A. (1994). Biodata phenomenology: Recruiters' perceptions and use of biographical information in resume screening. *Journal of Applied Psychology*, *79*, 897–908.

Bruns, A. (2013). Faster than the speed of print: Reconciling "Big Data" social media analysis and academic scholarship. *First Monday*, *18*, 1–14.

Bry, F., & Schaffert, S. (2002). *A gentle introduction into Xcerpt, a rule-based query and transformation language for xml*. Munich, Germany: Computer Science Institute. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.63.5184&rep=rep1&type=pdf

Burns, G. N., Christiansen, N. D., Morris, M. B., Periard, D. A., & Coaster, J. A. (2014). Effects of applicant personality on resume evaluations. *Journal of Business and Psychology*, *29*, 573–591.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*, 191–235.

Carlile, P. R. (2002). A pragmatic view of knowledge and boundaries. *Organization Science*, *13*, 442–455.

Carlson, R. E. (1972). The current status of judgmental techniques in industry. In W. C. Byham & D. Bobin (Eds.), *Alternatives to paper and pencil personnel testing*. Pittsburgh, PL: University of Pittsburg.

Carraher, S. M., Sullivan, S. E. E., & Crocitto, M. M. (2008). Mentoring across global boundaries: An empirical examination of home- and host-country mentors on expatriate career outcomes. *Journal of International Business Studies*, *39*, 1310–1326.

Carter, N. M., Gartner, W. B., Shaver, K. G., & Gatewood, E. J. (2003). The career reasons of nascent entrepreneurs. *Journal of Business Venturing*, *18*, 13–39.

Cascio, W. F. (1975). Accuracy of verifiable biographical information blank responses. *Journal of Applied Psychology*, *60*, 767–769.

Castells, M. (2011). *The rise of the network society: The information age: Economy, society, and culture*. Chichester, United Kingdom: John Wiley & Sons.

Cattuto, C., Benz, D., Hotho, A., & Stumme, G. (2008). Semantic grounding of tag relatedness in social bookmarking systems. In A. Sheth, S. Staab, M. Paolucci, D. Maynard, T. Finin, & T. Krishnaprasad (Eds.), *Semantic Web – ISWC* (Vol. 5318, pp. 615–631). Berlin: Springer-Verlag.

Centraal Bureau voor de Statistiek. (2013). *Standaard bedrijfsindeling 2008*. Voorburg, The Netherlands: Author.

Cerdin, J.-L., & Pargneux, M. L. (2009). Career and international assignment fit: Toward an integrative model of success. *Human Resource Management*, *48*, 5–25.

Chen, C.-C., Huang, Y.-M., & Lee, M.-I. (2011). Test of a model linking applicant résumé information and hiring recommendations. *International Journal of Selection and Assessment*, *19*, 374–387.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From Big Data to big impact. *MIS Quarterly*, *36*, 1165–1188.

Chesñevar, C., Mcginnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., … Willmott, S. (2006). Towards an argument interchange format. *The Knowledge Engineering Review*, *21*, 293–316.

Cheung, K.-H., Samwald, M., Auerbach, R. K., & Gerstein, M. B. (2010). Structured digital tables on the semantic web: Toward a structured digital literature. *Molecular Systems Biology*, *6*, 1–13.

Christen, P., & Goiser, K. (2007). Quality and complexity measures for data linkage and deduplication. In F. J. Guillet & H. J. Hamilton (Eds.), *Quality Measures in Data Mining* (pp. 127–151). Berlin, Germany: Springer-Verlag.

Ciancarini, P., Di Iorio, A., Nuzzolese, A. G., Peroni, S., & Vitali, F. (2013). Semantic annotation of scholarly documents and citations. In *AI* IA 2013: Advances in Artificial Intelligence* (Vol. 8249, pp. 336–347). Cham, Switzerland: Springer International.

Clark, D. E. (2004). Practical introduction to record linkage for injury research. *Injury Prevention*, *10*, 186–191.

Codd, E. F. (1974). *Recent investigations in relational data base systems*. Armonk, NY: IBM Thomas J. Watson Research Division.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cohen, W., Ravikumar, P., & Fienberg, S. (2003). A comparison of string metrics for matching names and records (Vol. 3, pp. 73–78). Presented at the Knowledge Discovery and Data Mining, Washington, DC: American Association for Artificial Intelligence.

Cole, M. S., Feild, H. S., & Giles, W. F. (2004). Interaction of recruiter and applicant gender in resume evaluation: A field study. *Sex Roles*, *51*, 597–608.

Cole, M. S., Feild, H. S., Giles, W. F., & Harris, S. G. (2004). Job type and recruiters' inferences of applicant personality drawn from resume biodata: Their relationships with hiring recommendations. *International Journal of Selection and Assessment*, *12*, 363–367.

Cole, M. S., Feild, H. S., Giles, W. F., & Harris, S. G. (2008). Recruiters' inferences of applicant personality based on resume screening: Do paper people have a personality? *Journal of Business and Psychology*, *24*, 5–18.

Cole, M. S., Feild, H. S., & Stafford, J. O. (2005). Validity of resumé reviewers' inferences concerning applicant personality based on resumé evaluation. *International Journal of Selection and Assessment*, *13*, 321–324.

Cole, M. S., Rubin, R. S., Feild, H. S., & Giles, W. F. (2007). Recruiters' perceptions and use of applicant résumé information: Screening the recent graduate. *Applied Psychology: An International Review*, *56*, 319–343.

Connell, C., Bayliss, L., & Farmer, W. (2012). Effects of eBook readers and tablet computers on reading comprehension. *International Journal of Instructional Media*, *39*, 131–140.

Cook, P. F., Larsen, K. R., Sakraida, T. J., & Pedro, L. (2012). A novel quantitative approach to concept analysis: The InterNomological Network. *Nursing Research*, *61*, 369–378.

Cooper, H. M. (1989). *Integrating research: A guide for literature reviews*. Newbury Park, CA: SAGE.

Corbin, J. M., & Strauss, A. L. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*, 3–21.

Correll, S. J. (2004). Constraints into preferences: Gender, status, and emerging career aspirations. *American Sociological Review*, *69*, 93–113.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society, B*, *20*, 215–242.

Cox, M., & Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. In *Proceedings of the 8th Conference on Visualization* (pp. 235–245). Los Alamitos, CA: IEEE Computer Society Press.

Coyle, K. (2006). Mass digitization of books. *The Journal of Academic Librarianship*, *32*, 641–645.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Cunningham, J. W., Powell, T. E., Wimpee, W. E., Wilson, M. A., & Ballentine, R. D. (1996). Ability-requirement factors for general job elements. *Military Psychology*, *8*, 219–234.

Curseu, P. L., & Boros, S. (2008). Stereotypes and prejudice towards women managers: An experimental illustration using the Goldberg-paradigm in a Romanian sample. *Studia Psychologica*, *50*, 233–254.

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning*. Hoboken, NJ: John Wiley & Sons.

Date, C. J. (2003). *An introduction to database systems* (8th ed.). New York, NY: Pearson Higher Education.

Day, C., Elliot, B., & Kington, A. (2005). Reform, standards and teacher identity: Challenges of sustaining commitment. *Teaching and Teacher Education*, *21*, 563–577.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*, 391–407.

DeLisi, M., & Piquero, A. R. (2011). New frontiers in criminal careers research, 2000-2011: A state-of-the-art review. *Journal of Criminal Justice*, *39*, 289–301.

Derous, E., Ryan, A. M., & Nguyen, H.-H. D. (2012). Multiple categorization in resume screening: Examining effects on hiring discrimination against Arab applicants in field and lab settings. *Journal of Organizational Behavior*, *33*, 544–570.

de Solla Price, D. J. (1970). Citation measures of hard science, soft science, technology, and nonscience. In C. E. Nelson & D. K. Pollock (Eds.), *Communication among Scientists and Engineers* (pp. 3–22). Lexington, MA: D.C. Heath and Company.

Desposato, S. W. (2006). Parties for rent? Ambition, ideology, and party switching in Brazil's chamber of deputies. *American Journal of Political Science*, *50*, 62–80.

Devlin, A. S. (1997). Architects: Gender-role and hiring decisions. *Psychological Reports*, *81*, 667–676.

Dewatripont, M., Jewitt, I., & Tirole, J. (1999). The economics of career concerns, part I: Comparing information structures. *Review of Economic Studies*, *66*, 183–198.

De Witte, D., Van de Velde, J., Van Bel, M., Audenaert, P., Demeester, P., Dhoedt, B., … Fostier, J. (2013). Comparative motif discovery in the cloud. Presented at the Benelux Bioinformatics Conference, Brussels, Belgium.

Dickey, E. (2007). *Ancient Greek scholarship: A guide to finding, reading, and understanding scholia, commentaries, lexica, and grammatical treatises, from their beginnings to the Byzantine period*. New York, NY: Oxford University Press.

Dickmann, M., & Doherty, N. (2008). Exploring the career capital impact of international assignments within distinct organizational contexts. *British Journal of Management*, *19*, 145–161.

Dickmann, M., Doherty, N., Mills, T., & Brewster, C. (2008). Why do they go? Individual and corporate perspectives on the factors influencing the decision to accept an international assignment. *International Journal of Human Resource Management*, *19*, 731–751.

Diebold, F. X. (2012, November 26). *A personal perspective on the origin(s) and development of "Big Data": The phenomenon, the term, and the discipline*. Retrieved from http://papers.ssrn.com/abstract=2202843

Diesner, J., Frantz, T. L., & Carley, K. M. (2006). Communication networks from the Enron email corpus. *Computational & Mathematical Organization Theory*, *11*, 201–228.

Ding, L., Kolari, P., Ding, Z., & Avancha, S. (2007). Using ontologies in the semantic web: A survey. In R. Sharman, R. Kishore, & R. Ramesh (Eds.), *Ontologies* (Vol. 14, pp. 79–113). New York, NY: Springer.

DiNucci, D. (1999). Fragmented future. *Print*, *53*, 31, 221–222.

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology*, *32*, 271–297.

Dokko, G., Wilk, S. L., & Rothbard, N. P. (2009). Unpacking prior experience: How career history affects job performance. *Organization Science*, *20*, 51–68.

Duffy, R. D., Allan, B. A., & Dik, B. J. (2011). The presence of a calling and academic satisfaction: Examining potential mediators. *Journal of Vocational Behavior*, *79*, 74–80.

Duffy, R. D., Dik, B. J., & Steger, M. F. (2011). Calling and work-related outcomes: Career commitment as a mediator. *Journal of Vocational Behavior*, *78*, 210–218.

Duffy, R. D., & Sedlacek, W. E. (2007). The presence of and search for a calling: Connections to career development. *Journal of Vocational Behavior*, *70*, 590–601.

Dumbill, E. (2013). Making sense of Big Data. *Big Data*, *1*, 1–2.

Dunleavy, E. M., Mueller, L. M., Buonasera, A. K., Kuang, D. C., & Dunleavy, D. G. (2008). On the consequences of frequent applicants in adverse impact analyses: A demonstration study. *International Journal of Selection and Assessment*, *16*, 333–344.

Dutta, S., & Mia, I. (2009). *The global information technology report 2008–2009: Mobility in a networked world*. Geneva, Switzerland: World Economic Forum.

Dutton, J. E., & Ragins, B. R. (Eds.). (2007). *Exploring positive relationships at work: Building a theoretical and research foundation*. Mahwah, NJ: Lawrence Erlbaum.

Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., & DuBois, D. L. (2008). Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of Vocational Behavior*, *72*, 254–267.

Eby, L. T., Casper, W. J., Lockwood, A., Bordeaux, C., & Brinley, A. (2005). Work and family research in IO/OB: Content analysis and review of the literature (1980-2002). *Journal of Vocational Behavior*, *66*, 124–197.

Egghe, L. (2006). Theory and practice of the g-index. *Scientometrics*, *69*, 131–152.

Elangovan, A. R., Pinder, C. C., & McLean, M. (2010). Callings and organizational behavior. *Journal of Vocational Behavior*, *76*, 428–440.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*, 103–127.

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data analysis. *National Science Review*, *1*, 293–314.

Feito, J. A., & Donahue, P. (2008). Minding the gap: Annotation as preparation for discussion. *Arts and Humanities in Higher Education*, *7*, 295–307.

Feldman, D. C. (1989). Careers in organizations: Recent trends and future directions. *Journal of Management*, *15*, 135–156.

Fellegi, I. P., & Sunter, A. B. (1969). A theory for record linkage. *Journal of the American Statistical Association*, *64*, 1183–1210.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed-and random-effects methods. *Psychological Methods*, *6*, 161.

Fiorina, M. P. (1994). Divided government in the American states - a by-product of legislative professionalism. *American Political Science Review*, *88*, 304–316.

Fitzgerald, L. F., & Rounds, J. B. (1989). Vocational behavior, 1988: A critical analysis. *Journal of Vocational Behavior*, *35*, 105–163.

Floridi, L. (2009). Web 2.0 vs. the semantic web: A philosophical assessment. *Episteme*, *6*, 25–37.

FreshMinds, CapGemini, & EMC. (2015). *Big & fast data: The rise of insight-driven business*. Paris, France: CapGemini. Retrieved from https://www.capgemini.com/thought-leadership/big-fast-data-the-rise-of-insight-driven-business

Fuller, Jr., B., & Marler, L. E. (2009). Change driven by nature: A meta-analytic review of the proactive personality literature. *Journal of Vocational Behavior*, *75*, 329–345.

Gennaioli, N., & Rossi, S. (2010). Judicial discretion in corporate bankruptcy. *Review of Financial Studies*, *23*, 4078–4114.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*, 321–326.

Germain, M.-L., & Tejeda, M. J. (2012). A preliminary exploration on the measurement of expertise: An initial development of a psychometric scale. *Human Resource Development Quarterly*, *23*, 203–232.

Ghiselli, E. E. (1959). 2. The generalization of validity. *Personnel Psychology*, *12*, 397–402.

Ghiselli, E. E. (1974). Some perspectives for industrial psychology. *American Psychologist*, *29*, 80–87.

Gibbons, R., & Murphy, K. J. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *The Journal of Political Economy*, *100*, 468–505.

Giles, J. (2012). Computational social science: Making the links. *Nature*, *488*, 448–450.

Gilster, P. A., Davison, B., & Dickmeyer, W. (2001). Channel the resume flood with applicant tracking systems. *Workforce*, *80*, 30.

Gimpel, J. F. (1973). A theory of discrete patterns and their implementation in SNOBOL4. *Communications of the ACM*, *16*, 91–100.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.

Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology*, *55*, 178.

Glover, I., Xu, Z., & Hardaker, G. (2007). Online annotation – Research and practices. *Computers & Education*, *49*, 1308–1320.

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*, 198–208.

Google. (2015). Interest over time in Big Data. Retrieved December 21, 2015, from https://www.google.com/trends/explore#q=big%20data

Grannis, S. J., Overhage, J. M., Hui, S., & McDonald, C. J. (2003). Analysis of a probabilistic record linkage technique without human review. In *AMIA Annual Symposium Proceedings* (pp. 259–263). Washington, DC: AMIA.

Gravano, L., Ipeirotis, P. G., Jagadish, H. V., Koudas, N., Muthukrishnan, S., Pietarinen, L., & Srivastava, D. (2001). Using q-grams in a DBMS for approximate string processing. *IEEE Data Eng. Bull.*, *24*, 28–34.

Graves, L. M., & Karren, R. J. (1992). Interviewer decision processes and effectiveness: An experimental policy-capturing investigation. *Personnel Psychology*, *45*, 313.

Greenhaus, J. H., & Beutell, N. J. (1985). Sources of conflict between work and family roles. *Academy of Management Review*, *10*, 76–88.

Griskevicius, V., Delton, A. W., Robertson, T. E., & Tybur, J. M. (2011). Environmental contingency in life history strategies: The influence of mortality and socioeconomic status on reproductive timing. *Journal of Personality and Social Psychology*, *100*, 241–254.

Grosof, B. N., Horrocks, I., Volz, R., & Decker, S. (2003). Description logic programs: Combining logic programs with description logic (pp. 48–57). Presented at the WWW2003, Budapest, Hungary: ACM Press.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*, 199–220.

Gruber, T. R. (2007). Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems, 3*, 1–11.

Gruber, T. R. (2009). Ontology. In *Encyclopedia of Database Systems* (Vol. 23, pp. 1963–1965). New York, NY: Springer-Verlag.

Guion, R. M. (1965). Synthetic validity in a small company: A demonstration. *Personnel Psychology*, *18*, 49–63.

Gunz, H. P., & Jalland, R. M. (1996). Managerial careers and business strategies. *Academy of Management Review*, *21*, 718–756.

Gunz, H. P., & Peiperl, M. (2007). Introduction. In *Handbook of Career Studies* (pp. 1–10). Thousand Oaks, CA: SAGE.

Hagmaier, T., & Abele, A. E. (2012). The multidimensionality of calling: Conceptualization, measurement and a bicultural perspective. *Journal of Vocational Behavior*, *81*, 39–51.

Hallberg, U. E., & Schaufeli, W. B. (2006). "Same same" but different? *European Psychologist*, *11*, 119–127.

Hall, D. T. (2004). The protean career: A quarter-century journey. *Journal of Vocational Behavior*, *65*, 1–13.

Hall, D. T., & Chandler, D. E. (2005). Psychological success: When the career is a calling. *Journal of Organizational Behavior*, *26*, 155–176.

Hambrick, D. C. (2007). The field of management's devotion to theory: Too much of a good thing? *Academy of Management Journal*, *50*, 1346–1352.

Hammer, L. B., Allen, E., & Grigsby, T. D. (1997). Work-family conflict in dual-earner couples: Within-individual and crossover effects of work and family. *Journal of Vocational Behavior*, *50*, 185–203.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology*, *94*, 562–575.

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., … Hilf, E. R. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, *30*, 310–314.

Harold, C. M., McFarland, L. A., & Weekley, J. A. (2006). The validity of verifiable and non-verifiable biodata items: An examination across applicants and incumbents. *International Journal of Selection and Assessment*, *14*, 336–346.

Hartley, J., Bartlett, S., & Branthwaite, A. (1980). Underlining can make a difference: Sometimes. *The Journal of Educational Research*, *73*, 218–224.

Harvey-Cook, J. E., & Taffler, R. J. (2000). Biodata in professional entry-level selection: Statistical scoring of common format applications. *Journal of Occupational and Organizational Psychology*, *73*, 103–118.

Harzing, A.-W. (2013). Journal quality list. Retrieved July 16, 2013, from www.harzing.com/jql.htm

Haslberger, A., & Brewster, C. (2009). Capital gains: Expatriate adjustment and the psychological contract in international careers. *Human Resource Management*, *48*, 379–397.

Hedges, L. V., & Olkin, I. (1985). *Statistical method for meta-analysis*. Orlando, FL: Academic Press.

Hendrickson, A. R. (2003). Human resource information systems: Backbone technology of contemporary human resources. *Journal of Labor Research*, *24*, 381–394.

Herzog, T. N., Scheuren, F. J., & Winkler, W. E. (2007). *Data quality and record linkage techniques*. New York, NY: Springer Science & Business Media.

Heslin, P. A. (2005). Conceptualizing and evaluating career success. *Journal of Organizational Behavior*, *26*, 113–136.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557–560.

Higgins, M. C. (2005). *Career imprints: Creating leaders across an industry*. San Francisco, CA: Jossey-Bass.

Hillmert, S., & Jacob, M. (2010). Selections and social selectivity on the academic track: A life-course analysis of educational attainment in Germany. *Research in Social Stratification and Mobility*, *28*, 59–76.

Hindman, A. H., Skibbe, L. E., Miller, A., & Zimmerman, M. (2010). Ecological contexts and early learning: Contributions of child, family, and classroom factors during Head Start, to literacy and mathematics growth through first grade. *Early Childhood Research Quarterly*, *25*, 235–250.

Hirschi, A. (2011). Callings in career: A typological approach to essential and optional components. *Journal of Vocational Behavior*, *79*, 60–73.

Hitzler, P., & Harmelen, F. van. (2010). A reasonable semantic web. *Semantic Web*, *1*, 39–44.

Hogenboom, A., Hogenboom, F., Frasincar, F., Schouten, K., & Meer, O. van der. (2012). Semantics-based information extraction for detecting economic events. *Multimedia Tools and Applications*, *64*, 27–52.

Holland, J. L. (1973). *Making vocational choices: A theory of careers*. Englewood Cliffs, NJ: Prentice Hall.

Hong, H., Kubik, J. D., & Solomon, A. (2000). Security analysts' career concerns and herding of earnings forecasts. *The RAND Journal of Economics*, *31*, 121–144.

Hoque, M. A., Siekkinen, M., & Nurminen, J. K. (2014). Energy efficient multimedia streaming to mobile devices - A survey. *IEEE Communications Surveys Tutorials*, *16*, 579–597.

Hunter, I., Dik, B. J., & Banning, J. H. (2010). College students' perceptions of calling in work and life: A qualitative analysis. *Journal of Vocational Behavior*, *76*, 178–186.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology*, *75*, 28–42.

Hutchinson, K. L. (1984). Personnel administrators' preferences for résumé content: A survey and review of empirically based conclusions. *Journal of Business Communication*, *21*, 5–14.

Hutchinson, K. L., & Brefka, D. S. (1997). Personnel administrators' preferences for resume content: Ten years after. *Business Communication Quarterly*, *60*, 67–75.

IBM. (2011). *The clouds are rolling in... is your business ready?* Retrieved from https://www.ibm.com/developerworks/community/blogs/ff67b471-79df-4bef-

9593-
4802def4013d/entry/2011_ibm_tech_trends_report_the_clouds_are_rolling_in_is_your_business_ready5?lang=en

Inkson, K., Gunz, H., Ganesh, S., & Roper, J. (2012). Boundaryless careers: Bringing back boundaries. *Organization Studies*, *33*, 323–340.

Inkson, K., & Savickas, M. L. (Eds.). (2013a). *Careers as human experience* (Vol. 3). Thousand Oaks, CA: SAGE.

Inkson, K., & Savickas, M. L. (Eds.). (2013b). *Careers in context* (Vol. 2). Thousand Oaks, CA: SAGE.

Inkson, K., & Savickas, M. L. (Eds.). (2013c). *Careers in practice* (Vol. 4). Thousand Oaks, CA: SAGE.

Inkson, K., & Savickas, M. L. (Eds.). (2013d). *Foundations of career studies* (Vol. 1). Thousand Oaks, CA: SAGE.

Jacobs, A. (2009). The pathologies of Big Data. *Communications of the ACM*, *52*, 36–44.

Jaro, M. A. (1989). Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *Journal of the American Statistical Association*, *84*, 414–420.

Jassawalla, A. R., & Sashittal, H. C. (2009). Thinking strategically about integrating repatriated managers in MNCS. *Human Resource Management*, *48*, 769–792.

Jeanneret, P. R. (1992). Applications of job component/synthetic validity to construct validity. *Human Performance*, *5*, 81.

Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, *7*, 238–257.

Johnson, L. L. (1988). Effects of underlining textbook sentences on passage and sentence retention. *Reading Research and Instruction*, *28*, 18–32.

Johnson, T. E., Archibald, T. N., & Tenenbaum, G. (2010). Individual and team annotation effects on students' reading comprehension, critical thinking, and meta-cognitive skills. *Computers in Human Behavior*, *26*, 1496–1507.

Jokinen, T., Brewster, C., & Suutari, V. (2008). Career capital during international work experiences: Contrasting self-initiated expatriate experiences and assigned expatriation. *International Journal of Human Resource Management*, *19*, 979–998.

Justeson, J. S., & Katz, S. M. (1995). Technical terminology: Some linguistic properties and an algorithm for identification in text. *Natural Language Engineering*, *1*, 9–27.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. In *46th International Conference on System Sciences (HICSS)* (pp. 995–1004). Maui, HI: IEEE.

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, *71*, 589–617.

Kanter, R. M. (1990). *When giants learn to dance*. New York, NY: Free Press.

Kaplan, S. H., Sullivan, L. M., Dukes, K. A., Phillips, C. F., Kelch, R. P., & Schaller, J. G. (1996). Sex differences in academic advancement - results of a national study of pediatricians. *New England Journal of Medicine*, *335*, 1282–1289.

Karren, R. J., & Barringer, M. W. (2002). A review and analysis of the policy-capturing methodology in organizational research: Guidelines for research and practice. *Organizational Research Methods*, *5*, 337–361.

Kelchtermans, G. (2009). Who I am in how I teach is the message: Self-understanding, vulnerability and reflection. *Teachers and Teaching*, *15*, 257–272.

Kemery, E. R., Dunlap, W. P., & Griffeth, R. W. (1988). Correction for variance restriction in point-biserial correlations. *Journal of Applied Psychology*, *73*, 688–691.

Khapova, S. N., & Arthur, M. B. (2011). Interdisciplinary approaches to contemporary career studies. *Human Relations*, *64*, 3–17.

Kim, T.-Y., Hon, A. H. Y., & Crant, J. M. (2009). Proactive personality, employee creativity, and newcomer outcomes: A longitudinal study. *Journal of Business and Psychology*, *24*, 93–103.

King, L. A., & Napa, C. K. (1998). What makes a life good? *Journal of Personality and Social Psychology*, *75*, 156–165.

Knouse, S. B. (1994). Impressions of the resume: The effects of applicant education, experience, and impression management. *Journal of Business and Psychology*, *9*, 33–45.

Knublauch, H., Oberle, D., Tetlow, P., & Wallace, E. (Eds.). (2006, September 3). A semantic web primer for object-oriented software developers. W3C. Retrieved from http://www.w3.org/TR/sw-oosd-primer/

Krueger, D. C., Stone, D. L., & Stone-Romero, E. (2014). Applicant, rater, and job factors related to weight-based bias. *Journal of Managerial Psychology*, *29*, 164–186.

Landon, B. E., Reschovsky, J., & Blumenthal, D. (2003). Changes in career satisfaction among primary care and specialist physicians, 1997-2001. *Journal of the American Medical Association*, *289*, 442–449.

Laney, D. (2001). Application delivery strategies. META Group. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Larsen, K. R., Lee, J., Li, J., & Bong, C. H. (2010). A transdisciplinary approach to construct search and integration. In *AMCIS 2010 Proceedings* (pp. 1–9). Lima, Peru: AMCIS.

Larsen, K. R., & Monarchi, D. E. (2004). A mathematical approach to categorization and labeling of qualitative data: The latent categorization method. *Sociological Methodology*, *34*, 349–392.

Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, *84*, 575–603.

Laub, J. H., Nagin, D. S., & Sampson, R. J. (1998). Trajectories of change in criminal offending: Good marriages and the desistance process. *American Sociological Review*, *63*, 225–238.

Lawshe, C. H. (1952). What can industrial psychology do for small business (a symposium). *Personnel Psychology*, *5*, 31–34.

Lee, C. I. S. G., Felps, W., & Baruch, Y. (2014a). Mapping career studies: A bibliometric analysis. *Academy of Management Proceedings*, *2014*. doi:10.5465/AMBPP.2014.284

Lee, C. I. S. G., Felps, W., & Baruch, Y. (2014b). Toward a taxonomy of career studies through bibliometric visualization. *Journal of Vocational Behavior*, *85*, 339–351.

Lee, D. M. S., Trauth, E. M., & Farwell, D. (1995). Critical skills and knowledge requirements of IS professionals - a joint academic-industry investigation. *MIS Quarterly*, *19*, 313–340.

Le, H., Schmidt, F. L., Harter, J. K., & Lauver, K. J. (2010). The problem of empirical redundancy of constructs in organizational research: An empirical investigation. *Organizational Behavior and Human Decision Processes*, *112*, 112–125.

Leigh, J. P., Kravitz, R. L., Schembri, M., Samuels, S. J., & Mobley, S. (2002). Physician career satisfaction across specialties. *Archives of Internal Medicine*, *162*, 1577–1584.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*, 79–122.

Lent, R. W., Brown, S. D., & Hackett, G. (2000). Contextual supports and barriers to career choice: A social cognitive analysis. *Journal of Counseling Psychology*, *47*, 36–49.

Lent, R. W., Brown, S. D., Sheu, H. B., Schmidt, J., Brenner, B. R., & Gloster, C. S. (2005). Social cognitive predictors of academic interests and goals in engineering: Utility for women and students at historically black universities. *Journal of Counseling Psychology*, *52*, 84–92.

Lentz, E., & Allen, T. D. (2009). The role of mentoring others in the career plateauing phenomenon. *Group & Organization Management*, *34*, 358–384.

Lerner, J., & Tirole, J. (2002). Some simple economics of open source. *Journal of Industrial Economics*, *50*, 197–234.

Levine, E. L., & Flory, A. (1975). Evaluation of job applications - A conceptual framework. *Public Personnel Management*, *4*, 378–385.

Liakata, M., & Soldatova, L. (2008). *Guidelines for the annotation of general scientific concepts*. Aberystwyth, United Kingdom: Department of Computer Science, Aberystwyth University. Retrieved from http://www.cl.cam.ac.uk/~yg244/10crab/S3_CoreSC.pdf

Li, J., & Larsen, K. R. (2011). Establishing nomological networks for behavioral science: A natural language processing based approach. In *ICIS 2011 Proceedings*. Shanghai, China: AIS.

Lin, H., Davis, J., & Zhou, Y. (2009). An integrated approach to extracting ontological structures from folksonomies. In L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvonen, … E. Simperl (Eds.), *Semantic Web: Research and Applications* (Vol. 5554, pp. 654–668). Berlin: Springer-Verlag.

Linzer, M., Konrad, T. R., Douglas, J., McMurray, J. E., Pathman, D. E., Williams, E. S., … Rhodes, E. (2000). Managed care, time pressure, and physician job satisfaction: Results from the physician worklife study. *Journal of General Internal Medicine*, *15*, 441–450.

Lips-Wiersma, M., & Hall, D. T. (2007). Organizational career development is not dead: A case study on managing the new career during organizational change. *Journal of Organizational Behavior*, *28*, 771–792.

Lopatin, L. (2006). Library digitization projects, issues and guidelines: A survey of the literature. *Library Hi Tech*, *24*, 273–289.

Luukkonen, T., Tijssen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, *28*, 15–36.

Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*, *1*, 1–13.

Maclean, M., Harvey, C., & Chia, R. (2012). Sensemaking, storytelling and the legitimization of elite business careers. *Human Relations*, *65*, 17–40.

Madden, S. (2012). From databases to Big Data. *IEEE Internet Computing*, *19*, 4–6.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology*, *44*, 763–792.

Manyika, J., Chui, M., Brad, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. New York, NY: McKinsey & Company. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Manzoni, A., Härkönen, J., & Mayer, K. U. (2014). Moving on? A growth-curve analysis of occupational attainment and career progression patterns in West Germany. *Social Forces*, *92*, 1285–1312.

Maranda, M.-F., & Comeau, Y. (2000). Some contributions of sociology to the understanding of career. In A. Collin & R. A. Young (Eds.), *The future of career* (pp. 37–52). Cambridge, United Kingdom: Cambridge University Press.

Marshall, C. C. (1997). Annotation: From paper books to the digital library. In *Proceedings of the Second ACM International Conference on Digital Libraries* (pp. 131–140). New York, NY: ACM Press.

Marshall, C. C. (1998). Toward an ecology of hypertext annotation. In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia* (pp. 40–49). New York, NY: ACM Press.

Masinter, L., Berners-Lee, T., & Fielding, R. T. (2005). Uniform resource identifier (URI): Generic syntax. The Internet Society. Retrieved from http://tools.ietf.org/html/rfc3986%22%3ERFC

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston, MA: Houghton Mifflin Harcourt.

Mazen, A. M. (1990). The moderating role of social desirability, age, and experience in human judgment: Just how indirect is policy capturing? *Organizational Behavior and Human Decision Processes*, *45*, 19–40.

McAndrew, D. A. (1983). Underlining and notetaking: Some suggestions from research. *Journal of Reading*, *27*, 103–108.

McCloy, R. A. (1994). Predicting job performance scores without performance data. In B. F. Green Jr. & A. S. Mavor (Eds.), *Modeling Cost and Performance for Military Enlistment: Report of a Workshop* (pp. 61–100). Washington DC: National Academy Press.

McCloy, R. A. (2001). Predicting job performance scores in jobs lacking criterion data. Presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA: Academic Press.

McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. (1972). A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). *Journal of Applied Psychology*, *56*, 347–368.

McCrory, M., & Mueller, D. (2000). Nebraska “whole picture” recruitment: From paperwork to “people work” featuring the persoic applicant tracking system. *Public Personnel Management*, *29*, 505–510.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988). Job experience correlates of job-performance. *Journal of Applied Psychology*, *73*, 327–330.

McKinney, A. P., Carlson, K. D., Mecham, R. L., D’Angelo, N. C., & Connerley, M. L. (2003). Recruiters’ use of GPA in initial screening decisions: Higher GPAs don’t always make the cut. *Personnel Psychology*, *56*, 823–845.

Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*. Gaithersburg, MD: National Institute of Standards and Technology. Retrieved from http://faculty.winthrop.edu/domanm/csci411/Handouts/NIST.pdf

Menger, P. M. (1999). Artistic labor markets and careers. *Annual Review of Sociology*, *25*, 541–574.

Miles, A., & Bechhofer, S. (2009). SKOS simple knowledge organization system reference. W3C. Retrieved from http://www.w3.org/TR/skos-reference/

Ministerie van Social Zaken en Werkgelegenheid. (2015). *Monitor arbeidsmarkt april 2015*. The Hague, The Netherlands: Author. Retrieved from https://www.rijksoverheid.nl/documenten/rapporten/2015/05/19/monitor-arbeidsmarkt-april-2015

Mitkov, R. (2005). *The Oxford Handbook of Computational Linguistics*. New York, NY: Oxford Univeristy Press.

Moreland, J. L., Dansereau, D. F., & Chmielewski, T. L. (1997). Recall of descriptive information: The roles of presentation format, annotation strategy, and individual differences. *Contemporary Educational Psychology*, *22*, 521–533.

Mossholder, K. W., & Arvey, R. D. (1984). Synthetic validity: A conceptual and comparative review. *Journal of Applied Psychology*, *69*, 322–333.

Neal, D. A., & Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *The Journal of Political Economy*, *104*, 869–895.

Nemanick, J. Richard C., & Clark, E. M. (2002). The differential effects of extracurricular activities on attributions in résumé evaluation. *International Journal of Selection and Assessment*, *10*, 206–217.

Newcombe, H. B., Kennedy, J. M., Axford, S. J., & James, A. P. (1959). Automatic linkage of vital records computers can be used to extract follow-up statistics of families from files of routine records. *Science*, *130*, 954–959.

Ng, T. W. H., & Feldman, D. C. (2008). The relationship of age to ten dimensions of job performance. *Journal of Applied Psychology*, *93*, 392–423.

Noll, M. G., & Meinel, C. (2008). Exploring social annotations for web document classification. In *Proceedings of the 2008 ACM Symposium on Applied Computing* (pp. 2315–2320). New York, NY: ACM Press.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*, 101–115.

Oberlander, J., & Nowson, S. (2006). Whose thumb is it anyway? Classifying author personality from weblog text. In *Proceedings of the COLING/ACL Main Conference Poster Sessions* (pp. 627–634). Stroudsburg, PA: Association for Computational Linguistics.

O'Brien, K. E., Biga, A., Kessler, S. R., & Allen, T. D. (2010). A meta-analytic investigation of gender differences in mentoring. *Journal of Management*, *36*, 537–554.

OECD. (2013). *The internet economy on the rise: Progress since the Seoul Declaration*. Paris, France: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/science-and-technology/the-internet-economy-on-the-rise_9789264201545-en

Oliphant, van N., & Alexander, E. R. I. (1982). Reactions to resumes as a function of resume determinateness, applicant characteristics, and sex of raters. *Personnel Psychology*, *35*, 829–842.

O'Reilly, T. (2005). What Is Web 2.0. Retrieved from http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html

Osborne, J. W. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation*, *15*, 1–9.

Özbilgin, M., & Tatli, A. (2011). Mapping out the field of equality and diversity: Rise of individualism and voluntarism. *Human Relations*, *64*, 1229–1253.

Özmen Uysal, Ö. (2010). Business ethics research with an accounting focus: A bibliometric analysis from 1988 to 2007. *Journal of Business Ethics*, *93*, 137–160.

Parasuraman, S., Purohit, Y. S., Godshalk, V. M., & Beutell, N. J. (1996). Work and family variables, entrepreneurial career success, and psychological well-being. *Journal of Vocational Behavior*, *48*, 275–300.

Parker, P., Hall, D. T., & Kram, K. E. (2008). Peer coaching: A relational process for accelerating career learning. *Academy of Management Learning & Education*, *7*, 487–503.

Penev, L., Agosti, D., Georgiev, T., Catapano, T., Miller, J., Blagoderov, V., … others. (2010). Semantic tagging of and semantic enhancements to systematics papers: ZooKeys working examples. *ZooKeys*, *50*, 1–16.

Penev, L., Kress, W. J., Knapp, S., Li, D.-Z., & Renner, S. (2010). Fast, linked, and open–the future of taxonomic publishing for plants: Launching the journal PhytoKeys. *PhytoKeys*, *1*, 1–14.

Pérez, J., Arenas, M., & Gutierrez, C. (2006). Semantics and complexity of SPARQL. In *International Semantic Web Conference Proceedings* (Vol. 4273, pp. 30–43). Athens, GA: Springer.

Peroni, S., Dutton, A., Gray, T., & Shotton, D. (2015). Setting our bibliographic references free: Towards open citation data. *Journal of Documentation*, *71*, 253–277.

Peroni, S., & Shotton, D. (2012). FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *Web Semantics: Science, Services and Agents on the World Wide Web*, *17*, 33–43.

Peters, H. P. F., & van Raan, A. F. J. (1993). Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. *Research Policy*, *22*, 23–45.

Peterson, N. G., & Bownas, D. A. (1982). Skill, task structure, and performance acquisition. *Human Performance and Productivity*, *1*, 49–105.

Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., Fleishman, E. A., Levin, K. Y., … Dye, D. M. (2001). Understanding work using the Occupational Information Network (O*NET): Implications for practice and research. *Personnel Psychology*, *54*, 451–492.

Peterson, N. G., Wise, L. L., Arabian, J., & Gene, R. (2001). Synthetic validation and validity generalization: When empirical validation is not possible. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 411–451). Mahwah, NJ: Lawrence Erlbaum.

Piotrowski, C., & Armstrong, T. (2006). Current recruitment and selection practices: A national survey of fortune 1000 firms. *North American Journal of Psychology*, *8*, 489–496.

Piquero, A. R., Farrington, D. P., & Blumstein, A. (2003). The criminal career paradigm. In M. H. Tonry (Ed.), *Crime and justice: A review of research*. Chicago, IL: University of Chicago Press.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539–569.

Porter-O'Donnell, C. (2004). Beyond the yellow highlighter: Teaching annotation skills to improve reading comprehension. *English Journal*, *93*, 82–89.

Press, G. (2013). A very short history of Big Data. *Forbes*. Retrieved from http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/

Primoff, E. S. (1955). *Test selection by job analysis: The J-coefficient, what it is, how it works*. Washington, DC: U.S. Civil Service Commission, Standards Division.

Primoff, E. S. (1959). 4. Empirical validations of the J-coefficient. *Personnel Psychology*, *12*, 413–418.

Probert, B. (2005). "I just couldn't fit it in": Gender and unequal outcomes in academic careers. *Gender Work and Organization*, *12*, 50–72.

Putka, D. J., & Oswald, F. L. (2015). Implications of the Big Data movement for the advancement of I-O science and practice. In S. Tonidandel, E. King, & J. Cortina

(Eds.), *Big data at work: The data science revolution and organizational psychology*. New York, NY: Routledge.

Quińones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytic review. *Personnel Psychology*, *48*, 887–910.

Rabinowitz, H. K., Diamond, J. J., Markham, F. W., & Paynter, N. P. (2001). Critical factors for designing programs to increase the supply and retention of rural primary care physicians. *Journal of the American Medical Association*, *286*, 1041–1048.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy*, *41*, 1262–1282.

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, *61*, 1871–1887.

Rahwan, I., Zablith, F., & Reed, C. (2007). Laying the foundations for a World Wide Argument Web. *Artificial Intelligence*, *171*, 897–921.

Ramos-Rodríguez, A.-R., & Ruíz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: A bibliometric study of the Strategic Management Journal, 1980–2000. *Strategic Management Journal*, *25*, 981–1004.

Richardson, J., McBey, K., & McKenna, S. (2008). Integrating realistic job previews and realistic living conditions previews: Realistic recruitment for internationally mobile knowledge workers. *Personnel Review*, *37*, 490–508.

Rijksoverheid. (2005). *Het onderwijssysteem in Nederland 2005*. The Hague, The Netherlands: Ministerie van Onderwijs, Cultuur en Wetenschappen. Retrieved from http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/rapporten/2005/10/14/het-onderwijssysteem-in-nederland-2005/onderwijs-doc-2005-eurydice-nl.pdf

Rip, A., & Courtial, J. P. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, *6*, 381–400.

Rizzi, F., van Eck, N. J. P., & Frey, M. (2014). The production of scientific knowledge on renewable energies: Worldwide trends, dynamics and challenges and implications for management. *Renewable Energy*, *62*, 657–671.

Robusto, C. C. (1957). The cosine-haversine formula. *American Mathematical Monthly*, *64*, 38–40.

Rodrigues, R. A., & Guest, D. (2010). Have careers become boundaryless? *Human Relations*, *63*, 1157–1175.

Ross, C. M., & Young, S. J. (2005). Resume preferences: Is it really "business as usual"? *Journal of Career Development*, *32*, 153–164.

Roulin, N., & Bangerter, A. (2013). Extracurricular activities in young applicants' résumés: What are the motives behind their involvement? *International Journal of Psychology*, *48*, 871–880.

Rubin, R. S., Bommer, W. H., & Baldwin, T. T. (2002). Using extracurricular activity as an indicator of interpersonal skill: Prudent evaluation or recruiting malpractice? *Human Resource Management*, *41*, 441–454.

Ruiz-Iniesta, A., & Corcho, O. (2014). A review of ontologies for describing scholarly and scientific documents. In *Proceedings of 4th Workshop on Semantic Publishing* (Vol. 1155). Aachen, Germany: CEURWS. Retrieved from http://www.researchgate.net/profile/Oscar_Corcho/publication/263495382_A_review_of_ontologies_for_describing_scholarly_and_scientific_document/links/53d8bc880cf2a19eee836c97.pdf

Sackett, P. R. (1991). Exploring strategies for clustering military occupations. In A. K. Widgor & B. F. Green (Eds.), *Technical issues* (Vol. 2). Washington, DC: National Academy Press.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, *88*, 1068–1081.

Sarcevic, A., Palen, L., White, J., Starbird, K., Bagdouri, M., & Anderson, K. (2012). "Beacons of hope" in decentralized coordination: Learning from on-the-ground medical twitterers during the 2010 Haiti earthquake. In *Proceedings of the ACM*

2012 Conference on Computer Supported Cooperative Work (pp. 47–56). New York, NY: ACM Press.

Savickas, M. L. (2002). Career construction: A developmental theory of vocational behavior. In D. Brown (Ed.), *Career choice and development* (4th ed., pp. 149–205). San Francisco, CA: Jossey-Bass.

Savickas, M. L. (2005). The theory and practice of career construction. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 42–70). Hoboken, NJ: John Wiley & Sons.

Savickas, M. L., & Porfeli, E. J. (2012). Career Adapt-Abilities Scale: Construction, reliability, and measurement equivalence across 13 countries. *Journal of Vocational Behavior*, *80*, 661–673.

Scherbaum, C. A. (2005). Synthetic validity: Past, present, and future. *Personnel Psychology*, *58*, 481–515.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, *50*, 835–850.

Schneider, J., Groza, T., & Passant, A. (2013). A review of argumentation for the Social Semantic Web. *Semantic Web*, *4*, 159–218.

Schröder, T., Rogers, K. B., Ike, S., Mell, J. N., & Scholl, W. (2013). Affective meanings of stereotyped social groups in cross-cultural comparison. *Group Processes & Intergroup Relations*, *16*, 717–733.

Schwab, D. P. (1980). Construct validity in organizational behavior. *Research in Organizational Behavior*, *2*, 3–43.

Scopus. (2014). Content coverage guide. Elsevier. Retrieved from http://www.scopus.fecyt.es/SiteAssets/Pages/info_editores/sccg0510.pdf

Seibert, S. E., & Kraimer, M. L. (2001). The five-factor model of personality and career success. *Journal of Vocational Behavior*, *58*, 1–21.

Seibert, S. E., Kraimer, M. L., & Liden, R. C. (2001). A social capital theory of career success. *Academy of Management Journal*, *44*, 219–237.

Shaffer, G. S., Saunders, V., & Owens, W. A. (1986). Additional evidence for the accuracy of biographical data: Long-term retest and observer ratings. *Personnel Psychology*, *39*, 791–809.

Shepitsen, A., Gemmell, J., Mobasher, B., & Burke, R. (2008). Personalized recommendation in social tagging systems using hierarchical clustering. In *Proceedings of the 2008 ACM Conference on Recommender Systems* (pp. 259–266). New York, NY: ACM Press.

Shirky, C. (2005). Ontology is overrated – Categories, links, and tags. Retrieved from http://shirky.com/writings/ontology_overrated.html

Shotton, D. (2009). Semantic publishing: The coming revolution in scientific journal publishing. *Learned Publishing*, *22*, 85–94.

Shotton, D. (2015, January 22). FaBiO, the FRBR-aligned bibliographic ontology. University of Oxford Vocabularies. Retrieved from http://vocab.ox.ac.uk/fabio

Simperl, E. (2009). Reusing ontologies on the Semantic Web: A feasibility study. *Data & Knowledge Engineering*, *68*, 905–925.

Simpson, M. L., & Nist, S. L. (1990). Textbook annotation: An effective and efficient study strategy for college students. *Journal of Reading*, *34*, 122–129.

Singer, M. S., & Bruhns, C. (1991). Relative effect of applicant work experience and academic qualification on selection interview decisions: A study of between-sample generalizability. *Journal of Applied Psychology*, *76*, 550–559.

Singh, D., & Reddy, C. K. (2014). A survey on platforms for Big Data analytics. *Journal of Big Data*, *2*, 1–20.

Singh, J. (1991). Redundancy in constructs: Problem, assessment, and an illustrative example. *Journal of Business Research*, *22*, 255–280.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: Visualizing classifier performance in R. *Bioinformatics*, *21*, 3940–3941.

Specia, L., & Motta, E. (2007). Integrating folksonomies with the semantic web. In E. Franconi, M. Kifer, & W. May (Eds.), *Semantic Web: Research and Applications, Proceedings* (Vol. 4519, pp. 624–639). Berlin: Springer-Verlag.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, *60*, 950–958.

Spitters, M., Bonnema, R., Rotaru, M., & Zavrel, J. (2010). Bootstrapping information extraction mappings by similarity-based reuse of taxonomies. In *Proceedings of the EKAW 2010 Workshop on Reuse and Adaptation of Ontologies and Terminologies* (pp. 1–13). Lisbon, Portugal: Springer-Verlag.

Spurk, D., Abele, A. E., & Volmer, J. (2011). The Career Satisfaction Scale: Longitudinal measurement invariance and latent growth analysis. *Journal of Occupational and Organizational Psychology*, *84*, 315–326.

Steel, P. D. G., Huffcutt, A. I., & Kammeyer-Mueller, J. (2006). From the work one knows the worker: A systematic review of the challenges, solutions, and steps to creating synthetic validity. *International Journal of Selection and Assessment*, *14*, 16–36.

Steel, P. D. G., & Kammeyer-Mueller, J. (2009). Using a meta-analytic perspective to enhance job component validation. *Personnel Psychology*, *62*, 533–552.

Stokes, G. S., Hogan, J. B., & Snell, A. F. (1993). Comparability of incumbent and applicant samples for the development of biodata keys: The influence of social desirability. *Personnel Psychology*, *46*, 739–762.

Stone, D. L., Deadrick, D. L., Lukaszewski, K. M., & Johnson, R. (2015). The influence of technology on the future of human resource management. *Human Resource Management Review*, *25*, 216–231.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, *100*, 255–270.

Strohmeier, S. (2007). Research in e-HRM: Review and implications. *Human Resource Management Review*, *17*, 19–37.

Sturman, M. C. (2003). Searching for the inverted U-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management*, *29*, 609–640.

Sullivan, S. E. E. (1999). The changing nature of careers: A review and research agenda. *Journal of Management*, *25*, 457–484.

Sullivan, S. E. E., & Baruch, Y. (2009). Advances in career theory and research: A critical review and agenda for future exploration. *Journal of Management*, *35*, 1542–1571.

Suutari, V., & Brewster, C. (2003). Repatriation: Empirical evidence from a longitudinal study of careers and expectations among Finnish expatriates. *International Journal of Human Resource Management*, *14*, 1132–1151.

Sweeney, M. M. (2002). Two decades of family change: The shifting economic foundations of marriage. *American Sociological Review*, *67*, 132–147.

Tang, N. (2014). Big data cleaning. In L. Chen, Y. Jia, T. Sellis, & G. Liu (Eds.), *Web technologies and applications* (pp. 13–24). Cham, Switzerland: Springer International Publishing.

Tesser, A., & Krauss, H. (1976). On validating a relationship between constructs. *Educational and Psychological Measurement*, *36*, 111–121.

Thomson, K. S. (1984). Marginalia: The literature of science. *American Scientist*, *72*, 185–187.

Thomson Reuters. (2015). Citation report Big Data. Retrieved December 21, 2015, from http://apps.webofknowledge.com/CitationReport.do

Thoms, P., McMasters, R., Roberts, M. R., & Dombkowski, D. A. (1999). Resume characteristics as predictors of an invitation to interview. *Journal of Business and Psychology*, *13*, 339–356.

Tijdens, K. G., Dragstra, A. E., Dragstra, D., Klaveren, M., Osse, P., Wetzels, C. M. M. P., & Zorlu, A. (2002). *Loonwijzers 2001/(2002): Werk, lonen en beroepen van mannen en vrouwen in Nederland*. Amsterdam, The Netherlands: Amsterdam Institute for Advances Labor Studies. Retrieved from http://dare.uva.nl/record/1/204208

Tirole, J. (1994). The internal organization of government. *Oxford Economic Papers*, *46*, 1–29.

Tokar, D. M., Fischer, A. R., & Subich, L. M. (1998). Personality and vocational behavior: A selective review of the literature, 1993-1997. *Journal of Vocational Behavior*, *53*, 115–153.

Tonidandel, S., & LeBreton, J. M. (2010). Determining the relative importance of predictors in logistic regression: An extension of relative weight analysis. *Organizational Research Methods*, *13*, 767–781.

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology*, *26*, 1–9.

Tonidandel, S., LeBreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological Methods*, *14*, 387–399.

Troester, M. (2012). *Big data meets Big Data analytics: Three key technologies for extracting real-time business value from the Big Data, that threatens to overwhelm traditional computing architectures*. Cary, NC: SAS Institute.

Turnbull, J., Lea, D., Parkinson, D., Phillips, P., Francis, B., Webb, S., … Ashby, M. (Eds.). (2010). *Oxford advanced learner's dictionary* (8th ed.). Oxford, United Kingdom: Oxford University Press. Retrieved from http://www.oxforddictionaries.com/definition/learner/big-data

Tyson, K., Castellino, D. R., & Darity, W. (2005). It's not "a black thing": Understanding the burden of acting white and other dilemmas of high achievement. *American Sociological Review*, *70*, 582–605.

U.S. Bureau of Labor Statistics. (2010). Record unemployment among older workers does not keep them out of the job market. *Issues in Labor Statistics*, *10*, 1–3.

U.S. Bureau of Labor Statistics. (2012). Young adult employment during the recent recession. *Issues in Labor Statistics*, *12*, 1–11.

Valcour, M., Ollier-Malaterre, A., Matz-Costa, C., Pitt-Catsouphes, M., & Brown, M. (2011). Influences on employee perceptions of organizational work–life support: Signals and resources. *Journal of Vocational Behavior*, *79*, 588–595.

van Eck, N. J. P., & Waltman, L. (2007). Bibliometric mapping of the computational intelligence field. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *15*, 625–645.

van Eck, N. J. P., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, *60*, 1635–1651.

van Eck, N. J. P., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*, 523–538.

van Eck, N. J. P., & Waltman, L. (2011). Text mining and visualization using VOSviewer. Retrieved from http://arxiv.org/abs/1109.2058

van Eck, N. J. P., Waltman, L., Dekker, R., & Van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology*, *61*, 2405–2416.

van Eck, N. J. P., Waltman, L., van den Berg, J., & Kaymak, U. (2006). Visualizing the computational intelligence field. *IEEE Computational Intelligence Magazine*, *1*, 6–10.

Waaijer, C. J. F., van Bochove, C. A., & van Eck, N. J. P. (2010). Journal editorials give indication of driving science issues. *Nature*, *463*, 157–157.

Waaijer, C. J. F., van Bochove, C. A., & van Eck, N. J. P. (2011). On the map: Nature and Science editorials. *Scientometrics*, *86*, 99–112.

Wajda, A., & Roos, L. L. (1987). Simplifying record linkage: Software and strategy. *Computers in Biology and Medicine*, *17*, 239–248.

Waltman, L., van Eck, N. J. P., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, *4*, 629–635.

Wang, M., & Hanges, P. J. (2011). Latent class procedures: Applications to organizational research. *Organizational Research Methods*, *14*, 24–31.

Wang, M., Zhan, Y., Liu, S., & Shultz, K. S. (2008). Antecedents of bridge employment: A longitudinal investigation. *Journal of Applied Psychology*, *93*, 818–830.

Warr, M. (1998). Life-course transitions and desistance from crime. *Criminology*, *36*, 183–216.

Watkins, L. M., & Johnston, L. (2000). Screening job applicants: The impact of physical attractiveness and application quality. *International Journal of Selection and Assessment*, *8*, 76–84.

Wenzel, R., & van Quaquebeke, N. (in press). The double-edged sword of Big Data in organizational research. *Organizational Research Methods*.

Werbel, J., & Others, A. (1989). Is prescreening biased? *Journal of Career Planning and Employment*, *49*, 41–43.

Western, B. (2002). The impact of incarceration on wage mobility and inequality. *American Sociological Review*, *67*, 526–546.

Whiston, S. C., & Keller, B. K. (2004). The influences of the family of origin on career development: A review and analysis. *The Counseling Psychologist*, *32*, 493–568.

Williams, E. S., Konrad, T. R., Linzer, M., McMurray, J., Pathman, D. E., Gerrity, M., … Douglas, J. (1999). Refining the measurement of physician job satisfaction results from the physician worklife survey. *Medical Care*, *37*, 1140–1154.

Winkler, W. E. (1990). String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In *Proceedings of the Section on Survey Research* (pp. 354–359). Washington, DC: ASA.

Winkler, W. E. (1999). *The state of record linkage and current research problems*. Washington, DC: Statistical Research Division, US Census Bureau.

Wolfe, J. L. (2000). Effects of annotations on student readers and writers. In *Proceedings of the Fifth ACM Conference on Digital Libraries* (pp. 19–26). San Antonio, TX: ACM Press.

Woo, S. E. (2011). A study of Ghiselli's hobo syndrome. *Journal of Vocational Behavior*, *79*, 461–469.

Yan, L., Zhang, Y., Yang, L. T., & Ning, H. (2008). *The Internet of things: From RFID to the next-generation pervasive networked systems*. New York, NY: CRC Press.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*, 665–670.

Yarnold, P. R., & Soltysik, R. C. (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC: APA Books.

Yeung, C. A., Noll, M. G., Gibbins, N., Meinel, C., & Shadbolt, N. (2011). SPEAR: Spamming-resistant expertise analysis and ranking in collaborative tagging systems. *Computational Intelligence*, *27*, 458–488.

Young, I. P., Young, K. H., & Oto, T. (2011). Superintendents' assessment of Hispanic, Latino, and white job candidates for a high school principalship: Effects of cultural subdivisions and student characteristics. *Leadership and Policy in Schools*, *10*, 207–233.

Young, R. A., Marshall, S. K., Foulkes, K., Haber, C., Lee, C. S. M., Penner, C., & Rostram, H. (2011). Counseling for the transition to adulthood as joint, goal-directed action. *Journal of Vocational Behavior*, *79*, 325–333.

Zavrel, J., & Daelemans, W. (2003). Feature-rich memory-based classification for shallow NLP and information extraction. Text Mining, Theoretical Aspects and Applications, 33–54.

Zebrowitz, L. A., Tenenbaum, D. R., & Goldstein, L. H. (1991). The impact of job applicants' facial maturity, gender, and academic achievement on hiring recommendations. Journal of Applied Social Psychology, 21, 525–548.

Zickuhr, K., & Rainie, L. (2014). *E-reading rises as device ownership jumps*. Washington, DC: Pew Research Center. Retrieved from http://pewinternet.org/Reports/2014/E-Reading-Update.aspx

Zikopoulos, P., deRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D., & Giles, J. (2013). *Harness the power of Big Data: The IBM Big Data platform*. New York, NY: McGraw Hill Professional.

Zitt, M., Bassecoulard, E., & Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics*, *47*, 627–657.

Summary

Digital computers entered our homes, landed on our desktops, slipped into our pockets, and have seemingly become ubiquitous. At an ever faster pace, these devices have become highly interconnected and interoperable. Consequently, our archives, our work, our actions, and our interactions are increasingly digitalized and stored in databases or made accessible via the Internet. This data, generally characterized by high volume, variety, and velocity (i.e., accumulation rate), has come to be called "Big Data". As of yet, Big Data has seldom been utilized in management research. Therefore, this dissertation explores the opportunities that Big Data brings for management scholars and describes three distinct projects that show how Big Data can be utilized in management research.

The first project demonstrates how science mapping can be used to provide a systematic review of an academic field. We focused specifically on the field of career studies – a field characterized by diversity and fragmentation. Using the key topics from the titles and abstracts of peer-reviewed publications, two science maps were created. The first map depicts the "local" (i.e., management) literature on careers (3,141 articles). The second the "global" (i.e., across the social sciences) literature (16,146 articles). We use the maps to describe the field and compare and contrast the local and global literature to identify avenues that career scholars in the field of management have yet to explore and reveal potentially understudied topic areas for career scholars in general.

The second project describes an innovative and powerful approach to making academic literature machine-readable. The project is geared specifically to the management literature, an area known for its high degree of concept proliferation. We created a dual-purpose software platform called "ReNotate". For individuals this platform consists of a software application that helps in reading and reviewing academic articles. It enables highlighting and annotating (i.e., "tagging") and these tags can subsequently be queried relatively easily. For the collective the platform consists of a large database of validated and aggregated data from the tags of individuals. Dependent upon the tag structure used, the platform could allow publications to be found based on study characteristics, concepts and their definitions to be traversed to their respective publications, and the relationships between concepts to be queried.

The final project employs Big Data to find out what determines whether an individual gets invited to a job interview. We use data from the Applicant Tracking Systems (ATSs) of 48 companies (N = 441,769 applicants) and apply synthetic validity and relative weight analysis to predict whether the individual is invited and to identify the relative importance of 18 different factors relating to the applicant (i.e., demographic factors, work and educational experience), vacancy (i.e., relevance of applicant education, skills, and prior experience), company (i.e., external or internal applicant), applicant pool (i.e., number of applications received), and labor market (i.e., occupation vacancy rate). The accuracy of the prediction, established on a holdout sample, was moderately high with 69.45% of the applicants invited and 68.78% not-invited correctly classified. For vacancies that did not seem to require a cover letter 82.51% of those invited and 80.27% of those not invited were correctly classified. Age and experience emerged as the most relevant in the selection of candidates for the second round of the application process. Results further indicate that hiring professionals' are relatively inconsistent in how they adjust the weighting of selection criteria with respect to demographics and biodata to the occupation and industry.

Samenvatting (Dutch Summary)

Computers zijn onze huizen binnengedrongen, hebben zich op onze bureaus genesteld, zijn in onze zakken beland en lijken inmiddels alom aanwezig. Op een nog hoger tempo zijn onze digitale apparaten met elkaar en met het internet verbonden geraakt. Bijgevolg worden onze archieven, ons werk, onze interacties en zelfs onze bewegingen in toenemende mate gedigitaliseerd. Deze data, veelal gekenmerkt door zijn grote omvang, verscheidenheid en accumulatiesnelheid, noemen we "Big Data". Big data wordt nog weinig benut in managementonderzoek. Daarom verkent dit proefschrift de verschillende mogelijkheden die Big Data met zich meebrengt en beschrijft het drie uiteenlopende projecten die laten zien hoe Big Data kan worden gebruikt in managementonderzoek.

Het eerste project toont aan hoe science mapping gebruikt kan worden om een systematisch verkregen overzicht te bieden van een academisch veld. We hebben ons specifiek gericht op onderzoek naar carrières – een onderzoeksgebied gekenmerkt door diversiteit en fragmentatie. Met behulp van de kernwoorden uit de titels en samenvattingen van wetenschappelijke artikelen hebben wij twee wetenschapsplattegronden gecreëerd. Een geeft de "lokale" (d.w.z., management) literatuur over carrières weer (3,141 artikelen), de ander de "globale" (d.w.z., sociale wetenschappen) literatuur (16,146 artikelen). We gebruiken de plattegronden om het veld te beschrijven en vergelijken het lokale met het globale veld om nieuwe onderzoeksmogelijkheden te identificeren en onderbelichtte onderzoeksgebieden te onthullen.

Het tweede project beschrijft een innovatieve methode waarmee de belangrijkste informatie uit wetenschappelijke publicaties aanwendbaar kan worden gemaakt voor computers. Dit is vooral van belang voor de managementliteratuur, aangezien deze bekend staat om een wildgroei van academische concepten. Aan de basis van onze methode staat een tweeledig softwareplatform genaamd "ReNotate". Voor individuen bestaat dit platform uit een softwareprogramma waarmee academische artikelen gelezen kunnen worden, de tekst kan worden gemarkeerd en geannoteerd (d.w.z., "taggen") en vervolgens makkelijk kan worden doorzocht. Voor het collectief bestaat het platform uit een grote database van gevalideerde en geaggregeerde tags van individuen. Afhankelijk van de gebruikte tag structuur maakt het "ReNotate" platform het mogelijk om publicaties te vinden op basis van de eigenschappen van de studie, concepten en definities terug te leiden naar de publicaties waarin ze zijn genoemd of relaties tussen concepten op te vragen.

In het laatste project wordt Big Data ingezet om inzicht te krijgen in de factoren die bepalen of een sollicitant zal worden uitgenodigd voor een sollicitatiegesprek. We hebben gebruik gemaakt van data uit Applicant Tracking Systems (ATSs) van 48 verschillende bedrijven (N = 441,769 sollicitanten) en passen gespecialiseerde analysetechnieken ("synthetic validity" en "relative weight analysis") toe om te voorspellen welke sollicitanten worden uitgenodigd en wat het relatieve belang is van verschillende voorspellende factoren. De nauwkeurigheid was middelmatig, met 69.45% van de kandidaten die daadwerkelijk waren uitgenodigd en 68.78% van de kandidaten die in werkelijkheid niet waren uitgenodigd correct geclassificeerd. Voor vacatures waar geen sollicitatiebrief voor nodig leek te zijn werd 82.51% van de uitgenodigde en 80.27% van de niet uitgenodigde kandidaten correct geclassificeerd. Leeftijd en ervaring kwamen naar voren als de meest bepalende factoren. Resultaten lijken aan te tonen dat recruitment- en selectieprofessionals relatief inconsistent zijn in hoe zij hun selectiecriteria met betrekking tot demografische kenmerken en biodata aanpassen aan het beroep en de industrie.

About the Author

Colin Lee (1985) began his studies in Interdisciplinary Social Science at Utrecht University. After obtaining a Master's degree in Social Policy he moved to the Rotterdam School of Management for a second graduate degree as Master of Philosophy in Business Research and went on to pursue a PhD.

He conducts research on topics at the intersection of organizational behavior, human resources, and careers studies. Reflecting his interdisciplinary background, he makes use of recent developments in data extraction, normalization (i.e., standardization), storage, and processing in order to provide insight into how people match to work and offer tools that help improve the dissemination of academic research. His work has been published in the Journal of Vocational Behavior and Academy of Management Best Paper Proceedings. In 2014 one of his papers was nominated for the best student paper award of the Career Division and in 2015 he received the best student paper award for the Human Resource Division at the Academy of Management.

Before starting his PhD, Colin enjoyed a successful sporting career in Field Hockey, representing the Netherlands in the U16 youth squad and playing and coaching at a semi-professional level thereafter.

As of March 15th, 2016, Colin will be working as a Postdoctoral scholar at the Haskayne School of Business, at Calgary University in Canada.

Author Portfolio

Publications & Proceedings

Markoulli, M., Lee, C.I.S.G., & Felps, W. (2015). Mapping Human Resource Management: A Structural Analysis of the Field. *Best Paper Proceedings of the Academy of Management Annual Meeting*, Vancouver, Canada.

Best Student Convention Paper Award, Human Resources Division

Lee, C.I.S.G., Felps, W., & Baruch, Y. (2014). Toward a Taxonomy of Career Studies Through Bibliometric Visualization. *Journal of Vocational Behavior, 85*(3), 339-351.

Lee, C.I.S.G., Felps, W., & Baruch, Y. (2014). Mapping Career Studies: A Bibliometric Analysis. *Best Paper Proceedings of the Academy of Management Annual Meeting*, Philadelphia, PA, U.S.A.

Arnon Reichers Best Student Paper Award Finalist

Lee, C.I.S.G. (2012). Research Note: Trending Topics in Career Research of 2011. *The Career Forum, 31*(2), 8-12. [Special Entry in the Newsletter of the AOM Career Division]

Selected Work in Progress

Markoulli, M., Lee, C.I.S.G., & Felps, W. *Mapping Human Resource Management: A Structural Analysis of the Field*. Status: R&R Human Resource Management Review.

Byington, E., Felps, W., & Lee, C.I.S.G. *A Big Picture View of the Organizational Behavior Field*. Status: Under Review at Journal of Applied Psychology.

Lee, C.I.S.G. & Felps, W. *Selecting Applicants from CV Data: Where do we Stand?* Status: In preparation for Personnel Psychology.

Lee, C.I.S.G., Felps, W., & Frasincar, F. *Collaborative Tagging of Academic Articles: Making Management Publications Machine-Readable*. Status: In preparation for Organizational Research Methods.

Lee, C.I.S.G., Felps, W, & Biemann, T. *The Optimal Rate of Job Hopping*. Status: Analyzing data, targeting Human Relations.

Finch, R., Byington, E., Felps, W., & Lee, C.I.S.G. *Groups and Teams: The State of the Field*. Status: Analyzing data, targeting Journal of Management.

Focal Conference Presentations

Markoulli, M., Lee, C.I.S.G., & Felps, W. (2014, September). *Mapping HRM Research: A Bibliometric Analysis of the HRM Field*. 4th Global TechMining Conference, Leiden, Netherlands.

Lee, C.I.S.G. & Felps, W. (2013, July). *Mapping Careers: Research in Management and Beyond*. 29th EGOS Colloquium, Montreal, Canada.

Lee, C.I.S.G. & Felps, W. (2013, March). *Sailing Across Boundaries: Conceptualizing and Measuring Intellectual Travel*. New Avenues for Careers Research, London, United Kingdom.

Touburg, G. & Lee, C.I.S.G. (2013, March). *Tacit Knowledge Transfer Through Global Narratives*. 5th International Conference on Rhetoric and Narratives in Management Research, Barcelona, Spain.

Teaching Experience

2013	Lecturer (32.5 hrs) and Coordinator Undergraduate course “Cross-Cultural Management”
2013	Lecturer (4 hrs) and Coordinator Try out for High School students “IBA Experience”
2012	Workshop Lecturer (18 hrs) Undergraduate course “Cross-Cultural Management”
2011 - 2013	Workshop Lecturer (2 x 12 hrs; 1 x 16 hrs) Undergraduate course “Organizational Behavior”
2011	Guest Lecturer (1.5 hrs) Graduate course “Philosophy of Science”
2011	Guest Lecturer (3 hrs) Graduate course “Strategic and International HRM”
Since 2011	(Co-)supervision of M.Sc. theses (14x)

Grants and Awards

2016-2017	Co-writer of Australian Research Council (ARC) Discovery Project. Felps, W. & Lee, T. (When) should organizations employ over/underqualified individuals? (AU $173,596)
2015	Best Student Convention Paper Award, Human Resources Division, Academy of Management Meeting
2015	Submission featured in the Best Paper Proceedings, Academy of Management Meeting (top ~10% of submissions)
2014	Submission featured in the Best Paper Proceedings, Academy of Management Meeting (top ~10% of submissions)
2014	Arnon Reichers Best Student Paper Award Finalist, Academy of Management Meeting
2014	Travel grant from the "Erasmus Trustfonds" for a research visit to the Australian School of Business, University of New South Wales
2010-2011	Research Assistantship (personally obtained the funding)

Selected Skills & Languages

Computer Skills:	Python, Django, PyQt, SPSS, R, SQL, VBA, Javascript, NVivo, AMOS
Languages:	Dutch (native), English (near native)

Memberships & Service

Academy of Management (AOM)

European Group for Organizational Studies (EGOS)

International Association of Cross-Cultural Competence and Management (IACCM)

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ERIM PH.D. SERIES RESEARCH IN MANAGEMENT

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://repub.eur.nl/pub. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics at the Erasmus University Rotterdam (EUR).

DISSERTATIONS LAST FIVE YEARS

Abbink, E.J., *Crew Management in Passenger Rail Transport*, Promotor(s): Prof.dr. L.G. Kroon & Prof.dr. A.P.M. Wagelmans, EPS-2014-325-LIS, http://repub.eur.nl/pub/76927

Acar, O.A., *Crowdsourcing for Innovation: Unpacking Motivational, Knowledge and Relational Mechanisms of Innovative Behavior in Crowdsourcing Platforms*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2014-321-LIS, http://repub.eur.nl/pub/76076

Akin Ates, M., *Purchasing and Supply Management at the Purchase Category Level: strategy, structure and performance*, Promotor(s): Prof.dr. J.Y.F. Wynstra & Dr. E.M. van Raaij, EPS-2014-300-LIS, http://repub.eur.nl/pub/50283

Akpinar, E., *Consumer Information Sharing*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2013-297-MKT, http://repub.eur.nl/pub/50140

Alexander, L., *People, Politics, and Innovation: A Process Perspective*, Promotor(s): Prof.dr. H.G. Barkema & Prof.dr. D.L. van Knippenberg, EPS-2014-331-S&E, http://repub.eur.nl/pub/77209

Almeida e Santos Nogueira, R.J. de, *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*, Promotor(s): Prof.dr.ir. U. Kaymak & Prof.dr. J.M.C. Sousa, EPS-2014-310-LIS, http://repub.eur.nl/pub/51560

Bannouh, K., *Measuring and Forecasting Financial Market Volatility using High-frequency Data*, Promotor(s): Prof.dr. D.J.C. van Dijk, EPS-2013-273-F&A, http://repub.eur.nl/pub/38240

Ben-Menahem, S.M., *Strategic Timing and Proactiveness of Organizations*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2013-278-S&E, http://repub.eur.nl/pub/39128

Benning, T.M., *A Consumer Perspective on Flexibility in Health Care: Priority Access Pricing and Customized Care*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2011-241-MKT, http://repub.eur.nl/pub/23670

Berg, W.E. van den, *Understanding Salesforce Behavior using Genetic Association Studies*, Promotor(s): Prof.dr. W.J.M.I. Verbeke, EPS-2014-311-MKT, http://repub.eur.nl/pub/51440

Betancourt, N.E., *Typical Atypicality: Formal and Informal Institutional Conformity, Deviance, and Dynamics*, Promotor(s): Prof.dr. B. Krug, EPS-2012-262-ORG, http://repub.eur.nl/pub/32345

Bliek, R. de, *Empirical Studies on the Economic Impact of Trust*, Promotor(s): Prof.dr. J. Veenman & Prof.dr. Ph.H.B.F. Franses, EPS-2015-324-ORG, http://repub.eur.nl/pub/78159

Blitz, D.C., *Benchmarking Benchmarks*, Promotor(s): Prof.dr. A.G.Z. Kemna & Prof.dr. W.F.C. Verschoor, EPS-2011-225-F&A, http://repub.eur.nl/pub/22624

Boons, M., *Working Together Alone in the Online Crowd: The Effects of Social Motivationsand Individual Knowledge Backgrounds on the Participation and Performance of Members of Online Crowdsourcing Platforms*, Promotor(s): Prof.dr. H.G. Barkema & Dr. D.A. Stam, EPS-2014-306-S&E, http://repub.eur.nl/pub/50711

Brazys, J., *Aggregated Marcoeconomic News and Price Discovery*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2015-351-F&A, http://repub.eur.nl/pub/78243

Burger, M.J., *Structure and Cooptition in Urban Networks*, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.R. Commandeur, EPS-2011-243-ORG, http://repub.eur.nl/pub/26178

Byington, E., *Exploring Coworker Relationships: Antecedents and Dimensions of Interpersonal Fit,Coworker Satisfaction, and Relational Models*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-292-ORG, http://repub.eur.nl/pub/41508

Camacho, N.M., *Health and Marketing: Essays on Physician and Patient Decision-Making*, Promotor(s): Prof.dr. S. Stremersch, EPS-2011-237-MKT, http://repub.eur.nl/pub/23604

Cancurtaran, P., *Essays on Accelerated Product Development*, Promotor(s): Prof.dr. F. Langerak & Prof.dr.ir. G.H. van Bruggen, EPS-2014-317-MKT, http://repub.eur.nl/pub/76074

Caron, E.A.M., *Explanation of Exceptional Values in Multi-dimensional Business Databases*, Promotor(s): Prof.dr.ir. H.A.M. Daniels & Prof.dr. G.W.J. Hendrikse, EPS-2013-296-LIS, http://repub.eur.nl/pub/50005

Carvalho, L. de, *Knowledge Locations in Cities: Emergence and Development Dynamics*, Promotor(s): Prof.dr. L. Berg, EPS-2013-274-S&E, http://repub.eur.nl/pub/38449

Cox, R.H.G.M., *To Own, To Finance, and To Insure - Residential Real Estate Revealed*, Promotor(s): Prof.dr. D. Brounen, EPS-2013-290-F&A, http://repub.eur.nl/pub/40964

Deichmann, D., *Idea Management: Perspectives from Leadership, Learning, and Network Theory*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2012-255-ORG, http://repub.eur.nl/pub/31174

Deng, W., *Social Capital and Diversification of Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2015-341-ORG, http://repub.eur.nl/pub/77449

Desmet, P.T.M., *In Money we Trust? Trust Repair and the Psychology of Financial Compensations*, Promotor(s): Prof.dr. D. de Cremer, EPS-2011-232-ORG, http://repub.eur.nl/pub/23268

Dollevoet, T.A.B., *Delay Management and Dispatching in Railways*, Promotor(s): Prof.dr. A.P.M. Wagelmans, EPS-2013-272-LIS, http://repub.eur.nl/pub/38241

Doorn, S. van, *Managing Entrepreneurial Orientation*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch, & Prof.dr. H.W. Volberda, EPS-2012-258-STR, http://repub.eur.nl/pub/32166

Douwens-Zonneveld, M.G., *Animal Spirits and Extreme Confidence: No Guts, No Glory?* Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2012-257-F&A, http://repub.eur.nl/pub/31914

Duca, E., *The Impact of Investor Demand on Security Offerings*, Promotor(s): Prof.dr. A. de Jong, EPS-2011-240-F&A, http://repub.eur.nl/pub/26041

Duyvesteyn, J.G. *Empirical Studies on Sovereign Fixed Income Markets,* Promotor(s): Prof.dr P.Verwijmeren & Prof.dr. M.P.E. Martens, EPS-2015-361-F&A, hdl.handle.net/1765/79033

Duursema, H., *Strategic Leadership: Moving Beyond the Leader-Follower Dyad*, Promotor(s): Prof.dr. R.J.M. van Tulder, EPS-2013-279-ORG, http://repub.eur.nl/pub/39129

Eck, N.J. van, *Methodological Advances in Bibliometric Mapping of Science*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2011-247-LIS, http://repub.eur.nl/pub/26509

Elemes, A, *Studies on Determinants and Consequences of Financial Reporting Quality,* Promotor: Prof.dr. E.Peek, EPS-2015-354-F&A, http://hdl.handle.net/1765/79037

Ellen, S. ter, *Measurement, Dynamics, and Implications of Heterogeneous Beliefs in Financial Markets*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2015-343-F&A, http://repub.eur.nl/pub/78191

Eskenazi, P.I., *The Accountable Animal*, Promotor(s): Prof.dr. F.G.H. Hartmann, EPS-2015-355-F&A, http://repub.eur.nl/pub/78300

Essen, M. van, *An Institution-Based View of Ownership*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. G.M.H. Mertens, EPS-2011-226-ORG, http://repub.eur.nl/pub/22643

Evangelidis, I., *Preference Construction under Prominence*, Promotor(s): Prof.dr. S.M.J. van Osselaer, EPS-2015-340-MKT, http://repub.eur.nl/pub/78202

Faber, N., *Structuring Warehouse Management*, Promotor(s): Prof.dr. MB.M. de Koster, Prof.dr. Ale Smidts, EPS-2015-336-LIS, http://repub.eur.nl/pub/78603

Fernald, K., *The Waves of Biotechnological Innovation in Medicine: Interfirm Cooperation Effects and a Venture Capital Perspective,* Promotor(s): Prof.dr. E.Claassen, Prof.dr. H.P.G.Pennings & Prof.dr. H.R. Commandeur, EPS-2015-371-S&E, http://hdl.handle.net/1765/79120

Fourne, S.P., *Managing Organizational Tensions: A Multi-Level Perspective on Exploration, Exploitation and Ambidexterity*, Promotor(s): Prof.dr. J.J.P. Jansen & Prof.dr. S.J. Magala, EPS-2014-318-S&E, http://repub.eur.nl/pub/76075

Gharehgozli, A.H., *Developing New Methods for Efficient Container Stacking Operations*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2012-269-LIS, http://repub.eur.nl/pub/37779

Gils, S. van, *Morality in Interactions: On the Display of Moral Behavior by Leaders and Employees*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2012-270-ORG, http://repub.eur.nl/pub/38027

Ginkel-Bieshaar, M.N.G. van, *The Impact of Abstract versus Concrete Product Communications on Consumer Decision-making Processes*, Promotor(s): Prof.dr.ir. B.G.C.
Dellaert, EPS-2012-256-MKT, http://repub.eur.nl/pub/31913

Gkougkousi, X., *Empirical Studies in Financial Accounting*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. E. Peek, EPS-2012-264-F&A, http://repub.eur.nl/pub/37170

Glorie, K.M., *Clearing Barter Exchange Markets: Kidney Exchange and Beyond*, Promotor(s): Prof.dr. A.P.M. Wagelmans & Prof.dr. J.J. van de Klundert, EPS-2014-329-LIS, http://repub.eur.nl/pub/77183

Hekimoglu, M., *Spare Parts Management of Aging Capital Products,* Promotor: Prof.dr.ir. R. Dekker, EPS-2015-368-LIS, http://hdl.handle.net/1765/79092

Heij, C.V., *Innovating beyond Technology. Studies on how management innovation, co-creation and business model innovation contribute to firm's (innovation) performance*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-370-STR, http://repub.eur.nl/pub/78651

Heyde Fernandes, D. von der, *The Functions and Dysfunctions of Reminders*, Promotor(s): Prof.dr. S.M.J. van Osselaer, EPS-2013-295-MKT, http://repub.eur.nl/pub/41514

Heyden, M.L.M., *Essays on Upper Echelons & Strategic Renewal: A Multilevel Contingency Approach*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-259-STR, http://repub.eur.nl/pub/32167

Hoever, I.J., *Diversity and Creativity*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2012-267-ORG, http://repub.eur.nl/pub/37392

Hogenboom, A.C., *Sentiment Analysis of Text Guided by Semantics and Structure,* Promotor(s):Prof.dr.ir. U.Kaymak & Prof.dr. F.M.G. de Jong, EPS-2015-369-LIS, http://hdl.handle.net/1765/79034

Hogenboom, F.P., *Automated Detection of Financial Events in News Text*, Promotor(s): Prof.dr.ir. U. Kaymak & Prof.dr. F.M.G. de Jong, EPS-2014-326-LIS, http://repub.eur.nl/pub/77237

Hollen, R.M.A., *Exploratory Studies into Strategies to Enhance Innovation-Driven International Competitiveness in a Port Context: Toward Ambidextrous Ports,* Promotor(s) Prof.dr.ing. F.A.J. Van Den Bosch & Prof.dr. H.W.Volberda, EPS-2015-372-S&E,
hdl.handle.net/1765/78881

Hoogendoorn, B., *Social Entrepreneurship in the Modern Economy: Warm Glow, Cold Feet*, Promotor(s): Prof.dr. H.P.G. Pennings & Prof.dr. A.R. Thurik, EPS-2011-246-STR, http://repub.eur.nl/pub/26447

Hoogervorst, N., *On The Psychology of Displaying Ethical Leadership: A Behavioral Ethics Approach*, Promotor(s): Prof.dr. D. de Cremer & Dr. M. van Dijke, EPS-2011-244-ORG, http://repub.eur.nl/pub/26228

Hout, D.H. van, *Measuring Meaningful Differences: Sensory Testing Based Decision Making in an Industrial Context; Applications of Signal Detection Theory and Thurstonian Modelling*, Promotor(s): Prof.dr. P.J.F. Groenen & Prof.dr. G.B. Dijksterhuis, EPS-2014-304-MKT, http://repub.eur.nl/pub/50387

Houwelingen, G.G. van, *Something To Rely On*, Promotor(s): Prof.dr. D. de Cremer &

Prof.dr. M.H. van Dijke, EPS-2014-335-ORG, http://repub.eur.nl/pub/77320

Hurk, E. van der, *Passengers, Information, and Disruptions*, Promotor(s): Prof.dr. L.G. Kroon & Prof.mr.dr. P.H.M. Vervest, EPS-2015-345-LIS, http://repub.eur.nl/pub/78275

Hytonen, K.A., *Context Effects in Valuation, Judgment and Choice: A Neuroscientific Approach*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2011-252-MKT, http://repub.eur.nl/pub/30668

Iseger, P. den, *Fourier and Laplace Transform Inversion with Applications in Finance*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2014-322-LIS, http://repub.eur.nl/pub/76954

Jaarsveld, W.L. van, *Maintenance Centered Service Parts Inventory Control*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2013-288-LIS, http://repub.eur.nl/pub/39933

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2011-222-LIS, http://repub.eur.nl/pub/22156

Kappe, E.R., *The Effectiveness of Pharmaceutical Marketing*, Promotor(s): Prof.dr. S. Stremersch, EPS-2011-239-MKT, http://repub.eur.nl/pub/23610

Karreman, B., *Financial Services and Emerging Markets*, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.P.G. Pennings, EPS-2011-223-ORG, http://repub.eur.nl/pub/22280

Khanagha, S., *Dynamic Capabilities for Managing Emerging Technologies*, Promotor(s): Prof.dr. H.W. Volberda, EPS-2014-339-S&E, http://repub.eur.nl/pub/77319

Kil, J., *Acquisitions Through a Behavioral and Real Options Lens*, Promotor(s): Prof.dr. H.T.J. Smit, EPS-2013-298-F&A, http://repub.eur.nl/pub/50142

Klooster, E. van 't, *Travel to Learn: the Influence of Cultural Distance on Competence Development in Educational Travel*, Promotor(s): Prof.dr. F.M. Go & Prof.dr. P.J. van Baalen, EPS-2014-312-MKT, http://repub.eur.nl/pub/51462

Koendjbiharie, S.R., *The Information-Based View on Business Network Performance: Revealing the Performance of Interorganizational Networks*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.mr.dr. P.H.M. Vervest, EPS-2014-315-LIS, http://repub.eur.nl/pub/51751

Koning, M., *The Financial Reporting Environment: The Role of the Media, Regulators and Auditors*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. P.G.J. Roosenboom, EPS-2014-330-F&A, http://repub.eur.nl/pub/77154

Konter, D.J., *Crossing Borders with HRM: An Inquiry of the Influence of Contextual*

Differences in the Adoption and Effectiveness of HRM, Promotor(s): Prof.dr. J. Paauwe & Dr. L.H. Hoeksema, EPS-2014-305-ORG, http://repub.eur.nl/pub/50388

Korkmaz, E., *Bridging Models and Business: Understanding Heterogeneity in Hidden Drivers of Customer Purchase Behavior*, Promotor(s): Prof.dr. S.L. van de Velde & Prof.dr. D. Fok, EPS-2014-316-LIS, http://repub.eur.nl/pub/76008

Kroezen, J.J., *The Renewal of Mature Industries: An Examination of the Revival of the Dutch Beer Brewing Industry*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens, EPS-2014-333-S&E, http://repub.eur.nl/pub/77042

Kysucky, V., *Access to Finance in a Cros-Country Context*, Promotor(s): Prof.dr. L. Norden, EPS-2015-350-F&A, http://repub.eur.nl/pub/78225

Lam, K.Y., *Reliability and Rankings*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-230-MKT, http://repub.eur.nl/pub/22977

Lander, M.W., *Profits or Professionalism? On Designing Professional Service Firms*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2012-253-ORG, http://repub.eur.nl/pub/30682

Langhe, B. de, *Contingencies: Learning Numerical and Emotional Associations in an Uncertain World*, Promotor(s): Prof.dr.ir. B. Wierenga & Prof.dr. S.M.J. van Osselaer, EPS-2011-236-MKT, http://repub.eur.nl/pub/23504

Legault-Tremblay, P.O., Corporate Governance During Market Transition: Heterogeneous responses to Institution Tensions in China, Promotor: Prof.dr. B. Krug, EPS-2015-362-ORG, http://repub.eur.nl/pub/78649

Lenoir, A.S. *Are You Talking to Me? Addressing Consumers in a Globalised World,* Promotor(s) Prof.dr. S. Puntoni & Prof.dr. S.M.J. van Osselaer, EPS-2015-363-MKT, , http://hdl.handle.net/1765/79036

Leunissen, J.M., *All Apologies: On the Willingness of Perpetrators to Apologize*, Promotor(s): Prof.dr. D. de Cremer & Dr. M. van Dijke, EPS-2014-301-ORG, http://repub.eur.nl/pub/50318

Li, D., *Supply Chain Contracting for After-sales Service and Product Support*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2015-347-LIS, http://repub.eur.nl/pub/78526

Li, Z., *Irrationality: What, Why and How*, Promotor(s): Prof.dr. H. Bleichrodt, Prof.dr. P.P. Wakker, & Prof.dr. K.I.M. Rohde, EPS-2014-338-MKT, http://repub.eur.nl/pub/77205

Liang, Q.X., *Governance, CEO Identity, and Quality Provision of Farmer Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2013-281-ORG, http://repub.eur.nl/pub/39253

Liket, K., *Why 'Doing Good' is not Good Enough: Essays on Social Impact Measurement*, Promotor(s): Prof.dr. H.R. Commandeur & Dr. K.E.H. Maas, EPS-2014-307-STR, http://repub.eur.nl/pub/51130

Loos, M.J.H.M. van der, *Molecular Genetics and Hormones: New Frontiers in Entrepreneurship Research*, Promotor(s): Prof.dr. A.R. Thurik, Prof.dr. P.J.F. Groenen, & Prof.dr. A. Hofman, EPS-2013-287-S&E, http://repub.eur.nl/pub/40081

Lovric, M., *Behavioral Finance and Agent-Based Artificial Markets*, Promotor(s): Prof.dr. J. Spronk & Prof.dr.ir. U. Kaymak, EPS-2011-229-F&A, http://repub.eur.nl/pub/22814

Lu, Y., *Data-Driven Decision Making in Auction Markets*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.dr. W. Ketter, EPS-2014-314-LIS, http://repub.eur.nl/pub/51543

Manders, B., *Implementation and Impact of ISO 9001*, Promotor(s): Prof.dr. K. Blind, EPS-2014-337-LIS, http://repub.eur.nl/pub/77412

Markwat, T.D., *Extreme Dependence in Asset Markets Around the Globe*, Promotor(s): Prof.dr. D.J.C. van Dijk, EPS-2011-227-F&A, http://repub.eur.nl/pub/22744

Mees, H., *Changing Fortunes: How China's Boom Caused the Financial Crisis*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2012-266-MKT, http://repub.eur.nl/pub/34930

Mell, J.N., *Connecting Minds: On The Role of Metaknowledge in Knowledge Coordination*, Promotor: Prof.dr.D.L. van Knippenberg, EPS-2015-359-ORG, http://hdl.handle.net/1765/78951

Meuer, J., *Configurations of Inter-firm Relations in Management Innovation: A Study in China's Biopharmaceutical Industry*, Promotor(s): Prof.dr. B. Krug, EPS-2011-228-ORG, http://repub.eur.nl/pub/22745

Micheli, M.R., *Business Model Innovation: A Journey across Managers' Attention and Inter-Organizational Networks*, Promotor(s): Prof.dr. J.J.P. Jansen, EPS-2015-344-S&E, http://repub.eur.nl/pub/78241

Mihalache, O.R., *Stimulating Firm Innovativeness: Probing the Interrelations between Managerial and Organizational Determinants*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch, & Prof.dr. H.W. Volberda, EPS-2012-260-S&E, http://repub.eur.nl/pub/32343

Milea, V., *News Analytics for Financial Decision Support*, Promotor(s): Prof.dr.ir. U.

Kaymak, EPS-2013-275-LIS, http://repub.eur.nl/pub/38673

Naumovska, I., *Socially Situated Financial Markets: A Neo-Behavioral Perspective on Firms, Investors and Practices*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. A. de Jong, EPS-2014-319-S&E, http://repub.eur.nl/pub/76084

Nielsen, L.K., *Rolling Stock Rescheduling in Passenger Railways: Applications in short term planning and in disruption management*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2011-224-LIS, http://repub.eur.nl/pub/22444

Nuijten, A.L.P., *Deaf Effect for Risk Warnings: A Causal Examination applied to Information Systems Projects*, Promotor(s): Prof.dr. G.J. van der Pijl, Prof.dr. H.R. Commandeur & Prof.dr. M. Keil, EPS-2012-263-S&E, http://repub.eur.nl/pub/34928

Osadchiy, S.E., *The Dynamics of Formal Organization: Essays on bureaucracy and formal rules*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens, EPS-2011-231-ORG, http://repub.eur.nl/pub/23250

Ozdemir, M.N., *Project-level Governance, Monetary Incentives, and Performance in Strategic R&D Alliances*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2011-235-LIS, http://repub.eur.nl/pub/23550

Peers, Y., *Econometric Advances in Diffusion Models*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-251-MKT, http://repub.eur.nl/pub/30586

Peters, M., *Machine Learning Algorithms for Smart Electricity Markets*, Promotor(s): Prof.dr. W. Ketter, EPS-2014-332-LIS, http://repub.eur.nl/pub/77413

Porck, J., *No Team is an Island: An Integrative View of Strategic Consensus between Groups*, Promotor(s): Prof.dr. P.J.F. Groenen & Prof.dr. D.L. van Knippenberg, EPS-2013-299-ORG, http://repub.eur.nl/pub/50141

Porras Prado, M., *The Long and Short Side of Real Estate, Real Estate Stocks, and Equity*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2012-254-F&A, http://repub.eur.nl/pub/30848

Poruthiyil, P.V., *Steering Through: How organizations negotiate permanent uncertainty and unresolvable choices*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. S.J. Magala, EPS-2011-245-ORG, http://repub.eur.nl/pub/26392

Pourakbar, M., *End-of-Life Inventory Decisions of Service Parts*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2011-249-LIS, http://repub.eur.nl/pub/30584

Pronker, E.S., *Innovation Paradox in Vaccine Target Selection*, Promotor(s): Prof.dr. H.J.H.M. Claassen & Prof.dr. H.R. Commandeur, EPS-2013-282-S&E, http://repub.eur.nl/pub/39654

Protzner, S. *Mind the gap between demand and supply: A behavioral perspective on demand forecasting,* Promotor(s): Prof.dr. S.L. van de Velde & Dr. L. Rook, EPS-2015-364-LIS, *http*://repub.eur.nl/pub/79355

Pruijssers, J.K., *An Organizational Perspective on Auditor Conduct*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2015-342-S&E, http://repub.eur.nl/pub/78192

Retel Helmrich, M.J., *Green Lot-Sizing*, Promotor(s): Prof.dr. A.P.M. Wagelmans, EPS-2013-291-LIS, http://repub.eur.nl/pub/41330

Rietveld, N., *Essays on the Intersection of Economics and Biology*, Promotor(s): Prof.dr. A.R. Thurik, Prof.dr. Ph.D. Koellinger, Prof.dr. P.J.F. Groenen, & Prof.dr. A. Hofman, EPS-2014-320-S&E, http://repub.eur.nl/pub/76907

Rijsenbilt, J.A., *CEO Narcissism: Measurement and Impact*, Promotor(s): Prof.dr. A.G.Z. Kemna & Prof.dr. H.R. Commandeur, EPS-2011-238-STR, http://repub.eur.nl/pub/23554

Rösch, D. *Market Efficiency and Liquidity,* Promotor: Prof.dr. M.A. van Dijk, EPS-2015-353-F&A, http://hdl.handle.net/1765/79121

Roza-van Vuren, M.W., *The Relationship between Offshoring Strategies and Firm Performance: Impact of innovation, absorptive capacity and firm size*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2011-214-STR, http://repub.eur.nl/pub/22155

Rubbaniy, G., *Investment Behaviour of Institutional Investors*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2013-284-F&A, http://repub.eur.nl/pub/40068

Schoonees, P. *Methods for Modelling Response Styles*, Promotor: Prof.dr P.J.F. Groenen, EPS-2015-348-MKT, http://repub.eur.nl/pub/79327

Shahzad, K., *Credit Rating Agencies, Financial Regulations and the Capital Markets*, Promotor(s): Prof.dr. G.M.H. Mertens, EPS-2013-283-F&A, http://repub.eur.nl/pub/39655

Sousa, M.J.C. de, *Servant Leadership to the Test: New Perspectives and Insights*, Promotor(s): Prof.dr. D.L. van Knippenberg & Dr. D. van Dierendonck, EPS-2014-313-ORG, http://repub.eur.nl/pub/51537

Spliet, R., *Vehicle Routing with Uncertain Demand*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2013-293-LIS, http://repub.eur.nl/pub/41513

Staadt, J.L., *Leading Public Housing Organisation in a Problematic Situation: A Critical Soft Systems Methodology Approach*, Promotor(s): Prof.dr. S.J. Magala, EPS-2014-308-ORG, http://repub.eur.nl/pub/50712

Stallen, M., *Social Context Effects on Decision-Making: A Neurobiological Approach*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2013-285-MKT, http://repub.eur.nl/pub/39931

Tarakci, M., *Behavioral Strategy: Strategic Consensus, Power and Networks*, Promotor(s): Prof.dr. D.L. van Knippenberg & Prof.dr. P.J.F. Groenen, EPS-2013-280-ORG, http://repub.eur.nl/pub/39130

Teixeira de Vasconcelos, M., *Agency Costs, Firm Value, and Corporate Investment*, Promotor(s): Prof.dr. P.G.J. Roosenboom, EPS-2012-265-F&A, http://repub.eur.nl/pub/37265

Troster, C., *Nationality Heterogeneity and Interpersonal Relationships at Work*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2011-233-ORG, http://repub.eur.nl/pub/23298

Tsekouras, D., *No Pain No Gain: The Beneficial Role of Consumer Effort in Decision-Making*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2012-268-MKT, http://repub.eur.nl/pub/37542

Tuijl, E. van, *Upgrading across Organisational and Geographical Configurations*, Promotor(s): Prof.dr. L. van den Berg, EPS-2015-349-S&E, http://repub.eur.nl/pub/78224

Tuncdogan, A., *Decision Making and Behavioral Strategy: The Role of Regulatory Focus in Corporate Innovation Processes*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch, Prof.dr. H.W. Volberda, & Prof.dr. T.J.M. Mom, EPS-2014-334-S&E, http://repub.eur.nl/pub/76978

Uijl, S. den, *The Emergence of De-facto Standards*, Promotor(s): Prof.dr. K. Blind, EPS-2014-328-LIS, http://repub.eur.nl/pub/77382

Vagias, D., *Liquidity, Investors and International Capital Markets*, Promotor(s): Prof.dr. M.A. van Dijk, EPS-2013-294-F&A, http://repub.eur.nl/pub/41511

Veelenturf, L.P., *Disruption Management in Passenger Railways: Models for Timetable, Rolling Stock and Crew Rescheduling*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2014-327-LIS, http://repub.eur.nl/pub/77155

Venus, M., *Demystifying Visionary Leadership: In search of the essence of effective vision communication*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-289-ORG, http://repub.eur.nl/pub/40079

Vermeer, W., *Propagation in Networks:The impact of information processing at the actor level on system-wide propagation dynamics,* Promotor: Prof.mr.dr. P.H.M.Vervest, EPS-2015-373-LIS, http://repub.eur.nl/pub/79325

Visser, V.A., *Leader Affect and Leadership Effectiveness:How leader affective displays influence follower outcomes*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-286-ORG, http://repub.eur.nl/pub/40076

Vlam, A.J., *Customer First? The Relationship between Advisors and Consumers of Financial Products*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-250-MKT, http://repub.eur.nl/pub/30585

Waltman, L., *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*, Promotor(s): Prof.dr.ir. R. Dekker & Prof.dr.ir. U. Kaymak, EPS-2011-248-LIS, http://repub.eur.nl/pub/26564

Wang, T., *Essays in Banking and Corporate Finance*, Promotor(s): Prof.dr. L. Norden & Prof.dr. P.G.J. Roosenboom, EPS-2015-352-F&A, http://repub.eur.nl/pub/78301

Wang, Y., *Information Content of Mutual Fund Portfolio Disclosure*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2011-242-F&A, http://repub.eur.nl/pub/26066

Wang, Y., *Corporate Reputation Management: Reaching Out to Financial Stakeholders*, Promotor(s): Prof.dr. C.B.M. van Riel, EPS-2013-271-ORG, http://repub.eur.nl/pub/38675

Weenen, T.C., *On the Origin and Development of the Medical Nutrition Industry*, Promotor(s): Prof.dr. H.R. Commandeur & Prof.dr. H.J.H.M. Claassen, EPS-2014-309-S&E, http://repub.eur.nl/pub/51134

Wolfswinkel, M., *Corporate Governance, Firm Risk and Shareholder Value*, Promotor(s): Prof.dr. A. de Jong, EPS-2013-277-F&A, http://repub.eur.nl/pub/39127

Yang, S., *Information Aggregation Efficiency of Prediction Markets*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck, EPS-2014-323-LIS, http://repub.eur.nl/pub/77184

Zaerpour, N., *Efficient Management of Compact Storage Systems*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2013-276-LIS, http://repub.eur.nl/pub/38766

Zhang, D., *Essays in Executive Compensation*, Promotor(s): Prof.dr. I. Dittmann, EPS-2012-261-F&A, http://repub.eur.nl/pub/32344

Zwan, P.W. van der, *The Entrepreneurial Process: An International Analysis of Entry and Exit*, Promotor(s): Prof.dr. A.R. Thurik & Prof.dr. P.J.F. Groenen, EPS-2011-234-ORG, http://repub.eur.nl/pub/23422

Erasmus Research Institute of Management - ERIM

BIG DATA IN MANAGEMENT RESEARCH

EXPLORING NEW AVENUES

Digital computers entered our homes, landed on our desktops, slipped into our pockets, and have seemingly become ubiquitous. At an ever faster pace, these devices have become highly interconnected and interoperable. Consequently, our archives, our work, our actions, and our interactions are increasingly digitalized and stored in databases or made accessible via the Internet. This data, generally characterized by high volume, variety, and velocity (i.e., accumulation rate), has come to be called "Big Data". As of yet, Big Data has seldom been utilized in management research. Not without cause, the discussion in the management literature has barely surpassed deliberation on privacy risks. Nevertheless, there are many ways in which Big Data can contribute to management science in a responsible fashion. This dissertation explores the opportunities that Big Data brings for management scholars and describes three distinct projects that show how Big Data can be utilized in management research.

The first project demonstrates how science mapping, when applied to digital repositories of academic journals, can be used to provide a systematic review of an academic field. The second project describes an innovative and powerful platform called "Re", which uses the highlights and annotations of individuals reading academic articles to make those articles machine-readable and thus highly searchable. The final project uses data from the Applicant Tracking Systems of 48 different companies (N = 441,769 applicants) to find out what determines whether an individual gets invited to a job interview.

ERIM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERIM PhD Series
Research in Management

Erasmus Research Institute of Management - ERIM
Rotterdam School of Management (RSM)
Erasmus School of Economics (ESE)
Erasmus University Rotterdam (EUR)
P.O. Box 1738, 3000 DR Rotterdam,
The Netherlands

Tel.	+31 10 408 11 82
Fax	+31 10 408 96 40
E-mail	info@erim.eur.nl
Internet	www.erim.eur.nl

Design & layout: B&T Ontwerp en advies (www.b-en-t.nl) Print: Haveka (www.haveka.nl)