Chapter 4
Preparing the Next Generation of Education Researchers for Big Data in Higher Education

David C. Gibson and Dirk Ifenthaler

Abstract Research in social science, education, psychology, and humanities is still dominated by research methodologies that primarily divide the world into either *qualitative* or *quantitative* approaches. This relatively small toolkit for understanding complex phenomena in the world limits the next generation of education researchers when they are faced with the increased availability of big data. In this chapter, we are calling attention to data mining, model-based methods, machine learning, and data science in general as a new toolkit for the next generation of education researchers and for the inclusion of these topics in researcher preparation programs. A review of the state of the art in research methodology courses and units shows that most follow a traditional approach focusing on quantitative and/or qualitative research methodologies. Therefore, this chapter makes a case for a new data science foundation for education research methodology. Finally, benefits and limitations of computationally intensive modeling approaches are critically reviewed.

Keywords Learning analytics • Big data • Research methodology • Machine learning • Higher education research • Computational modeling

Introduction

Given the increasing availability of data from vast interconnected and loosely coupled systems of administrative, academic, and personal information flowing within and across organizations and businesses, the challenge of data management, analysis, visualization, and interpretation, which is integral to advancing knowledge and understanding in the arts and sciences, is constantly evolving. The situation highlights two concepts at the heart of our argument, *complexity* and the *role of large amounts of dynamic evolving data* in scientific modeling and theory formation. We argue that the current tools and processes for the preparation of researchers in many fields are inadequate for facing both complexity and big data, and we propose a new

D.C. Gibson (✉) • D. Ifenthaler
Curtin University, Bentley, WA, Australia
e-mail: david.c.gibson@curtin.edu.au

conceptual foundation for research preparation and support courses and units in higher education.

Concerning complexity, the phenomena studied in all fields are often complicated as well as highly varied, with many overlapping elements and parts interconnected in a variety of ways. The phenomena are often set inside of other systems or are themselves made of many subsystems, with numerous relationships. As important as being *complicated*, both the systems and their environments are *dynamically evolving* in time, often with many self-referential influences and flows that can lead to chaotic and surprising behavior such as self-organizing and adaptive capabilities in natural and living systems (Bar-Yam, 1997; Holland, 1995; Liu, Slotine, & Barabási, 2011; Rockler, 1991).

In relation to the challenges brought on by dynamic data in knowledge creation, the large amount of data now available for social scientists, for example, is far too complex for conventional database software to store, manage, and process. In addition to the huge volume of data, the data accumulates in real time, with a requisite need to analyze and use the information to make timely decisions. Finally, the source and nature of this enormous and quickly accumulating data is highly diverse (Gibson & Webb, 2015; Ifenthaler, Bellin-Mularski, & Mah, 2015). Hence the next generation of researchers across all fields of arts and sciences face new challenges for identifying valuable information from big data and understanding multilayered interactions of complex phenomena.

But research in social science, education, psychology, and humanities is still dominated by research methodologies that primarily divide the world into either *qualitative* or *quantitative* approaches (see Creswell, n.d. for a 40-year history retrospective shaped by Sage Publications). For the most part, the two approaches are treated as philosophically and operationally disconnected and capable of being bridged only by "mixing" the methods. This has led to a simplistic view of research that is hampering understanding of complex phenomena in many fields. We believe that there is a new "third way" to approach research and this article outlines its main features as part of a call for higher education research preparation programs to invest in up-skilling faculty and redesigning research methodologies units and courses. In this article, we concentrate the examples on the social sciences, education, humanities, and arts, but the same argument holds as well for many of the sciences.

To illustrate the problem of the traditional divide and the lack of resolution of the twentieth-century debates (see, e.g., Caporaso, 1995; Onwuegbuzie & Leech, 2005; Rihoux & Grimm, 2006; Shah & Corley, 2006; Tarrow, 2010) that are now frozen into the research preparation programs in higher education, we offer a brief example. On the *quantitative* side of research preparation, null hypothesis significance testing (NHST) is the dominant analytical strategy taught to each successive generation of researchers. However, NHST is a limited way to interpret data because it refers primarily to the question of whether there is a *significant effect* or not (Cumming, 2012) and whether to support or discredit a priori speculations about some aspect of a population (Kachigan, 1991). This approach, a mainstay of doctoral dissertations, leaves unaddressed the questions of in what ways data are related, within what structures, and with what specific predictable (or approximately

predictable) bounded as well as changing sequences and sets of relationships. Currently, higher education research preparation courses and processes are educating the next generation of researchers without an adequate toolkit for understanding complex models and enabling them to participate in the benefits of a computational mindset to theory and knowledge building.

By "computational mindset" we mean to differentiate a capacity for research conceptualization that differs from applying a specific set of skills, such as whether one can program a computer, solve an equation, or build an operational model of a mechanism. We refer instead to a capacity of "awareness plus literacy" (where all those specific skills are highly welcomed on the team!) concerning the role of algorithms in transforming the nature of scientific inquiry in the late twentieth century (Chaitin, 2003). The next generation of researchers needs to understand this change and its implications for significance in new research and embrace a "third way" of thinking about the integration and new empowerment of both qualitative and quantitative perspectives of a research program through computationally intensive modeling, visualization, and exploratory data analytic methods.

Therefore, in this chapter, we are calling attention to data mining, model-based methods, machine learning, and data science in general as part of a new toolkit needed in higher education research. The next section provides a discipline-based example that focuses on new challenges for education researchers. The third section reviews the state of the art in higher education unit and course-based research methodology offerings in order to note the absence of knowledge about big data analytics. The fourth section proposes key elements of a framework for preparing the next generation of researchers for the era of big data analytics. The chapter concludes by asking for integrating alternative analytics methodologies into existing curricula, which will better enable a new generation of researchers to participate in big data research.

Challenges for a New Era of Education Researchers

One of the promises of big data in educational settings is to enable a new level of evidence-based research into learning and instruction and make it possible to gain highly detailed insight into student performance and their learning trajectories as required for personalizing and adapting curriculum and assessment (Shum & Ferguson 2011). Being accountable for student success, higher education institutions that analyze and create new interventions and actions based on data analytics in their contexts may enhance their institutional effectiveness. Furthermore, if developed as an organizational capacity, the ongoing analysis of big data can provide insights into the design of learning environments and inform decisions about how to manage educational resources on all levels (Ifenthaler et al., 2015).

Educational data mining (EDM) describes techniques and tools to analyze all kinds of data on different hierarchical levels in educational settings (Berland, Baker, & Berland, Baker, & Bilkstein, 2014; Romero, Ventura, Pechenizkiy, & Romero, Ventura, Pechenizkiy, & Baker, 2011). In addition to the nested hierarchical character

of much educational data (e.g., answer level, session level, student level, teacher and institutional level), the performance time, sequence of actions, and evolving elements of the learning context are also important features of relevant data in educational settings. EDM is interdisciplinary and draws on machine learning, artificial intelligence, computer science, and classical test statistics to analyze data collected during learning and teaching. Although closely related to learning analytics, which focuses on improving learning and performance with feedback loops to the learner and instructor (Ferguson, 2012; Ifenthaler, 2015; Long & Siemens, 2011), EDM focuses on exploring new patterns in data and on developing new models at all levels of an educational system. Some of the common goals of current EDM practices are (1) predicting academic performance and student success for recruitment, retention, and work readiness, (2) evaluating student learning within course management systems and improving instructional sequences, as well as (3) evaluating different kinds of adaptive and personalized support. Additionally, EDM is advancing research about modeling student, domain, and software characteristics.

EDM involves five methods: (1) prediction, (2) clustering, (3) relationship mining, (4) distillation of data for human judgment, and (5) discovery via models. Prediction includes models about academic performance of students, for example, by analyzing their behavior in an online learning environment. Clustering methods can be used to group students according to specific characteristics, e.g., preference or performance patterns to recommend actions and resources to similar users. Relationship mining, which is perhaps the most often applied method in EDM, refers to identifying relationships among variables, like classroom activities, student interaction or student performance, and pedagogical strategies. The fourth technique, distillation of data for human judgment, aims to depict data in a way that enables researchers to quickly identify structures in the data. The last method, discovery via models, uses a preexisting model that is then applied to other data and used as a component in further analysis.

Accordingly, the next generation of education researchers need to be equipped with a new set of fundamental competencies that encompass areas needed for such computationally intensive research (e.g., data-management techniques for big data, working with interdisciplinary teams who understand programming languages, as well as cognitive, behavioral, social, and emotional perspectives on learning) and the fundamental principles of the computational mindset, by which we mean a bedrock of professional knowledge (including heuristics) that inclines a researcher toward computational modeling when tackling complex research problems.

State of the Art in Research Methods Units and Courses

Since the nineteenth century, debates among education researchers have focused on the differences between *quantitative* and *qualitative* approaches to research (Gage, 1989). However, the two methodologies entail more than different ways of gathering data; they also express different, often opposing and conflicting, assumptions

about the purpose of research and phenomena in the world (Bryman, 1988). An in-depth analysis of research literature reveals several common dichotomies, such as qualitative–quantitative, subjective–objective, inductive–deductive, hermeneutics–positivism, understanding–explanation, and descriptive–predictive (McLaughlin, 1991). Only recently, this dichotomous view on education research has faded as more and more research studies combine qualitative and quantitative features of inquiry through the "mixed methods" approach (Creswell, 2008). The mixed methods approach primarily alternates between the two methods, places them in sequences, or interleaves the various perspectives. In the approach we are advocating here, there is a tighter connection that operationalizes the qualitative aspects of both content and process via algorithmic integration with computational resources as a coadjutant (mutually assisting cocreator) in theory formation. For example, active visualization is not viewed as a representation of what is known or an illustration of what has been found in data but is instead used to explore, discover, and in multiple ways present the possible relationships among data points, assisting in the search for patterns rather than performing a role as a display of knowledge. The proposed stance we are introducing and discussing here is thus an active, interactive coproducer of knowledge, with algorithms and algorithmic agents working alongside human thought and action.

The current state of the art in preparing education researchers for the future is the research dissertation project, often supported by a research methodology course. Differences exist across the world as well as across the university in terms of the specificity of that preparation, for example, preparation for research in physics is quite different from that in education or psychology. Numerous textbooks have been published to support research methodology courses, mostly focusing on classical research practices including (1) linear steps in the process of conducting research, (2) restricted number of possible research designs, and (3) limited number of accepted analytics strategies (e.g., Bortz & Döring, 1995; Cohen, Manion, & Morrison, 2011; Creswell, 2008; Denzin & Lincoln, 2000).

Only recently, researchers in education have started to bridge between standard research practices in the humanities-oriented educational, social, and psychological research fields and the scientifically oriented cognitive science, computer science, and artificial intelligence fields. However, most research preparation courses in higher education still follow a traditional approach focusing on quantitative, qualitative, or mixed methods research designs. The following four examples provide evidence for the absence of alternative computationally intensive modeling methodologies required for the analysis of big data in educational settings.

Example 1: Short Certificate Course on Research Methods

A 3-month online course created by a consortium of partners of the Alexis Foundation aims at preparing researchers to develop the most appropriate methodology for their research studies. The short certificate course includes four modules. The first module deals with types of research and the research process. The second

module focuses primarily on hypothesis-driven research (identifying a hypothesis, gathering and making sense of data that test it, and interpreting the data). The third module explores alternative models of research but with much less weight as alternatives to hypothesis testing. The fourth module supports the "write-up" with a range of scholarly reports and mentions ethics fairly close to the advice on footnoting and citations (http://www.ccrm.in/syllabus.html).

Example 2: Research Methods in a Faculty of Education

An introductory research methods course in education at the University of Freiburg, Germany, is taught over two semesters (32 weeks) and has a strong emphasis on quantitative analytical strategies including descriptive and inferential statistics. The course uses a research-based learning approach (Freeman, Collier, Staniforth, & Smith, 2008; Healey, 2005) by integrating a research project conducted by the students as the driver of the overall course experience. The lecturer introduces a current research problem (e.g., teacher's perception of school development) at the beginning of a semester, and students are asked to form small research groups (approximately four students per group). After a self-guided in-depth literature review, students are asked to identify research problems within the larger context of the research project (e.g., what factors hinder teachers from active participation in school development?). In the next step, students develop the research methodology including instruments and procedures. Depending on the status of the overall research project, instruments are provided by the lecturer or are developed as pilot instruments by the students. The lecturer and teaching assistants help in organizing the sample for the data collection (including necessary permissions, etc.). The data analysis is performed within groups in the tutorials, while problems and outcomes are addressed in the lectures to enable students to develop a broader understanding of the issues emerging across all the projects. As a final outcome of the course, students produce a research project report following scientific guidelines (Ifenthaler & Gosper, 2014).

Example 3: Research Methods in a Faculty of Information

A research methods course in the faculty of information at the University of Toronto has an emphasis on qualitative methods. First, the course offers an overview of different approaches, considerations, and challenges involved in social research. Second, the course reviews core human research methods such as interviews, ethnographies, surveys, and experiments. Third, it explores methods used in critical analysis of texts and technologies (discourse/content/design analysis, historical case studies), with an emphasis on digital information (e.g., virtual worlds, videogames, and online ethnographies). Fourth, it also discusses mixed methods approaches, case studies, participatory and user-centered research, as well as research involving minors (http://current.ischool.utoronto.ca/course-descriptions/inf1240h).

Example 4: Research Methods in a Faculty of Business Education

The last example stems from a postgraduate fully online course with a duration of 13 weeks that is provided by the Australian Catholic University through Open Universities Australia (www.open.edu.au). The course includes a range of concepts and techniques associated with both qualitative and quantitative methods of research that are applicable for business and/or information systems. The syllabus includes sessions focusing on (1) types of research, (2) design, (3) defining the research question, (4) search and reviewing the literature, (5) methods and instrument in quantitative research, (6) methods and instruments in qualitative research, (7) sampling and data collection, (8) presenting and describing quantitative data, (9) inference for quantitative data, (10) qualitative data analysis, (11) mixed methods (quantitative and qualitative), and (12) writing a research report (http://www.open.edu.au/courses/business/australian-catholic-university-research-methods--mgmt617-2015).

Summary of Current State of Research Preparation

These four examples are emblematic of the current state in higher education research preparation courses and offer evidence of the absence of awareness of the transformation of the leading edge of research and practice driven by computational science methods. In spite of the rise of "computational" as a prefix to new fields in biology, chemistry, political science, modeling, architecture, neuroscience, and elsewhere, the basic research preparation experiences in the arts, humanities, and social sciences have, for the most part, remained rooted in late-nineteenth- and early-twentieth-century epistemology. In the next section, we outline why the current state of research preparation is inadequate for the era of big data and some of the key ideas central to the third way which deeply integrates the traditions to better advance knowledge as well as research practice via computational science approaches to understanding complex systems.

Preparing the Next Generation of Education Researchers for the Era of Big Data

A new foundation for research methodology in multidisciplinary research extends the traditional quantitative and qualitative approaches with complex systems understandings that entail and require new data-management and analysis techniques for big data. Big data in higher education is driven and enabled primarily by interactive technologies such as user tracking on web sites, user actions and products in highly interactive digital learning and assessment platforms, and large-scale data collection

in projects at increasing scale sizes and complexity (diversity of data sources) as well as resolutions (data records per user). Therefore, the next generation of researchers must be able to demonstrate competencies in the fast-changing technological field of big data analytics and be able to apply new tools, algorithms, and analytic platforms to various scenarios in education, social sciences, humanities, business, health systems, leadership, policy, and many other areas of application.

Limitations of Regression Models for Big Data

A major analytic strategy in education and other social science research is regression modeling or prediction analysis (Kachigan, 1991). In this section, we provide a simple example that illustrates a linear regression, with one or multiple criteria and predictors. However, linear regression algorithms are limited for application in big data analytics because they assume that rules and data are independent. Specifically, there are three assumptions about the population of data that must be met in order for linear regression to be an adequate model of the phenomenon under study: (1) there needs to be a probability distribution of *independent* values of each criterion, (2) the variances of all the distributions have to be *equal* to one another, and (3) the means of the distributions must all *fall on the regression line*. But in a complex data environment with dynamic interdependencies, these assumptions are almost never met. Worse still, to make research-based predictions and discuss findings as though these conditions *have* been met when they are often unstated and assumed and that the phenomenon under study *is therefore reasonably represented* as linear, independent, and well behaved creates inaccurate models and understandings. The emergent qualitative traditions that matured in the late twentieth century noticed and reacted to this shortcoming (e.g., Guba, 1985; Lincoln, Guba, Lincoln, & Guba, 1985) but did not extend the computational toolkit. Instead a whole new branch of methods and traditions arose which did not depend upon or take advantage of computational resources. Data science methods now emerging have reintroduced the possibility of a scienfitically defensible bridge between the two worlds of qualitative and quantitative methods for those who wish to unify the divide by discovering and modeling nonlinear and complex relationships.

For example, to identify nonlinear and complex parameter relationships in data, one successful approach uses support vector machines (Cortes & Vapnik, 1995). A support vector machine (SVM) is a binary classification technique based on supervised machine learning in the broad area of artificial intelligence (Drucker, Burges, Kaufman, Smola, & Vapnik, 1997). The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making the SVM a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to

belong to a category based on which side of the gap they fall on. SVMs can efficiently perform a nonlinear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. This example illustrates how an alternative analysis method can help overcome one of the limitations that students trained in a traditional research methodologies course will be confronted with when they are requested to approach a research problem in big data.

Big Data

Big data is often referred to with three "v" aspects: volume, variety, and velocity (Romero et al., 2011). There are usually a large number of records representing by "orders of magnitude" more information than in past research practices. The data typically streams in from a wide network of sources at varying timescales, resolutions, and levels of semantic import. Finally, the data builds up in near real time and must be analyzed rapidly if timely decisions are to be made, so new forms of filtering, patterning, and saving aggregate information on the fly are needed to assist in the rapid analysis and decision-making process (Ifenthaler & Widanapathirana, 2014). A complex adaptive education system, in comparison, has a large number of possible state-spaces (volume), systems, and subsystems that are actively contributing to the system's evolution while remaining open to an ever-changing outside environment (variety) and multiple time scales that depend on the fastest subsystem (velocity).

Both data science and complex adaptive systems are unique fields and are evolving separate terms, tools, practices, and communities, but there is a remarkable alignment, as we might well expect, since our knowledge of systems is often created by sensor networks that feed our best-fit and ever-changing models in the form of computational representations. That is, the computer-based models that are now the common architecture of the sciences (e.g., astronomy, chemistry, biology, medicine, physics, sustainable ecosystem models) are both a result of and a creator of big data. As a result, a new worldview has emerged in which data science integrated with a conception of evolutionary algorithms is now the applied mathematics of empirical science. This change in worldview has been chronicled by writers from many fields: political and economic (Beinhocker, 2006; Friedman, 2005; Radzicki, 2003), philosophical and practical (Manning, 1995; Newman, 1996; Putnam, 1992; Tetenbaum, 1998), scientific and mathematical (Holland, 1995; Prigogine, 1996), and historical and sociological (Diamond, 2005; McNeill, 1998; Wicks, 1998).

Six Key Ideas for a New Conception

We hold that research preparation in many fields needs to catch up to the rest of science and move quickly to incorporate complexity and data science ideas into research methods courses in all fields. A rebalancing is needed to shift practice from its roots

Table 4.1 Six key ideas for a new conception of a research methodology course

Complex system concept	Definition
Nonlinearity	A nonlinear system is one in which the output is not directly proportional to its input; the cause of some response by the system is not the simple sum of the stimuli, as it is in linear systems
Feedback loops	Information is recycled, connecting the current state to past states of the system
Openness	The system accepts "inputs from" and "outputs to" a larger external environment
Memory	Impacts on the current state of the system are carried forward into future states of the system
Nested relationships	Components of the system may themselves be complex systems
Emergent properties	Properties of the whole system depend upon the nonlinear nested relationships of the components and often need a new level of analysis and representation from that of the components

and current practices and into innovative exploratory new arenas. Table 4.1 shows six key ideas which were outlined in Gibson and Knezek (2011) and could form the backbone of a new conception of a research methodology course to begin the process of acquainting researchers with complexity ideas. These are not offered as an exhaustive list but as a set of key ideas underpinning the new analysis methods.

The concepts presented in Table 4.1 imply the use of new computational, representational, and epistemological tools and methods that help connect complex systems knowledge with the knowledge created via traditional qualitative and quantitative methods.

The comfort zone of researchers starts with the tools and processes they already know and must add to that knowledge base incrementally when the need arises. If a research team sees that there is nobody on the team with the knowledge and skills to deal with the above in both a qualitative and quantitative sense, then the team needs to expand its capacity to include a trained data scientist who can help fill the gaps.

Big Data Analytics in Education

A new foundation for research methodology in education research needs to provide people with practical hands-on experience on the fundamental platforms and analysis tools for linked big data, introduce several data storage methods and how to distribute and process them, introduce possible ways of handling analytics algorithms on different platforms, and highlight visualization techniques for big data analytics. Additional competencies include large-scale machine learning methods as foundations for human–computer interaction, artificial intelligence, and cognitive networks.

An example of key topics for a course focusing on big data analytics in education can be found at Columbia University (http://www.ee.columbia.edu/~cylin/course/bigdata). An introductory unit could focus on big data analytics, platforms, data storage, and data processing. A second unit could introduce different big data analytics algorithms, such as recommender, clustering, and classification. A final unit could introduce key concepts of data visualization and graph computing.

In addition to the abovementioned course content, the following elements could be integrated into a course focusing on big data analytics in education:

- Distributed and cloud-based data management, data cleaning, and data integration
- Using metadata
- Harvesting and extraction of unstructured data
- Probabilistic and predictive modeling
- Pattern recognition
- Data, text, and image mining
- Network analyses (social relationships, structural implications, information flows)
- Semantic web and ontologies
- Sentiment analysis

Conclusion

The key idea here is that we are comparing how knowledge emerges from exploratory analytics versus from hypothesis testing. Both approaches can lead to a model, but the first approach invents the model where the second approach validates it.

What we are advocating is a balancing of the creative impulse with external validation, both with increased global professional community engagement and the establishment of research that is more open, transparent, and amenable to scientific scrutiny, meeting the criteria of reproducibility and generalizability.

We may be criticized that some science cannot be made reproducible or generalizable and that insisting on these criteria might lose something. So we reply that what we are proposing does not have to replace current methods that are subjective, opaque, and incommensurable; we only need to allow the new methods and knowledge to take their place among the current practices as quickly as possible so that a new generation of scientists in all fields will have the option of participating in big data research and analysis.

Benefits include:

- Open data (anonymous data sets that form a new benchmark community)
- Open data transformation processes (no black box data transformations)
- Open algorithms and algo-sequences (fully transparent processing)
- Reproducible results (might lead to new forms of "meta-analysis" where the concerns are NOT commensurability and confidence in ES)

- Generalizable results (might lead to actual models in the scientific sense of the word)

    Limitations include:

- Key teaching and learning information may not be captured (this applies to all methodologies, not just computational approach).
- Difficulties in translating computational models developed into actions to improve teaching and learning (especially where relationships are not linear/curvilinear).
- Variables identified within models may be indicators rather than the causal variables (e.g., the number of books borrowed from library is an indicator; causal variable may be amount of time spent reading. The act of taking books out of the library does not in itself promote learning and teaching). Without a theory driving the analysis, it is difficult to distinguish between the two.

Researching big data is a new and fast-growing field with numerous career opportunities for people with the curiosity, knowledge, and skill to collaborate on teams that solve complex problems with computational methods. Researching big data is most often a collaborative process, because the problems and attendant solutions are complex, entailing overlapping fields of expertise. Solutions often call for computational and discipline knowledge including mathematics, systems thinking, and educational, psychological, and organizational theory. Therefore, a new data science foundation for education research methodology and preparing the next generation of education researchers for big data in higher education is inevitable.

References

Bar-Yam, Y. (1997). *Dynamics of complex systems*. Reading, MA: Addison-Wesley.

Beinhocker, E. (2006). *The origin of wealth: Evolution, complexity and the radical remaking of economics*. Boston, MA: Harvard Business School Press.

Berland, M., Baker, R. S., & Bilkstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning, 19*(1-2), 205–220. doi:10.1007/s10758-014-9223-7.

Bortz, J., & Döring, N. (1995). *Forschungsmethoden und Evaluation*. Berlin: Springer.

Bryman, A. (1988). *Quantity and quality in social research*. London: Unwin Hyman.

Caporaso, J. A. (1995). Research design, falsification, and the qualitative-quantitative divide. *American Political Science Review, 89*(2), 457–460. doi:10.2307/2082441.

Chaitin, G. (2003). *Algorithmic Information Theory Third Printing*. Computer.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). New York: Routledge.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. doi:10.1007/bf00994018.

Creswell, J. W. (2008). *Educational research. Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson.

Creswell, J. (n.d.). How Sage has shaped research methods: A forty year history.

Cumming, G. (2012). *Understanding the new statistics. Effect sizes, confidence intervals, and meta-analysis*. New York: Taylor & Francis Group.

Denzin, N. K., & Lincoln, Y. S. (2000). *Handbook of qualitative research*. Thousand Oaks: Sage.

Diamond, J. (2005). *Collapse: How societies choose to fail or succeed*. New York: Viking Penguin.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems 9* (pp. 155–161). Cambridge, MA: MIT Press.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317. doi:10.1504/IJTEL.2012.051816.

Freeman, J. V., Collier, S., Staniforth, D., & Smith, K. J. (2008). Innovations in curriculum design: A multi-disciplinary approach to teaching statistics to undergraduate medical students. *BMC Medical Education, 8*, 28. doi:10.1186/1472-6920-8-28.

Friedman, T. (2005). *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Straus & Giroux.

Gage, N. L. (1989). The paradigm wars and their aftermath: A "historical" sketch of research on teaching since 1989. *Educational Researcher, 18*(7), 4–10.

Gibson, D., & Knezek, G. (2011). Game changers for teacher education. In P. Mishra & M. Koehler (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2011* (pp. 929–942). Chesapeake, VA: AACE.

Gibson, D., & Webb, M. (2015). Data science in educational assessment. *Education and Information Technologies, 20*(4), 697–713.

Guba, E. G. (1985). The context of emergent paradigm research. In Y. S. Lincoln (Ed.), *Organizational theory and inquiry* (pp. 79–104). Newbury Park, CA: Sage.

Healey, M. (2005). Linking research and teaching exploring disciplinary spaces and the role of inquiry-based learning. In R. Barnett (Ed.), *Reshaping the university: New relationships between research, scholarship and teaching* (pp. 30–42). Maidenhead, UK: McGraw-Hill International.

Holland, J. (1995). *Hidden order: How adaptation builds complexity*. Cambridge, MA: Perseus Books.

Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *Encyclopedia of educational technology*. Thousand Oaks, CA: Sage.

Ifenthaler, D., Bellin-Mularski, N., & Mah, D.-K. (2015). Internet: Its impact and its potential for learning and instruction. In J. M. Spector (Ed.), *Encyclopedia of educational technology*. Thousand Oaks, CA: Sage.

Ifenthaler, D., & Gosper, M. (2014). Research-based learning: Connecting research and instruction. In M. Gosper & D. Ifenthaler (Eds.), *Curriculum models for the 21st Century. Using learning technologies in higher education* (pp. 73–90). New York: Springer.

Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning, 19*(1-2), 221–240. doi:10.1007/s10758-014-9226-4.

Kachigan, S. K. (1991). *Multivariate statistical analysis: A conceptual introduction*. New York: Radius Press.

Lincoln, Y. S., Guba, E. G., Lincoln, E., & Guba, Y. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature, 473*(7346), 167–173. doi:10.1038/nature10011.

Long, P. D., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 31–40.

Manning, P. K. (1995). The challenges of postmodernism. In J. Van Maanen (Ed.), *Representation in ethnography*. Thousand Oaks, CA: Sage.

McLaughlin, E. (1991). Oppositional poverty: The quantitative/qualitative divide and other dichotomies. *Sociological Review, 39*(2), 292–308.

McNeill, W. (1998). History and the scientific worldview. *Hisotry and Theory, 37*(1), 1–13.

Newman, D. V. (1996). Emergence and strange attractors. *Philosophy of Science, 63*(2), 245–261.

Onwuegbuzie, A. J., & Leech, N. L. (2005). Taking the "q" out of research: Teaching research methodology courses without the divide between quantitative and qualitative paradigms. *Quality and Quantity, 39*(3), 267–296. doi:10.1007/s11135-004-1670-0.

Prigogine, I. (1996). *The end of certainty: Time, chaos, and the new laws of nature*.

Putnam, H. (1992). *Renewing philosophy*. Cambridge, MA: Harvard University Press.

Radzicki, M. J. (2003). Mr. Hamilton, Mr. Forrester, and a Foundation for Evolutionary Economics. *Journal of Economic Issues, 37*(1), 133–173.

Rihoux, B., & Grimm, H. (2006). Innovative comparative methods for policy analysis: Beyond the quantitative-qualitative divide. *Innovative Comparative Methods for Policy Analysis: Beyond the Quantitative-Qualitative Divide*. doi:10.1007/0-387-28829-5

Rockler, M. J. (1991). Thinking about chaos: Non-quantitative approaches to teacher education. *Action in Teacher Education, 12*(4), 56–62.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. J. D. (Eds.). (2011). *Handbook of educational data mining*. Boca Raton, FL: CRC Press.

Shah, S. K., & Corley, K. G. (2006). Building Better Theory by Bridging the Quantitative – Qualitative Divide. *Journal of Management Studies, 43*(8), 1821–1835. doi:10.1111/j.1467-6486.2006.00662.x.

Shum, S. B., & Ferguson, R. (2011). *Social learning analytics* (pp. 1–26). Knowledge Media Institute & Institute of Educational Technology. doi:10.1145/2330601.2330616.

Tarrow, S. (2010). Bridging the quantitative-qualitative divide. *Rethinking Social Inquiry* (pp. 101–110). Retrieved from http://books.google.com/books?hl=en&lr=&id=1VQK7EGohB4C&oi=fnd&pg=PA171&dq=Bridging+the+Quantitative-+Qualitative+Divide&ots=Z7fT2VDyFr&sig=Sg2h7poL6RrbxUkVcJGNwNuFtKs

Tetenbaum, T. J. (1998). Shifting paradigms: from Newton to chaos. *Organizational Dynamics, 26*(4), 21–32.

Wicks, D. (1998). Organizational structures as recursively constructed systems of agency and constraint: compliance and resistance in the context of structural conditions. *The Canadian Review of Sociology and Anthropology, 35*(3), 369–390.