BIG DATA LEARNING

Buy This Book

By Elliott Masie
With Bob Baker & Learning CONSORTIUM Members
www.biglearningdata.com (for resources and updates)
# TABLE OF CONTENTS

**Dedication** ................................................. v  
**Elliott Masie**

**Foreword** ................................................. vii  
**Tony Bingham**

**Introduction** .............................................. 1  
**Bob Baker**

## SECTION I

**Big Learning Data and Data Analysis Are Important** ............................................. 7

- **Ch1** On Big Learning Data: Thoughts From Elliott Masie ................................ 9  
  **Elliott Masie**

- **Ch2** Why Are Big Learning Data and Data Analytics Important? ...................... 17  
  **Nigel Paine**

- **Ch3** The Skills and Mindset Required for Big Learning Data ......................... 23  
  **Donald H. Taylor**

## SECTION II

**Impacts on People in the Learning Field** ......................................................... 29

- **Ch4** Big Learning Data and Training Programs:  
  Start Small to Go Big and Go Big to Go Long ............................................. 31  
  **Tom King**

- **Ch5** Big Learning Data: Three Roles of the Learning Leader .......................... 45  
  **Coley O’Brien**

- **Ch6** Stakeholder Perspectives and Needs for Big Learning Data .................... 49  
  **Rahul Varma, Dan Bielenberg, Dana Alan Koch**
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ch7</td>
<td>Avoiding the Dangers of Big Learning Data</td>
<td>55</td>
</tr>
<tr>
<td>A.D. Detrick</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ch8</td>
<td>Big Learning Data Risks: Privacy, Transparency, Culture, and Silly Data</td>
<td>67</td>
</tr>
<tr>
<td>Elliott Masie</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SECTION III</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approaches to Big Learning Data</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Ch9</td>
<td>Case Study: It’s Bigger Than Big Data— Metrics and Measurement Without a Strategy Is Just Data</td>
<td>75</td>
</tr>
<tr>
<td>Nickole Hansen, Peggy Parskey, Jennifer O’Brien</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ch10</td>
<td>Case Study: Big Data, Big Difference in Training and Development</td>
<td>85</td>
</tr>
<tr>
<td>Jeff Losey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ch11</td>
<td>Case Study: Points for the Training Scoreboard</td>
<td>97</td>
</tr>
<tr>
<td>Ben Morrison</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ch12</td>
<td>Case Study: Big Learning Data Sources and Analytics</td>
<td>107</td>
</tr>
<tr>
<td>Doug Armstrong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ch13</td>
<td>A Perspective From K-12: A View From the U.S. Department of Education on Big Learning Data in K-12</td>
<td>123</td>
</tr>
<tr>
<td>Epilogue: Moving Forward</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>Bob Baker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>About the Authors</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>Index</td>
<td>143</td>
<td></td>
</tr>
</tbody>
</table>
DEDICATION

Big data can help create big learning. Let’s approach big data as learners, embracing these two perspectives:

*The price of light is less than the cost of darkness.*

—Arthur C. Nielsen

*Not everything that can be counted counts, and not everything that counts can be counted.*

—Albert Einstein
Nate Silver loves data. And he knows how to use it. Silver, who left a job as an economic consultant with KPMG and nurtured his love for statistics and baseball, developed a system for projecting player performance and careers. That system, Player Empirical Comparison and Optimization Test Algorithm (PECOTA), is used by baseball business professionals to predict the performance and value of major league players.

Nate Silver is also the guy who accurately predicted the election outcomes in 49 of 50 states in the 2008 U.S. presidential election, and improved his own stats by accurately predicting the election outcomes in all 50 states in the 2012 election. Data, demographics, behavior—and the correct analysis of all of it—are powerful, and always have been.

But in today’s digital landscape, one where people walk around with access to the world’s knowledge in their pockets and regularly interact with millions of pieces of information, data takes on new dimensions. We are only beginning to conceptualize what can be done with the mammoth amount of information to which we have access.

I have talked about the power of technology in the learning profession for years. It is revolutionizing the way learning and development practitioners do their work. Leveraging big data is the next logical step in this evolution. The outputs of technology—the data that we gather—provide learning professionals a new vantage point from which to view the work they do.
I am excited about this book that Elliott Masie and his team of collaborators have created. What I found in these pages was a comprehensive SWOT analysis of big data and its implications for the training and development profession. Most refreshing are the honest questions asked by the contributors who wonder whether the field is ready to embrace, understand, and apply the power of big data in their work. The writers identify the strengths and weaknesses, the opportunities and threats that exist as the learning profession gets its collective arms around the what, why, and how of slicing and dicing more information than ever before.

For me, one of the most powerful chapters in the book is chapter 6, “Stakeholder Perspectives and Needs for Big Learning Data,” because it talks about impact, the real heart of the issue. “Analytics and business go hand in hand,” the authors say. I believe that learning and business go hand in hand too. Training professionals are wise to make their value proposition to the organization in the language of business. This is something I talked about in depth in Presenting Learning. More importantly, for learning to really be the business driver it can be, it is incumbent on training professionals to know and understand the business, who the stakeholders are, and how learning can help those stakeholders achieve their objectives. Big data holds the promise of new insights and of exploring new ways to drive results. It is literally untapped potential at our fingertips.

In Silver’s 2012 book, The Signal and the Noise: Why So Many Predictions Fail but Some Don’t, he states, “The numbers have no way of speaking for themselves. We speak for them. We imbue them with meaning. Like Caesar, we may construe them in self-serving ways that are detached from their objective reality.” This is the exact challenge identified and addressed in the book you are now reading.

We have access to volumes of data but we must understand what it can tell us, what it does tell us, and as importantly what it can’t and doesn’t tell us. As training professionals, we take that information and layer it over the organizational goals we are seeking to support, the gaps we are trying to close, and the engagement and retention metrics we are trying to improve. And then we create courses, programs, initiatives, and processes that have sustained business impact.
Big Learning Data addresses all of this. From explaining why it’s important to the field, to identifying impacts on and cautions for practitioners, to offering case studies that frame the discussion, this book is an important and timely work that every learning professional should read. I believe we owe Elliott and his team a debt of gratitude for bringing this book to us at this critical time. Always forward thinking, Elliott is calling attention to the next transformative opportunity for our field.

Big learning data can empower us to develop the knowledge and skills of professionals around the world in ways that we’ve never been able to do before. It’s never been a more exciting time to be in the learning profession.

—Tony Bingham
September 2013
INTRODUCTION

We live in an extraordinary time in history when it comes to the volume of data that exists around us and the volume that is being created. Data are everywhere. Consider this:

- Intel Corporation estimates that the world generates 1 petabyte (1,000 terabytes) of data every 11 seconds, or the equivalent of 13 years of high-definition (HD) video (Finnan, 2013).

- The proliferation of devices such as PCs and smartphones worldwide, increased Internet access within emerging markets, and the boost in data from machines such as surveillance cameras or smart meters have all contributed to the doubling of the digital universe within the past two years alone. There is now a mammoth 2.8 ZB (zettabytes), according to a December report titled "IDC Digital Universe," which was sponsored by EMC Corp.

This has opened the door to the world of big data. Big data is generated and affects our lives on a daily basis:

- According to a Cisco report in June 2012, big data solutions could help reduce traffic jams or even eliminate them with predictive, real-time analysis on traffic flows. The data could feed immediate changes to traffic signals, digital signs, and routing—before backups begin. Paper receipts from retailers and banks that clutter one’s wallet are moving toward replacement by electronic records. Businesses could enrich these records through contextual and comparative information. The report also noted that individuals
could manage, share, monetize, and utilize the data through, for example, budget management and health advice applications.

* As of early 2012, the big data market stood at just more than $5 billion based on related software, hardware, and services revenue, according to market research firm Wikibon. The total big data market reached $11.4 billion in 2012, ahead of Wikibon’s 2011 forecast. The big data market is projected to reach $18.1 billion in 2013, an annual growth of 61 percent. This puts it on pace to exceed $47 billion by 2017, the report said.

What is big data, exactly? Definitions of big data vary. There are, however, several common characteristics in these definitions. The term generally describes three aspects of data:

* **Volume:** Big learning data enables an organization to access and analyze a volume of data for a richer perspective. *Volume* can mean information about thousands of learners taking a course or experience. *Volume* can mean you are looking at multiple data points, over time, about a single learner. *Volume* can provide data on a deeper and richer set of learning activities—even capturing the time a learner paused while answering a specific question. And, *volume* might someday bring together learning data from hundreds of organizations, providing a global perspective.

* **Velocity:** Big learning data enables learners and organizations to have rapid access to data—even in real time. Imagine a worker entering a wrong answer into an assessment exam. *Velocity* would instantly provide him with remedial and enrichment options based on his historical learning patterns and successful strategies from thousands of other learners who also failed that question. Finally, *velocity* would allow learning producers the ability to make adjustments to content delivery—based on rapid analysis of user experience—on a continual basis.

* **Variety:** Big learning data connects the dots, weaving together a wider variety of information from talent, performance, demographics, and business metrics. You can then see the correlations between learning performance and other behavior and background points. Imagine correlating performance reviews
with learning activities and hiring data, either for thousands of employees or drilled down to a single worker.

And with this have come new methods for working with data: data analytics. These approaches are required to handle the volume of data and to portray them in useful and powerful ways that result in new capabilities or significant improvements in existing ones. Analysis of big data also offers the potential for better predicting the future with predictive analytics.

**Big Learning Data**

So, what does big data mean for our workplace learning field: big learning data?

Quite simply, big learning data is big data that we apply to our learning field. But the implications of big learning data are far from simple. It will require us to think first of all about data in new ways, including why big learning data is important, as well as to develop new skills and mindsets in our field to deal with it. It also requires us to take a deeper look at which data are—and will be—available, not just in our learning functions, but in our organizations and with our learners, for example. At the same time, our organizational stakeholders will play key roles in how we move forward as learning functions with big learning data. The roles that learning leaders might play in leveraging big learning data are also significant.

Big learning data has the potential to play a substantial role in shaping the future of learning from various perspectives. For example, imagine how robust data and analytics might enable us to more deeply personalize the learning experience. Collecting data on the time between keystrokes by the learner might provide insights into how confident she is and may afford the opportunity to design in “enriching” experiences. Plus, big learning data might help us become more effective in an area that has been challenging for learning professionals: learning evaluation. It also has the potential to inform many more strategic decisions about how learning works in our organizations, including what technologies to invest in.

At the same time, there are numerous challenges and traps that we as learning professionals need to watch out for and guard against. Chief among them is the quality and value of the data itself. Some data is just “silly” data. And because there is a lot of it, not all has value or will have
impact. We need to be careful about how much we depend on data for making decisions. Lessons of experience teach that over-dependence on quantitative data without qualitative insights can be a trap. We also face challenges in how much data to share and collect. Along with this, there will no doubt be issues of ethics and transparency that will become both significant and problematic.

**What You Will Find in This Book**

In this book, we bring together multiple perspectives on big learning data for a practical look at what it means and the potential it offers in the world of workplace learning. We explore the topic from the point of view that in our organizations, the process of workforce learning generates an enormous amount of data that ranges from who takes which courses or consumes which learning objects, to the timing and impact of learning activities on performance or career retention, to how many seconds people watch certain videos or access key documents on the corporate server. Data are generated by the actions and decisions of learners, managers, customers, and others. Some of the data are meaningful, some are confusing, and some are intriguing. Assuming we had the learning systems and analysis models, what big learning data would we collect and use to improve the learning process?

The perspectives presented in this book focus on several important aspects of the answer to this important question. The opening chapters from Elliott Masie, Nigel Paine, and Donald H. Taylor present an overview of what big learning data is and why it is important. They foreshadow how the workplace learning field will be affected as big data’s influence increases, and how skills and mindsets will need to change. The next chapters by Tom King, Coley O’Brien, Rahul Varma, Dan Bielenberg, Dana Alan Koch, A.D. Detrick, and Elliott Masie provide more in-depth perspectives on key aspects of big learning data, including: sources of data and analytics; the role of learning leaders; how big data can affect learning design in training programs; the relationship with learning function stakeholders; and the potential dangers of big learning data. The book concludes with several case studies from Nickole Hansen, Peggy Parskey, Jennifer O’Brien, Jeff Losey, Ben Morrison, and Doug Armstrong that discuss how their organizations are implementing a big learning data approach. Thought leadership in
the education field extends beyond workplace learning, for example, also in the K–12 space. So we have added a summary of the U.S. Department of Education's recent report *Expanding Evidence Approaches for Learning in a Digital World*, which focuses on how big learning data might inform sound decisions, fuel innovation, and optimize technology-based learning resources.

**References**


SECTION I

BIG LEARNING DATA AND DATA ANALYSIS ARE IMPORTANT
To open this book, we share some perspectives on where we are and where we might be headed with big data in organizational learning. These perspectives are shaped by experience in working with more than 240 public and private sector organizations of various sizes that comprise the MASIE Learning CONSORTIUM.

Big Learning Data

Overall, we see big data as the ability to analyze, compare, and slice enormous streams of data—primarily by-products of the digital age. Therefore, what makes big data “big” is looking at a vast number of data elements across a volume of incidence—not just one person, but thousands of people, for example. It is also the phenomena of now having very large amounts of data from myriad sources: many transactions, computer movements, and aggregations of some noncomputer behaviors, including biological phenomena. However, there is also a lot of meaningless data. As
such, part of the big data model is figuring out “Where do I look at a vast volume from a value perspective?”

Big data opens the possibilities of understanding at a deeper level that, in most cases, can’t be achieved otherwise. For example, it can give an historical analysis: why did people vote that way at the poll; why did people go or not go to that course? It can also provide a predictive framework: how do I get more people to a course; or get more people to the poll; or even on some level, how do I use a design phenomenon to personalize for an individual experience based on his history and how a wider set of data is used.

Big data in learning provides learning professionals with new opportunities. Whether the learning professionals want to use big data or not, businesses in many cases are already leveraging big data for business intelligence and are inevitably going to draw the connection between learning and customer satisfaction. We believe that what learning folks can do, whether or not their organizations are pushing for business intelligence, is to use these data points to help them better design learning, better evaluate the impact of learning, better fuel an evidence-based approach to experimentation, and better create personalization.

Big learning data will ultimately come down to value. That is, what can we gain from big data? Benefits can be for the learner, the designer, the manager, or the organization, enabling each one to do something better, faster, cheaper, more strategically, and more persuasively.

**Sources of Big Learning Data**

The problem with data, historically, is that we’ve always gone for the low-hanging fruit. We, as learning professionals, have collected inexpensive, easily acquired data from people while they are in our domain, usually a classroom or program. In fact, data that may be the most impactful is higher up the tree. For example, it’s finding out six weeks after training from a learner’s manager whether there was a difference. Or if you go back to the learner a year after the program, or three years later, you can find out whether he is still with the company, or if there was any difference in performance.

We also may need to rethink where we get our data. We will need to look at learning in the broader context, and when we do, we will begin to see
many potentially valuable data sources, including many that already exist. One way to look is from a broader human resources perspective. The relationship between selection, training, and competency is very interesting. Often we evaluate the impact of a leadership program, for example, with the assumption that we did great things in the program. In reality, we know that a lot of it has to do with how well we select the people to go into that program from our pool, and how well we select people to join the organization. We might also look at what they did before they came into the program and what their manager did after they left the program. Another way to think about where to get data is in an element of design: the usability of learning content and resources. For example, every instructor will mention books, articles, and now TED Talks, every time they teach. Instructors typically mention them based on their opinions about those they enjoy. We might look at whether people went to them, or if they completed them, what they thought about them, and so on. So if I as an instructor was in a more data-rich environment, those dropped resources in my conversation, or in my e-learning module, or in my WebEx session, would be informed and maybe challenged by what the data puts out there.

We will also need to consider alternative approaches to collecting data, which have some important implications. Some areas where our approaches to data sources might change include:

* **Depth of measurement:** We have looked, for example, at whether learners passed an exam. But more valuable data might include not just the answer, but also characteristics of how learners answer the question—such as, how long it took them to answer and whether their mouse hovered over a wrong answer for a while.

* **Expense:** We have relied on inexpensive data. Some data that we will use in big learning data will be more expensive to get than what we have traditionally used. But what we easily collected tended to be modestly superficial. Collecting data through interviews with managers of learners, for example, costs much more, but yields much more data.

* **Types of data:** We have looked for how learners have answered a question. But more valuable would be their confidence in answering the question.
As an example of these points, early in my career I was involved with The Rockefeller Foundation, and I was funding some arts and education programs. We were trying to evaluate whether people went to see art that we hung in museums. We went through all sorts of ways of doing it and Robert Stake, who was one of the godfathers of evaluation, had a great comment: “What you need to do is buy a carpet-measuring device and roll it around the museum to see where the carpet was mashed down the most.” And sure enough, it was a measure of which pieces of art got the most attention, good or bad. Well, nobody would have ever said in a grant to The Rockefeller Foundation that we’re going to measure carpet. But in a big data model, we would measure carpet. Are there the equivalent of “carpet data” that we need to ferret out?

Shifts From Big Learning Data

As we move toward big learning data approaches, we will need to shift in the learning field in at least two ways.

1. We need to have an anthropological view of the learning process, to understand that there are many factors that may influence learning. We need to realize learning may influence or may support or destroy the impact of learning, thus broadening our view of potentially relevant data.

2. We need to have an analytical approach, which says that if we gather this data, we need to analyze that data with integrity and with a more sophisticated multivariable analysis. How do we display that data so it brings meaning to people? If I’m given this data, what do I do with it strategically, and how do I handle that?

In making these changes, it will be important to also consider that we want to be anthropologists who are analytical and have an ethical code. As we broaden our data and analysis aperture, this will become more important for learning professionals. For example, it would be interesting to ask every person in class what the level of stress in their life outside of work is. It may be the key indicator of success. But we don’t know how to ethically ask that question, so while there are data points that may be really powerful, we have to ask if they are within our ethical domain.
Some Challenges Ahead

There will be numerous challenges that learning organizations will confront as they increasingly move toward big learning data. One challenge is readiness—the extent to which individuals making decisions are ready to operate with a massively enhanced set of data. For example, designers pick a few elements that they want, but they aren’t limitless in that. At which point do we analyze what color background works best, or at which point do we analyze the font size? And on some level, they may all be key, but I don’t know if we have the design skills to do this. It’s not emotional intelligence; it’s the big data intelligence we need to do this.

A second challenge is the need at some point to upgrade, alter, or change learning systems. Most learning management systems (LMSs) are not big data ready. It doesn’t mean that they couldn’t be adapted or enhanced, but most of them are not big data ready. Also, there ultimately is an open model of big data in which we understand that data exist in many domains. We need to understand where and how and in what way it’s appropriate to share and use that data, because it could lead from a performance or personnel perspective to some pretty radical things happening. For example, if I suddenly realize that the 15 people who left the organization as managers were all hired by Claire, then we don’t have a learning problem, we have a Claire problem. So once you take away the boundaries, you’re going to end up with some perspectives that aren’t in your area of control, and that can be intriguing.

Another challenge is being able to distinguish between valuable and “silly” data. We can have very elaborate analysis of things that are meaningless. For example, we can have an infographic of the total number of people that have gone through training. It’s an interesting piece of data and we can get very granular, but we need to take a big data analysis approach to say: who they are; why they went to the training; what the outcome was; what happened to the people who didn’t go; or how the outcomes are variable. If we don’t analyze it from a big data perspective, then it’s lacking something.
Potential Impact

If you posit that big learning data will create an environment of rich information, it would suggest that the learner will be better informed and it would then allow for greater personalization. For example, say someone wants to take job B, having done job A for a year. Big data would indicate, first of all, the number of people who did A that then got to B. Of the people who got B, what preparation did they have? It would also indicate which learning programs were most effective, and what the timing was for when they attempted to change to job B. It might even get all the way down to if we could do it, and we could sort against the volumes, then we are not just targeting against three people. If it’s in a big organization, if it’s the federal government, I may be looking across a thousand people who did that. We might find that watching videos and spending three hours with a key manager once a week is more powerful than going away for a 28-day training program, so that the knowledge set could allow the learner to be a designer, a decider, or a targeter.

Big learning data also could be informative from a feedback and context arena, because very often somebody might fail at a topic but not know why they are failing. It becomes interesting when the learner can look not just at themselves, but at other people who have had that experience. They may certainly get an insight, either that would explain it so they are not frustrated or that they could use to correct it—so that they could succeed again.

Also, if you implemented big data in a comprehensive way, the learner potentially becomes very invested in inputting data to the process, because they see the impact of how it works. We would need to be concerned about whether learning can be micro-engineered by data, and I am always a combination of a behaviorist and a more gestalt approach. That said, ultimately we can program a better mix or selection, though we would want to be cautious if we thought that with enough data we can program everybody to a predicted success at a predictive rate, and at a predictive time, and have very few failures in that process.
Looking Ahead

When looking out some years, the question arises: How different will learning be in organizations that embrace a big learning data approach? We would say moderate, because we in learning organizations are still creatures of habit. A lot of learning is culturally habit-based and because it is so, contrary to most hype, we have never seen an instant revolution in our field. One might say webinars were a brand new invention, but they go back to the mid-1990s when the first webinar providers came out. So there aren’t true overnight successes.

We have some decisions to make that will affect the pace of adoption of a big learning data approach. We will need to decide, for example, whether we use big data and in which formats. Do we use it retroactively as analytical to evaluate what worked or did not work at a big level? Do we use it predictively to influence what we do going forward at a big level, or do we take the personalized side of big data where we use the data to target you? In some sense that’s the Amazon.com model, which uses a lot of data from a lot of people and then uses the data about what each individual has looked at, thought about, visited, and not bought. And so then we could target an advertisement about other learning opportunities they might need. I don’t know whether we call that last bit big data. I think we are too early in the process and some people would say no, because if you are talking about a unit of one it’s not big data to them. But you could do that level of personalization if somewhere you had collected big data.

We are in the early stages of considering the possibilities for big learning data in the organizational workplace. And while exciting, this means that we have a lot of work to do as learning professionals. Some early thought leaders are emerging and are beginning to define the dimensions of where big learning data may lead us. My hope for readers of this book is that you will be active learners, evidence-based experimenters, and will explore both the opportunities and challenges that big learning data presents to the learning and knowledge world. Big data is here. Our opportunity is to shape big learning data to help create big—personalized—effective learning for employees and organizations.