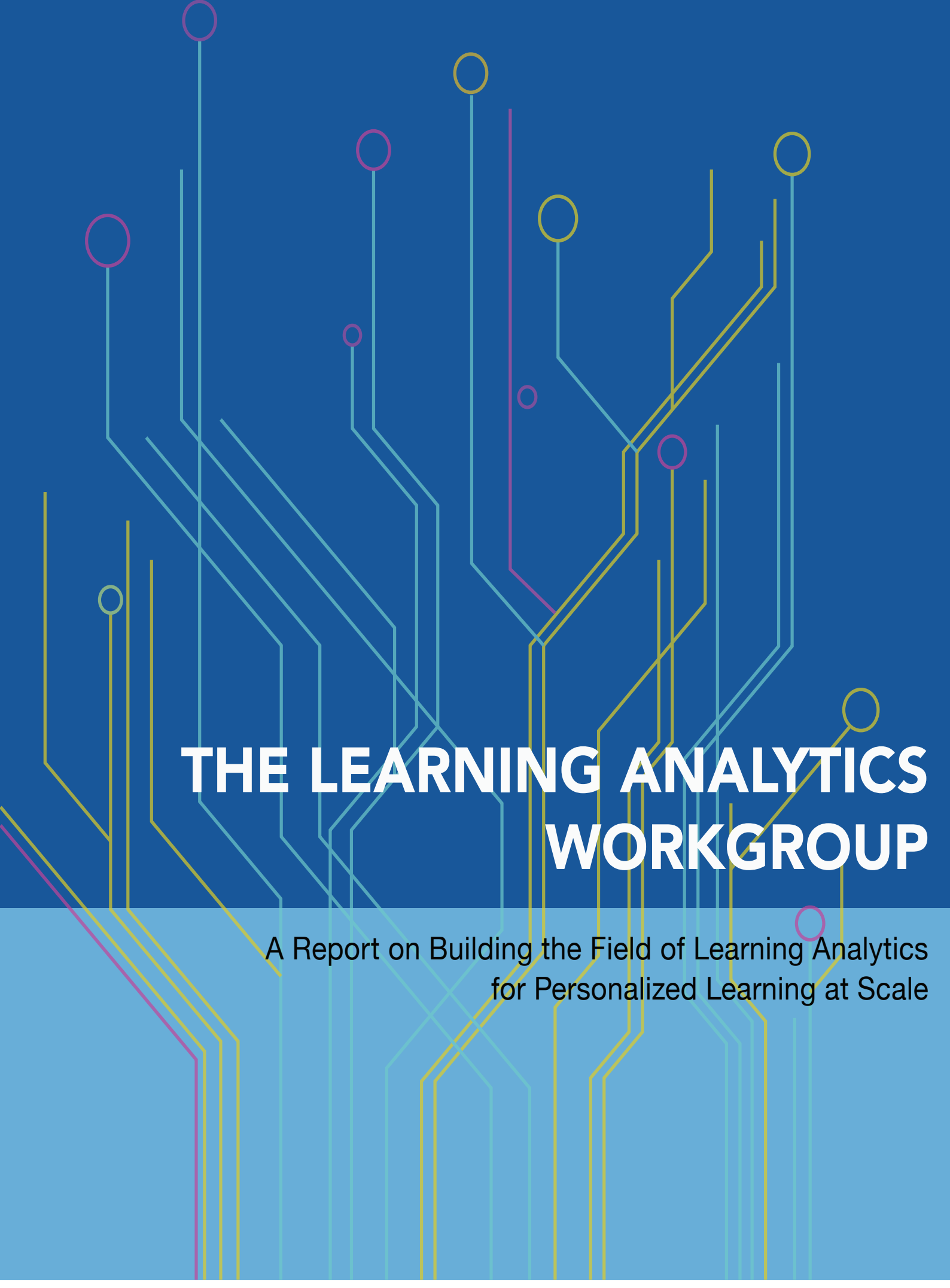


THE LEARNING ANALYTICS WORKGROUP

A Report on Building the Field of Learning Analytics for Personalized Learning at Scale

Submitted by:

Roy Pea, D.Phil., Oxon. David Jacks Professor of Education and Learning Sciences, Stanford University

Author Contact Information:

Roy Pea, D.Phil., Oxon.
David Jacks Professor of Education and Learning Sciences
Professor, Computer Science (Courtesy)
Stanford University, Graduate School of Education, and
Director: H-STAR Institute (Human Sciences and Technologies Advanced Research)
Wallenberg Hall
450 Serra Mall, Building 160
Stanford CA 94305-2055
roypea@stanford.edu
650.724.3720 Office

This report was made possible through a grant from the Bill & Melinda Gates Foundation and the MacArthur Foundation to the Board of Trustees of the Leland Stanford Junior University. Special thanks to LAW team member Associate Professor Taylor Martin, Dr. Sarah Brasiel, and the Active Learning Lab of Utah State University, for their exceptional support in the final period of report preparation.

Table of Contents

LAW Project: Background and Purpose 1

Section 1: A Conceptual Framework for Building the Field of Learning Analytics 10

Section 2: Critical Questions for Understanding How to Build the Field of Learning Analtytics.....23

Section 3: Articulating and Prioritizing New Tools, Approaches, Policies, Markets, and Programs Within the Field of Learning Analytics 42

Section 4: Determining Resources Needed to Address Priorities 47

Section 5: Road Map for How to Implement the Field-Building Strategy and How to Evaluate Progress 61

References 80

LAW PROJECT:
BACKGROUND AND PURPOSE

This brief section of the report sketches the motivating conditions for the creation of the proposed LAW Project and its current and planned activities.

The Call for Change –The Birth of the LAW Project

On August 3, 2011, the Bill and Melinda Gates Foundation convened a multisector group at the University of Chicago's Computation Institute.

The goal of the meeting, hosted by University of Chicago's professor Ian Foster, was to consider how to best build capacity in the field for creating an innovative and sustainable ecosystem dedicated to advancing the state of learning data and learning analytics on behalf of all children's college and career readiness. An emergent focus during the meeting was on the impending opportunities and challenges associated with understanding the orders of magnitude of data generated as millions of K-12 students and their teachers transition into digitally enabled teaching and learning compared to what is managed by today's school data systems in utilizing those data to improve education. The message was that the growth of data in education surpasses the capacity to make sense of it and to employ insights derivable from the data to guide practices.

September 2011, following this meeting, LAW Principal Investigator Roy Pea drafted and shared a white paper for the Gates Foundation on a strategy for building the field of learning analytics to address these challenges and opportunities.

In a related development, the National Academy of Education convened a summit on December 1-2, 2011 on adaptive educational technologies in Washington, DC to consider how to best move forward the research community's understanding of available data from these technologies and to understand their possible applications in research and for educational improvement.

Broader discussions of the learning analytics and personalized learning challenges and opportunities with Gates Foundation and other researchers and policy makers at the National Academy of Education summit made clear the importance of concerted action to develop a multi-sector plan for building the field of learning analytics in support of improving education. Accordingly, during the winter of 2011–2012, the proposal was developed at Stanford and in collaboration with the funders for what would become the LAW Project, with funding committed in July 2012 and the first workshop planned for September 21–22, 2012, at Stanford University. The LAW Project next would convene experts in the field through workshops and conference panels to understand the needs and current developments to inform the building of the field.

LAW PROJECT ACTIVITIES: SUMMARY

From September 2012 through July 2014, the LAW Project planned and conducted four different workshops, planned and presented three conference panels, led a crowd-sourcing campaign for soliciting learning analytics resources for the field, launched the first Learning Analytics Summer Institute (LASI-13), commissioned 11 white papers on a variety of vital issues for building the field of learning analytics, and completed this final report as a culmination of this work. This report would not have been possible without the 37 contributors or advisors to the LAW Project, who represent multiple interested sectors: 13 from the academy, 9 from for-profit companies, 7 from nonprofit organizations, 4 from government, and 4 from philanthropic foundations. We provide an overview of each of these key activities as an important context for this report.

Workshops

Four workshops were held between September 2012 and March 2013. The dates, locations, and brief information about each workshop are provided in Table 1. The attendees for each workshop and their affiliations are provided in Table 2.

Table 1. Four Workshops for the LAW Project

WORKSHOP NUMBER	LOCATION	DATE	AREA OF FOCUS
Workshop 1	Stanford University	September 21-22, 2012	Defining key clusters of issues for subgroups to develop as task forces
Workshop 2	Stanford University	October 17-18, 2012	Continued work in task force groups to define issues
Workshop 3	Washington, D.C.	November 3, 2012	Planned to overlap with the National Academy of Education's annual meeting to enlist a broader group of participants to contribute to the research work and field-building activities
Workshop 4	Seattle, WA Gates Foundation	March 14-15, 2013	Discussions of the LAW project findings and open issues with Gates Foundation Program Officers

Table 2. Workshop Attendees and Affliations

NAME	AFFILIATION	WORKSHOP #			
		1	2	3	4
Ryan Baker	*Teachers College, Columbia University*	X			X
John Behrens	*Pearson NSC*		X	X	X
Marie Bienkowski	*SRI International*	X	X	X	X
Paulo Blikstein	*Stanford University*	X	X	X	
Simon Buckingham Shum	*Knowledge Media Institute(KMI), Open University*		X		
John Henry Clippinger	*MIT Media Lab &Institute for Institutional Innovation & Data Driven Design (ID3)*		X		X
Stephen Coller	*Bill & Melinda Gates Foundation*	X	X		
Ed Dieterle	*Bill & Melinda Gates Foundation*	X		X	X
John Easton	*US Department of Education, Institute of Education Sciences(IES)*	X		X	
Michelle Elia	*CPSI*		X		X
Ian Foster	*University of Chicago*		X		
Paul Franz	*Stanford University*	X	X	X	X
Bernd Girod	*Stanford University*		X		
Wayne Grant	*Intel Corporation*		X		
Myron Gutmann	*National Science Foundation (NSF), Directorate for Social, Behavioral, and Economic Sciences (SBE)*	X			
Patricia Hammar	*P.K. Hammar Legal and PKH Enterprises*		X	X	X
Robert M. Hauser	*National Research Council*		X	X	
Tom Kalil	*White House OSTP (Office of Science and Technology Policy)*			X	
Kenneth R. Koedinger	*Carnegie Mellon University (CMU)*	X		X	
Jace Kohlmeier	*Khan Academy*		X	X	
Daphne Koller	*Coursera and Stanford University*	X			
L. Arthi Krishnaswami	*RyeCatcher(Past: The College Board)*		X		X
Taylor Martin	*Utah State University*	X	X	X	X
Robert J. Mislevy	*ETS*			X	
Joan Ferrini-Mundy	*ational Science Foundation (NSF), Directorate for Education and Human Resources (EHR)*	X		X	
Ari Bader-Natal	*Minerva*		X	X	X
David Niemi	*Kaplan*	X		X	X
Zachary A. Pardos	*Workshop on Personalizing Learning (WPL) and UC Berkeley*	X			
Roy Pea	*Stanford University*	X	X	X	X

Table 2. Workshop Attendees and Affliations (continued)

NAME	AFFILIATION	WORKSHOP #			
		1	2	3	4
Philip J. Piety	*Johns Hopkins University*			X	X
Prasad Ram	*Gooru*		X	X	X
Lauren Resnick	*University of Pittsburgh*			X	X
Paul Resnick	*University of Michigan School of Information*	X			
Martina A. Roth	*Intel Corporation*		X		
Steve Schoettler	*Junyo*	X	X	X	X
Harry Shah	*Gooru*		X		
John Stamper	*Human-Computer Interaction Institute, Carnegie Mellon University (CMU)*		X	X	X
Emily Schneider	*Stanford University*	X	X	X	X
Marcelo Worsley	*Stanford University*		X	X	
Molly Bullock	*Stanford University*	X	X	X	
Darrell M. West	*Brookings Institute*			X	
Bob Wise	*Alliance for Excellent Education*	X		X	

The participants at each workshop met in five task force groups to develop knowledge and recommendations in one of five areas. What follows is a summary of the work in each task force.

Task Force 1

Advancing Education Data

This task force focused on advancing education data. This task force discussed questions about the value proposition for stakeholders, required competencies for education data scientists, and the resources for different stakeholders. They made recommendations for the options and pathways to train and foster education data scientists. They also provided recommendations for an online community for learning analytics resources and peer learning opportunities.

Task Force 2

Adapting Learning Technologies to Education

This task force focused on questions about how to make learning technologies adapt to education. The goal is to evolve the infrastructure for learning analytics for K-12 digital curricula and assessments, e-texts and associated opportunities for big data education science. They focused on the content and method of assessment and data collection. They also considered how to improve the infrastructure according to target users and learning environments.

Task Force 3

Learning Science to Provide Personalized, Pervasive Learning at Scale

This task force focused on learning science related to providing personalized, pervasive learning at scale across contexts. The three primary aspects discussed were connected learning models, multiplicity of learning resources, and recommended engines for learning resources and experience. Some detailed topics about analysis of personalized learning were carefully discussed from different aspects like modes, outcome analysis, learning priorities, research methodology and scope, learner models, transparency, and feedback research.

Task Force 4

Creating Multimodal Learning Analytics

This task force focused on how to create multimodal learning analytics to ensure that the education data include contextual features of learning environments. Two main questions were discussed: What are the priority issues in providing the better theories, methods, and tools for multiple data streams? How can we achieve data privacy and anonymity when creating multimodal learning analytics?

Task Force 5

Data Privacy, Research IRB, and Ethical and Internet Protocol (IP) Considerations

This task force focused on how to manage data privacy, research IRB, and ethical and IP considerations for learning in a networked world. In order to research this topic, the participants analyzed this question from the perspectives of different stakeholders, and they tried to look for policy and practice, protection options for different learning contexts or demographics, and methods of data storage and sharing.

Conference Panels to Share Knowledge and Rally Participation

Next, three conference panels were developed for two purposes: ***knowledge sharing and rallying*** to attract talented data scientists, analysts, and developers to the cause. The sessions were planned for venues where learning scientists and educational researchers rarely go—the data science oriented trade conferences hosted by O'Reilly. This effort also included one keynote presentation (see Table 3).

Table 3. Conferences Related to the LAW Project

Conference	Location	Date	Keynote Panel Title	Keynote Speaker Panelists
Strata	New York, NY	October 24, 2012	"Data Exponential: K-12 Learning Analytics for Personalized Learning at Scale—Opportunities & Challenges"	Roy Pea, Kenneth Koedinger, Taylor Martin, Stephen Coller
Strata	Santa Clara, CA	February 27, 2013	Morning Keynote: "Using Data to Honor the Human Right to Education" Panel: "Learning's Clarion Call: Teaming to Improve U.S. Education With Big Data Science"	Keynote: Prasad Ram Panelists: Marie Bienkowski, Jace Kohlmeier, Zachary A. Pardos, Sharren Bates
SXSW-Edu	Austin, TX	March 6, 2013	"Building the Field of Learning Analytics"	Roy Pea (Chair), Stephen Coller, Ken Koedinger, Taylor Martin

The First Learning Analytics Summer Institute (LASI 2013)

This project was co-funded by the Bill and Melinda Gates Foundation. The event was co-organized by Roy Pea, Taylor Martin, Dragon Gasevic, and John Behrens. This was a strategic 5-day event, July 1–5, 2013, attended by people from all over the world in person or virtually. The objective was for participants to be equipped to actively engage in advancing the learning analytics field through their teaching and research activities.

Crowd-sourcing Campaign for Soliciting Learning Analytics Resources

The 11 categories for which we sought contributions were as follows (in alphabetical order):

- Companies and Non-Profits
- Conferences and Societies
- Courses and Syllabi
- Data Science-Other
- Degree/Certificate Programs
- Grant Opportunities
- News/Blogs/Reports
- Research Labs
- Research Publications
- Success Cases
- Tools and Algorithms

Another activity to build networks and elicit learning analytics field-building resources involved using the IdeaScale crowd-sourcing platform. Both the LAW membership and the LASI 2013 program applicants were invited to contribute resources of various kinds to the platform for our uses in developing the LAW final report (see *http://learninganalyticsworkinggroup.ideascale.com*).

Over 300 resources were contributed by over 100 individuals.

History of the Development of the Field of Learning Analytics

- 1990's Research in Intelligent Tutoring Systems
- 2008 International Educational Data Mining Society first conference is held
- 2009 First publication of the Journal on Education Data Mining
- 2011 First International Conference on Learning Analytics and Knowledge
- 2011 Founding of the Society for Learning Analytics Research
- 2013 Journal of Learning Analytics is established

—For more information, see Martin and Sherin's (2013) introduction in the special issue of *The Journal of the Learning Sciences*, "Learning Analytics and Computational Techniques for Detecting and Evaluating Patterns in Learning."

White Papers

Finally, over the initial period of the LAW Project, as we conducted the workshops and surveyed the available literature in the field, priorities were developed for a set of 11 white papers that we felt were needed to build out specific issues for which a concerted and focused research and synthesis effort was warranted. The titles and authors of the 11 white papers developed are provided in Table 4.

Table 4. White Paper Titles and Authors

AUTHOR(S)	WHITE PAPER TITLE
Ryan Baker, Kenneth Koedinger	*Towards Demonstrating the Value of Learning Analytics for K-12 Education*
John Behrens, Bob Mislevy, Phil Piety, Kristen diCerbo	*Inferential Foundations for Learning Analytics in the Digital Ocean*
Marie Bienkowski	*Putting the Learner at the Center: Exposing Analytics to Learning Participants*
Paulo Blikstein	*Multimodal Learning Analytics and Assessment of Open-Ended Artifacts*
John Clippinger	*InBloom: Building Trust for The Protected Sharing and Analysis of Student Data for Personalized Learning*
Trish Hammar	*Learning Analytics Ecosystem Could be Fostered by Institutional Review Board (IRB) Reform*
L. Arthi Krishnaswami	*User Experience Techniques for Advancing and Communicating Learning Analytics*
Taylor Martin	*Connected Learning, Education Data Science, and Big(gish) Data*
David Niemi, Richard Clark	*Using Learning Analytics to Improve Academic Persistence*
Lauren Resnick, Carolyn Rose, Sherice Clarke	*Conversation Analytics*
Mary Ann Wolf, Bob Wise	*Policy and Capacity Enablers and Barriers for Learning Analytics*

The result of these efforts of the LAW Project is the ***culmination of a final report***, which includes the sections that follow:

The purposes of the LAW Project have been advancing each of these objectives, which we hope to make clear in the remainder of this report.

SECTION 1

A Conceptual Framework for Building the Field of Learning Analytics

College and Career Success

Many students come to college unprepared, as evidenced by as many as 40% of 1st-year college students being placed in developmental courses, with fewer than 60% of students completing college within 6 years. With technological advances, there are new ways to provide support to students to improve college and career readiness. In order for students to develop adaptive expertise to succeed in their college and careers, they need to have opportunities to develop and implement scientific processes, inquiry, critical thinking, and creativity, to name a few such skills.

Teachers, school administrators, parents, and students need to track learning activity and progress to accomplish the goal of college and career success for all students. As teachers and administrators are responsible for tracking the progress of many students, there is a need to be able to visualize learning at different levels of aggregation and use that information to guide their decision making. Such visualizations of learning progress can guide further instructional interventions and provision of progressive learning resources and experiences. Unfortunately, we fall behind when it comes to providing teachers, administrators, and families with the tools that they need to track progress and to ensure all students achieve college and career success.

Where We Fall Behind

It is noteworthy that there are ingrained historical reasons underlying the nonpersonalized learning that is commonplace practice today in the United States. Education historian Patricia Graham (2005) has highlighted how, early in the 20th century, a period of massive assimilation due to America's growing immigrant population and the transition from farming to an industrialized society based in cities led to vast expansions in the quantity and size of schools and a rigid common curriculum, organized by grades. The industrial factory model was broadly adopted for education and developed mass production and distribution of curriculum materials and teaching techniques. This industrial era of instructional design tended to provide all students with uniform learning experiences, requiring them to adapt to how, when, and where instruction was provided. This instructional philosophy and practice, still prevalent today, conflicts with contemporary research into how people learn (Bransford, Brown, & Cocking, 2000), which reveals that, with enough access to "learning opportunities" (Moss, Pullin, Gee, Haertel, & Young, 2008), time, guidance, appropriate "mindset" about intelligence (Dweck, 2007) and persistence (Duckworth & Seligman, 2005), almost everyone can learn just about anything to a great extent, and yet almost no one learns exactly the same way, through the same pathways. This understanding and desire to make and change educational experiences so that they are appropriate for every learner defines the need for personalized learning.

Graham (2005) also observed how schools in the United States following WW I began to seek to customize education by accommodating to the needs of children of distinct abilities and interests. For example, John Dewey's pedagogical innovation in "learning by doing" was an inspiration for later cognitive and social learning theories. In addition, the ***Brown v. Board of Education*** Supreme Court decision of 1954 sought to increase access to education for diverse populations, including minority and low-income students and children with disabilities, which was followed by the Elementary and Secondary Education Act of 1965 to expand education opportunities.

Where modern instructional theories advocate for equalization of learning opportunities and personalized learning, an inability to leverage simultaneously ***(a) an economy of scale and (b) learning analytics and education data mining by data scientists at scale*** has limited and restricted the scope and freedom to implement personalized learning widely. Another limitation has been the splintered learning standards across the states. But now most states are employing the Common Core State Standards in Mathematics and English Language Arts, and over 40 states are in early implementation planning for the newly released Next Generation Science Standards. The LAW Project has been focused on making progress towards personalizing learning at scale by building the field of learning analytics to support and advance initiatives and multi-state or national technology infrastructure to support new ecosystems of personalized learning and teaching. Next, we discuss the importance of personalized learning and what learning analytics can contribute.

> "Failure to support this effort or delaying its initiation will hamper our country's ability to provide personalized learning at scale to all students, with corresponding losses to the intellectual diversity and value of our graduates to the workforce and society at large."

What is Personalized Learning?

The National Academy of Engineering (2008) identified 14 grand challenges humans must solve to sustain and improve the human condition for current and future generations. The proposed effort targets the challenge that they identified of advancing personalized learning at scale for all learners with varying needs, skill levels, interests, dispositions, and abilities, arguing that continuously capturing, deriving meaning from, and acting on the production of vast volumes of data produced by learners engaged with digital tools are activities fundamental to personalized learning. A similarly framed grand challenge on personalized learning was proposed in the 2010 National Education Technology Plan (U.S. Department of Education, 2010). To effectively address this challenge head on will require a deliberate, consorted, and sustained effort by the academy and nonprofits, industry, government, private foundations, and practitioners to build the field of learning analytics and the associated human capital,

tools, and scientific knowledge capable of processing for strategic application the data of digital learning, data that are now growing exponentially. Failure to support this effort or delaying its initiation will hamper our country's ability to provide personalized learning at scale to all students, with corresponding losses to the intellectual diversity and value of our graduates to the workforce and society at large.

The authors of the 2010 National Education Technology Plan sought to provide clarity to the concept of personalized learning vis-à-vis related concepts: ***individualization, differentiation, and personalization***. These terms have become buzzwords in education, but little agreement exists on what exactly they mean, beyond the broad concept that each is an alternative to the one-size-fits-all model of teaching and learning. For example, some education professionals use personalization to mean that students are given the choice of what and how they learn according to their interests, and others use it to suggest that instruction is paced differently for different students. Throughout this plan, we use the following definitions for these key terms.

PERSONALIZED LEARNING CONCEPTS

Individualization

Instruction that is paced to the learning needs of different learners. Learning goals are the same for all students, but students can progress through the material at different speeds according to their learning needs. For example, students might take longer to progress through a given topic, skip topics that cover information they already know, or repeat topics on which they need more help.

Differentiation

Instruction that is tailored to the learning preferences of different learners. Learning goals are the same for all students, but the method or approach of instruction varies according to the preferences of each student or what research has found works best for students like them.

Personalization

Instruction that is paced to learning needs, tailored to learning preferences, and tailored to the specific interests of different learners.

In an environment that is fully personalized, the learning objectives and content as well as the method and pace may all vary (so personalization encompasses differentiation and individualization).

Why does Personalized Learning Matter?

There are many important arguments for the importance of personalized learning for education. The ones we review here include:

- Supporting learning for all students
- Improving educational performance
- Facilitating cost efficiencies through educational productivity and organizational optimization
- Accelerating educational innovation

SUPPORTING LEARNING FOR ALL STUDENTS

The 2010 National Education Technology Plan (U.S. Department of Education, 2010) called out the importance of making progress on personalized learning so as to support learning for the full spectrum of U.S. students:

> "The always-on nature of the Internet and mobile access devices provides our education system with the opportunity to create learning experiences that are available anytime and anywhere. When combined with design principles for personalized learning and UDL [Universal Design for Learning], these experiences also can be accessed by learners who have been marginalized in many educational settings: students from low-income communities and minorities, English language learners, students with disabilities, students who are gifted and talented, students from diverse cultures and linguistic backgrounds, and students in rural areas. (p. 23) "

In the following box we provide an overview of universal design, UDL, and three important core principles to make learning accessible to all students.

UNIVERSAL DESIGN

"Making learning experiences accessible to all learners requires universal design, a concept well established in the field of architecture, where all modern public buildings, including schools, are designed to be accessible by everyone. Principles and guidelines have been established for universal design in education based on decades of research and are known as Universal Design for Learning (UDL)" (U.S. Department of Education, 2010, p. 19). The 2010 National Education Technology Plan (U.S. Department of Education, 2010) lists three core principles for UDL, as follows:

- Provide multiple and flexible methods of presentation of information and knowledge. Examples include digital books, specialized software and websites, text-to-speech applications, and screen readers.
- Provide multiple and flexible means of expression with alternatives for students to demonstrate what they have learned. Examples include online concept mapping and speech-to-text programs.
- Provide multiple and flexible means of engagement to tap into diverse learners' interests, challenge them appropriately, and motivate them to learn. Examples include choices among different scenarios or content for learning the same competency and opportunities for increased collaboration or scaffolding.

IMPROVED EDUCATIONAL PERFORMANCE

Two factors work together to facilitate improved performance.

FIRST

We can anticipate improved outcomes on traditional measures of learning, since we are teaching what we have always taught but more effectively by virtue of the value added by learning analytics. One prominent example is implementing mastery learning in American classrooms at scale with the Cognitive Tutor learning software and associated curriculum (e.g., Pane, McCaffrey, Slaughter, Steele, & Ikemoto, 2010; Ritter, Anderson, Koedinger, & Corbett, 2007; Ritter, Kulikowich, Lei, McGuire, & Morgan, 2007). Cognitive Tutor software was developed around an artificial intelligence model, where a student's weaknesses in terms of concept mastery are identified and then a set of problems are customized for the student along with customized prompts to focus on areas where

the student is having difficulty. Learning analytics makes possible the assessment of outcomes that are now visible for the first time: Process can become a measurable outcome, and inferences from process data can be used to adapt an educational intervention for greater effectiveness.

SECOND

New kinds of valued outcomes may be obtained because new types of outcomes become measurable. For instance, automated detection of engagement and affect can inform automated intervention within online learning systems to reduce the frequency of disengaged behavior and improve learning of domain content (Arroyo et al., 2007; Arroyo, Woolf, Cooper, Burleson, & Muldner, 2011; Baker et al., 2006). Teachers can also be provided with information on recent student engagement (Walonoski & Heffernan, 2006). The measurement challenges of gauging learner progress in developing 21st century competencies such as collaboration and communication should also be amenable to process analysis from learning analytics. Learning dashboards that keep students on track with their course requirements and earning credits for courses taken online also should lead to improved educational performances and college readiness.

FACILITATING COST EFFICIENCIES THROUGH EDUCATIONAL PRODUCTIVITY AND ORGANIZATIONAL OPTIMIZATION

One of the promises of personalized learning in technology-enhanced educational environments is that the redesigns it will enable should lead not only to better educational achievements for the full spectrum of learners but also to cost efficiencies associated with greater educational productivity. As argued in the 2010 National Education Technology Plan (U.S. Department of Education, 2010), tight economic times and basic fiscal responsibility demand that we get more out of each dollar we spend. The plan noted that while improving productivity has been a daily focus of most American organizations across all sectors, education has neither incorporated many of the practices other sectors regularly use to improve productivity and manage costs nor used technology to enable or enhance them. Implicit in the personalized learning vision is releasing the traditional "seat-time" measures of educational attainment and focusing on mastery of standards-based competencies, allowing for learners in a class to diverge in the content they are studying. The new learners focus on their own learning sequence. For different arrangements of classroom size and disciplinary focus, school leaders can organize personnel differently than they have in the past—from blended learning enabling online learning outside of school to radically redesigned schools that restructure the provision of education with intensive personalization of learning with technology. Furthermore, educational institutions (at national, state, district, institutional, departmental, and course levels) are "driving blind," with weak feedback loops to evaluate the impact of ongoing practices or changes that are implemented in their practices. As with business intelligence in the corporate sector, it is becoming increasingly possible to see what is going on in a system, with the responsibility to act on that intelligence to optimize organizational processes and productivity. Closing feedback loops creates the opportunity to establish more efficient organizational structures.

ACCELERATING EDUCATIONAL INNOVATION

A final argument for personalized learning is that as research-validated learning analytics spreads through an open innovation commons and goes to market as part of educational products and services (e.g., interactive data visualization techniques, recommendation algorithms, learning maps), the platforms that are already in front of teachers and students become channels that expose them to new measures of effective learning. As one example, Desire2Learn's new dashboard has social learning analytics. In using this product educators will encounter social network analytics for their students for the first time but in doing so will be made aware of a new "outcome," thus catalyzing reflection on how to use such learning analytics for their purposes. For example, using Desire2Learn, students can share their presentations with their peers or publicly to receive feedback and engage in meaningful conversations. Teachers have access to social learning analytics where they can track student activity and progress within this type of learning outcome.

The Need to Build a Field of Learning Analytics

People of all ages are learning in face-to-face and online courses, in games, through peer-peer collaboration, and in the full learning ecology where their mobile devices accompany them; the data captured for learning analytics will need to encompass this panoply (Herr-Stephenson, Rhoten, Perkel, & Sims, 2011; NSF Cyberlearning Task Force, 2008; Shute & Ventura, 2013; Williamson, 2013). The endgame is personalized cyberlearning at scale for everyone on the planet for any knowledge domain.

There are urgent and growing national and global needs for the development of human capital, research tools and strategies, and professional infrastructure in the field of learning analytics and education data mining, made up of data scientists (straddling statistics and computer science) who are also learning scientists and education researchers. As the interactions and transactions that contribute to education at all levels and learning all the time, anywhere go "deeply digital," mediated by cyberinfrastructure, enormous opportunities exist to make sense of the vast quantities of data that are generated from these learning and teaching processes.

> "The endgame is personalized cyberlearning at scale for everyone on the planet for any knowledge domain."

In the following box we provide an overview of ***cyberinfrastructure and cyberlearning***, two key concepts in building the field of learning analytics.

CYBERINFRASTRUCTURE AND CYBERLEARNING

» **Cyberinfrastructure** is the coordinated aggregate of software, hardware, and other technologies, as well as human expertise, required to support current and future discoveries in science and engineering. The challenge is to integrate relevant and often disparate resources to provide a useful, usable, and enabling framework for research and discovery characterized by broad access and "end-to-end" coordination. Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high-performance networks to improve research productivity and enable breakthroughs not otherwise possible. Like the physical infrastructure of roads, bridges, power grids, telephone lines, and water systems that support modern society, cyberinfrastructure refers to the distributed computer, information, and communication technologies combined with the personnel and integrating components that provide a long-term platform to empower the modern scientific research endeavor.

» **Cyberlearning** is "learning that is mediated by networked computing and communications technologies," as defined in a seminal NSF Task Force on Cyberlearning (2008) report, later leading to the current NSF Cyberlearning grant program.

The field of learning analytics should strive to support incremental progress on the grand challenge problem of personalized learning by continuously capturing, deriving meaning from, and acting on data generated in digitally enhanced learning environments by students of diverse demographics with varying needs, skill levels, interests, dispositions, and abilities. Incremental levels of challenge, increased growth of understanding and expertise, and ongoing opportunities for success for every learner characterize personalized learning. These will be the primary benefits of knowledge gained from ongoing experiments and analysis of learning and teaching activities completed in the cyberinfrastructure. To improve the learning experience and the quality of the data scientists have to work with, digital learners make progress through important and relevant content topics cohesively (as compared to piecemeal) that encourage doing with understanding (Barron et al., 1998), beyond practicing isolated skills, in what is increasingly being called "deeper learning" (National Research Council, 2013). Learners can, for example, regularly complete explicit and "stealth" assessment opportunities embedded in the learning environment to demonstrate what they know and can do within cyberlearning activities, as compared to the typically summative assessments completed outside of cyberlearning (Shute, 2011).

The exponential growth of education data to be generated by digitally enhanced learning environments requires education data scientists and people with sense-making talent able to bring these datasets into productive interactive systems so that the various stakeholders—from teachers to learners, principals to parents—can visualize learning at different levels of aggregation and use it to guide their decision making. This would allow for further instructional interventions and provision of progressive learning resources and experiences. Personalized learning as a vision indicates how we need ***sensing systems for learning***, and the hosts of issues pertaining to any large-scale environmental or medical/health sensor systems apply here as well—data privacy and fine-grained access privileges to individuals with specific roles will need to be developed through iterative programs of continuous improvement oriented design research. As sensors are becoming a part of everything we interact with (e.g., wearable technology), we can now have a broader definition of what learning is and where data on learning can come from.

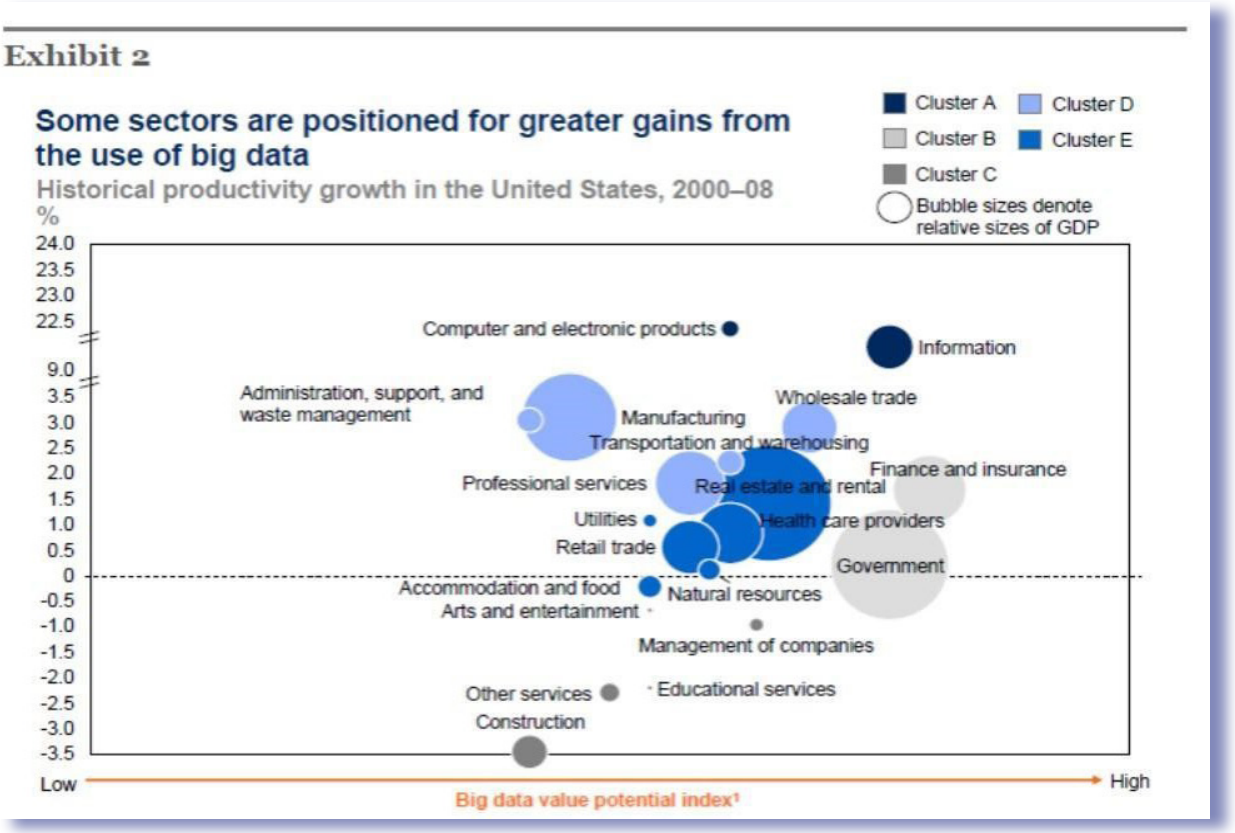
Data science, as a distinct professional specialization, is in its infancy (Hey, Tansley, & Tolle, 2009; though see Cleveland, 2001, and the 2003 debut of *The Journal of Data Science*). What we are calling out for is an even newer specialization, ***education data science*** (Pea, Childress & Yowell, 2012). Although data science had its foundations in computational statistics, in many ways it came to pre-eminence in 2008 when online communities such as Facebook and LinkedIn started to introduce applications on their sites that took advantage of the analysis of very large sets of user data in order to anticipate and predict user preferences (Patil, 2011). Data products are now essential to the value of our modern social networks and consumer sites (Manyika et al., 2011). They range from employing Hadoop to analyze enormous datasets with parallel distributed computing designs to Amazon's pioneering work on A/B testing to optimize web page layout, to any of the recommendation systems employed by Amazon, Apple, Netflix and other web retailers. (A/B testing is a quantitative method, developed by marketing researchers, that allows researchers to complete single-variable tests rapidly among sample populations in order to improve response rates and user experiences.) People staffing these teams come from a wide array of academic disciplines that initially did not have to do with "data science," but all of which involved dealing with and managing enormous datasets: business intelligence, oceanography, meteorology, particle physics, bioinformatics, proteomics, nuclear physics, fluid and thermal dynamics, and satellite imagery data.

> "Education is a sector far behind the curve in taking advantage of the advances being made in data science in adjacent sectors of the economy."

What all of these people have in common today is their lack of affiliation to any school of education or education industry. Education is a sector far behind the curve in taking advantage of the advances being made in data science in adjacent sectors of the economy. Cyberlearning infrastructures will need to link their efforts to expand the open platform for creating a vibrant, multi-state or nationwide learning-technology ecosystem with parallel efforts to evangelize the exponential education data opportunity with relevant leaders and practitioners in the data science community. The LAW Project has sought to advance this latter aim with panels on Learning Analytics and Education Data Science at Strata New York, Strata Santa Clara, and SXSW-Edu, though there is much work ahead.

A clear signal of the depth of the challenges before us in bringing data science to education is a recent McKinsey Global Institute report (Manyika et al., 2011) indicating that educational services in its present state of development is among the least likely of all societal sectors to benefit from the innovation acceleration that big data promise (see Figure 1). The report noted, "The public sector, including education, faces higher hurdles because of a lack of data-driven mind-set and available data" (Manyika et al., 2011, p. 9). An accelerated learning analytics field potentially would transform this woeful state.

Figure 1. Position of the education sector to gain from use of big data



From Big Data: *The Next Frontier for Innovation, Competition, and Productivity,* by J. Manyika et al., 2011, p. 9, New York, NY: McKinsey Global Institute. Reprinted with permission.

An additional emphasis in the McKinsey Global Institute Big Data report recognizes the role of policy making in these needed transformations:

> "Policy makers need to recognize the potential of harnessing big data to unleash the next wave of growth in their economies. They need to provide the institutional framework to allow companies to easily create value out of data while protecting the privacy of citizens and providing data security. They also have a significant role to play in helping to mitigate the shortage of talent through education and immigration policy and putting in place technology enablers including infrastructure such as communication networks; accelerating research in selected areas including advanced analytics; and creating an intellectual property framework that encourages innovation. Creative solutions to align incentives may also be necessary, including, for instance, requirements to share certain data to promote the public welfare . "
>
> ***(Manyika et al., 2011, p. 13)***

The McKenzie report also forecasts an enormous shortage of deep analytical talent with the existing labor pool and higher education pipeline of those now being trained in data sciences (Manyika et al., 2011, p. 109). This shortage will impact several aspects of adoption of learning analytics systems:

- **Developing an understanding of and rationale for learning analytics as both an instructional and a policy tool**
- **Building capacity for implementation of learning analytics systems and solutions at the school, district, and state levels**
- **Identifying and developing policies that will support and enable effective learning analytics**
- **Developing funding models to support learning analytics**
- **Conducting learning analytics research in formal K-12 settings**

LAW Project: A Multi-sector Approach to Building a Field

The LAW Project established the Learning Analytics Workgroup, comprised of representatives from multiple sectors and representing disparate fields. The members of the LAW were selected for their exceptional subject matter expertise, for vital contributions to fields constituting the emerging interdisciplinary field of learning analytics, and for their representation and leadership in relation to different societal sectors (academy, nonprofits, industry, government, philanthropy). The topics for which we sought to bring expertise to the project are shown in Table 5.

Table 5. Learning Analytics Workgroup Project Areas of Expertise

EDUCATION AND LEARNING EXPERTISE	OTHER EXPERTISE
Assessment	Computer science
Connected learning	Data privacy policies and technologies
Digital media and learning	Data science
Education research	Law and society
Education data mining	Recommendation systems
Education technology	Social statistics
Education policy	Technology policy and strategy
EdTech entrepreneurship	Machine learning
Learning sciences	Artificial intelligence
Learning analytics	

To build the field of learning analytics that can meet the challenge of personalized learning through cyberlearning infrastructures will require leveraging the talents, skills, and other resources from the following:

- The academy
- Nonprofits
- Industry
- Private foundations
- Governmental agencies

Our work together over this past year in workshops and other activities sought to bring together the strategies and insights from these different fields and sectors to identify key challenges and enablers for building the field of learning analytics towards personalized K-12 learning at scale. In the remainder of this report, the results of these deliberations, conversations and associated research work are represented, before we outline a set of recommendations for warranted next steps.

SECTION 2

Critical Questions for Understanding How to Build the Field of Learning Analytics

What Are the Key Challenges and Enablers?

In our discussions throughout the work of the LAW Project, we have repeatedly returned to the organizing heuristic of characterizing key challenges and key enablers. Given the complexities of tackling personalized learning at scale, and the diverse stakeholders who need to be involved in coordinated efforts to make deliberative progress on the different aspects of these challenges, it is perhaps not surprising that we have developed a sizeable framework for this report. Furthermore, there is an inevitable interconnectedness to the different aspects of the challenges and enablers, so clean separations in the topics below are not possible.

1

First, we consider it vital to ***foreground the challenges of educators*** in relation to the prospects of personalized learning—what do they need, and what may better enable their practices to achieve the personalized learning vision?

2

We recognize that different educational stakeholders will have different ***success metrics for learners***. What outcomes should we care about in the development of personalized learning? Which do we care about that need further research and development for use in personalized learning systems?

3

For personalized learning, a pre-eminent objective is ***creating a model of the learner***. What characteristics are important as predictors for what is appropriate to support the learner's personalized progress? What are the classes of variables and data sources for building a learner model of the knowledge, difficulties, and misconceptions of an individual? How can those models be comprehensible to students to support both their intrapersonal and interpersonal learning goals?

4

A broad set of topics is encompassed in the question of ***how to establish a well-functioning, personalized-learning research infrastructure***. This question about research infrastructure transcends learning analytics and is of concern for any scientific discipline, such as astronomy, environmental sciences, biology, and physics. It involves the end-to-end individual and community workflow of the science, from the planning of research to community-wide issues of data standards and interoperability, to data collection and processing, to data analysis, to using analytics for evidence-based decisions and actions. Technologies are needed throughout the full continuum. And we are hopeful for learning analytics that we can learn from best practices about research infrastructure from other disciplines in order to avoid reinventing the wheel. And because education data science is a human science, we need to exercise great care in our data privacy policies and frameworks and in how informed consent and other provisions of federal laws for protecting human subjects in research are actualized in our domain of science and practice.

5

The transformations of educational systems that personalized learning, when actualized, will bring about have important consequences for ***the preparation and professional development of teachers and educational leaders*** of schools, districts, and states.

We now survey each of these issues in turn before moving forward to consider how to accelerate progress in the following areas:

- Tackling research challenges
- Catalyzing human capital development to supply the diverse sectors of our society who will need education data scientists trained in learning analytics
- Making recommendations about next steps for funders and other educational stakeholders

HOW DO WE FOREGROUND THE CHALLENGES OF EDUCATORS FOR PERSONALIZED LEARNING?

Underlying the possibility of personalized learning is the creation of a set of shared technology services for providing information and tools that teachers and students can use together throughout their students' learning careers. For our report, it is valuable to contribute insights from the teacher voices that have informed the priorities of the InBloom Project (also known as the Shared Learning Collaborative; 2012): 60 focus groups were conducted, and 790 teachers and administrators were interviewed (Colorado, Illinois, Georgia, Massachusetts, New York, and North Carolina) about how their educational practices function today and the challenges they encounter. These data and InBloom's subsequent work with focus groups ranking elements in scenarios informed 10 different categories of scenarios for services or solutions that the marketplace could enable to support their practices and address their challenges.

Nine "opportunity areas" were identified for technology innovation: high-level topics within the world of education that hold the greatest potential to deliver value in improving current offerings that suffer from lack of data integration, incomplete feature sets, or require improvements in user experience.

OPPORTUNITY AREAS

- Training, professional development, and networking
- Communication and collaboration with education stakeholders
- Technology selection, management, and usage
- Budgeting, human resources, and performance management
- Supporting lesson planning, instruction, and assessment
- Creating learning maps to track education progress
- Viewing student profiles and history and managing their transitions
- Course, career, and college planning
- Learning intervention flagging, action guidance, and measurement

Four scenarios were most highly ranked across all districts where research was conducted.

- **Student view of individual learning map.** Learning maps are defined as graphical, data-driven views through a curriculum. They include skills mastered, courses completed, and adaptive paths to mastery of new skills. Students can use learning maps to understand their personal progress compared to local, state, and national standards; plan for next actions; and receive motivation and guidance from teachers, counselors, and parents. A student's view of this map is the most highly-rated scenario by teachers and administrators alike.

- **Interventions, flagging action needed, and measurement topics**. These are important to educators, who want not only to see which interventions work for their students but also help identify students who show early signs of difficulty in their classes.

- **Training and professional development courses are often difficult for educators to locate and make time to take**. This scenario outlines easier ways for educators to find and leverage accredited professional development resources and enables them to use online tools to collaborate with peers on best practices for instruction.

- **"See the whole student."** Student profiles hold the promise of more than just biographical and attendance records. Teachers asked for more comprehensive student information, including intervention history, and for the ability to sort and filter profiles across a variety of data points.

Beyond these highly ranked scenarios, teachers called out these elements from other scenarios as very important:

- Student profile shows history of interventions and results of those interventions
- System can automatically flag students who show early signs of struggle or opportunity
- Teachers can see a student's profile, along with family biographical information
- Teachers can modify lesson plans to suit the needs of their class
- The system automatically records which tasks the student completed and how he or she did

DEFINING SUCCESS METRICS FOR PERSONALIZED LEARNING

While the notion of 'success metrics' for K-12 personalized learning may seem initially straightforward (e.g., mastery of the competencies that are the intent of the instruction or learning resources utilized), there are opportunities to wish for far more in this new world of education data science and personalized learning informed by the sciences of learning. High grades and completed courses that are on the pathway to college readiness are important as success metrics. Retention in school and in challenging subjects is integral if learners are to become college-ready. But increasingly, we are learning that there are other measurable aspects of learning processes that may serve as key drivers of learning and which may also be subject to intervention as a type of outcome.

Among the topics of special attention today are the so-called "**non-cognitive factors**" in learning such as academic persistence/perseverance (aka 'grit'), self-regulation, and engagement or motivation (Dweck, Walton, & Cohen, 2013; Farrington et al., 2012).

Diverse disciplines inform this work, and there is special promise indicated from research that finds that such factors are both malleable and instrumental in helping learners stay on track to educational achievements and course completions (Yeager & Walton, 2011).

Another topic of central interest is the notion of "**mindset**"—how a learner conceives of the nature of mind and intelligence—either as an entity given as a trait or as incremental and improvable by appropriate practices (Dweck, 2007; PERTS, 2014).

As mindset researchers observe, many students have "toxic beliefs" about their abilities or the value of school that depress their motivation or derail their attention and that predict lower academic achievement, resilience with difficulties, and likelihood of cheating. An increasing body of evidence indicates that short interventions can shift the mindset of K-12 learners in these matters, with corresponding reduction in toxic beliefs and their associated negative outcomes.

In our LAW workshops, we also considered other success metrics, each of which has its own rationale for its importance:

Collective intelligence

This is contributing to the learning of groups in which one participates (Woolley et al., 2010). This metric takes on increasing importance as one of the most vetted 21st century competencies (National Research Council, 2013). It soon will be assessed in the 2015 Programme for International Student Assessment (PISA).

Innovation and creativity

In addition to learning academic content, students need to know how to keep learning and applying what they know in innovative and creative ways throughout their life. There is a vital differentiating nature of these characteristics in an increasingly competitive global economy in which jobs that can be routinized and replaced by machines will be (Brynjolfsson & McAfee, 2011; Levy & Murnane 2012). A growing proportion of jobs will require innovative and expert thinking or complex communication and problem solving.

Preparation for future learning

There is a difference between learning what one needs to know to master a subject in school and learning how to learn on the job and to adapt to change. Bransford and Schwartz (1999) proposed a dynamic assessment format to measure just this, which they termed preparation for future learning (PFL). Using PFL assessments, the goal is to tease out whether one approach or another is better preparing students for future learning, rather than simply applying what they learned in instruction during a static test (Schwartz & Arena, 2012).

It is beyond our purpose to survey them here, but we do observe that success metrics for K-12 personalized learning such as these should evolve as the needs for educated competencies evolve with society—as argued in the recent report from the National Research Council (2013) Education for Life and Work.

Creating Learner Models

At the heart of personalized learning is a continuously evolving ***model of the learner***. It is in relation to that learner model (aka "learner profile") that learning activities become ***personalized***; recommendations for learning resources or activities are aligned to that model. Inferences about risks associated with struggling during learning can provide early warning signals that would recommend teacher attention.

- How are such models developed?
- What characteristics are important to build up through either explicit or tacit measures to serve as predictors in personalized learning analytics?
- How important are learner data aggregation and interoperability across digital platforms provided by multiple publishers and vendors in building the learner model?

One can imagine very simple and also increasingly complex models of the learner. To take an example from a different domain, music recommendation engines seek to recommend music choices based on an inferred model of the music listener's preferences, based on either explicitly offered or, more commonly, tacitly developed measures of their choices in a digital music environment. But such a model would not be sensitive to the context for the music listener: What is his or her mood? What other activities is the listener involved in? The appropriate musical choice is likely to follow from a better sensing of the situation in which he or she is listening to the music, and more subtle aspects of a model of the music listener would need to be developed. Yet one can imagine a more sophisticated music recommendation system that includes in its listener model data associated with the mood of the listener at that point in time, or other contextual variables that matter for the desirability of specific music experiences.

> "The increasing use in educational systems of such digital learning environments means that far broader data will be brought to bear in developing models of the learner in the near future. "

Similar issues apply for models of the learner. For many years, school information systems have been representing many details about the backgrounds or histories of learners. These include demographic information, free or reduced-price lunch status, designated categories of special learning needs such as English language learners or specific disabilities, behavioral incidents, home conditions in terms of parents or other caregivers, as well as very thin data for performance in courses (typically a letter grade each semester, rarely anything more fine grained). In digital learning environments minute steps of learners' activities are collected as they work on problems and tasks, and other forms of context sensing are being developed (e.g., videos of teacher–learner interactions and surveys of learners' experiences with their teacher; Kane & Staiger, 2012). The increasing use in educational systems of such digital learning environments means that far broader data will be brought to bear in developing models of the learner in the near future.

"In digital learning environments minute steps of learners' activities are collected as they work on problems and tasks, and other forms of context sensing are being developed."

There is increasing recognition that human learning is a complex, multisensory affair of embodied semiotic interactions and that the production and perception of meaning in context engage the full range of sensory modalities. This is important because many challenges are associated with understanding how learning is occurring within and across formal and informal settings, as learners and educational systems exploit increasingly pervasive mobile learning devices and online educational applications and resources such as massive open online courses (MOOCs), open educational resources, Wikipedia, web searches, digital curricula, games, and simulations. Yet most research on learning in education has minimal sensing of the contexts in which learning processes are being enacted and in which learning outcomes are developed, since classroom studies dominate. A variety of technologies make possible new inquiries for advances on these issues.

In this section, we characterize the variety of learning experiences that may be used to develop a model of the learner. Our characterization is in terms of sources of evidence that are used to build the model. Like any data collected to form a model, there are likely to be issues associated with data quality and the strength of the signal in the data with respect to the inferences that these data are intended to warrant. In other words, the scientific merits of the learner model need to be borne out in its practical value for making predictions and yielding actionable results that improve educational processes and outcomes for that learner. Consider the sources of evidence that span a range from census-type data to digital learning data to biometric data, shown in Table 6.

Table 6. Sources of Evidence to Build a Learner Model

SOURCE OF EVIDENCE	DESCRIPTION	EXAMPLE
Interaction with educational resources	Metrics for interaction during learning activities that include navigation behaviors, answers provided to exercises and quizzes, types of errors made, and timing data associated with student performances during learning activities	Use of a learning aid within a software program (e.g., calculator or virtual manipulative for mathematics)
Social metrics	Metric for how the learner interacts with other learners and the teacher during learning activities, or with speech that is recorded (with its various properties such as semantic content, prosody, etc.)	Student use of text chat or forum posts
Mindset	Survey or self-report data concerning a student's mindset about the relationships between his or her strategic effort during learning and the development of competencies, and as a function of domain of learning.	Student report of relationship between effort and ability in mathematics
Past performance	Historical indicators from a learner's past performances that represent the attainment of concepts, skills, or competencies to date	Prior achievement test performance
Preferences for learning media or genre	Historical indicators about the learner's preference for learning media or genre, where choices were made available	Student preference for visual aids versus narrative text when given content selection options
Perseverance or persistence	Historical indicators about the learner's perseverance or persistence with learning activities when experiencing challenges as indexed by errors and timing data	Time spent completing activities in content where errors are made
Administrative data	Distal context indicators about the teacher, school, district, community, or state based on administrative data	The school and teacher associated with a student
Demographic information	Distal context indicators providing demographic information about the learner	Gender, age, race, ethnicity, language background, family characteristics (e.g., single parent)
Temporal history	Proximal context indicators representing the temporal history of a learner's actions for which data are available on a given day	Time of day
Emotional state	Proximal indicators that relate to learning, such as emotional state or recent sleep and nutrition status	Facial expressions detected by a computer webcam while learning
Social network	Proximal context indicator such as social relational and social network data	Participation in a chat session with one or more other learners
Classroom disruptions	Proximal context and distal indicators for classroom disruptions from records concerning behavioral incidents reported in the learner's classroom on a given day and over time	Behavior incident report

In our discussions of building the learner model, we made several important additional observations:

» *Model data quality*

Without reliable, valid, efficient, and fair measures collected from multiple sources, and analyzed by trained researchers applying methods and techniques appropriately, the entire value of a research study or a program evaluation is questionable, even with otherwise rigorous research designs and large sample sizes.

» *Expand data types and model complexity*

Previous research has tended to focus on ease of data collection and easier-to-build models, versus weighing the value of hard-to-collect data used for creating hard-to-build models. For example, a great majority of research articles in education data mining come from a small set of research questions, such as predicting course dropout rates and knowledge modeling—things that are relatively easy to think about. It is harder to think about changes in English proficiency in an English language learner contributing to learning outcomes, or differences in biological responses to stressful situations in school, among other significant data and modeling issues.

» *Make learner models transparent*

Currently, the concept of learner model tends to be restricted to a model of the learner that is built up by a provider of educational products or services and employed opaquely behind the experiences of the learner, teacher, or other educational stakeholder, without the ability to be inspected or modified by an individual. We should critically consider the potential importance of making the learner model transparent to these stakeholders. In the case of credit scores for adults, one has the authority to both ask for the information that is used to derive the score and to challenge faulty data that may be associated with the score.

A range of important topics is therefore worth investigating concerning learner model transparency. They include the following:

- ***Disclose learner model properties to the learner, parent, or other stakeholder.*** The challenges here include representing the state of a model of a learner in terms of purported knowledge state, competency levels, topical interests, mindset, and other properties in ways that are understandable by those agents to whom it is revealed. For some characteristics of the learner model, the learner may have elected to make a "badge" for acquiring a competency public in a digital medium such as Facebook or other social media website. But many others may not wish for it to be made open or may not understand the potential consequences if it is made public. We also must consider the risks associated with second-order, unintended effects of capture and use of learner model information. For example, stereotyping risks are associated with labeling learners by categories, such as attention-deficit disorder. This concern interacts significantly with the intertwined nature of learning new competencies (learning how and learning that); the development of personal identity (learning to be); and how identities are socially constructed, managed by learners, and may constrain or advance learning. This is because stereotyping and positioning opportunities for learners open and close as a function of how others view them and how they view their possible future selves.

- ***Deal with errors or faulty profile characteristics in learner models.*** As in the case of demographic data that could be falsely represented in school records, there may well be errors in learner models. How would a provider respond to the challenge of the veracity of the models if the complaint came from a learner, teacher, or parent? What if a learner were given the opportunity to engage in new learning too soon as a result of an inaccurate estimation of knowledge? At a minimum, there should be some prospect of an independent audit of the quality of the data upon which inferences are made that end up in the learner models constructed by the provider of an educational product or service.

As we consider research and development activities concerning how to build powerful learner models that enable improved educational processes and outcomes fundamental to education data science, we must weigh the benefits against the risks.

For example, data privacy issues that are raised in collecting many of these data types are a real issue that will be discussed in the following research infrastructure section. Although some commercial companies have developed learner modeling technologies for personalizing learning and associated learning analytics (e.g., Knewton and DreamBox), given the proprietary nature of the algorithms and associated research that is being conducted for product development and improvements, research using these platforms is not available for review to understand lessons that might inform the broader field of education data science.

How important are ***learner data aggregation and interoperability*** across digital platforms provided by multiple publishers and vendors in building the learner model? In digital learning environments, learners are likely to be engaged with multiple educational technologies simultaneously—e-texts, learning games, a learning curriculum from a commercial publisher, and open educational resources. How will education data science efforts effectively investigate progress for individual learners across this panoply of experiences? Only instrumenting what a student did within a specific course or learning application, and not across other learning environments that the student is engaged in, creates data islands, and the connected data vision of personalized learning is left unfulfilled. Another issue associated with the user experiences for teachers and learners of multiple learning environments is the challenge associated with multiple sign-ons versus a single sign-on.

An issue which foregrounds other challenges is the goal of ***creating connected learning models***. How can we connect learning that is networked across and within the formal and informal learning and educational activities in which any learner participates (see Figure 2)?

Figure 2. Formal and Informal Learning across the Lifespan

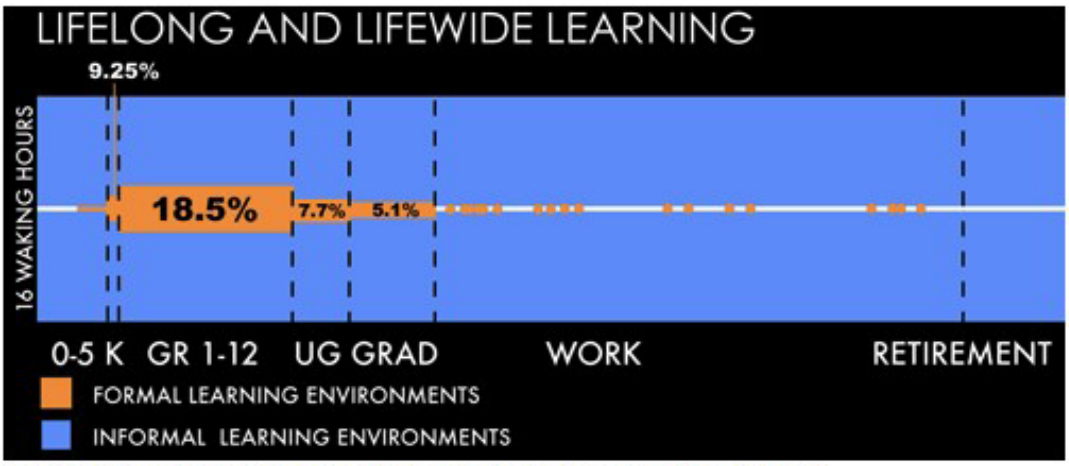


Figure: The LIFE Center Lifelong and Lifewide Learning Diagram

Many scenarios associated with current research and development on personalized learning are restricted to school-based digital learning and do not encompass the full spectrum of learner experiences in their connected lives. How might we bring learning data from learning outside traditional channels into a student's learning model—such as interests pursued in learning that is mediated by online community websites or learning games or pursued in informal settings such as science museums? It is noteworthy as we raise this prospect to observe the very different challenges associated with expanding the bounds of data collected to out of school settings and the special data privacy needs that surface in this context, including family privacy in addition to child privacy. What are the best practices and tools to advance this vision (e.g., Open Badging), and what technology and policy affordances and constraints will need to be considered?

Uses of Learner Models for Recommendations

A major use of the learner models is for either recommending or providing learning experiences with resources or activities that ***are predicted to be appropriate for the learner to make educational progress***. Personalized learning is characterized by incremental levels of challenge, increased growth of understanding and expertise, and ongoing opportunities for success customized for every learner. Visions of personalized learning invoke the utility of recommendation engines for learning resources that are adaptive to learner models or profiles (data stores of preferences, performances, demographic characteristics, etc.).

Recommendations include many different dimensions for focus. At issue is the fundamental question: *What's the next thing that a learner should do? What shape should that experience take?* The focus of recommended or provisioned learner experience could include the following:

- What next hint should be offered for enabling a learner to achieve success with a problem (scaffolding)?
- What learning resources should be used, or what choices of learning resources should be provided, and what challenge should be attempted (next steps)?
- What instructional pace should be sustained (pace)?
- What pedagogical intervention is needed by the teacher next (teaching)?
- What student learning groups should be formed and why (social learning)?

"Socializing" personalized learning is not a contradiction in terms. This is both about modeling the learner in terms of social relations and about making recommendations for personalized learning that include social, collaborative learning activities. Personalization is often described as individual learning, but for teachers forming learning groups, what learning data support could guide peer assignment to more effective learning groups? How can students engaged in peer-to-peer learning be matched up as co-learning associates?

Establishing a Functional Personalized-Learning Research Infrastructure

In order for learning analytics to reach its potential as a field significant efforts must be made to establish a shared research infrastructure. In this section we provide an overview of the needed data sharing, analysis and visualization tools, collaboration practices, data-management policies, and IRB reforms that will enable development of learning analytics as a field and implementation of personalized learning at scale. What are the components of this research infrastructure that need to be designed, funded, created from scratch, or adapted from available technologies and existing systems.

Making research questions in learning analytics answerable is in part a matter of developing a data sharing infrastructure.

Platforms like Globus—a big data software infrastructure for distributed computing by researchers—provide a model for a possible solution to the problem of data sharing in learning analytics research. What kind of data sharing infrastructure is needed in order for researchers to pose and answer the questions of learning analytics? Meeting the needs of educational researchers entails answering questions about how to integrate a data sharing infrastructure with existing school policies and infrastructures. For example, what are the key elements of that infrastructure for K-12 digital curricula and assessments, which would allow for rapid data collection, sharing, and iteration of curricular designs? What productive mechanisms would enable big data, the learning sciences, educators, and industry communities to productively coordinate their efforts to improve learning both in-school and out-of-school?

Data analysis and visualization tools are of vital importance to any big data infrastructure.

Systems like Hadoop, R, and other software solutions have become prominent in the data science and business intelligence communities, with the DataShop for education data mining engaging some educational researchers. Which components are already available and can be leveraged from existing software systems to be employed in learning analytics? Which software packages and particular libraries are used to address the kinds of questions that learning analytics researchers will pose? For example, Python's ggplot2 library, the iPython Notebook, and NumPy all have affordances for researchers working with big data, which could be part of a shared language for learning analytics researchers. Or perhaps R's many graphing and visualization libraries will better serve these researchers. Regardless, establishing at least a general collection of accepted and standard tools, if not a specific set of software applications and libraries, will help researchers in learning analytics better communicate with each other

about their work, allowing for better data sharing and more rapid progress. However, one challenge to this idea is that these tools are rapidly evolving, which makes it imperative that researchers be able to adapt to these changes.

For any data-sharing infrastructure for learning analytics to be functional, it will need to incorporate visualization and data report systems like learning dashboards, which will be accessible to decision makers, including teachers, learners, administrators, and policy makers.

An important part of developing a shared-data infrastructure is ensuring that data sharing incorporates the range of stakeholders in the educational ecosystem. What kinds of design issues—such as data and visualization literacies—must be addressed in creating interactive data visualization systems that effectively provide feedback over different time scales and suggest next-step relevant resources and choices? How can such systems provide sufficient and actionable data while avoiding data overload? How should dashboard use by learners, teachers, and parents be related to use of recommendation engines for these same audiences?

Any shared-data infrastructure relies in large part upon data quality and interoperability.

The learning analytics community will need to adopt a shared understanding of what constitutes high-quality data, including what kinds of metadata ought to be present for any given dataset. Data scientists of all kinds need to know what they are working with when doing their analysis, so ensuring that data quality standards are agreed upon and met is essential to shared-data infrastructures. As datasets increasingly become connected, as in learner models, more metadata need to represent the limitations and assumptions associated with those data, such as collection issues, non-representativeness, measurement error sources, or other data biases. In an age of big data, having the right data is more important than having the most data. An important element of the right data is that data must be of high quality and able to address the research question, while having transparent limitations and biases. Otherwise learner models and other results of learning analytics research will not be trustworthy. Any shared data infrastructure in learning analytics will need to address key questions about measurement and trust. How do we get enough high-quality data? Can we trust models that have been developed? How do we know that we are measuring correctly? Are we measuring the right things?

Data-management policies of IRBs, school permissions, and data sharing agreements are a key part of developing a shared-data infrastructure.

The IRB is an essential part of academic research on human subjects, but its current policies were not developed with current data-rich digital environments in mind. It is not that IRBs cannot handle new research paradigms; rather, the policies and standards that would allow them to make informed decisions about the risks of research in learning analytics simply do not currently exist. It will be important to consider what kinds of benefit and risk frameworks will be adequate to protect the privacy and equal-opportunity rights of learners in digital learning

systems when researchers engage in learning analytics inquiries.

An important step will be to develop procedures for consultation with experts on data protection and security.

Separating data protection recommendations from the IRB evaluation process will ensure that the IRB can focus on its primary job of protecting human subjects from harm, while a more specialized and standardized process can be used to evaluate data issues. Most IRBs do not have a privacy expert, and so the perceived risk of data-intensive research is easy to misestimate. Whereas data issues are a part of potential harm for human subjects, ensuring that this aspect of review is handled by data experts is the only way to ensure that it is done rigorously across institutions. Similarly, standardizing consent language for data-intensive projects will expedite and simplify the IRB's review process.

Standardization of data privacy and information technology security systems also would streamline the review process, as researchers would know exactly what kinds of safeguards need to be in place before submitting a research proposal.

Institutions need to be aware that data security is a changing landscape, which means that standards will need to be revisited and revised.

The speed with which this landscape is changing is also a major reason for the need for standards. Data analysis and collection can now be done at a scale that was unthinkable only a decade ago. In fact, computer and Internet users provide web companies a range of sensitive data. Researchers working in digital environments with large data sets and data exhaust produced by interaction with web-based portals need to differentiate between appropriate and inappropriate use of data. IRBs need to have standard guidance for these issues.

In building a field with a shared research infrastructure, developing standards for privacy and security, informed consent, anonymity, and data sharing is a project not only for scholars and industry partners but also for review boards.

It is important that data sharing be done with a conscience and in a way that is ethical and respectful of student rights. However, it is equally important that such data sharing be allowed to occur under the guidance of collaboratively developed and respected standards, or else research in learning analytics will be substantively curtailed.

Professional Development of Teachers and Educational Leaders

We now survey the implications of the development of learning analytics for teacher education programs and the training of educational leaders broadly. Data literacy will become an important skill for teachers, as making data-enhanced decisions in the classroom will depend upon the ability of a teacher to quickly make sense of data visualizations presented in learning dashboards. Teachers will not need to be data analysts but will need to be trained in interpreting the visual presentation of data from their classrooms in a way that will effectively inform their instructional decisions. Far from being eliminated, teacher agency will be vital in a personalized-learning-enabled educational system. Teachers can make nuanced judgments about learners based on many dimensions of information that are not going to be represented in the data of personalized learning systems. A key skill of the teacher in this setting will be synthesizing his or her personal understanding of the classroom context with the data presented through the learning dashboard.

"Teachers can make nuanced judgments about learners based on many dimensions of information that are not going to be represented in the data of personalized learning systems."

Teachers will likely better motivate students to engage in learning activities than recommendation systems. A teacher can engage a student in a dialogue around next steps in the student's learning in a way that even the most advanced recommendation systems cannot. Teachers thus can help learners choose from available options and can inspire them to take their efforts seriously. Without the teacher as intermediary, it is not clear that learners will be able to benefit as intended from personalized learning systems.

Although this section was filled with a multitude of questions, we believe it was important to document from a variety of sources the questions related to the key challenges and enablers. After reviewing these questions, we considered what the priorities should be for future research in learning analytics. At the same time we realized that new tools, approaches, and policies will be needed to address these questions. The next section outlines our recommendations in these areas.

Revisions to teacher training programs to reflect the new skills and roles of the teacher in a personalized learning system nevertheless raise a series of questions that will be important to consider. How will decision-making roles be distributed across stakeholders in a learner's education or for an educational institution? Who makes which decisions with access to which data and when? For example, who will be responsible for setting up A/B experiments in digital learning environments and measuring results for tracking progress and guiding improvements? Here, answers could range from technology companies to district or school leaders to individual teachers (though at a small enough scale, there may be too few students for such experiments to produce meaningful results). Further questions concerning a range of stakeholders are posed below.

- **Will students in a personalized learning system make choices, or will the system make choices for them?**
- **How should the role of the teacher and personalized learning technologies be distributed? What data should teachers have access to and what kinds of decisions should they be able to make? What data should learning systems employ? What kinds of recommendations should those systems be allowed to make directly to the student, and which should be mediated by the teacher?**
- **To what degree should school leaders appraise teacher effectiveness and assign professional development based on data from student personalized learning systems?**
- **How do we know if results from one school, district, or state will generalize to other schools, districts, or states? This question is important at the level of individual student learning, as recommendation systems may not be transferable across institutional boundaries, and also at the level of systematic policies, as teachers and administrators in one setting may have very different institutional needs than their counterparts in another setting.**

SECTION 3

Articulating and Prioritizing New Tools, Approaches, Policies, Markets, and Programs Within the Field of Learning Analytics

Priorities for Research

There are several priority areas on which early research in learning analytics should focus. Overall, the motivating question for the field is how to develop personalized learning systems. For which learners does a learning intervention work or not, under what conditions, and why? Emphases within this broad question will include reducing achievement gaps for equal opportunity, fostering a diversity of curricular and pedagogical approaches to meet the needs of diverse learners, developing clear metrics for effectiveness, increasing efficiency by lowering cost and time of learning interventions, reducing dropout rates, and increasing learner interest and persistence in challenging and complex subjects.

To further explore these research priorities, we will present three "grand challenges" for research in learning analytics. We see these grand challenges as areas where early success could demonstrate the value of education data sciences. These challenges could be supported by competitions to create predictive learner models that get the greatest percentage of learners to competency in the shortest time at the lowest cost. Of particular value will be interventions and content that focus on "gating concepts," material with which many learners struggle, which prevents them from further advancing in a topic (for example, ratio in mathematics and energy transformation in the sciences). Research, however, should not focus solely on content but also on important so-called "non-cognitive outcomes" and developmental milestones. Learning analytics systems presumably will allow researchers and educators to identify early warning indicators when learners struggle with key developmental phases like pre-algebraic thinking prior to their enrollment in early algebra classes.

GRAND CHALLENGE 1

Learning progressions and the Common Core

How can learning analytics help refine our understanding and practices involving learning progressions in digital learning environments for Common Core State Standards in mathematics and language arts and the Next Generation Science Standards? Researchers could mine one or more of the strands in the standards, mapping knowledge components with large education datasets with robust instruments that treat standards as an initial assertion and then test whether these competencies are correctly described as Knowledge Components (as tracked in various Pittsburgh Science of Learning Center projects; see, for example, Ritter, Anderson, et al., 2007 and Pavlik, Cen, and Koedinger, 2009). It is useful to start with the Common Core, both to ensure that the learning progressions suggested therein are valid and to provide alternative assessment systems to purely content- and outcome-based tests that are currently prominent.

> "How can learning analytics help refine our understanding and practices involving learning progressions in digital learning environments for Common Core State Standards in mathematics and language arts and the Next Generation Science Standards?"

Note that there are particular challenges in English and language arts standards as the core competencies for learners repeat themselves at various grades while text complexity increases. Drawing on existing research on literacy (Wolf & Katzir-Cohen, 2001), researchers could address these issues by appropriating existing theoretical and conceptual frameworks, which can be applied at each grade band and then instrumented in analytics systems. Such work would necessitate multimodal inputs for analytics systems.

It seems likely that a learning registry or associated indices, such as the Learning Registry Index (www.learningregistry.org) will have a role to play as a clearinghouse for assertions related to alignment or fit of curricula to Common Core standards.

GRAND CHALLENGE 2

Standards-based assessments for digital learning

How can we systematize the mapping of standards onto a bank of formal and informal assessment questions, with the goal of assessing content mastery and making recommendations for teacher practice in response to evaluation of learners' competencies? What type of valid, reliable, and engaging assessments should we use to capture core competencies? Can these be indexed to existing learning registries, to help ensure fit with standards? What kinds of tools will teachers need in order to create assessments that follow these strategies, or else to select effectively from available assessments in a way that meets the needs of their particular classrooms?

Possibilities for assessment are not limited to ex post facto exams. Assessments can be used to direct instructional practices in a formative way. Understanding student choices in a learning platform or a game may help uncover and represent misconceptions. Choice-based formative assessments can provide a significant amount of data from a limited number of questions, and when embedded in learning games they also help engage students. It also should be possible to assess aspects of student learning strategies and dispositions like motivation and perseverance. When capturing data about student learning processes for assessment development, one should keep in mind both cognitive and metacognitive processes that manifest in student choice patterns. It should be noted that ensuring an appropriate level of challenge for a given learner will help reveal more about that learner's knowledge state and competency with the material, as well as that learner's disposition, attitude, and strategies.

GRAND CHALLENGE 3

Creating multimodal learning analytics

Traditional assessments tend to focus on a limited set of data generated by student activity on isolated, high-stakes exams. Expanding education data to capture contextual features of learning environments will allow assessment to focus not only on student demonstrations of knowledge on pre-designed assessment tasks, but also to capture aspects of learners interacting with each other and their learning environment. As networked learning technologies become pervasive, the possibilities of data collection to enhance learning for all opens up substantial and significant new opportunities for learning analytics which move beyond students' solo interactions with computer software and websites to include contextual data on learning environments. These contextual data sources include gesture, speech, spatial position, affect, and other variables that can be gleaned from video records and other sensors like eye trackers in learning environments.

What are the priority issues in providing better theories, methods, and tools required to conceptualize, instrument, and temporally coordinate the multiple data streams available during technology-enhanced learning by individuals and groups beyond traditional clickstream data, such as discourse, gesture, and emotional expression during classroom and group interactions? How can data privacy and anonymity be best achieved in relation to multimodal learning analytics?

A particular benefit of solving data collection and analysis problems in multimodal learning analytics is the potential to expand educational opportunities that are currently "boutique" methods such as robotics classes and "maker labs". These kinds of opportunities are difficult to assess using traditional methods. New data streams from multimodal analytics could provide evidence to make better warrants about learning achieved to encourage broader use of these kinds of instructional settings in education.

"Expanding education data to capture contextual features of learning environments will allow assessment to focus not only on student demonstrations of knowledge on pre-designed assessment tasks, but also to capture aspects of learners interacting with each other and their learning environment."

Investment in data sources like video, eye tracking, and skin temperature and conductivity is likely impractical at a large scale today. Yet, implemented at a small scale in parallel with easier-to-collect big data sources like clickstreams, it may be possible to develop points of contact that allow for better inferences from traditional data sources. This raises an important question for multimodal analytics. At what kind of scale should we gather multimodal data? A lot of data about a few students can still be "big data," depending on the granularity and precision of the instruments that are being used.

There are further orienting questions for this challenge. What are the costs and benefits of developing more expansive single-sign-on learning systems? What would the cost socially and economically be of developing a single-sign-on technical infrastructure and culture, and would that cost be worth the payoff, considering the ability to integrate data from across a range of traditional platforms with multimodal data sources? How might we support multimodal open access sharing frameworks that are flexible enough to change with technology and developments in data science methods? How might we ensure support for research into questions that are not necessarily the obvious ones to ask of big data, but that multimodal analytics make possible? How do we keep in mind the bigger picture and fundamental questions about learning, environments, and learner contexts given the prevalence of highly data-driven mentalities and methodologies in the data science community, not because availability-driven inquiries are bad to pursue, but because we don't want to abandon more difficult questions that could lead to new analysis and collection methods? How can we define questions that are sufficiently broad to admit multimodal data but sufficiently specialized to belong to learning analytics? In the next section we discuss the resources that will be needed to address the priorities outlined in this section.

"What would the cost socially and economically be of developing a single-sign-on technical infrastructure and culture, and would that cost be worth the payoff, considering the ability to integrate data from across a range of traditional platforms with multimodal data sources? "

SECTION 4

Determining Resources Needed to Address Priorities

Training Programs and Field Building

Technology has run ahead of the readiness and human capital in the emerging field of learning analytics. Demand is ahead of supply and will continue to be without a systematic effort at capacity building in the form of training programs and field building. Where will learning analytics and education data science specialists come from? What does a specialist in this field need to know and be able to do? Which current roles in academic and professional settings hold the potential education data scientists, and what do the people in those roles need to learn? Through what career pathways might education data scientists be developed? What programs of study need to be developed, and what associated resources will complement these programs? Mutually beneficial partnerships with industry should inform both professional data science training in learning analytics and prospective education data science applications and platforms for fundamental research and learning improvements.

We identify several competencies for education data science and learning analytics, based on the ongoing work in this emerging field by current scholars and industry experts:

- Computational and statistical tools and inquiry methods, including traditional statistics skills like multiple regression as well as newer techniques like machine learning, network analysis, natural language processing, and agent-based modeling
- General educational, cognitive science, and sociocultural principles in the sciences of learning, including specific educational content and awareness of key issues and debates around science, technology, engineering, and math (STEM); literacy; English language learners; and cultural, ethnic, and gender diversity in learning
- An ability not only to perform data analysis but also to recognize and evaluate data quality
- Principles of human–computer interaction, user experience design, and design-based research
- An appreciation for the ethical and social concerns and questions around big data, for both formal educational settings and non-school learning environments

We now survey the range of training programs that will enable learning analytics to grow as a field. These include programs for current faculty as well as training in both professional and research settings for future students.

FACULTY CROSS-TRAINING

Bringing current education faculty—especially those who study psychometrics and educational measurement—into learning analytics is an important goal. These scholars have significant expertise in many of the important areas of learning analytics. What kinds of training and recruitment efforts need to be made to encourage their participation in the field? Similarly, computer science, human–computer interaction, and statistics researchers are already contributing to education data-mining societies and journals. How can we further support their acquisition of key insights from the learning sciences? Not only faculty from computer science, statistics, and education but also faculty from a range of fields might contribute to new research on learning analytics. What do faculty from fields like bioinformatics or digital media studies need to learn and know in order to contribute to learning analytics, and what expertise do scholars in these areas already have which will be useful to the field? In recruiting existing faculty to the field, it will be important to establish opportunities for interested scholars to learn from more experienced learning analytics researchers in the form of, for example, summer institutes similar to the "Brain Camp" model used by the early cognitive neurosciences. Furthermore, such faculty should be encouraged to share course materials and jointly create courses, especially courses that reflect the core competencies of the field. Can such curriculum development efforts be combined with summer institutes?

Brain Camp—The Beginning of Cognitive Neuroscience as a Field

According to a 2012 interview with Michael S. Gazzaniga who was at Dartmouth at the time of the first Brain Camps in the late 1980's, "From the beginning, what we were trying to do was to gather together 70 of the brightest kids from around the world to come together and let them see how this field is going to work, and what the topics are going to be . . . to let them participate in the making of the field as well" (Bardin, 2012).

The full Chronicle of Higher Education article is available at http://chronicle.com/article/What-They-Built-at-Brain-Camp/134016

POSTDOCTORAL CROSS-TRAINING

What kinds of grant sources and partnerships will need to be created in order to encourage recent graduates from a variety of experiential foundations to encompass analytics techniques and questions into their future research? Graduates from computer science, data science, learning and educational sciences, computational statistics, computational linguistics, and others are all potential fits for learning analytics postdoctoral training. How can we best support linking young researchers with these interests with experienced faculty?

DEGREE AND CERTIFICATE OPTIONS

A range of certification options will need to be developed, including full degree programs at a variety of educational levels, certification programs, summer institutes, and courses (both traditional and online, specialized seminars and survey courses). Work has begun developing these courses, both within various institutions and at a broader scale. The Society for Learning Analytics runs an annual MOOC, and recently Coursera completed the first run of Ryan Baker's Education Data Mining MOOC. More development needs to be done to incorporate authentic research experience and apprenticeship into ongoing curricular developments. Further discussion of specific degree and certification options follows.

Doctorate degrees and training grants

The field would benefit from ***co-designed degrees*** offered in new programs across departments and schools. Schools of education will need to link with other fields that are data intensive and already have more coursework in place for the preparation of data scientists. Institutions should consider the development of modules, specializations, and certificates that can be elected or required for all students in doctoral programs in the learning sciences, economics or policy-based educational research, and other disciplinary research in education. An important question here is how to integrate industry concerns and opportunities into doctoral training. Will industrial datasets be available to doctoral students? What incentives will need to be put in place so that industry will create internships and affiliate programs for students?

If universities are going to help meet the need for greater numbers of education data scientists, how can they foster the creation of the kinds of ***interdisciplinary programs***, centers, and institutes that bridge computational statistics, computer science, and the learning sciences? Universities must be aware of existing opportunities and competitions that could help spur development of programs. They also will need to leverage existing resources without getting locked into the existing concerns of, for example, psychometrics, statistics, or machine learning. Resources exist in these fields, but learning analytics cannot merely recreate or reinvent prior research focused on a subset of its own interests (such as computational methods). Nevertheless, learning by analogy from existing fields will be beneficial, as data-intensive fields like bioinformatics, fluid

mechanics, genomics, and the geosciences are themselves relatively newly developed data sciences. Sketching their development out of their core fields will provide insight into how learning analytics can grow out of educational research.

How might university structures ***integrate with state and district-level educational systems*** to create feedback loops at a local or cluster level? Universities could work with collaboratives like the State Educational Technology Directors Association to direct progressive PhD candidates into universities and to broker match-ups that forward research both at the university and in those collaboratives. This kind of effort would be part of a broader learning-by-doing infrastructure whereby students and faculty can work on industry or school data analytics projects, including design, implementation, and evaluation of A/B tests, as well as secondary analysis and data mining with existing datasets.

Finally, what kinds of ***competency tracks*** will doctoral programs need to create? Can the multifaceted knowledge and skills of learning analytics be represented at various levels of competency, from mastery to familiarity? Which skills will be essential to all education data scientists, and which will be the particular expertise of a small subset of the field? How can the training of researchers in these various competencies reflect the demand for various skills and subskills in industry, school districts, states, and other organizations?

Master's degrees

What is achievable and desirable in a 1- or 2-year program? What kinds of professional skills can be developed without focusing on academic research as in a doctoral program? It may make sense to house master's degrees in learning analytics as joint offerings across statistics or computer science and education. Although no formal PhD programs have been developed in education data science, several institutions have explored professional training in the field at the master's level. The following institutions have implemented or are planning to implement master's degrees in education data science: Athabasca University, Carnegie Mellon, Worcester Polytechnic Institute, Stanford University, and Columbia University Teachers College.

Certificate programs

These programs will target industry professionals and others with education data science savvy. Key issues include the understanding the backgrounds of experts in data science and big data analytics, in order to expand their expertise to include core ideas from the learning sciences and learning analytics. Asking the right questions is a key ability in all data science, and doing so in an educational context requires an understanding of the unique issues and complexities of educational research. Are there specific certifications that industry would value for emerging job categories and that would be incentivized with fellowships? Furthermore, should certificate programs also be developed to help prepare professionals savvy in education data science to help support and develop infrastructure, teachers and teacher training programs, curriculum, special education, and assessment design?

Undergraduate programs

How do we attract a new generation of scholars and builders? Various strategies present themselves, including new majors and minors, as well as design competitions, research experiences for undergraduates, and specializations within existing programs. Whereas building the field will focus primarily on the kinds of advanced training that postgraduate studies can provide, the habits of mind and orientation towards inquiry that are essential to data science should be cultivated in programs that will attract high-potential students to enter the field.

KNOWLEDGE NETWORKING AND ONLINE COMMUNITIES

In addition to training and degree programs, building the field of learning analytics will require knowledge networking and online community building encompassing both training programs and industry professionals.

Recognizing and developing indicators of quality and establishing reputations for courses and programs will help establish a trusting relationship between stakeholders in learning analytics. Building a community around key grand challenges and questions will require a tactical effort to model collaborative practices, which encompass a range of professionals, researchers, and graduate students. Furthermore, industry can provide datasets for professional training. Several examples exist, ranging from the Pittsburgh Science of Learning Center's DataShop to Khan Academy's datasets, which are being studied under a Gates Foundation competition. Educational game companies with significant user-bases like Motion Math provide an opportunity to study informal learning and could offer a shared dataset for a community-building grand challenge.

Identifying potential industry partners from among the variety of companies that do work in education will be important.

Companies ranging from content publishers like Pearson to assessment developers like the Educational Testing Service to more focused learning-analytics startups like Junyo are all potential matches. LearnLab's Corporate Partners is a relevant resource for finding and researching interested parties. Such partnerships could involve sharing of data; funding for training programs and trainees; and data fellowships or internships, in which individual students or teams work closely with an organization over time to analyze a data corpus as part of their degree program.

Finally, another important component to field building will be funders.

How can funders best foster new educational programs and professions from PhD and master's programs in learning analytics and education data science? A potential model is the Institute of Education Science's competitive support of training programs. Direct funding of students through training fellowships is also a possibility.

Data Privacy and Information Protection

As highlighted in the 2010 National Education Technology Plan, there is also an urgent need to examine and bring into a contemporary era "the practices, policies and regulations to ensure privacy and information protection while enabling a model of assessment that includes ongoing gathering and sharing of data on student learning for continuous improvement" (U.S. Department of Education, 2010, p. xvii). As part of these changes, the 2010 National Education Technology Plan recommends the following:

> "Every parent of a student under 18 and every student 18 or over should have the right to access the student's own assessment data in the form of an electronic learning record that the student can take with them throughout his or her educational career. At the same time, appropriate safeguards, including stripping records of identifiable information and aggregating data across students, classrooms, and schools, should be used to make it possible to supply education data derived from student records to other legitimate users without compromising student privacy. "
>
> ***(U.S. Department of Education, 2010, pp. xvii–xviii)***

These topics have come to attract substantial attention by LAW participants, and several white papers for the project address them and make associated recommendations.

Funding Recommendations for Building the Field of Learning Analytics

What can be achieved in different periods within the next 5 years for building the field of learning analytics? Successfully building the field of learning analytics and education data science in the long term, will result in personalized learning for all students that regularly access large-scale digital learning and teaching platforms and the commercial and noncommercial providers of educational services and solutions that leverage its infrastructure. Towards this end, we will break our recommendations into the pre- and post-18-month time periods. We will also discuss research projects that should take place in the next 1–3 years, as well as 5-year research centers. These activities are outlined in Tables 7–10.

Table 7. Recommended Activities for the Next 18 Months

TOPIC	DESCRIPTION
Data standards	Develop standards for data ownership, privacy, sharing, and access.
Branding and communication	Brand the field and communicate through sponsored talks at key conferences and messaging around privacy protections.
Funding fellowships and internships	Prepare a next generation of education data scientists by funding the creation of predoc, postdoc, industry, and faculty fellowships and internships.
Review of existing resources and projects	Conduct a deep review of evidence and promising resources, products, and projects in order to maximize efficiency of new efforts.
Competitions and prizes	Convene a team to identify the top 5–10 grand challenges to be solved by education data science (with the goal of seeding future data competitions based on the recommendations of the team).
Pilot personalized learning management	Pilot a personalized learning management recommendation and reporting system with a key group of schools.
Researcher and EdTech startup connector	Identify educational technology startups that are willing to work with academic researchers who can collaborate in the context of an Imagine K-12 startup incubator.
Field-building event	Continue to hold a field-building event (e.g., similar to LASI-2013) and related activities with funding from foundations, government agencies, and industry partners.

TOPIC	DESCRIPTION
Teacher professional development	Identify and recruit regional hubs in which multi-sector groups of personalized learning system providers, teacher educators, teachers, and education data scientistists are willing to work together to more effectively use data from technology-enhanced learning and teaching to support teachers' decision-making processes. Identify seed funding resources to support workshops for fostering the development of such partnerships, whose work can coalesce for informing teacher education programs in a next stage of field development.

Table 8. Recommended Activities Beyond 18 Months

TOPIC	DESCRIPTION
Data ownership, privacy, sharing, and access	Collaborate with the US Department of Health & Human Services as the appropriate governance body for learning data issues.
Branding and communicating	Brand and communicate by working with Strata/O'Reilly to develop a StrataEdu.
Competitive awards for data science programs	Prepare the next generation of education data scientists through competitive awards for establishing university PHD programs in education data science.
Review research-based guide to resources	Conduct a deep review of evidence and promising resources, products and projects in order to maximize efficiency of new efforts.
Competitions	Hold competitions to tackle grand challenges based on prior recommendations.
Personalized Learning Management Recommendation and Reporting system	Establish a prototype of a personalized learning management recommendation and reporting system. Conduct iterative refinement of the system based on user feedback, teacher interviews, and feature requests. Research funding for projects to discover, validate, and bring to scale best practices and how best to represent these data to be useful for various stakeholders.
Researcher and ed tech startup connector	In addition to continuing the incubator program, disseminate success stories and create a social networking engine.
Field-building event	Continue to hold field-building events (e.g., similar to LASI-2013), and related activities such as Hackathons with funding from foundations, government agencies, and industry partners.
Teacher professional development	A network is developing of multi-sector hubs which establish data-driven improvement processes in associated educational systems that employ technology-enhanced teaching and personalized learning guided by education data science. Best practices and known challenges in data-guided educational decision-making by teachers become documented for pilot integration into teacher education programs.

Table 9. Research Projects, 1–3 Years

TOPIC	DESCRIPTION
State and district case studies	Conduct case studies of states and districts making significant progress with learning analytics for personalized learning. Analyze state and district examples to specify policies that enable learning analytics.
Toolkit	Develop toolkit of strategies, tools, and sample policies to disseminate widely to districts and states working to implement learning analytics.
Measure development	Develop methods for measurement of multimodal outcomes and how to integrate various types of data streams.
Mastery metrics	Define mastery by various metrics and expand to understanding mastery in unstructured as well as structured learning environments.
Personalized learning strategies	Optimize personalized learning strategies for different individuals, developmental levels, disciplines, and other dimensions.

Table 10. Five-Year Research Centers

CENTER	DESCRIPTION
Data Science Resource Center	This center would provide a "data marketplace" (with clarifying caveat that data are not being sold) and tools and services to help people use those tools to achieve their goals with big data. For education data, the Data Science Resource Center also would need to develop the trust frameworks and tools needed for data sharing and privacy protection.
Start-up Accelerator Center	This center would develop and run a cutting-edge startup accelerator for analytics-driven research on personalized learning and the teaching needed to support it.
Center for Learning at Scale	This center would focus on understanding personalized, contextualized learning at scale using analytics. This center's primary responsibility would be to conduct a longitudinal study that follows a group of eighth or ninth graders to college in a connected, personalized learning setting (i.e., school district, state, etc.) and contributes to understandings of the education stakeholder needs and policies for fostering successful outcomes.

The Value Proposition for Different Stakeholders in Building Learning Analytics

A confluence of breakthroughs is moving us closer to personalized learning pathways, including advances in the science of learning, the development of the Common Core standards, more sophisticated measures of effective teaching, growth in data mining and analytics, personalized and blended learning models, digitally born learning innovations, new measures of learning, and shared learning collaboratives for cyberlearning infrastructure that enable a multisector learning-technology ecosystem of products and services for personalized learning. Philanthropic foundations and government granting agencies have substantial interests in enabling the work that will advance the sciences, technologies, and interdisciplinary field building to actualize the vision of personalized learning for all. The learning analytics community also needs to step forward with a plan to address the challenges and opportunities discussed in this report.

As we make our recommendations, we realize the importance for each stakeholder of communicating the value proposition in relation to their problems of practice. What actions need to be taken by the different parties? What value would be derived from those actions? Which sectors are the best champions of the different action fronts for building the field of learning analytics? In this section, we briefly review why each stakeholder needs to take action, and what that action should be.

HIGHER EDUCATION INSTITUTIONS

Learning analytics is an opportunity to be seized now. Institutions of higher education could show leadership in addressing the emerging market demand for education data scientists trained in learning analytics by developing educational programs that contribute to human capacity building in this field. This includes not only PhD and master's level programs but also certificates, minors, and even survey courses for future researchers, educators, and policy makers who will not do learning analytics research but will confront it daily. Training future learning analytics experts is particularly valuable to institutions now, as the field is emerging, as early researchers in the area will be able to help the institutions themselves adapt to a new educational ecosystem. Because learning analytics will impact not only K-12 education but also higher education, training experts and developing programs in the field will be doubly beneficial to those institutions that take early initiative.

GOVERNMENT AND PHILANTHROPY

We have argued how funding to advance training programs in learning analytics and associated funding for interdisciplinary research centers and research projects is a vital priority. The aim is to fund advancing research and training that will accelerate breakthroughs in learning analytics and in associated innovative technologies that can contribute to personalized learning services and solutions for improving educational practices and outcomes. Foundations and government agencies need to provide Requests for Proposals for programs of research funding to which researchers, universities, and industry (when appropriate as partners) can respond. Funding agencies can create powerful partnerships; some relevant NSF grants are for building research communities of the kind needed. For example, one of our task forces developed an idea for a partnership between funding agencies to support multimodal and context-sensitive learning analytics work in particular, which would involve a combination of support for development of new sensing technologies, refinement of data analysis process, and incorporation of existing research and theory on the importance of context from the learning sciences.

RESEARCHERS IN UNIVERSITIES AND NONPROFITS

The case needs to be made that compelling new research questions and powerful technologies can be advanced to make new discoveries that mine the new data made possible in such digital learning systems and to innovate in sensing aspects of the learning environment that could contribute to better learning and teaching. There is great promise in opening up the "black box" of instructional treatments from long-established habits of administering pre- and posttests in school-based research. New capabilities of A/B in-vivo experiments along comparative testing of different features or methods of instruction, as investigated by the Pittsburgh Science of Learning Center among others, may allow researchers to better understand not only educational outcomes but also the range of instructional and learning processes that lead to those outcomes. In order to make the most of these opportunities, however, researchers will need to propose foundational research projects that solve key problems in the fields of learning analytics and education data science (e.g., STEM learning progressions in digital games; success for highly diverse learners; emotion sensing). In solving fundamental issues in learning analytics, there will be exciting new opportunities for basic research in learning sciences, psychology, and other social scientific fields related to personalized learning.

As these fields adopt large-scale, data-driven methodologies and inferences, there will be increased opportunities (and needs) for multidisciplinary research teams. Educational researchers and technology-enhanced learning researchers will be as helpless without a collaboratively designed, constructed, and accessed analytics infrastructure as a theoretical physicist without CERN, or a genomics researcher without gene databases. Employing a personalized learning infrastructure will enable rapid iteration on intervention, interface, and instructional designs. A consequence of these developments will be the requirement of a new training and education paradigm for scholars in education data science.

As learning analytics methods become more established, the field will have a dual potential. Learning analytics (a) may be able to energize existing fields of inquiry within and beyond education research with the promise of enormous amounts of data to address the questions researchers already have and (b) may enable the study of questions that researchers could not previously have imagined being able to ask (as is true in large-scale, computational social science such as Facebook studies).

INDUSTRY

The more encompassing educational ecosystem promised by learning analytics, in which data are more widely available on a broader range of student activities, contexts, and dispositions, will allow companies to offer compelling products and services that meet increasingly varied learner needs. From e-texts to embedded assessments to learning games indexed to standards for learning in and out of school, learning analytics will provide fuel for data-driven design and rapid iteration and innovation of new technologies. In virtuous cycles of appropriation, expertise developed by academic researchers will filter into industry innovations, and in turn researchers will be able to use industry products and services in their research. As academic researchers develop new and better measures of learning processes and outcomes, integrating the insights of learning scientists and analytics experts into product and service design should become a natural, dialogic process.

The role of industry in a data-driven, learning-analytics ecosystem is one of technological development and practical application of the kinds of basic research that analytics will enable for academic researchers. Maintaining the dialogue between basic research and technological innovation is a key role for industry stakeholders. is also in a unique place to think carefully about how to enact the various data protection and privacy recommendations and policies developed by government and educational bodies. Ensuring an open dialogue with learners and a careful exchange with other stakeholders that places the learners' rights and concerns at the forefront is an important contribution of industry to the development of a trustworthy and effective learning analytics ecosystem.

EDUCATORS AND EDUCATION LEADERS

There is considerable value to be contributed by partnering with teams in advancing learning analytics and education data science. There is the prospect of data-driven curricula and better tools to improve learning for all and to provide feedback for enhancing teaching and school leadership. How can a teacher know what learners know and provide instruction responsive to their individual needs? How can a teacher better identify students who are struggling and support them better? What new teacher professional development is needed, and what new roles will bridge technology and teaching? Educational systems (states, districts) need to participate in co-design and co-study of the new learning and teaching ecosystems employing cyberinfrastructure to advance goals of college- and career-ready high school students. In addition, the White House Office of Science and Technology Policy needs to communicate the important priorities associated with public and private sector progress on the topics of learning analytics and education data science, as well as the promise of progress in these fields for improving learning for all.

It will be important to develop ways to help policy makers make informed decisions based on learning analytics. For example, the complex topic of teacher performance evaluation may become more tractable with the development of learning analytics systems that capture more about context and circumstance in the classroom. Ensuring that policy makers have some way to understand learning analytics results will be essential to maximizing the usefulness of new research in setting policy.

In summary, the dimensions of our recommendations are multifaceted, reflecting the diverse stakeholders in the education ecosystem, but the core goal that motivates every stakeholder and every development in the field of learning analytics is the opportunity to improve learning for students across the educational spectrum, in both formal and informal settings.

We envision new learning analytics systems and technologies becoming trusted metacognitive resources for learners through expanded data collection and improved design of instructional interventions. We hope not to prescribe learner pathways and circumscribe learner abilities, but rather to enable learners to reach their potential by better guiding their cognitive and metacognitive processes and by making accessible to every learner a more personally rewarding and meaningful learning experience.

SECTION 5

Road Map for How to Implement the Field-Building Strategy and How to Evaluate Progress

Introduction

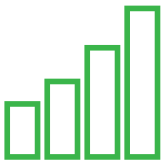
To develop a road map for building the field of learning analytics, we began by brainstorming four essentials to grow learning analytics as a field. We also considered how we could measure progress in growing the field. Then we identified areas in which this work has already been started, the necessary actions to move the field forward, and organizations to include as partners. We lay out a road map of activities to occur in three phases. The first phase needs to occur in the 1st year. The second phase occurs in Years 2 and 3. The final phase occurs in Years 4 and 5 and beyond. In order for the field to be built, it also will require visionary funding. These opportunities for advances will require funders from government agencies and private foundations to create priorities and associated funding streams; these programs will enable transformative research and development projects and foster networks to advance the promises and practices of this budding field.

Four Essentials for Learning Analytics to Grow as a Field

Learning analytics and education data science have tremendous potential for transforming the scientific understanding and practices of education and learning. For this promise to be realized, we have identified key catalysts that are essential to grow the field of learning analytics:

Human capital

We need universities enterprising enough to exploit the current developments in learning analytics; tackle the needs in education; and recognize the opportunities by creating new interdisciplinary and cross-department programs of study, research and training in education data science and learning analytics. We also need capacity development for educators (K-12) to understand how to improve data-based decision-making in their context.



Research

Industry should collaboratively engage in its own research and development along with partnerships with universities and other public sector organizations. This would bring strengths of scale and sustainability to the innovations in learning analytics and education data science that will be required to advance the science. For example, it would support the practices of personalized learning at scale and the sensing of the context of learning environments that can transcend the online-only limitations of digital learning, with the aim of enhancing education's effectiveness for all learners. A key to industry engagement is to make the process recognizable (e.g., leverage code+wiki on GitHub) and invest the time to make it understandable so that software developers in industry can extend the open source.

Tools

As in other societal domains such as predictive analytics for business and in big data science in astrophysics or genomics, education data science will need to create tools that are adapted to its questions, and that support the entire workflow of education data science, from study design, to experimentation and other forms of inquiries, to sense making and hypothesis testing of the data that are collected, to the community vetting of the science in order to improve the validities and utilities of the claims to knowledge that the scientific inquiries seek to establish.

Policy

To grow as a field, a vital priority is more open data sharing for multi-investigator studies than is traditional in the fields of education and the learning sciences. To echo a phrase and funding strategy that was dominant in National Science Foundation program funding in the 1980s and 1990s, "knowledge networks" will be important for accelerating the necessary advances in education data science and learning analytics.

Milestones for Measuring Progress

As we consider the road map to building the field of learning analytics, we also need to consider what milestones there would be to document that progress is being made. However, before considering milestones we had to consider the following questions.

- **What are the different activities that should be advanced over each of the next 5 years, and with what expected outcomes?**
- **What milestones can be used for tracking progress toward the diverse objectives associated with field-building and transformations of educational practices, research, and technologies?**
- **What resources would be required to establish these activities?**
- **What will happen AFTER the investments to enhance sustainability?**

Some examples of milestones for measuring progress are provided in Table 11.

Table 11. Example Milestones for Measuring Progress in Building the Field of Learning Analytics

MILESTONE AREA	DESCRIPTION OF MEASURABLE PROGRESS
Human capital	▶ An increase in percent of Carnegie-classified High and Very High Research university programs in learning analytics ▶ A decrease in the human capital gap as measured by an increase in percent of trained people in the field ▶ Improved decision-making on the part of districts, schools, and teachers to select products that are informed by learning analytics and have the greatest potential for improving outcomes for students ▶ Improved decision-making for teachers and administrators using data based on new understanding of learning analytics
Research	▶ An increase in the percent of learners engaged in personalized learning environments developed with information from the field of learning analytics ▶ Publication of case studies that inform capacity building with tangible models for districts to follow ▶ Publication of metrics for success and guidance for how to use learning analytics to apply these metrics
Tools	▶ An increase in the development and use of tools for learning analytics by members of the education community ▶ Publicly available toolkit for use by education researchers and districts for learning analytics
Policy	▶ Changes in policy related to data privacy and data sharing for education, corporations and universities that support learning analytics ▶ Publicly available templates for best data practices

Human Capital

PHASE 1

Preparing the next generation of education data scientists

In order to develop human capital in terms of the next generation of education data scientists, working with other training programs, such as IES and NSF training grants could provide new options to build this field. In addition, some universities are already starting master's programs in Learning Analytics that could be leveraged to further this effort. In Table 12 we provide some recommendations for types of training programs and related funding ideas to support these programs.

Table 12. Education Data Scientist Development Programs and Funding Recommendations

TRAINING PROGRAM	FUNDING SUPPORT RECOMMENDATIONS
Pre-doctoral fellowships	Support to sponsor PhD students to work with a professor at another university and complete one or more courses at that other university as part of their training (support to include travel costs).
PhD student internships	Support or development of 3-month summer or other internship programs within a company (e.g., Google) to learn a particular data analytic technique to be applied to a research project they are working on in their program.
Dissertation fellowships	Support PhD thesis work that advances education data science methodology and/or uses it to solve a critical problem for learning or teaching.
Faculty fellowships	Support faculty to spend a semester/quarter at another university to learn a specific technique or develop their competencies in data analytics methods.
Faculty travel grants	Support for faculty to attend data-science-related conferences or trainings to learn a specific technique or develop their competencies in data analytics methods.
Internships for ed tech professionals	Support for or development of a 3-to-6 month internship working on a research project related to the ed tech company's product in an academic lab as a visiting fellow. These could be funded by grants to the sponsoring universities.

LASI 2014 and beyond

LASI2013 was a very successful event; it exceeded the goals identified for attendees, and all reports from participants were extremely positive. We are currently working to keep the momentum going, and some of the projects included in this report came directly from new collaborations developed at LASI2013. It would be good for field-building to have LASI be both an annual meeting and include

related activities throughout the year. We recommend continuing this enabler event for up to 5 years with funding from foundations, government agencies, and industry partners. One area of growth would be to push beyond newcomers as the target audience for the workshops. In order to build human capital, the plan would be to find a balance between new participants and those who have shown specific gains from prior years, such as by building courses or degree programs, or developing collaboratives; specific gains should include demonstrable outcomes such as publications, new methods development, and cross-sector results. LASI 2014 took place at Cambridge at Harvard University June 30-July 2 with multiple sites around the globe organized for streaming and discussing workshop talks and panels.

Preparing education researchers

There are educational researchers with large datasets who are interested in learning analytic methods to apply to their data; however, they may not have funds to travel to conferences or to attend LASI. We recommend short, focused summer programs where educational researchers can bring their own data and support for analysis as they learn new areas of learning analytics. We recommend that the researchers bring a colleague so that the collaboration can continue in the long term. We recommend funding for faculty to attend the summer program along with funding for additional support from a learning analytics expert. Additional funding to support a graduate student to attend the summer program and continue work with the faculty person also could support capacity development. In addition, we recommend that a repository of examples of different analyses be developed and available online with related datasets for researchers to look at to see all of the potential analyses they could do with similar data. It would be helpful to develop a set of write-ups of learning analytic analysis methods for researchers to access as models for the research writing they are doing related to the analyses they have selected. One suggestion is to have a LASI type summer experience for mathematics and science education researchers and another one for general education researchers. Each year the summer experience could focus on different questions, such as longitudinal analyses of learning and engagement. This way people can network around methods and can collaborate with others studying similar questions.

Changing teacher and leader preparation

In order to support improved decision-making, we must change how we prepare K-12 education professionals. In order for schools to make effective instructional use of personalized learning, teachers and principals need to understand data and make data-based decisions; ideally, this should happen in teacher and school leader preparation programs and in programs for advancement of in-service education professionals. In Phase 1, we recommend identifying a group of quality school leaders and teachers who use data successfully in their decision-making process and teacher preparation providers interested in modifying their existing program to include attention to learning analytics (professors, university program staff) to create plans for integration of new methods into exiting preparation programs. First, we recommend that time is taken to observe and interview teachers and administrators to describe what they are doing that is effective. What do teachers and students do with the data they have? What data are helpful? What data are not helpful? How does data-based decision-making differ by

discipline? What do English teachers do with data differently than mathematics teachers? What do they do that is the same? This inquiry should not just include electronic data, but also paper data and paper systems that teachers may have developed over time that can also inform future work in this area.

Many universities offer certificates embedded within master's or EdD programs for in-service education professionals; providing resources for universities to offer a certificate focused on effective use of data and data-based decision-making, particularly one that fosters school-wide communication, would build needed capacity. In addition to programs, online resources should be created for both preservice and in-service teachers to access to learn more. One innovative idea would be to pair up a teacher wanting to learn more about analyzing his or her data with an educational researcher or data scientist who also can benefit from the experience of working with a new data set to work collaboratively to answer questions relevant to the teacher.

Many instructional resources on learning analytics can be drawn from, such as the ***Data, Analytics & Learning*** MOOC that George Siemens, Dragan Gaesevic, Carolyn Rose, and Ryan Baker are running through EdX in Fall 2014, which targets a very general audience, including teachers and administrators. Another resource is the Coursera MOOC, ***Big Data in Education,*** by Ryan Baker, which has now become an open textbook. Additional opportunities to provide content would be the ***Education Data Mining Conference*** and ***Applied Econometrics and International Development Journal***. There are also project outcomes from the NSF Building Community and Capacity for Data Intensive Research projects that could add valuable information to this process of changing how we prepare teachers.

PHASE 2

Start-up Accelerator Center

Recommended funding: $5 million total

This center would develop and run a cutting-edge startup accelerator for analytics-driven research on personalized learning and the teaching needed to support it. This effort would involve providing research teams with every tool to successfully launch a research proposal for funding from a variety of sources. Startup incubators provide help with everything from developing a grand vision and then narrowing interests to the mechanics of actually running a successful research lab within a university, to developing and sustaining research-industry partnerships. This center would run a startup accelerator program each summer to build capacity for growth in the Learning Analytics research field for the longer term. Another charge of this center is to determine the best way to train people in the field to use software applications (e.g., Python), determine what research questions to answer, and learn how to guide a research team. We really do not know how to train education data scientists. We recommend that this center focus on providing guidance in this area.

Integrate learning about data-based decision-making into educator preparation

In Phase 2, we recommend that a committee be formed to take what is learned in Year 1 and to work with universities to implement the planned integration of data-based decision-making into teacher and school leader preparation programs as well as into certificate programs for in-service teachers and leaders. We also recommend that documents and guides be created based on these new programs to share more broadly with other universities.

PHASE 3

Establishing university education data science programs

One recommendation is to provide competitive awards for establishing university education data science PhD programs encompassing departments of statistics, computer science, and education/ psychology. This could include some number of graduate fellowships (similar to the U.S. Department of Education, Institute of Education Sciences Training Grants in Education Sciences). The RFP could be developed and announced during 2014-2015, with funding provided in 2016. Three 4-year awards for $5 million each would support funding for initial faculty salaries and startup funds. We recommend that grantees provide sustainability plans within their response.

Worked examples for newcomers to the field

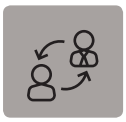
We recommend funding to support the creation of a resource for teaching students, for newcomers to the field who are self-teaching, and for reviewers. This resource would include a data set and worked examples that highlight the types of questions that can be asked and answered with different analytic techniques. The data set could include video data; standard quantitative data; and some sort of big data, such as log file data of participation. A set of worked analyses would assist in highlighting the nuanced questions that can be asked. Having video records linked to the log files of participation would help people critically explore the idea that big data is not the territory of learning, but more of a map. Finally, this resource could support people in thinking more broadly about the types of data that could be collected. Many people are unaware of the possibilities of types of data that could be collected in naturalistic environments in which non-deterministic, ill-structured, and generative learning occurs.

Research

PHASE 1

Research to prevent reinventing the wheel

We recommend development of ***The What Might Work and Why Clearinghouse,*** a project to identify what has been done and what research is out there to support the field of learning analytics. This resource could feed into the startup efforts discussed next.



Researcher and ed tech startup connector

In order to support the development of new university programs in learning analytics, quality learning resources (e.g., courseware and games) and assessments aligned with what we know about promoting learning must be developed. The best approach is to build on the incredible momentum that is happening in the ed tech sector. These companies have strong financial incentives to produce professional products. In Year 1, researchers from LASI will identify a few ed tech startups and academics interested in collaboration. Together, they will devise a strategy and design of the product and its back-end analytics as part of a Summer 2015 Imagine K12 startup session. The researchers will receive compensation for participating. It is important that we keep a focus on solving real world problems in education. Many connections already happening between industry and education can be leveraged in this area, such as capstone projects for master's students. This group also will be charged with developing a Lexicon of terms and examples to engage prospective participants to better understand learning analytics and education data science.

Case studies to inform capacity building and policy

The review of research and district and state examples point to several specific areas, outlined in the above recommendations, which should be addressed to maximize the potential of learning analytics. ***Identifying and developing case studies*** that demonstrate how to build capacity and policies will provide tangible models for other districts and states to follow. This could include development of sample apps to accompany case studies published on GitHub public code repository to catalyze the community. This will accelerate the potential of implementing learning analytics in more districts and states by garnering information from the early adopters and implementers. Research in this area should include the steps outlined in Table 13.

Table 13. Research to Support Capacity Building and Policies for Learning Analytics

RESEARCH STEP	DESCRIPTION OF NECESSARY ACTIONS
Select districts for case study	Locate one or more large districts that would welcome having a partner in this work and conduct a case study of how they make significant progress with learning analytics for personalized learning.
Analysis to identify strategies	Conduct analysis of data collected through the case study to identify specific strategies and tools to facilitate capacity building in the state, including a culture of informed decision-making, adequate infrastructure, human capital, and professional learning opportunities.
Analysis to identify policies	Analyze case examples to identify specific policies that enable learning analytics.
Development of a toolkit	Develop a toolkit of strategies, tools, and sample policies to disseminate widely to districts and states working to implement learning analytics. For example, tools like PSLC DataShop, EDM Workbench, and LightSIDE text mining tool bench.

Multimodal methods of measurement

Multimodal learning analytics techniques should allow researchers to examine unscripted, complex tasks in more holistic ways. By combining strong learning theory with techniques that have already been proven to be successful in the multimodal interfaces community, multimodal learning analytics should help researchers to analyze non-traditional learning environments at scale, as well as uncover new insights into the learning processes that students follow.

However, several steps are needed to move ahead in this area.

- Continued collaboration between computer scientists, linguistics researchers and education researchers, so that the appropriate techniques get applied to the correct learning situations
- Support for education researchers to become familiar with the tools of computation so they can play a larger role in designing the algorithms and tools used to analyze these rich multimodal data sets

As a starting point, education researchers who have a history of qualitatively working with multimodal datasets could help advance this space by partnering with data scientists to leverage more of the tools of computation in automating their analysis. Education researchers could benefit from increased awareness of the different available computational tools that could provide new insight into their research. Such opportunities will create more bridges between education and data mining, as well as providing education researchers with larger data sets to work with.

Measuring success

A critical research question is what does mastery or success look like in structured and less structured learning environments? For less structured environments, we need to figure out what counts as success and how we demonstrate it. For example, Drs. Taylor Martin and Deborah Fields are currently creating an analytics tool to automate analysis of students' code in Scratch (a graphic programming environment developed at MIT) and output progress reports for teachers translated from Scratch code blocks to specific computational thinking and programming concepts like abstraction and recursion. While it may seem more straightforward in more structured environments like intelligent tutoring systems, even there, what we include in measures of success is important, such as, persistence, preparation for future learning, or engagement. The more complex issue is measuring trajectories towards creating expertise when students are working in groups. In these contexts social factors may obscure the level of competency individual students within groups have attained, in some ways making them appear more capable than they are and in other cases less. Research is needed to figure out what to include as outcomes or measures for success beyond traditional attendance, behavior, and course grades. There are post-secondary outcomes related to college and career success that also need to be considered.

Example metrics for measuring success are included in Table 14.

Table 14. Potential metrics

Individual Metrics	Student Metrics	Teacher Metrics	School Metrics
Mastery Speed/pace of learning Creativity Persistence Motivation Engagement Affect Self-regulation Group collaboration	Attendance or Retention in school GPA Standardized test scores	Decrease in time spent on menial tasks Number of data-based decisions	Student achievement per dollar spent Graduation rates College-going rates

Prototype of personalized learning system

In order to foster the development of and adoption of personalized learning at scale, a robust system for reporting is required. In Phase 1 and 2 we recommend prototyping a personalized learning management recommendation and reporting system. In Phase 3, we recommend piloting and rolling out (free of charge) the refined system to a key group willing to participate in this pilot. This initial group could include schools participating in the longitudinal study with the Center for Learning at Scale.

Optimization of personalized learning

To inform the work going on in personalized learning, research is needed to understand measures that can serve as predictors for optimization of personalized learning systems. Across a student's academic year or academic career, a major challenge for implementing personalized learning is figuring out how to optimize for different individuals, developmental levels, and disciplines. In many ways, this is the advanced version of aptitude-treatment interactions. There are many potentially important characteristics established through either explicit or tacit measures to serve as predictors for optimization, and for each one there are associated costs and benefits.

- Geography (region)
- Cultural/linguistic/gender/racial diversity
- Preferences for media
- Pedagogies that have been particularly successful for the student
- Motivation/persistence and how it varies by domain or context
- Attributions of intelligence – epistemological beliefs and self-attribution
- Interests
- Social relations, including peer-peer, student-teacher, and others
- Parental involvement
- Prior education experience
- Prerequisite knowledge
- Response to negative feedback
- Preferred student strategies—proneness to gaming the system, carelessness, meta-cognitive strategies

PHASE 2

Center for Learning at Scale

Recommended funding: $10 million total

The primary responsibility of this center would be to conduct a longitudinal study that follows a group of eighth or ninth graders to college.

This study would have the following potential main goals:

- Test methods to boost college readiness for challenged students.
- Identify best practices for connected learning and blended learning
- Explore how to best collect longitudinal data that supplement classroom or learning data, taking advantage of archive data to conduct meaningful longitudinal studies
- Explore how to use data from multiple educational (formal and informal) experiences to understand lifelong learning
- Investigate teacher and administrator learning about data and how to use interfaces or dashboards to make decisions about instruction

PHASE 3

Guide to learning resources

Building on the findings from the basic 1-year research project to identify what has been done and what is out there, the What Might Work and Why Clearinghouse would provide a continually evolving research-based guide to learning resources following the newly released standards for evidence. The missing piece in websites like EdSurge, which report on a deluge of terrible and fantastic educational technologies, is a filtering mechanism. One recommendation is to fund a fellow at EdSurge for 5 years to analyze the research on all the promising new things they discover.

Refine and pilot personalized learning system

Building on what is learned in the Year 1 prototyping the personalized learning system, two efforts should continue through the 5-year period. First, we recommend piloting and rolling out (free of charge) the refined system to a key group willing to participate in this pilot. There should be continuous refinement of the system based on user feedback, teacher interviews, and feature requests. Second, significant research funding should be provided for projects to discover and validate or scale both (a) best practices for teachers, students, parents, and administrators for using the system and (b) how we best represent these data to be useful for these various stakeholders.

Researcher and ed tech startup connector-social network

We recommend continuing participation in the Imagine K12 start up incubator program. The core group will disseminate the success stories and create a social networking engine. Instead of building from the bottom up, it would be best to build off an existing one (e.g., Linked In, Yammer, etc.), but an obstacle may be that start-ups are unlikely to want to put their ideas in a public space prior to launch. This social network would link researchers and ed tech startups at any phase of a project. For example, those with data could connect with others for support in analysis, and newcomers could seek advice at the beginning of the design.

Tools

PHASE 1-3

Data Science Resource Center

Recommended funding:
$15 million total;
$5 million for Year 1,
$2.5 million continuing Years 2-5

There are two critical pieces to what this center should provide. The first is a data marketplace. The second is the provision of tools and services to help people use those tools to achieve their goals with big data. For education data, the Data Science Resource Center would also need to develop the trust frameworks tools needed for data sharing and privacy protection.

Data marketplace

It takes a lot of data of all kinds to add up to big data. The Data Science Resource Center will house a collection of datasets and streams for data scientists and developers to sample, experiment with, and use to create innovative analytics and applications. The focus of the center is on training people to analyze really big data, for example, creating a large data set that people could use to answer a variety of questions. This would be instrumental in assisting faculty with developing courses. In industry one of the challenges is aligning metadata. This effort requires curation, or it becomes impractical for the intended users to join across disparate datasets, regardless of how valuable those joins might be. Therefore, it is important that the Data Science Resource Center have as a central role the curation of metadata to allow for simple access to the data sets.

Toolkit and services

As in the data sciences more broadly, the goal of an analysis toolkit is to bring down the people cost of getting good answers from your big, interconnected data. Interested users can access the tools to develop their own applications or purchase services to support their research team or do the analytics (see Table 15).

Table 15. Sample Cloud Tools and Analytic Tools

TOOL	DESCRIPTION
Cloud tools	
Cloud Streams	*Streaming data and real-time analytics*
Cloud Queries	*NoSQL database and ad hoc, query-based analytics*
Cloud Hadoop	*Elastic Hadoop clusters and batch analytics*
Analytic tools	
Wukong	*Uses a simple scripting-based language (technically: a domain-specific language, or DSL) instead of complex MapReduce programming to rapidly develop and deploy Hadoop batch analytics and real-time stream analytics using Storm. Develop, test, and iterate in your local environment, then easily deploy to operational systems at full scale.*
Application templates	*Accelerate big data application development with application-specific templates containing prebuilt, domain-specific functionality.*
Cloud APIs	*Program against both native and abstracted APIs that let you tap into the power of Hadoop, Storm stream processing, NoSQL databases and more.*
Additional tools	
Trust frameworks	*Tools needed for companies, researchers, etc. to be able to share their data in this system.*
Data Pipeline management tools	*Tools that make it easier to integrate, process, store, and move data.*

Competitions

Another means to increase human capital and add to the research base in learning analytics would be to incentivize innovation by using industry models of competitions, such as X-Prize, Netflix Prize, and Kaggle. The goal would be to identify the top 5-10 problems or grand challenges to be solved in education data science. In Phase 1, we recommend that a team be convened to identify the problems upon which the challenges will be focused. Competitions would be held in subsequent phases. These grand challenges would be similar in spirit to the 2010 KDD Cup run by John Stamper and the EMNLP 2014 MOOC shared task organized by Carolyn Rose and George Siemens (*http://emnlp2014.org/workshops/MOOC/call.html*). We recommend that teams be incentivized to publish about their experiences during the competitions. For example, some of the best results from the two years of the Netflix Prize were not in the algorithm advancement, but rather in the candid remarks by competing teams who published papers about the prize afterwards.

Potentially, this center could sponsor hackathons to develop core tools that they need to accelerate the process. However, it is important that these hackathons be well sponsored. For example, Emirates Air sponsored a weekend hackathon with a $5,000 grand prize, plus a vendor contract to the winner. It was a huge success, which has been attributed to winning teams with ongoing relationships that included monetary and professional rewards.

Policy

PHASE 1

Templates for best data practices

The consensus is that changes to standards for data ownership, privacy, sharing, and access are foundational to enable the critical research to achieve the vision of personalized learning and to support policy changes. For the 1st year, the primary effort is developing a set of templates for best data practices based on use cases and approved by the appropriate governing body to be certified for IRBs across the United States. These templates and standards will expedite and shore up the quality of research proposal reviews. We also believe that it would be helpful to researchers to provide boilerplate language that researchers could use and cite regarding data standards, especially researchers who are trying to make their data open access.

K-12 data sharing and privacy standards

There is a need in Phase 1 to work with one or more states and a large district within each state to better understand how data sharing works. What are the issues? What is working? We recommend funding for deep engagement with the district and state leadership to understand data-sharing policies, data privacy, and methods for ensuring secure access of data. What is needed to enable successful implementation of a single login option in the future where all data are being collected and shared from all activity within the system (e.g., student benchmark testing, student daily performance and activity within digital learning experiences)?

PHASE 2-3

Trust frameworks

Trust frameworks are one of the long-term projects in this area. According to recommendations from the Aspen Institute Task Force on Learning and the Internet (2014), there are 6 characteristics of a trusted environment: (1) transparency and openness about what data is collected and how, (2) participation in decision- and policy-making, (3) data stewardship (de-identifying and/or deleting sensitive data), (4) technology innovation (e.g., privacy dashboards), (5) accountability (e.g., a code of conduct), (6) oversight and enforcement (e.g., regulatory arrangements). This task force also calls for funding to address new approaches, tools, and practices in this area.

Collaboration with governance body

The other long-term project includes collaborating with the US Department of Health and Human Services to manage and update templates and to create and manage trust framework recommendations. This will require engagement around three areas: (a) the framework of different types of educational research, (b) the privacy side (e.g., International Association of Privacy Professionals), and (c) the computer security side (e.g., an independent body that can use National Institute of Standards and Technology standards as guidelines). Socializing and building these elements will take coordination and going to conferences to get the word out.

Branding and getting the word out

There is a need for the field to identify a set of common messages that can be disseminated through conferences and other events. One example message would be around educating the public about data privacy protection. In Year 1, we recommend that talks be sponsored at conferences with a data science area of focus, such as the O'Reily Strata conference, the O'Reily OSCON Open Source Conventions, and SXSWedu conference. The sessions could be on learning analytics and how to get involved. As much as possible, these talks should get the word out about learning analytics (e.g., success and use cases) and contribute to the field of data science as a whole. Some example topics include the following:

- Managing your hidden data pipeline
- Novel data mining techniques and algorithms
- Examining how industry tools can be deployed to best bring about change in this new field

Additional ideas for getting the word out would be to leverage industry meet-ups to garner industry engagement and to fund MOOCs, tutorials, or summer programs on relevant topics. We recommend a collaboration with Strata and O'Reilly leaders to develop a StrataEdu conference.

References

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., . . . Woolf, B.P. (2007, June). Repairing disengagement with non-invasive interventions. In ***Proceedings of the 13th International Conference on Artificial Intelligence in Education***, 195-202.

Arroyo, I., Woolf, B. P., Corbett, A.T., Koedinger, K.R., Evenson, E., Roll, I., Wagner, A.z.,Naim, M., Raspat, J., K. (2011, July). The impact of animated pedagogical agents on girls' and boys' emotions, attitudes, behaviors and learning. In ***Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies (ICALT)***, 506-510.

Aspen Institute Task Force on Learning and the Internet. (2014). ***Learner at the Center of a Networked World.*** Washington, D.C.: The Aspen Institute, June 2014.

Baker, R.S.J.D., Corbett, A.T., Koedinger, K.R., Evenson, E., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., Beck, J. (2006) Adapting to When Students Game an Intelligent Tutoring System. ***Proceedings of the 8th International Conference on Intelligent Tutoring Systems***, 392-401.

Bardin, J. (2012, September 3). What they built at Brain Camp. ***The Chronicle of Higher Education.*** Available at chronicle.com/article/What-They-Built-at-Brain-Camp/134016.

Barron, B. J., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., . . . CTGV. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. ***The Journal of the Learning Sciences***, ***7 (3-4)***, 271-312.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. US Department of Education, Office of Educational Technology, 1-57. Downloadable from: http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). ***How people learn: Mind, brain, experience and school***. Washington, DC: National Academy Press.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. ***Review of Research in Education, 24***, 61-100.

Brynjolfsson, E., & McAfee, A. (2011). ***Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy***. Lexington, MA: Digital Frontier Press.

Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. ***International Statistical Review, 69(1)***, 21-26.

Duckworth, A.L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance in adolescents. ***Psychological Science, 16 (12)***, 939-944.

Dweck, C. (2007). ***Mindset: The new psychology of success—How we can learn to fulfill our potential***. New York, NY: Ballantine.

Dweck, C. S., Walton, G. M., & Cohen, G. L. (2013). ***Academic tenacity: Mindsets and skills that promote long-term learning***. Paper prepared for the Bill & Melinda Gates Foundation.

Farrington, C. A., Roderick, M., Allensworth, E. R., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). ***Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance: A critical literature review.*** Chicago, IL: The University of Chicago Consortium on Chicago School Research.

References

Graham, P. A. (2005). ***Schooling America: How the public schools meet the nation's changing needs***. New York, NY: Oxford University Press.

Herr-Stephenson, B., Rhoten, D., Perkel, D., & Sims, C. (2011). ***Digital media and technology in afterschool programs, libraries, and museums***. Cambridge, MA: MIT Press.

Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). ***The fourth paradigm: Data-intensive scientific discovery***. Redmond, WA: Microsoft Research.

Kane, T. J., & Staiger, D. O. (2012). ***Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*** (MET project policy and practice brief). Seattle, WA: Bill & Melinda Gates Foundation.

Levy, F., & Murnane, R. J. (2012). ***The new division of labor: How computers are creating the next job market***. Princeton, NJ: Princeton University Press.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). ***Big data: The next frontier for innovation, competition, and productivity***. New York, NY: McKinsey Global Institute. Retrieved from http://bit.ly/McKinseyBigDataReport

Martin, T., & Sherin, B. (2013). Learning analytics and computational techniques for detecting and evaluating patterns in learning: An Introduction to the special issue. ***Journal of the Learning Sciences, 22(4),*** 511-520.

Moss, P., Pullin, D. C., Gee, J. P., Haertel, E. H., & Young, L. J. (Eds.). (2008). ***Assessment, equity, and opportunity to learn***. New York, NY: Cambridge University Press.

National Academy of Engineering. (2008). ***Grand challenges for engineering.*** Retrieved from http://www.engineeringchallenges.org/

National Science Foundation Task Force on Cyberlearning. (2008). ***Fostering learning in the networked world—The cyberlearning opportunity and challenge: A 21st century agenda for the National Science Foundation***. Arlington VA: National Science Foundation.

National Research Council. (2013). ***Education for life and work***. Washington, DC: National Academy Press.

Pane, J. F., McCaffrey, D. F., Slaughter, M. E., Steele, J. L., & Ikemoto, G. S. (2010). An experiment to evaluate the efficacy of Cognitive Tutor Geometry. ***Journal of Research on Educational Effectiveness***, ***3,*** 254-281.

Patil, D. J. (2011). ***Building data science teams***. San Francisco, CA: O'Reilly.

Pavlik, P. I., Jr., Cen, H., & Koedinger, K. R. (2009, July). Performance factors analysis—A new alternative to knowledge tracing. In ***Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling*** (pp. 531-538).

Pea, R., Childress, S. M., & Yowell, C. M. (2012). ***The demand for education data scientists: To enable personalized learning, we need education data scientists capable of unlocking insights contained in a growing collection of digital learning and teaching activities***. Seattle, WA: The Bill & Melinda Gates Foundation.

PERTS. (2014). ***HS9—National Mindset Pilot.*** Retrieved from https://p3.perts.net/orientation/hs9

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. ***Psychonomic Bulletin & Review***, ***14 (2)***, 249-255.

Ritter, S., Kulikowich, J., Lei, P., McGuire, C. L., & Morgan, P. (2007). What evidence matters? A randomized field trial of Cognitive Tutor Algebra I. In T. Hirashima, U. Hoppe, & S. S. Young (Eds.), ***Supporting learning flow through integrative technologies*** (pp. 12–20). Amsterdam, the Netherlands: IOS Press.

References

Schwartz, D. L., & Arena, D. (2012). ***Measuring what matters most: Choice-based assessments for the digital age***. Cambridge MA: MIT Press.

Shared Learning Collaborative (2012, June 11). SLC Scenarios: Opportunities for application development. This SLC research report was available at www.slcedu.org until inBloom.org closed operations.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias, & J. D. Fletcher (Eds.), ***Computer games and instruction*** (pp. 503–524). Charlotte, NC: Information Age.

Shute, V. J., & Ventura, M. (2013). ***Stealth assessment: Measuring and supporting learning in video games***. Cambridge, MA: MIT Press.

U.S. Department of Education. (2010). ***Transforming American education: Learning powered by technology. National Education Technology Plan 2010.*** Retrieved from http://www.ed.gov/technology/netp-2010

Walonoski, J. A., & Heffernan, N. T. (2006, January). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. In M. Ikeda, K. Ashlay, & T.-W. Chan (Eds.), ***Proceedings of the 8th International Conference on Intelligent Tutoring Systems*** (pp. 382-391). Berlin, Germany: Springer-Verlag.

Williamson, B. (2013). ***The future of the curriculum: School knowledge in the digital age***. Cambridge, MA: MIT Press.

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. ***Scientific Studies of Reading***, ***5***, 211-239.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. ***Science***, ***330(6004),*** 686-688.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. ***Review of Educational Research, 81(2),*** 267-301.

The Learning Analytics Workgroup (LAW) Report

Building the Field of Learning Analytics for Personalized Learning at Scale

By Roy Pea

D.Phil., Oxon. David Jacks Professor of Education and Learning Sciences at Stanford University.