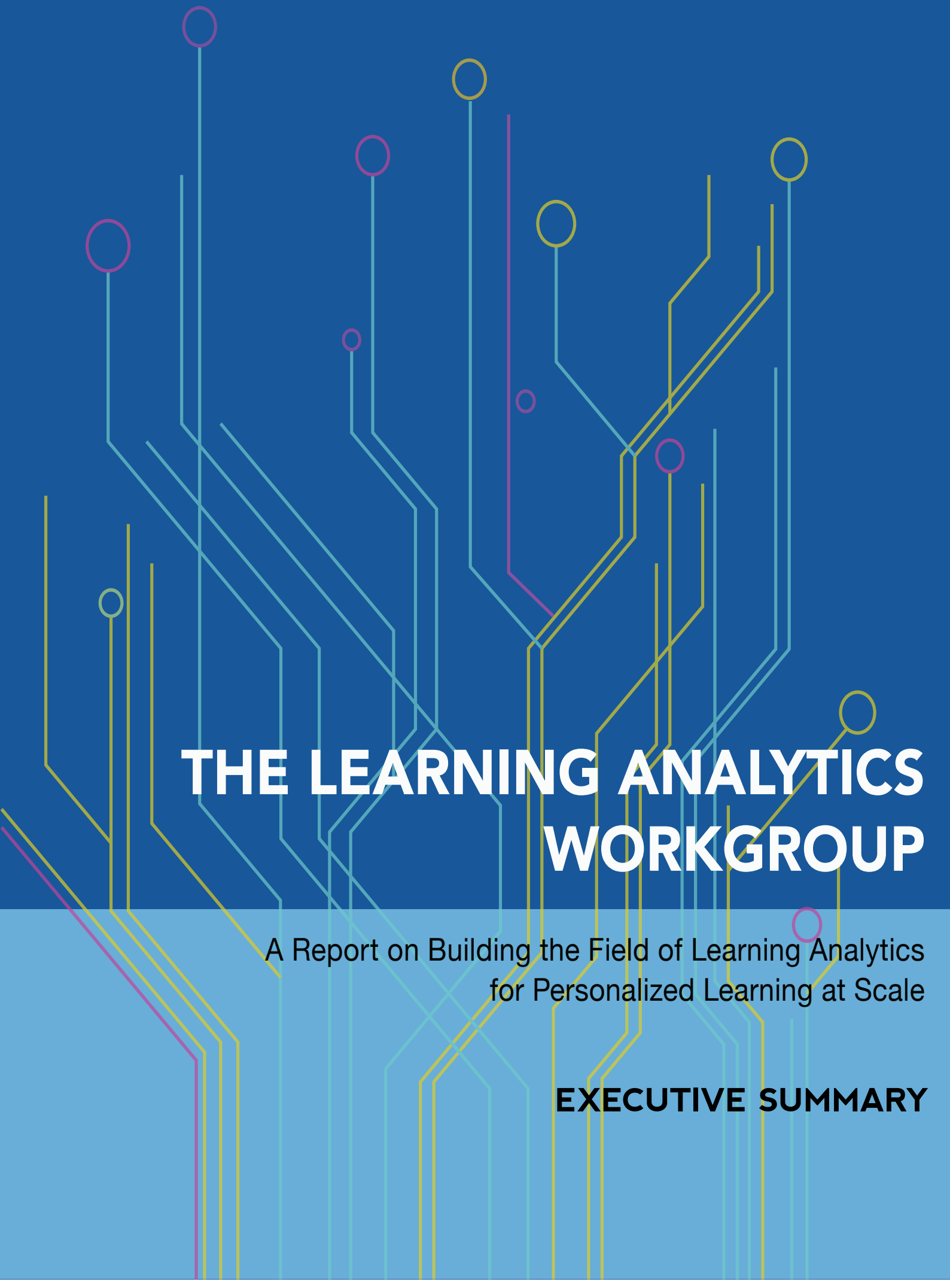


THE LEARNING ANALYTICS WORKGROUP

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

EXECUTIVE SUMMARY

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.

Executive Summary

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. Teachers, school administrators, parents, and students need to track learning activity and progress to accomplish the goal of college and career success for every student. 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 to 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 ensure that all students achieve college and career success.

The Learning Analytics Workgroup (LAW) Project was initiated at the convening of a multisector group by the Bill and Melinda Gates Foundation (BMGF) on August 3, 2011, at the University of Chicago's Computation Institute.

At this meeting our discussion began about 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. Impending opportunities and challenges will accompany our advancing understanding of the increasing orders of magnitude of data generated as millions of K-12 students and their teachers transition to digitally enhanced teaching and learning and to utilizing those data to improve education. The growth of data in education surpasses the capacity to make sense of such vital information and to employ insights derivable from the data to guide educational practices. The LAW Project has focused on making progress towards personalized learning at scale by building the field of learning analytics to support and advance initiatives and state-wide technology infrastructure to support new ecosystems of personalized learning and teaching.

The LAW Project convened experts in the field through workshops and conference panels to understand the needs and current developments to inform the building of the field. The result of these efforts is represented by this final report, which includes the following:

- A conceptual framework for building the field of learning analytics
- Critical questions for understanding how to build the field of learning analytics
- New tools, approaches, policies, markets, and programs of study associated with the field of learning analytics
- Resources needed to address priorities
- A road map of how to implement the field-building strategy and how to evaluate progress

The purpose of the LAW Project has been advancing each of these objectives, which we hope is made clear in the full report. However, for this executive summary we provide a brief summary of each of these important areas.

A Conceptual Framework for Building the Field of Learning Analytics

Personalized learning provides the following opportunities:

- Improving educational performance
- Facilitating cost efficiencies through educational productivity and organizational optimization
- Accelerating educational innovation

In building the field of learning analytics, we are targeting the challenge 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 is fundamental to personalized learning. 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.

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 exponential growth of education data to be generated by digitally-enhanced learning environments requires education data scientists and people with diverse sense-making talents to be able to bring these data sets 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 achievement would allow for further instructional interventions and provision of progressive learning resources and experiences. Personalized learning as a vision indicates that we need ***sensing systems for learning***. 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.

Executive Summary

Data science, as a distinct professional specialization, is in its infancy. What we are calling for is an even newer specialization, ***Education Data Science***. People with skills in this area currently 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. 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. However, to understand the vastness of learning that occurs we need people trained in understanding **Cyberlearning**, learning that is mediated by networked computing and communications technologies. 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 infrastructures will need to link their efforts to expand the open platform for creating a vibrant, multi-state or national learning technology ecosystem with parallel efforts to evangelize the exponential education data opportunity with relevant leaders and practitioners in the data science community.

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

The LAW Project established a workgroup 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, and philanthropy). This multisector approach is vital to building the field.

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

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 there are many questions to consider.

FIRST

- What do educators need?
- What may better enable their practices to achieve the personalized learning vision?

We consider it vital to foreground the challenges of educators in relation to the prospects of personalized learning.

Based on insights from close to 800 teachers and administrators interviewed across six states, the Shared Learning Collaborative identified nine opportunity areas for learning technology innovation. These hold the greatest potential to deliver value in improving current offerings that suffer from lack of data integration, have incomplete feature sets, or require improvements in user experience.

NINE OPPORTUNITY AREAS

AREA 1	*Supporting lesson planning, instruction, and assessment*
AREA 2	*Creating Learning maps to track education progress*
AREA 3	*Viewing student profiles and history, and managing their transitions*
AREA 4	*Course, career and college planning*
AREA 5	*Learning intervention flagging, action guidance, and measurement*
AREA 6	*Training, professional development, and networking*
AREA 7	*Communication and collaboration with education stakeholders*
AREA 8	*Technology selection, management, and usage*
AREA 9	*Budgeting, human resources, and performance management*

SECOND

- What outcomes should we care about in the development of personalized learning?
- Which of these outcomes need further research and development for use in personalized learning systems?

We recognize that different educational stakeholders will have different **success metrics for learners**.

While the notion of "success metrics" for K-12 personalized learning may seem initially straightforward, there are opportunities to wish for far more than just high grades and completed courses. Increasingly, we are learning that other measurable aspects of learning processes may serve as key drivers of learning and may be subject to intervention as a type of outcome. Among the topics of special attention today are the so-called "noncognitive factors" in learning, such as academic persistence or perseverance (aka "grit"), self-regulation, and engagement or motivation. 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. Other success metrics we have considered include collective intelligence (contributing to the learning of groups in which one participates), innovation and creativity, and preparation for future learning. Success metrics for K-12 personalized learning, such as these, should evolve as the needs for educated competencies evolve with society.

THIRD

- 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?

For personalized learning, a pre-eminent objective is **creating a 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.

We have identified a wide variety of evidence as sources for building a learner model: metrics of student interaction during learning activities, social metrics, data concerning student mindset, past performances, learning media or genre preference, perseverance and persistence, administrative data, demographic information, temporal history, and emotional state. 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 the data for research and evaluation purposes is questionable.

We should critically consider the potential importance of making the learner model transparent to teachers or other educational stakeholders, with the ability to be inspected or modified by an individual. Finally, we also need to consider the importance of learner data aggregation and interoperability across digital platforms provided by multiple publishers or vendors in building the learner model. One potential solution is the creation of connected learning models, networking across and within formal and informal learning, and educational activities in which any learner participates.

FOURTH

- 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?

A broad set of topics is encompassed in the question of how to establish a well-functioning, personalized-learning research infrastructure.

There are needs in the areas of data sharing, analysis and visualization tools, collaboration practices, data-management policies, and Institutional Review Board (IRB) reforms that will enable development of learning analytics as a field and implementation of personalized learning at scale. 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.

Data-management policies of IRBs, school permissions, and data-sharing agreements are key parts of developing a shared data infrastructure. For example, 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. Standardization of data privacy and information technology security systems would also streamline the review process, as researchers would know exactly what kinds of safeguards need to be in place before submitting a research proposal. In building a field with a shared research infrastructure, developing standards is a project not only for scholars and industry partners but also for review boards.

FIFTH

- How should the roles of the teacher and personalized learning technologies be distributed?
- What data should teachers have access to?
- What kinds of decisions should they be able to make?

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

Data literacy will become an important skill for teachers, as making data-enhanced decisions in the classroom will depend upon the ability of teachers 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. 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 will likely better motivate students to engage in learning activities than technology-based recommendation systems alone. 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. Revisions to teacher training programs will be needed to reflect the new skills and roles of the teacher in a personalized learning system.

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."

Articulating and Prioritizing New Tools, Approaches, Policies, Markets, and Programs

Early research in learning analytics should focus on several priority areas. 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? To further explore these research priorities, we 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. Learning analytics systems presumably will allow researchers and educators to identify early warning indicators when learners struggle with key developmental phases like prealgebraic 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, which treat standards as an initial assertion and then test whether these competencies are correctly described as Knowledge Components. 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. Drawing on existing research, researchers could address important issues by applying existing theoretical and conceptual frameworks, which can be applied at each grade band and then instrumented in analytics systems.

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 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. 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 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 not only to focus on student demonstrations of knowledge on predesigned 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 open 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.

Determining Resources Needed to Address the Priorities

TRAINING PROGRAMS

There is a need to develop training programs to develop capacity for learning analytics in education. 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? We identify several competencies for education data science and learning analytics specialists, such as computational and statistical tools and inquiry methods; general educational, cognitive science, and sociocultural principles of learning; in addition to an ability not only to perform data analysis but also to recognize and evaluate data quality, among other important competencies.

FIRST
We recommend a **faculty cross-training approach**

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. Faculty not only from computer science, statistics, and education but also from a range of fields may 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 that 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, for example, during summer institutes.

SECOND
There is a need to develop **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 other areas are all potential fits for learning analytics postdoctoral training. How can we best support linking young researchers with these interests with experienced faculty?

THIRD

There is a need for **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, as well as specialized seminars and survey courses). Work has begun developing these courses, both within various institutions and at a broader scale; however, more development needs to be done to incorporate authentic research experience and apprenticeship into ongoing curricular developments. 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 degree training. How might state and district-level educational systems create feedback loops at a local or cluster level, in what are being called Regional Education Innovation Clusters (http://www.ed.gov/edblogs/technology/innovation-clusters/)? The goal would be a broader learning-by-doing infrastructure whereby students and faculty can work on industry or school data analytics projects, including design, implementation, and evaluation, as well as secondary analysis and data mining with existing data sets. While 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 undergraduate programs that will attract high-potential students to enter the field.

FINALLY

Building the field of learning analytics will require **knowledge networking and online community** building which encompasses 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 that encompass a range of professionals, researchers, and graduate students. Furthermore, industry can provide datasets for professional training. Identifying potential industry partners from among the variety of companies that do work in education will be important. 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. This strategy includes networking with 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.

FUNDING RECOMMENDATIONS

What can be achieved in different periods within the next 5 years for building the field of learning analytics? We have made recommendations for what should happen in the next 18 months, the following 18 months, in research projects for 1–3 years, and for the development of 5-year research centers. General short-term activities include recommendations for data standards, communication, funding of fellowships and internships, review of existing resources and projects, competitions, the piloting of personalized learning management systems, and teacher preparation. The 1- to 3-year recommended research projects include state and district case studies, development of a toolkit of strategies, measure development, definition of mastery metrics, and optimized personalized learning strategies.

Successfully building the field of learning analytics and education data mining will result, in the long term, in personalized learning for all students who regularly access large-scale digital learning and teaching platforms and the commercial and noncommercial providers of educational services and solutions that leverage its infrastructure.

Three 5-year research centers are proposed:

Center 1 A Data Science Resource Center to provide a data marketplace of tools and services

Center 2 A Start-up Accelerator Center to develop a cutting-edge startup accelerator for analytics-driven research

Center 3 A Center for Learning at Scale to focus on understanding personalized, contextualized learning at scale using analytics

THE VALUE PROPOSITION FOR DIFFERENT STAKEHOLDERS

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 are waiting in the wings to determine if we can draft a plan that is worthwhile. The learning analytics community 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 of communicating the value proposition in relation to problems of practice for each stakeholder.

- 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?

We briefly review why each stakeholder needs to take action and what that action should be.

Executive Summary

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 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 institutions.

"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 institutions."

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. ***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 National Science Foundation (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 processes, and incorporation of existing research and theory on the importance of context from the learning sciences. 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 services and solutions for improving educational practices and outcomes.

"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 services and solutions for improving educational practices and outcomes."

"Learning analytics 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 it may enable the study of questions researchers could not previously have imagined being able to ask (as is true in large-scale computational social science such as renowned Facebook studies)."

In order to make the most of current opportunities, ***university and nonprofit researchers*** will need to propose foundational research projects that solve key problems in the fields of learning analytics and education data science. This will allow for exciting new opportunities for basic research in learning sciences, psychology, and other social scientific fields related to personalized learning. 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. As learning analytics methods become more established, the field will have a dual potential. Learning analytics 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 it may enable the study of questions researchers could not previously have imagined being able to ask (as is true in large-scale computational social science such as renowned Facebook studies).

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 ***industry*** 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. 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. 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. Industry 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.

"Industry 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."

Finally, there is considerable value to be contributed by partnering with teams of ***educators and education leaders*** in advancing learning analytics and education data science, including 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 teachers know what their 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 are new roles in bridging 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 and the promise of progress in these fields for improving learning for all.

- How can teachers know what their 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?
- What are new roles in bridging technology and teaching?

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

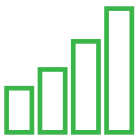
We envision new learning analytics systems and technologies becoming trusted metacognitive resources for learners through continued expanded data collection aligned with best practices and policies in data privacy, and improved design of instructional interventions. We hope not to prescribe learner pathways 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.

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

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 determined the necessary actions and identified potential areas where work is already being done and organizations to include as partners in this work. The four essentials are: 1) Human Capital, 2) Research, 3) Tools, and 4) Policy.

Human capital

We need universities enterprising enough to exploit the current developments in learning analytics, tackle the needs in education, and to 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.

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.

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. What follows is an overview of the funding recommendations around these four essentials to build the field of learning analytics.

Human Capital

Phase 1: Year 1

University degree, certificate programs, fellowships	• Network to develop training programs • Provide dissertation fellowships, pre-doctoral fellowships, PhD student internships • Provide faculty fellowships
Industry internships	• Internships for Ed Tech Professionals • PhD Students
Annual event: LASI	• Organize annual capacity building activity • Similar to Learning Analytics Summer Institute (LASI2013/LASI2014)
Preparing education researchers	• Develop short focused summer programs • Education researchers bring data and get support for analysis as they learn
Changing teacher and leader preparation	• Identify a group of quality teachers who use data successfully • Identify teacher preparation programs • Create plan for integrating Learning Analytics into programs

Phase 2: Year 2-3

Start-up accelerator center	• Develop a cutting-edge startup accelerator for analytics-driven research • Determine the best way to train people in the field
Integrate data-based decision-making into educator preparation	• Organize a committee to work with universities on implementation • Integrate learning analytics in teacher and school leader preparation • Integrate learning analytics in programs for in-service teachers and leaders

Phase 3: Year 4-5

Establishing university education data science programs	• Provide competitive awards for establishing university Education Data Science PhD programs • Develop programs that encompass departments of statistics, computer science, and education/psychology
Start-up accelerator center: worked examples	• Create a resource for teaching or for newcomers to the field • Provide a data set and worked examples that highlight the types of questions that can be asked and answered with different analytic techniques

Research

Phase 1: Year 1

Research to prevent reinventing the wheel	• Develop the What Might Work and Why Clearinghouse • Identify what has been done and what research is out there
Researcher & ed tech startup connector	• Identify a few Ed Tech startups to design competitions around solving real problems in education • Provide strong financial incentives to produce professional products
Case studies to inform capacity building and policies for learning analytics	• Identify and develop case studies that demonstrate how to build capacity and policies • Use case studies as tangible models for other districts and states to follow
Multimodal methods of measurement	• Develop multimodal learning analytics techniques • Researchers use multimodal methods to examine unscripted, complex tasks in more holistic ways
Measuring success	• Determine what mastery or success looks like in structured and less structured learning environments • Use new measures of success as outcomes
Prototype of personalized learning system **Optimization of personalized learning**	• Develop a prototype of a personalized learning management recommendation and reporting system • Selected schools/districts pilot the system and provide feedback • Determine explicit and/or tacit measures to serve as predictors of success • Use predictors in future analyses

Phase 2: Year 2-3

Center for learning at scale	• Develop a center that will conduct a longitudinal study that follows a group of eighth or ninth graders to college • Explore how to best collect longitudinal data that supplements classroom/learning data and take advantage of archival data

Phase 3: Year 4-5

Research to prevent reinventing the wheel	• Continue to identify what has been done and what is out there • The What Might Work and Why Clearinghouse to provide a continually evolving research-based guide to learning
Personalized learning pilot	• Refine the system based on user feedback and roll it out as a free pilot • Conduct projects to discover and validate/scale best practice for usage and visualization of data
Researcher & ed tech startup connector	• Continue participation in the Imagine K12 start up incubator program • Disseminate success stories • Create social network to link researchers and Ed tech startups

Tools

Phases 1,2 & 3: Years 1-5

Data science resource center

- Develop a center to create and provide a "Data Marketplace" or collection of datasets and streams for data scientists and developers
- Provide "Tools and Services" and help people use those tools to achieve their goals with big data

Competitions

- Incentivize innovation using industry models of competitions
- Identify the top 5-10 problems or grand challenges to be solved
- Competitions held in phases 2 and 3

Policy

Phase 1: Year 1

Templates

- Develop a set of templates for best data practices based on use cases
- Collaborate with the U.S. Department of Health and Human Services as the appropriate governance body for learning data policy issues

K-12 Data sharing and privacy standards

- Work with one or more states and a large district within each state to better understand how data sharing work
- Determine the issues, what is working, and effective methods for ensuring secure access to data

Phases 2 & 3: Years 2-5

Trust frameworks

- Develop and disseminate trust frameworks
- Develop tools, approaches, and practices for data sharing and privacy protection

Governing body

- Continue collaboration with U.S. Department of Health and Human Services to manage templates and create trust frameworks

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 along these funding lines. Some examples of milestones for measuring progress are provided in the following table.

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 ▶ Establishment of a governance body ▶ Publicly available templates for best data practices

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. Funding would be needed to sponsor talks at conferences with a data science area of focus, such as the O'Reily Strata conference, the O'Reily Open Source Conventions (OSCON), and SXSWedu conference. We recommend a collaboration with Strata and O'Reilly leaders to develop a StrataEdu conference.

Realizing the Promise:
IT TAKES A COMMUNITY OF FUNDERS TO BUILD A FIELD

We hope to have made the case with the LAW Report for the needs and the transformative potentials of making K-12 personalized learning a reality continuously improved by education data science and learning analytics. We encourage funders to consider an essential area aligned with their mission and interest and one or more activities to financially support in order to aid in this field building effort throughout the three phases of this work over the next five to ten years.

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.