Chapter 19: Data Literacy and Learning Analytics

Ellen B. Mandinach,¹ Lisa M. Abrams²

¹ WestEd, Scottsdale, USA
² Virginia Commonwealth University, Richmond, USA

DOI: 10.18608/hla22.019

ABSTRACT

Data use, whether through traditional methods in education or more sophisticated techniques such as learning analytics and educational data mining, has emerged as an important part of educational practice. Foundational to the use of data is data literacy; that is, educators’ ability to use data effectively and responsibly. A construct called data literacy for teachers has been operationalized and differs from assessment literacy to include the many diverse sources of data that educators now encounter. However, an issue, even with traditional data use is the extent to which educators have sufficient data literacy. The introduction of learning analytics presents the need for even more sophisticated data use capacity that may or may not be practical in most K-12 educational settings. This chapter explores the intersection of data literacy and learning analytics, and in doing so draws parallels between data use in the K-12 and post-secondary education settings, where data-driven decision making and learning analytics have traditionally been positioned. It provides a review of data literacy and the technologies that support data use. It discusses the practical challenges and constraints to transforming more traditional data use to include learning analytic strategies and how data literacy applies. The chapter then looks toward the opportunities and possibilities made possible by the sophisticated data use in learning analytics.

Keywords: Data literacy, accountability, continuous improvement, practical implications, challenges, opportunities

This chapter provides a link between data literacy and learning analytics (LA). It is our perspective that data literacy is fundamental to LA and educational practice, with LA being a sophisticated form of data-driven decision making (DDDM). The chapter provides a brief introduction to DDDM and data literacy and then a link to LA. It outlines the technology that supports DDDM, including implementation issues and challenges. It concludes with opportunities for DDDM and LA for research and practical next steps.

We first provide a foundation for the chapter by defining three key concepts.

- **DDDM** – the systematic collection and analysis of different types of data to inform decisions that will enhance students and schools [18].
- **Data literacy or data literacy for teachers (DLFT)** – “is the ability to transform information into actionable instructional knowledge and practices by collecting, analyzing, and interpreting all types of data (assessment, school climate, behavioral, snapshot, longitudinal, moment-to-moment, etc.) to help determine instructional steps. It combines an understanding of data with standards, disciplinary knowledge and practices, curricular knowledge, pedagogical content knowledge, and an understanding of how children learn” [17, p.2].
- **LA** – the use of measurement, data collection, analysis, and reporting to understand student learning and the learning environment through digital learning tools and technologies, intelligent data, and analytic models [3, 24].

Much has been written about DDDM, especially data use for accountability for schools and districts to measure progress as mandated by state and federal agencies. More recently the focus of DDDM has been on continuous improvement, although critics view accountability inextricably linked to data use. The purposes of DDDM is to provide an evidentiary base from which educators can make factual decisions to inform practice. Regardless of purpose, it is essential that educators have the skills and knowledge to use data effectively and responsibly; that is, they must be data literate [28, 29, 30]. Mandinach and Gummer developed a construct, DLFT that defines the skills, knowledge, and dispositions educators need to be data literate.

It is our position that data literacy must become a foun-
data collection and acquisition, storage, cleaning, integration, analysis, representation and visualization, and action(s).

Figure 1 illustrates some of the differences between the inquiry model in DDDM and the LA approach. The Figure represents an amalgam of models that exist so that they capture the essence of the many existing processes. However, the standard DDDM model infers that the impetus of the DDDM process is identifying an issue or posing an educational question. Although many of these steps can be found in DLFT, they are focused on high-level data skills and predictive models, and less so on the translation into just-in-time interpretations, actions, and decision-making.

Third, and by extrapolation, the statistical and technical skills needed for DDDM and LA differ as well as their level of complexity. Bowers [7] defines general categories of skills and topics for four job categories. There is little overlap with the traditional skills, even for administrators whose preparation involve more applied quantitative methods courses [7]. Although there has been no specific definition of the data skills for leadership in the DDDM literature, one can extrapolate from DLFT by modifying the fourth component, which focuses on pedagogical action to create administrative actions instead. The focus then would be on the decision-making skills, not research methods. Take for example what Bowers calls the data scientist, the focus is on educational data mining, LA, programming, design-based research, and technology and instruction. Only the last category, instruction, overlaps with DLFT. It is safe to say that data literacy in DDDM differs in how it is viewed and emphasized in LA and therefore leaves open opportunities for future research and development.

Fourth, DDDM and LA differ in a heavy reliance on technologies. The kind of analyses required in LA necessitates sophisticated technologies. Educational settings do have technologies [32, 50] such as learning management systems (LMS), data warehouses, assessment systems, data dashboards, and early warning indicator systems, but for many schools, even these technologies are too big, too expensive, or not practical. There is a push toward personalized learning environments that do have many technologies but again, this is impractical for many schools [9, 14] (Pane et al., 2017).

REVIEW

We documented central differences between LA and DDDM. This section reviews areas common to LA and DLFT practices to examine the extent to which LA theory and practice can be applied to and extend current DDDM and DLFT strategies.

Shared Purpose

LA and DDDM focus on the learner and learning. They share similar purposes to apply analytic strategies to improve learning. In LA, “analytics” refers to “software tools, machine learning techniques, and algorithms used for capturing, processing, indexing, storing, analyzing, and visualizing data” with the aim of improving learning and the learning environment [11, p.19]. Fundamental to DDDM is the use of diverse data, including achievement and behavior, to inform decisions about instruction to improve student outcomes. LA and DDDM function at the teaching and learner level to improve student outcomes. Both can involve decisions at all system levels [25]. For example, typical DDDM practices also can address: identifying student learning challenges; determining appropriate instructional responses; using parent and climate survey data to identify service needs; examining attendance, behavorial, and academic data to identify students at risk of being retained; and examining student course requests to refine instructional program offerings; [2, 33]. In this regard, the application of DDDM to multiple system levels
integrates LA with academic analytics and institutional analytics [11, 24]. LA and DDDM are intended to guide action to achieve the desired effect on student learning but also inform policies and decisions at the systems level.

Reliance on Data Management Systems

Both LA and DDDM rely on the use of data systems to support the analytic process. LMS collect and house information about student learning activities that are central to LA. The widespread implementation of course management tools provides a range of information about student learning. These data often serve as a proxy for student engagement. LMSs also include test results, discussion board postings, group interaction, frequency and duration of access, and overall progress in a course. These data are used to understand learners’ behaviors, engagement, and needs to improve student learning. Macfadyen [24] describes the benefits resulting from LA, including increased feedback to learning, enhanced student agency, better instructional coherence across courses; greater curriculum alignment, improved assessment of learning, and evaluation of teaching.

Likewise, the use of integrated data systems is common in K-12 settings. These systems enable educators to develop and administer assessments of student learning aligned with their instructional goals and content standards. Additional functionality can vary across school systems to include longitudinal data and predict future performance [52]. Data systems provide immediate feedback to teachers about student performance at grade, classroom, student levels according to overall or subgroup characteristics. Results can be used to support varied uses including identifying students’ strengths and weaknesses, grouping students according to ability levels, and determining appropriate remediation or re-teaching strategies [38, 44]. According to Farley-Ripple et al. [16], the DDDM literature organizes educator responses along conceptual and functional approaches.

Implementation Issues

Implementing LA and DDDM practices often requires shifts in resource allocation, increased capacity, and institutional cultures that promote inquiry-focused mindsets. The LA and DDDM literatures suggest that similar issues and potential barriers exist. Both require the investment of considerable human and infrastructure resources. First, technology and software tools are required to capture student learning data, administer assessments, and support the varying statistical analyses. Training is needed to learn how to use the systems to support data-informed instruction. Staff need increased capacity and technical skills to analyze, interpret, and apply information to learning issues. These technical skills are coupled with the need to develop in-depth understanding of data and time to engage in this work. These challenges are well-documented [10, 19, 23, 29].

Developing human capacity is a key component in DDDM [22, 42, 51]. Daniel [11] notes the lack of capacity and need for professional development in higher education that addresses both the technical and pedagogical knowledge needed to support different LA stakeholder groups. Daniel extends this discussion to the importance of institutional culture in ways that address potential resistance and privacy concerns about LA practices.

The cultural shifts required at large, post-secondary institutions are also necessary in K-12 settings. The DDDM literature documents the importance of context and school administration in encouraging effective data use. Supportive administrative actions include articulating clear and shared goals, establishing cultures and environments that value inquiry, and structuring time for discussing and analyzing data [36, 42, 51]. Institutional messaging about DDDM can have a powerful influence on the extent to which educators can realize its potential. For example, in schools and districts with a heightened focus on accountability, learning and improvement becomes lost. This fundamental dichotomy in how educators approach DDDM...
is characterized by key differences in culture - improvement versus compliance [1]. DDDM in compliance-based cultures is frequently highly prescriptive, involves the identification of students often on the cusp of passing (e.g., bubble kids) or located in a specific performance range, may be rife with inappropriate data use, and reflects superficial forms of inquiry, often detached from instructional practice. Alternatively, DDDM in settings with improvement or inquiry-focused cultures is closely tied to teaching, embedded in instructional planning conversations, and is related to professional mindsets concentrated on improvement and learners [37].

CHALLENGES AND COMPLEXITIES

This section explores the challenges and complexities of DDDM in classrooms and schools, providing a reality check to implementation and the data literacy needed for effective data use. We explore several relevant topics including the data needed for DDDM, data displays, the constraints of real-world educational settings and link each topic to data literacy and applications to LA.

Data

When most educators think of data, they think of test results. These are quantifiable measures that can reside in technologies. However, data are much more diverse than test scores. Educational data can be qualitative or quantitative and they extend beyond student performance. Mandinach and Gummer [29] have advocated for a broad definition of data that extends to demographics, socio-emotional, motivation, behavior, health, justice, special status (i.e., homelessness, foster care, military family, language learner, disability), to understand the whole child. With an increasing emphasis on data use and equity, Datnow and Park [12] have stressed the need to adopt an asset model that is based on understanding students’ strengths, interests, and contextual background, rather than a deficit model aligned with accountability. The whole child perspective not only broadens the notion of data but impacts how data are collected and where they are stored and accessed.

The broad perspective on data closely aligns with a foundational principle of data literacy, to use multiple sources of data to inform decisions. Educators must understand the importance of not just data triangulation but the need to examine contextual data to gain a comprehensive understanding of the student. According to the DLFT construct [29], data skills are informed by other sources of knowledge which include knowledge of learner and knowledge of context [46, 45], essential to understanding the whole child.

The use of multiple data sources is one area where DDDM practices lag behind the application and promise of LA. The LA literature includes a number of existing systems (learning management systems, student information systems) and tools that rely on and capture a wide range of student data, including demographic, behavioral, and academic information. Different tools have been developed that integrate with LMS to support students’ academic progress. For example, the Degree Compass system (Austin Peay State University) can assist in course planning; other technologies provide learners with feedback on their use of the LMS relative to peers according to class performance (University of Maryland Baltimore County, Check My Activity tool), and receive feedback about course performance, including alerts signaling a potential risk of failure. The latter is based on the Purdue University’s Course Signals system. Signals was frequently cited as an example of learning analytics. This “early alert system” relied on a predictive algorithm that included pre- and post-college admission data: high school grade point average, standardized test scores, socio-economic status, college course grades, frequency of advising appointments, and student use of the LMS to produce an indicator of risk, or potential, for failing a course [4, 24]. The system then alerted instructors and students as a form of early intervention. Even though initial outcomes of the Signals were promising, additional evaluation efforts revealed mixed impacts on student outcomes that when combined with implementation challenges lead to the closing of the program. A case study of Signals identified a number of factors important to future implementation of similar technologies including the: need for capacity and infrastructure to support timely integration of data; clarity of messages across courses and instructors; attention to timing and frequency of communications and impact on student motivation and learning outcomes; and the role of different institutional departments in education focused systems implementation and deployment [43].

Data Displays

More sophisticated technologies to support DDDM continue to emerge and have a long history [48, 49, 50, 53]. They are apparent in personalized learning environments [14, 39]. However, one concern raised is that there is little integration among the technologies that make the triangulation of data difficult for educators [34, 39]. Teachers have a difficult enough time with the data overload and triangulation [18], and personalized learning presents a larger challenge.

In contrast to the plethora of data from personalized learning, a recent trend is the creation of data dashboards and early warning indicator systems that present to educators targeted data [13]. Instead of bombarding educators with too much data, these systems streamline the data being presented to make them more readily interpretable.

To further complicate matters, there are critics of data systems more generally, commenting that the typical presentation format dumbs down the interpretation process and thereby misrepresents the data (Penuel & Shepard in [34]). The criticism is that many systems display data as a stop light with red indicating failure, yellow as cautionary, and green as passing and that this format distorts the meaning of the data that educators interpret the data in a cursory manner, and fail to be grounded in a theory of learning.

In terms of data literacy, understanding how to use data
technologies and the ability to understand trends and patterns are part of DLFT. But DLFT, even in taking the skill set to the most expert, does not likely extend to the level of sophistication required in LA [5].

Constraints

Many constraints exist in extending data use to the sophisticated level required of LA in terms of data literacy and educational realities. We raise several, but each deserves its own chapter. Thus, our goal is to raise the issues and accompanying questions. There are no easy answers.

First, what do we do with non-quantifiable data that do not readily fit into data systems, given the need for diverse data sources and how can LA accommodate such complexities? For example, how are data observed from the formative assessment process collected? How do we teach educators and data scientists to effectively use such data in their practice? Because of the need for diverse data sources, how can the firewalls across data silos for justice and health be overcome? What are the implications for the acquisition and protection of data from virtual learning environments?

Second, what can the field do to address interoperability issues and technology more generally? LA requires sophisticated data systems, whereas in most classrooms, such applications are not feasible. Cost is an issue. There is a knowledge barrier. The sophisticated skills and knowledge required of LA are not part of traditional educator preparation. Without denigrating educators, the more complex systems that exist in schools today may be beyond the grasp of many practitioners. Introducing the kinds of systems required by LA is even more of a stretch.

Third, how should the field handle the institutional diversity, considering that many districts can barely afford simple technologies, especially those that are small, rural, and charters? These schools must rely on more simplistic technologies and the ability to understand trends and patterns are part of DLFT. But DLFT, even in taking the skill set to the most expert, does not likely extend to the level of sophistication required in LA [5].

Fourth, how can we attain a sufficient level of data literacy among educators? As the National Forum on Educational Statistics [13] notes, educators need to know how to examine learner profiles, gain detailed knowledge of their students, and use diverse data sources with real-time, not just static data to understand student progress. They need to understand structured and unstructured data. Educators need to understand what data are needed for what purposes. Educators need to know how to discern trends and how to use the technologies to support data use. According to Bowers [7], educators need statistical knowledge, empirical reasoning, applied quantitative methods, and data visualization to personalize learning, and analyze performance patterns. Although some statistics are part of DLFT, most educators do not have the statistical literacy required of complex analytics. This is a major impediment. One could argue that data literacy is role dependent and that some educators may instead need to be good consumers of information, rather than hands-on with data. What is a sufficient level of data literacy? Should the field strive toward the level of expertise required of classifications Agasisti and Bowers [3] outline?

And if so, from where will the training come, given the dearth of DDDM being addressed in colleges of education [27] and the different foci from the best professional development providers, even with the emergence of data science courses at some universities there are fundamental questions around capacity building at both the pre-service and in-service levels, as well as the priorities of districts, given funding limitations.

Fifth, there are other general issues that exist in schools that may create challenges, what Jimerson et al. [21, 20] refer to as the enablers and challenges. Teacher time is an issue. Many think that DDDM is an add-on, not an integrated part of practice. DDDM requires too much time that could be devoted elsewhere. Educators need to be convinced of the value-added of DDDM and that it may not be just another passing fad. Thus, teacher beliefs play a role [15, 40]. Enculturation is important. Does a school have a data team and a data coach? Is data practice enculturated? Is there strong leadership that supports DDDM? Is there dedicated time for data work? All these factors make a difference [18].

Sixth, what are the ethical issues that surround the use of LA in DDDM and how do we prepare educators to use the data responsibly? With the large amounts of data and the technologies that support the data, there are ethical issues and threats to privacy that must be addressed. Wang [47] raises ethical issues around the use of artificial intelligence in DDDM that include unintended bias, a lack of humanism in decision-making, and moral values such as equity. Wang implicitly argues for the need for balance between the accuracy and efficiency of AI and the human considerations.

Finally, will educators use LA or know how to use it? How will the sophistication of LA translate to actual practice? Will educators know how to transform these data into decisions? This brings the issue of data literacy full circle. How do we prepare educators to use such data in a way that can effectively impact their practice?

These constrains are not trivial and should be considered thoughtfully about the implications for development and practice. Additional research is needed with full consideration for the state of current practice but with an eye to the potentials for future practice.

opportunities

Mandinach [26] discussed the challenges and opportunities (CHOPs) to DDDM in which the challenges far outweigh the opportunities but the opportunities far outweigh the challenges. The same situation should apply to data literacy and LA. We conclude with a forward-looking examination of the opportunities and consideration of what is possible in terms of building data literacy capacity for LA. We play off the challenges enumerated above and lead with specific topics.
How to Enhance Data Literacy

Data literacy will continue to be an issue for current and future educators. Capacity building is a highly systemic issue [28, 29], one that must be addressed by professional organizations, educator preparation programs, professional development providers, and local and state education agencies. The accumulation of data literacy skills, knowledge, and dispositions should be an ongoing process across the entire trajectory of educators’ careers, beginning during pre-service and reinforced through professional development, in-service training, and graduate courses. Data literacy, both basic and more advanced for LA, provides several kinds of opportunities. For colleges of education, it provides an opportunity to integrate DLFT into their courses, and perhaps in LA and data science. For professional development providers, it creates new opportunities for trainings. For professional organizations, it provides opportunities to reconsider the skills sets that are necessary parts of educators’ repertoires. For research and development staff, there is a need to create materials that can be used to build capacity, something for which the first author has advocated for years [35].

A Vision for Better Data Displays and More Effective LA

As Bowers et al. note [8], there is not only a need to build research and analytic capacity around data use in schools, but also to develop innovative data products that can help educators extract meaning from data displays and interpret data. LA requires sophisticated data displays. They must go beyond the stop light form of presentation and incorporate the diverse data sources we have discussed. They must make the data easily accessible, understandable, analyzable, and interpretable and provide reports that can be implemented and readily translated into actionable steps to inform practice. These characteristics require thoughtful design considerations that make the technologies attractive to and useable for educators, without sacrificing complexity. Such design constraints provide opportunities for the development of both sophisticated and easy to use technologies that will facilitate effective use of data.

How to Capitalize on Diverse Data Sources

Educators need rich and diverse data to address the complexities of the whole child. As noted above, interoperability and cost are issues. Systems need to adapt to both qualitative and quantitative data. We advocate for the broadest possible use of the diverse data. Such rich data provide untold opportunities for educators to gain a more comprehensive understanding of their students and educational situations. The diverse data also provide the means of moving the needle from a strict accountability focus to one that focuses on the whole child, continuous improvement, and understanding context beyond the school walls that impact students. If LA can provide the expertise to explore the full range of data sources, it would benefit the field.

A Challenge to the Field

According to Bowers [7], LA is a sophisticated form of DDDM that can enhance the use of evidence in education. The question remains whether the level of sophistication required in terms of data literacy, the needed technologies, and other skill sets such as statistical literacy, are realistic in educational settings. Mandinach [26] questioned, what is the least amount of data literacy that is acceptable for educators. The discussion at hand falls at the far end of the continuum of expertise in terms of whether educators should aspire to the data expertise required of LA and the roles and responsibilities of such individuals in typical educational settings. With increasing complexities come certain risks, over-analysis, and potential ethical and moral problems, as noted by Wang [47]. The challenge for the LA and DDDM fields is how to harness the potential value of LA and ensure that educators not only know how to access, analyze, and interpret complex data, but more importantly, how to transform those data into actionable educational practices. This is the essence of data literacy. Fundamental questions remain, whether all educators need to have a high level of sophistication, and what are the practicalities of adopting LA approaches in DDDM educational practice.

REFERENCES


