

LEARNING ANALYTICS CONSIDERED HARMFUL

Laurie P. Dringus

Graduate School of Computer and Information Sciences
Nova Southeastern University

ABSTRACT

This essay is written to present a prospective stance on how learning analytics, as a core evaluative approach, must help instructors uncover the important trends and evidence of quality learner data in the online course. A critique is presented of strategic and tactical issues of learning analytics. The approach to the critique is taken through the lens of questioning the current status of applying learning analytics to online courses. The goal of the discussion is twofold: (1) to inform online learning practitioners (e.g., instructors and administrators) of the potential of learning analytics in online courses and (2) to broaden discussion in the research community about the advancement of learning analytics in online learning. In recognizing the full potential of formalizing big data in online courses, the community must address this issue also in the context of the potentially “harmful” application of learning analytics.

KEYWORDS

learning analytics, online learning, evaluation, effectiveness, participation patterns, threaded discussions

I. INTRODUCTION

A. Background

Heightened interest about “big data” and learning analytics is building among educators, institutions, government sectors (e.g., The U.S. Department of Education), and accrediting groups. The *Chronicle of Higher Education’s 2011 Special Section on Online Learning* [1] reported these constituents are increasingly calling for big data to show clear patterns of learning and persistence in online courses and degree programs. Big data will soon, if not already, be used to make important decisions about student federal aid regulation, student persistence, graduation rates, quality learning, and institutional policy-making on effective online teaching and learning practices. Kelderman [1] reported that accreditors are attempting to keep pace with new federal regulations to provide tighter oversight on online programs, “requiring colleges to prove that students learn as much in distance courses as in face-to-face courses” (p. 4). These requirements increase the pressure on educational institutions to respond to new regulations and provide clear assessments of quality online education. Moreover, the instructor has a need to know what is happening in the online course; the use of learning analytics would produce information about student progress and the instructional process. Looking at student progress, the instructor may have to produce a record of a “data trail” to inform school administration of students’ progress or persistence in the course. Looking at the instructional process, the instructor may see patterns of poor class performance on a newly implemented assignment, which may lead to timely instructional intervention or “alteration of course content” [2].

The entire issue of *EDUCAUSE Review* (September/October, 2011) [3-7] focused on awareness building of the value of learning analytics in education, giving an overall impression that standard, measurable, reliable, and meaningful methods have not been formalized for managing and interpreting big data in online courses. Predictions from the *Horizon Report* [8] included that learning analytics will be a new

priority of focus in higher education in the next five years. Siemens [9] indicated that in its broader context, learning analytics is concerned with curriculum mapping, action, prediction, intervention and other processes; these areas are also important to formalize in online courses. In this regard, the online learning community needs to guide the direction as to how learning analytics are used in defining and evaluating big data in online courses. This guidance includes the need for defining data, developing learning analytics methodologies and tools, visualizing and sharing the nature of learning analytics output, and informing effective process and practice that leads to meaningful decision-making about learner performance.

This essay is written to present a prospective stance on how learning analytics, as a core evaluative approach, **must** help instructors uncover the important trends and evidence of quality learner data in the online course. The overarching must statement (above, in bold) is predicated on the basis that the online instructor must know what is going on in the online course [10, 2]. The context is in reaching the trenches — getting down to the course level [2] where online courses accumulate hundreds, if not thousands, of data points – data points that leave a history and “data trail” of participant activity from discussion postings, emails, chat or synchronous classroom recordings, assignment submissions, instructor feedback, and from many other artifacts. The data trail from the artifacts of online course production is the stuff that brings life to the online course. In phenomenology, the study of “essences” is meaning derived from “the description of the lived-through quality of lived experience and the description of meaning of the expressions of lived experience” [11, p. 25]. The data trail, across the online course, may be viewed as the culmination of the “lived experience” of the learner and the instructor. That experience is hidden behind the online environment and the technology. The lived experience of the online course participant – student and instructor – needs to be visualized as a data trail that is traceable and interpretable.

Currently, the ability to capture the full potential of the lived experience in online courses is limited in that the learning management system (LMS) does not present data in meaningful output for the instructor, nor is the instructor given usable and meaningful tools [10] to manage the data. The inability of the instructor to manage abundant data hinders the instructor to assess, in various visual and flexible ways to a full extent, a learners’ progress in the online course. The challenge is onerous for the online instructor when searching for data that either is not there or is very difficult to access and retrieve efficiently. The large and unwieldy “data dump” of transaction level data offers little value for the instructor if the yield of data persists in its raw form.

The problem is significant in that online courses produce a large and unmanageable amount of data [12] for which the instructor cannot work flexibly with to assess a learner’s progress in an online course. The learning management system is insufficient [10] and inefficient in producing a meaningful and traceable data trail of online course participants. The lack of transparency [7] and undefined data in the online course may result in poor decision-making about student progress and performance.

This paper brings into focus a critique of strategic and tactical issues of learning analytics in online courses. The critique is limited to issues related to information retrieval, transparency and usable data. Some intervention strategies are addressed, although the purpose of the paper is to build awareness of what data are needed in online courses. The main context is contained to one aspect of the online course: student progress in an online course and evidence of quality measures in online discussion forums. The approach to the critique is taken through the lens of questioning the current status of applying learning analytics to online courses. The goal of the discussion is twofold: (1) to inform online learning practitioners (e.g., instructors and administrators) of the potential of learning analytics in online courses and (2) to broaden discussion in the research community about the advancement of learning analytics in online learning. In recognizing the full potential of formalizing big data in online courses, the community must address this issue also in the context of the potentially “harmful” application of learning analytics.

B. On ‘considered harmful’

The popular phrase “...‘*considered harmful*’” in the title comes from Dijkstra’s (1968) [13] *Communications of ACM (CACM)* letter *Go To Statement Considered Harmful*, in which he criticizes the

excessive use of the GOTO statement in programming languages and also builds a case for a change towards structured programming. Technology articles have included this “considered harmful” popular phrase in the title as an attention grabber to offer an opportunity for debate or criticism of something that needs consideration. One such article that greatly inspires this current paper is a “considered harmful” essay by Greenberg and Buxton [13] on usability evaluation. They questioned the status of usability evaluation and challenged the Human-Computer Interaction (HCI) community, by calling for a change in how usability evaluation is viewed and practiced. They indicated that, “The phrase ‘considered harmful’ signals a critical essay that advocates change” (p. 111).

Although Greenberg and Buxton explored usability evaluation, a topic unrelated to learning analytics, there is a general statement from their article that could be highly relevant to apply to learning analytics, if we reframe and explore the general statement in this new context. Greenberg and Buxton’s statement (abstract, p. 111) resonates: **“Yet, evaluation can be ineffective and even harmful if naively done ‘by rule’ rather than by ‘thought’.”** This profound statement resonates for the online learning research community in that the statement could be re-addressed through raising issues and questions about learning analytics and its application and value to online learning. What learning analytics can do or how it may help or hinder the assessment and evaluation process is important to ascertain at this point in online education development. Habrowski, Suess, and Fritz [5, p. 16] stressed the need for “higher education to transform its own culture.” Information technology should be used to apply rigorous approaches to analytics in “supporting evidence-based decision-making and management” [7, p. 4]. In similar context, the online learning research community must bring transparency to effective practice of learning analytics to deter potentially wrongful uses of big data in online courses. Greenberg and Buxton’s [13] general evaluation statement is strategic and tactical in nature and is worth exploring and bringing to the context of learning analytics, to build awareness and spark discussion in the online learning research community about evaluation practices.

II. EXPLORING THE EVALUATION STATEMENT

To explore Greenberg and Buxton’s evaluation statement, each element of their statement is broken down into core words and is explored in the context of applying learning analytics to the online course. The “online course” is addressed broadly through discussion points that are most often framed through “If” or “Would” or “Could” discussion points. Reflections on discussion points follow with related work to support the general nature of the problem and some examples from the online discussion forum to explore the extent of the problem. The problem is highlighted through the lens of the following five “must” statements pertaining to effective learning analytics in online courses. These must statements correspond to the current status of data in the online course or the online discussion forum and what is needed in learning analytics to advance the understanding of students’ data trail and their progress. The must statements are geared to practitioners and researchers as a challenge to address an array of strategic and tactical issues that are currently germane to big data in online courses.

Must statements for learning analytics in online courses:

1. Effective learning analytics in online courses **MUST** develop from the stance of getting the right data and getting the data right. What is meaningful data?
2. Effective learning analytics in online courses **MUST** have transparency. What do we see?
3. Effective learning analytics in online courses **MUST** yield from good algorithms. What are we looking for?
4. Effective learning analytics in online courses **MUST** lead to *responsible* assessment and *effective use* of the data trail. What do we do with the data?
5. Effective learning analytics in online courses **MUST** inform *process* and *practice*. How do we improve the online experience?

A. Getting the Right Data and Getting the Data Right

Effective learning analytics MUST develop from the stance of *getting the right data and getting the data right*. What is meaningful data?

“Yet, evaluation can be ineffective and even harmful if naively done ‘by rule’ rather than ‘by thought’”

Discussion Points — Learning Analytics (in the context of evaluation in online courses) can be ineffective:

- If the data that captures what learners and instructors “throw off” (data trail) in the online course has *no impact* on improving or changing instructional design or practice.
- If the data that captures what learners “throw off” in the online course or in the online forum produces *no meaningful evidence* of learning or non-learning.

Brown and Diaz [3] stated the need to understand “evidence” and “impact.” The discussion points relate to the problem of heavy reliance on production information, which is most often in the form of raw data – in the discussion forum there are frequencies, number of postings, time stamps, subject lines, and message identification numbers. There is an abundance of this type of temporal data. This “low hanging fruit” level data is accessible [10], but is only partially useful and extractable from the LMS. Results of a 2009 EDUCAUSE Center for Applied Research (ECAR) survey [5, p. 22] of IT staff and educators indicated that, data-wise in approaching learning analytics, analysis is at “stage one of extraction and reporting of transaction level data.” To address evidence and impact, analysis should reach “stage five automatic triggers or alerts” (p. 22). Stage five would inform meaningful evidence of learning or non-learning; there would be strong visible evidence for “success,” “at-risk,” or “failure” in determining the data trail of students’ progress.

Evidence and impact as described by Brown and Diaz requires extraction of semantic information [10], as this is much more useful in determining a need to change course content or to invoke instructor intervention. In the discussion forum, social networking analysis (SNA) may map social connections and discussions, but the utility of this bears on relevant knowledge construction evidence that shows there are either alarming gaps in students’ focus on a topic discussion or that there are positive nodes of being on target in the discussion. Gaps inform the instructor that intervention or change may be needed in the online discussion. Nodes or groupings of postings that show students are on-target help to validate the instructional design aspect of the assignment or the learning content. De Liddo, Shum, Quinto, Bachler, and Cannavacciuolo [14] emphasized discourse-centric learning analytics showing conceptual and social network patterns at the level of individual learners and groups. Zhang and Sun [15] identified a comprehensive set of measures for online knowledge building discourse (p. 74), including social interaction patterns, content measures, and lexical measures. Information relevant to knowledge construction and gaps could help improve or change instructional design and practice. As knowledge construction evidence is more meaningful to describe students’ progress, production information alone, even in big numbers, reveals very little. Hrabrowski, Suess, and Fritz [5] stated the essence of the problem: “Finding the meaning, significance, or even a recurring pattern or trend in a mountain of data at our disposal is difficult” (p. 22).

Perhaps the starting place for considering the effectiveness of learning analytics in online courses would best develop from the stance of *getting the right data and getting the data right*. This is a play on words from Greenberg and Buxton [13, p. 115] in their stance on ‘getting the right design and getting the design right’, in the context of usability evaluation and user interface design. The premise is that more data is not better data. Sometimes the lack of data is worse than having too much data, and worse still, is having data and not knowing what the data means or how to manage it. Duval [16] indicated that a problem around learning analytics is the “lack of clarity about what exactly should be measured to get a deeper understanding of how learning is taking place” (p. 14). What are data in an online course [17]? What are

data in an online forum [18]? How shall progress indicators that flow from the online course or online forum be defined and measured? Will the indicators be recognizable in transparent, traceable, and extractable forms? These questions lead to considering transparency of data and what is visible in the data trail.

B. Transparency

Effective learning analytics MUST have *transparency*. What do we see?

“Yet, evaluation can be ineffective and even harmful if naively done ‘by rule’ rather than ‘by thought’”

Discussion Points — Learning Analytics (in the context of evaluation in online courses) can be ineffective:

- If the data does *not show* a data trail about learners as a way to *visualize* signs of their “success,” “at-risk,” or “failure” in the online course or online forum.
- If the data output from the forum is *fragmented* – the output lacks synergy or is static in nature.
- If the data produces extensive patterns of *usage* and *activity*, but there is no *flexibility* for the online instructor to query a specific pattern.
- If the dashboard (the user view of data output) is considered usable, but the information presented is *useless*.

The LMS does not provide an adequate data trail in the form of identifiable and actionable information about learners so as to visualize signs of their “success,” “at-risk,” or “failure” in the online course or online forum. For example, in the discussion forum, last access data that shows current or inactive participation is insufficient in revealing the true status of a student’s progress. For example, a student may be inactive in a forum, but may be active (off-line) in the course. Hrabowski, Suess, and Fritz [5] suggested that students should be given their own “compass” in that they should be able to follow their own plan and their path of progress. A meaningful compass for students is practically non-existent in the LMS. Shea et al. [19] suggested that identifying patterns of learning presence behaviors exhibited by students (e.g., forethought and planning, monitoring, and strategy use) would be beneficial in following students’ progress across the online course. Figure 1 shows a mock-up of an extended dashboard of online course activity. In its dynamic form, each activity section of the dashboard would be selected to further query and generate reporting about a certain activity.

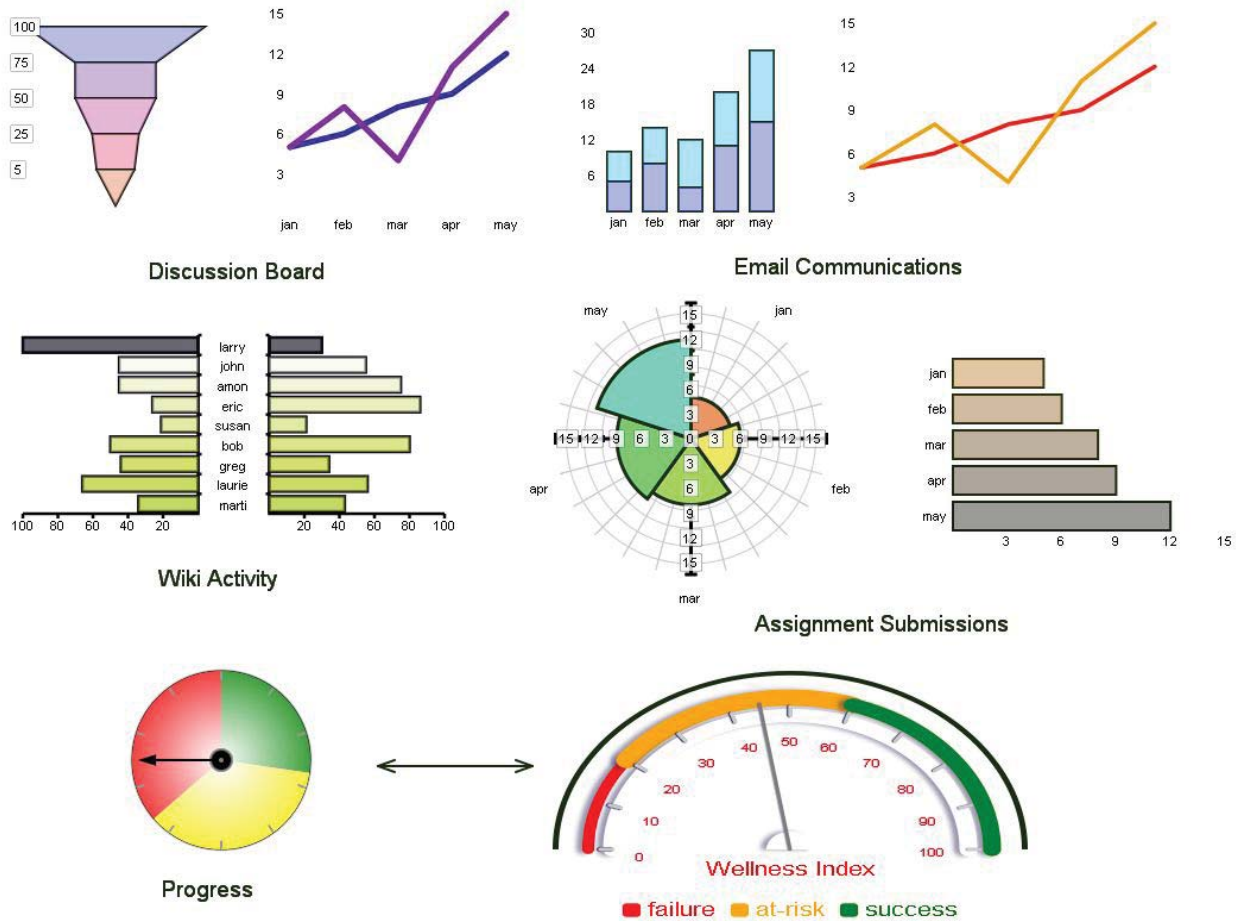


Figure 1. An extended dashboard of online course activity

Ali et al. [10] stated: “Educators need comprehensive and informative feedback about the use of their online courses” (p. 470). The data output from the discussion forum is most often fragmented. Fragmented output lacks synergy as the data are raw in form and in raw form do not compare and contrast with other data. Poor information visualization results in output that lacks a clear and evolving view of students’ progress. For example, in Blackboard, the static nature of pie charts that show nice colors and small-to-large pie slices of participation by topic is information of limited value. Another view (a table) reveals how many postings a student made per month. Another view (a table) reveals posting frequencies across topics. Together none of these views are synthesized, and in single views, none show important pieces of students’ progress. In the discussion forum, data can be compiled by the instructor’s grouping of selected postings (i.e., by checking on boxes associated with individual postings). Standing alone, basic analytics on technology usage from a page view [10] lacks comprehensive and informative feedback. That data cannot be viewed in different forms or meaningful contexts unless the instructor imports that data into an external software analysis program, such as a social networking analysis tool [14], and forms queries using that software. Poor information visualization and data fragmentation hinder the ability to view data differently and in meaningful patterns.

If the instructor is able to query the LMS for a student’s data trail of progress, there must be flexibility for the online instructor to query a specific pattern of progress, beyond superficial patterns of usage and activity. For example, an instructor may want to follow a student’s activity in terms of timeliness of postings and responses, but the output from LMS does not report both time and response level patterns

within a certain thread of discussion or across many threaded discussions. Wise, Marbouti, Speer, and Hsiao [20] identified several meaningful variables that measure students' interactions in discussions (e.g., percent of sessions with posting actions). Other more comprehensive patterns related to critical inquiry (e.g., the process for reaching the resolution stage) [21, 22] and metacognition [23] may involve several codes and extensive queries. A range of simple to complex variables is of interest to query in the discussion forum.

The Blackboard dashboard offers basic parameters: last name, first name, username, role, date/time of last login, days since last login, review status, adaptive release, and view grades. Dashboards are giving some information in transparent form, but suffer from superficial and limited parameters. Hrabowski, Suess, and Fritz [5] indicated that dashboards have become “window-dressing” as dashboard indicators lack key factors of assessment value. Visualization and log data capabilities are limited in today's LMS. New tools that work in conjunction with LMS output in discussion forums, such as SNAPP, (Social Networks Adapting Pedagogical Practice) provide some value in viewing a portion of a topic discussion; at least in beginning to provide a sociogram of temporal participation interaction. The goal of tools like SNAPP is to automate learning activity by combining content analysis and social network analysis [24]. LOCO-Analyst [10] is a tool developed to provide educators with feedback on students' learning context data, including feedback from learning objects and students' interactions. These and other tools seek to advance meaningful and usable coding tools integrated into the online course or discussion forum, but much more work is needed to extend transparency, usability, and the scope of content analysis activity from these tools.

Transparent coding tools need to advance to help online instructors manage and use the information from a discussion forum. Beyond getting the data, how the data are used for intervention is an important aspect of transparency as well. On the intervention side, “nudge” analytics, (i.e., subtle interactions that influence students to be on target) [25], may be one approach to expand transparency. There are also data mining techniques and tools that are under development to filter learning and interaction patterns visually to promote intervention or student-directed feedback [10, 26, 27, 28].

Transparency is significant in that students' progress must be revealed in real time or from any point of time in the instructional and learning process. This requires on demand, highly visible, and flexible views to work from. What should a dashboard give us? A meter of progress (or gauge for wellness) showing the momentum of a discussion forum [18] may reveal how things are going in topic discussion. On a higher level of knowledge construction, when is social presence, specifically in context of interactive indicators such as acknowledgement, agreement and approval [29], at its high point or low point? Duval [16, abstract, p. 9] stated that information visualization techniques are needed in the dashboard so that learners and teachers “no longer need to drive blind.” The dashboard of today's LMS is insightful about some level of participant activity, is less insightful about student progress, and is even less insightful about knowledge construction or level of critical inquiry. The ability to perform a meaningful content analysis is lacking in the LMS; also lacking in the LMS is the transparent meter of student progress, the data trail, from which content analysis can be performed.

C. Good Algorithms

Effective learning analytics MUST yield from *good algorithms*. What are we looking for?

“Yet, evaluation can be ineffective and even **harmful if naively done ‘by rule’** rather than ‘by thought’”

Discussion Points — Learning Analytics can be harmful if naively done by rule:

- If the rule is “show time stamps and number of postings” as evidence of student participation—the “low hanging fruit” data from the forum can easily become “direct measures” and “evidence of persistence.”
- If the rule is “capture all presence activity of the instructor,” in what ways will the visual presence capture the breadth of activity?

It is unclear what direct measures and evidence of persistence exists in the online course beyond the “low hanging fruit” data of access and posting activity. Accreditation boards and federal constituents may impose new rules or require certain progress indicators to show quality is taking place in online courses. For example, these groups may impose mandates for reports of students’ level of participation in an online course for students who may be receiving financial aid. What data are being reported to higher education administrators as a quality online learning experience? Does data in accreditation and assessment reports primarily consist of superficial direct and indirect measures of quality? Long and Siemens [6] raised an important question: “How can the potential value of the data be leveraged without succumbing to the dangers associated with tracking students’ learning options based on deterministic modeling?” (p. 38).

Direct measures in online courses must include creative indicators as evidence codes that actually describe engagement, not interaction activity alone. There is a need to match ‘low hanging fruit’ transaction data with high-level data. Also, there is a need to evaluate engagement across several collaborative tools inside and outside the LMS. Long and Siemens [6] stated that “most LMS analytic models do not capture activity by online learners outside the LMS (e.g., Facebook, Twitter, other),” p. 36. Evidence-based indicators for Twitter activity may be different contextually from indicators for wiki activity, forum activity, and other collaborative initiatives. Suthers and Rosen [30] addressed learning analytics in the context of distributed learning and fragmentation across multiple media and sites and the need to work out disassociations of activities that are fragmented across multiple logs. Largely, the emphasis on content analysis has been on text-based discussions, but with the increased use of audio, video, and graphics embedded in online discussions, in addition to text mining from the discussion forum, for example, how are data mined from an audio posting?

The focus of the data trail and activity has been mainly on the student. What is the visible and meaningful presence activity of the instructor and how is that data trail interpreted by administration as effective performance? What data in the online course represents a true capture of the effort of the online instructor [31]? Time-on-task and breadth of activity cannot be gauged within the LMS alone. Vatrappu, Teplov, Fujita, and Bull [32] recommended that teachers adopt “teaching analytics” for just-in-time teaching and assessment, requiring an intuitive and visually powerful representation. If rich information is available to the instructor, it is uncertain how data, in revealing the instructor’s effort and performance in an online course, will lead to interpretations and decisions about class sizes, course loads, and facilitation practices.

The potential “harm” in using big data for evaluating “naively by rule” impacts instructors and students not only from a production level, but also impacts the quality of the online learning experience. The stability of online learning depends on sustaining a quality experience. Sandelowski [33] described the partiality of interview transcription and how partiality alters reality. She described “reality” in that, “what ends up on the printed page – the raw data – is actually already partly cooked,” having already changed the original experience from what it is originally intended and what “was supposed to be preserved” (p. 312). In the online discussion forum, most often the breadth of the lived experience, if to capture fully, cannot be solely quantified. Big data must also be evaluated qualitatively. Perhaps there is no such thing as raw data, but that through transforming written words from the online forum, the challenge is how the transcription will capture the real story of the experience.

D. Responsible Assessment and Use

Effective learning analytics MUST lead to *responsible* assessment and *effective* use of the data trail. What do we do with the data?

“Yet, evaluation can be ineffective and even **harmful if naively done ‘by rule’** rather than ‘by thought’”

Discussion Point — Learning Analytics can be harmful if naively done by rule:

- If the data that shows *absence* from the online course or the discussion forum is interpreted *only* as “drop” or “no show” or “inactive.”

The essence of the data trail is the revealing of human behavior [2]. Siemens stated: “An integrated system should be able to track my physical and online interactions...” (www.elearnspace.org/blog). Learning trails can have a surveillance aspect that can be harmful. Brown and Diaz [3] raised an essential question: “who will consume our evidence”? (p. 50). Privacy is a concern when “detailed data of learner interactions are tracked” [16, p. 14].

Absence is a vague and undefined term in online courses. Absence in the online course does not necessarily equate to inactivity, non-performance or no progress by the student. In the online discussion forum, for example, an online learner is a reader as well as a contributor. Lurking or reading behavior is probably one of the most prominent activities in the discussion board, yet there is no reliable way to measure it meaningfully. Lurking is bundled with inactivity data; lurking data needs to be dealt with productively. Dennen [34] stated that learning is an invisible practice, but lurking can be a productive practice. Dennen described one way lurking data can be identified is through “encouraging students to document their reflective acts” (p. 216). Is there a way to cull silent activity from blogs that the students keep for the course? Wise et al. [20] characterized student interaction patterns through students self-monitoring and their listening behaviors. Absence needs to be more concretely defined while also acknowledging the breadth of listening, reading, or lurking behaviors, or the behind-the-scenes activities that exist both inside and outside the LMS. (That invisible activity is an essential part of describing the lived experience in the online course.)

Responsible assessment and effective use of the data trail are essential to the advancement of understanding how learning transforms in an online course. The potentially harmful aspect of learning analytics in evaluating progress in online learning is that poor decisions will derive from what data are visible and extractable in the LMS and from ill-defined indicators of progress. Dzuiban [4] stated the core of the problem in simple and profound terms: “uncollected data cannot be analyzed” (p. 48).

E. Process and Practice

Effective learning analytics **MUST** inform *process* and *practice*. How do we improve the online experience?

“Yet, evaluation can be ineffective and even harmful if naively done ‘by rule’ **rather than ‘by thought’**”

Discussion Points — Learning Analytics by thought:

- Could lead to data approaches and tools that *support creative ways to reflect* on the dynamics of the online experience.
- Could lead to *definable codes and specific descriptions* from indicators identified in the research literature that can be built into the LMS database for extraction and content analysis.
- Could provide *good algorithms* for big data showing a flexible and meaningful *wellness index* (i.e., the general “health” of the forum).
- Could lead to *good self-reflection*: Meaningful and big data across courses and across time frames may reveal to the online instructor, his or her body of work or process in the online course environment. (What am I doing? How am I doing?)
- Could lead to *traceable Communities of Practice*: – a way to collect social capital through conversation routing and knowledge profiles.

Creative and innovative approaches and tools in learning analytics may also transform understanding of the value of online learning. To embark on expansive data collection of the data trail, there is a need for a data warehouse to manage a wide net of indicators that can be applied to evaluate contextual data from various artifacts in online courses. In basic terms [35], a data warehouse provides benefits that include maintaining data history, consistent codes, integration of data from multiple source systems and the ability to restructure data for extended query performance. With indicators available from a data warehouse, new approaches can be applied to evaluating contextual and metadata from the online course. For example, Dennen [34] suggested microethnography as a potential methodology; this requires extensive data collection. Learning analytics with microethnography could be useful in examining online discussions and a wide variety of contextual data (from interviews and field notes), to provide thick descriptions of both the discussion and the contextual surroundings (p. 215). Fournier, Kop, and Sitlia [12] recommended that what researchers might consider rich data would be better analyzed through qualitative methods. They suggested that virtual ethnography would bring analysis to a level where the reflection is as close as possible to “what is happening in the chosen setting” (p. 105).

The online learning research community has advanced in the past decade in constructing meaningful constructs, models, and methods for evaluating online learning progress. Using the Community of Inquiry (CoI) [22] as an example, indicators derived from the research literature in studying the presence factors (cognitive (c), teaching (t), and social presence (s)) offer an array of meaningful and measurable qualities of productive learning and communication in online learning environments. These qualities (such as facilitation (t), group cohesion (s), and resolution (c), etc.) need transposing into essential data types or codes in the database, i.e., the information that gets parsed as data in code form and is ‘mined’ to draw out coherent patterns. Supporting rubrics or instruments (e.g., the *Community of Inquiry Survey Instrument* [36]) that measure the presence factors should be built into the LMS. The need to federate the data [37] and provide normalization in defining the variables [37] is essential in turning raw data into coherent and usable data.

Broad interpretations of student progress result from not having definable data types or codes in the LMS database. The solution would require having common or standard codes or tags that define value of the content in an online discussion. For example, Dringus and Ellis [18] suggested useful codes for the discussion forum may include a code for “persuasive argument,” a code for “topic shift,” a code for “off-topic,” or a code for “chit chat.” The potential for defining codes or tags is unlimited in terms of how quality discussion is assessed or evaluated by the instructor. The lack of data types in the LMS prevents good decision-making in determining to what extent an instructional practice is of value to continue. As there are potentially hundreds or thousands of data types that can be parsed for content analysis, there is an urgent need for the online learning research community to derive evidence from the research and provide definable codes and specific descriptions that can be built into the LMS database for extraction and content analysis.

Could big data help gauge a flexible and meaningful wellness index (i.e., the general “health” of the forum)? If the dashboard, such as “check my activity tool” [6, p. 36] or a wellness index [18] produces a meter of progress, in real time, that shows the momentum of a forum discussion, such could help students gauge their own level of progress as well. Dringus and Ellis [18] suggested a wellness index could reveal interesting time pattern analytics of student participation: duration, time-on-task, flow consistency, gaining momentum (discussion), losing momentum (discussion), and latency of topic discussion. How is the lived experience captured from the online discussion? What is the life of a forum? What is the life of a topic? What is the life of a topic over time?

Big data and learning analytics could lead to *good self-reflection*, showing instructors their own impact in the online course. “What does productive instructor effort look like in an online course” [31, abstract]? Self-reflection requires good data to sort out these questions: How am I doing as a facilitator? How well is the forum progressing? Do I provide frequent feedback? Is my work of satisfactory quality? What changes do I need to make to be a better online instructor? These questions of self-reflection require longitudinal data across time and across courses. Shea, Hayes, and Vickers [31] discussed formative and

summative assessment must go beyond the threaded discussion. Assessment must go across the course quantitatively and qualitatively—in discussions, assignments, course design, and various learning activities. Long and Siemens [6, p. 38] stated: “Analytics in education must be transformative.” Swan [38] stated that actionable reporting is warranted.

Communities of Practice (CoPs) are the advanced online learning environments of the future. CoPs offer a way to collect social capital through conversation routing and knowledge profiles [39]. The success of a CoP success is characterized by several integral elements—a *domain* identity by shared interests, *community* with joint activities, discussions, helping one another, sharing information, and *practice* – a shared resource of experiences and expertise [40]. In moving from single online courses to online learning communities in domain areas, big data will be needed to help track and coordinate successful and sustainable collaborative online learning experiences.

The last part of Greenberg and Buxton’s general statement, ‘rather than by thought,’ opens the dialogue for recognizing the benefits of learning analytics in online learning. Learning analytics must inform process and practice, giving the research community the opportunity to determine the value of learning analytics and to address tough questions: How are we doing in online learning? What has improved in online learning? How do we describe our reality of the online experience? In the phenomenological context, to produce lived experience descriptions, capturing the essence of the reality involves “describing the experience from the inside in” [11, p. 66].

III. CONCLUSIONS/DISCUSSION

A. Learning Analytics ‘by thought’

Returning to the title, “Learning Analytics Considered Harmful,” learning analytics can be considered harmful *if* a strong effort is not made in procuring responsible assessment and effective use in online learning. Priority attention is needed from the online learning research and practitioner community to consider ways learning analytics will influence or change the nature of academic output in the online course. Several ‘**by thought**’ priority issues come into focus. Stakeholders should be mindful that:

1. Our data trail depends on the expanding nature or wide net of indicators (i.e., our inquiry base) that educators will need to normalize to assess engagement and learning. Large data sets are already being derived from activities from within the online course and the participants’ data trail from these activities. Our definition of ‘data trail’ may dictate (in a big way) what we study and understand about the online experience.
2. There will be ever increasing large volumes of data evolving from online courses. The benefits of big data can only be obtained through good learning management, reliable data warehousing and management, flexible and transparent data mining and extraction, and accurate and responsible reporting.
3. Big data will influence interpretations and decisions about student progress and about the quality of online learning in general, and
4. Those interpretations and decisions may impact (potentially) our ‘control’ over the quality of the online experience.

B. Final Comments On the Evaluation Statement and the Must Statements

“Yet, evaluation can be ineffective and even harmful if naively done ‘by rule’ rather than ‘by thought’.”

Returning once more to the overarching evaluation statement and the related *must* statements, the breadth of this essay is indicative of the need for an on-going dialogue to review and re-review the current status of big data in online courses. Change is inevitable as is expanded understanding of persistence, sustainability, scalability, community, engagement, and the variety of lived experience that culminates as effective practice in online learning. There is an immediate need to drive advocacy and guide the direction for responsible assessment and effective use of learning analytics in online learning across the

field and in online courses. Practitioners and researchers in online learning must take the lead in “getting the right data and getting the data right.” Not doing so may result in wrongful or inadequate uses of data and analytics in online learning by a number of constituents including educators, institutions, government sectors, and accrediting groups. For example, institutions may be interested in interventions for academic achievement and retention, but the understanding of what goes on in the online course involves far more complex constructs and meanings beyond this.

A start towards advocacy and guidance involves building awareness of the issues and challenges of the use of learning analytics in online learning. A few awareness-building suggestions are offered as a starting point. First, a standards committee (perhaps initiated by the Sloan or EDUCAUSE groups) is needed to address the gamut of issues and gaps in learning analytics and to provide strong advocacy in representing the constituency of researchers, practitioners, and participants in online learning. Second, recognizing practical significance of learning analytics is important. The data trail from the artifacts of online course production must be measurable, visible and transparent in real time (as it happens), and facilitative to inform process and practice. This includes awareness building of what goes on and what we are looking for in the online course. Seeing the data is critical in examining performance, progress, and effective practice. Intervention is the other critical half of transparency. With data made visible, what is the intervention strategy? Will garden variety approaches and tools like nudge analytics [25], LOCO-Analyst [10], and SNAPP [24] bring real time intervention to student participation and performance? Last, but equally critical, is that awareness-building of learning analytics starts with good questions to drive out good data, leading to responsible assessment and effective use of the data trail in learning analytics. Perhaps these three questions are a good start to continue the dialogue: There will be more and more data to manage from online courses, but will we know what to do with it? What is quality performance data in online courses? Who will use that data and for what purpose(s)?

IV. ABOUT THE AUTHOR

Laurie P. Dringus is Professor of Information Systems in the Graduate School of Computer and Information Sciences at Nova Southeastern University. Her research interests include human-computer interaction, information design, and online learning environments. She has published widely several articles and presentations related to the research, development, and evaluation of online learning environments. Since 1998, Laurie has served as Editor-in-Chief of *The Internet and Higher Education*, a refereed international journal published quarterly by Elsevier (<http://www.elsevier.com/locate/iheduc>).

V. ACKNOWLEDGMENTS

The author would like to thank Dr. Eric Ackerman for his graphical work on the dashboard figure.

VI. REFERENCES

1. **Kelderman, E.** Online programs face new demands from accreditors. *The Chronicle of Higher Education: Online Learning* 4-5 (November 11, 2011).
2. **Siemens, G.** Learning and knowledge analytics. (August 5, 2011). <http://www.learninganalytics.net>.
3. **Brown, M.B., and Diaz, V.** Seeking evidence of impact: Opportunities and needs. *EDUCAUSE Review* 46(5): 43-54 (September/October 2011).
4. **Dziuban, C.** Impact on a transformational environment. *EDUCAUSE Review* 46(5): 48 (September/October 2011).
5. **Hrabowski, F.A., Suess, J., and Fritz, J.** Assessment and analytics in institutional transformation. *EDUCAUSE Review* 46(5): 15-28 (September/October 2011).
6. **Long, P., and Siemens, G.** Penetrating the analytics in learning and education. *EDUCAUSE Review* 46(5): 31-40 (September/October 2011).
7. **Oblinger, D.G.** Doing better with less. *EDUCAUSE Review* 46(5): 4-6 (September/October 2011).
8. Horizon Report. Four to five years: learning analytics (2011).

- <http://wp.nmc.org/horizon2011/sections/learning-analytics/>.
9. **Siemens, G.** What are learning analytics? *Elearnspace* (2010).
<http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/>.
 10. **Ali, L., Hatala, M., Gasevic, D., and Jovanovic, J.** A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education* 58: 470-489 (2012).
 11. **van Manen, M.,** *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy*, New York: SUNY Press, 1990.
 12. **Fournier, H., Kop, R., and Sitlia, H.,** The value of learning analytics to networked learning on a personal learning environment. In: *LAK 11*, 104-109, Banff, AB: Canada, 2011.
 13. **Greenberg, S., and Buxton, B.,** Usability evaluation considered harmful (some of the time). In: *CHI 2008*, 111-120, Florence, Italy: ACM Press, 2008.
 14. **De Liddo, A., Shum, S.B., Quinto, I., Bachler, M., and Cannavacciuolo L.,** Discourse-centric learning analytics. In: *LAK 11*, 23-33, Banff, AB: Canada, 2011.
 15. **Zhang, J., and Sun, Y.** Quantified measures of online discourse as knowledge building indicators. *International Conference in Computer-Supported Collaborative Learning (CSCSL 2011)* 72-79 (2011).
 16. **Duval, E.,** Attention please! Learning analytics for visualization and recommendation. In: *LAK 11*, 9-17, Banff, AB: Canada, 2011.
 17. **Dringus, L.P., and Ellis, T.** Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education* 45: 141-160 (2005).
 18. **Dringus, L.P., and Ellis, T.** Temporal transitions in participation flow in an asynchronous discussion forum. *Computers & Education* 54: 340-349 (2010).
 19. **Shea, P., Hayes, S., Uzuner, S., Vickers, J., Bidjerano, T., Pickett, A., Gozza-Cohen, M., Wilde, J., and Jian, S.** Learning presence: Additional research on a new conceptual element within the Community of Inquiry (Co) framework. *The Internet and Higher Education* 15(2): 89-95 (2012).
 20. **Wise, A.F., Marbouti, F., Speer, J., and Hsiao, Y-T.,** Towards an understanding of “listening” in online discussions: A cluster analysis of learners’ interaction patterns. In: *International Conference in Computer-Supported Collaborative Learning (CSCL 2011)*, 88-95, 2011.
 21. **Garrison, D.R., and Arbaugh, J.B.** Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and Higher Education* 10(3): 157-172 (2007).
 22. **Garrison, D.R., Anderson, T., and Archer, W.** Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education* 2-3: 87-105 (2000).
 23. **Akyol, Z., and Garrison, D.R.** Assessing metacognition in an online community of in inquiry. *The Internet and Higher Education* 14(3): 183-190 (2011).
 24. **Bakharia, A., and Dawson, S.,** SNAPP: A bird’s-eye view of temporal participant interaction. In: *LAK’11*, Banff, AB: Canada, 168-173, 2011.
 25. **Carmean, C., and Mizzi, P.** The case for nudge analytics. *EDUCAUSE Quarterly* 33(4) (2010).
 26. **Baker, R., and Yacef, K.** The state of educational data mining in 2009: a review and future visions. *JEDM – Journal of Educational Data Mining* 1(1): 3-17 (2010).
 27. **Mazza, R., and Dimitrova, V.** CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies* 65(2): 125-139, (2007).
 28. **Romero, C., Gutierrez, S., Freire, M., and Ventura, S.,** Mining and visualizing visited trails in web-based educational systems. In: *Proceedings of the 1st International Conference on Educational Data Mining*, Montreal: Canada, 2008.
 29. **Swan, K.,** Developing Social Presence in Online Course Discussions. In: S. Naidu (Ed.) *Learning & Teaching with Technology: Principles and practices*, Sterling, VA: Kogan Page, 2003.
 30. **Suthers, D., and Rosen, D.,** A unified framework for multi-level analysis of distributed learning. In: *LAK 11*, Banff, Alberta: Canada, 64-74, 2011.

31. **Shea, P., Vickers, J., and Hayes, S.** Online instructional effort measured through the lens of teaching presence in the Community of Inquiry framework: A re-examination of measures and approach. *The International Review of Research in Open and Distance Learning* 11(3) (2010).
32. **Vatrapu, R., Teplovs, C., Fujita, N., and Bull, S.** Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making. In: *LAK 11*, Banff, AB: Canada, 93-98, 2011.
33. **Sandelowski, M.** Focus on qualitative methods: Notes on transcription. *Research in Nursing & Health* 17: 311-314 (1994).
34. **Dennen, V.** Looking for evidence of learning: Assessment and analysis of methods for online discourse. *Computers in Human Behavior* 24: 205-219 (2008).
35. Data warehouse [definition]. Benefits of a data warehouse. *Wikipedia*. http://en.wikipedia.org/wiki/Data_warehouse.
36. **Garrison, D.R.** The community of inquiry survey instrument. Communities of inquiry (CoI). (2011). <http://communitiesofinquiry.com/methodology>.
37. **Ice, P., Boston, W.E., Swan, K., and Diaz, S.** Data changes everything: Introduction to learning analytics. *Pre-conference Workshop at Sloan-C 17th International Conference on Online Learning*, Orlando, Florida, November 9-11, 2011.
38. **Ice, P., Swan, K., Picciano, A., Shea, P., and Dringus, L.** Because we can: The era of big data. *Panel session at Sloan-C 17th International Conference on Online Learning*, Orlando, Florida November 9-11, 2011.
39. **Wenger, E.**, *Communities of practice: Learning, Meaning, and Identity*, Cambridge, UK: Cambridge University Press, 1999.
40. **Ram, A., Hua, A., Ram, P., and Sahay, S.**, Open social learning communities. In: *WIMS 11*, Sogndal, Norway: ACM, 2011.