Proceedings of the 10th International Conference on LA and Knowledge (LAK '20), March 23–27. ACM, 2020.
https://doi.org/10.1145/3375462.337554

From Theory to Action: Developing and Evaluating Learning Analytics for Learning Design

Korah J. Wiley
University of California-Berkeley
Berkeley, CA, USA
korah.wiley@berkeley.edu

Yannis Dimitriadis
Universidad de Valladolid
Valladolid, Spain
yannis@yllera.tel.uva.es

Allison Bradford
University of California-Berkeley
Berkeley, CA, USA
allison_bradford@berkeley.edu

Marica C. Linn
University of California-Berkeley
Berkeley, CA, USA
mclinn@berkeley.edu

ABSTRACT

The effectiveness of using learning analytics for learning design primarily depends upon two concepts: grounding and alignment. This is the primary conjecture for the study described in this paper. In our design-based research study, we design, test, and evaluate teacher-facing learning analytics for an online inquiry science unit on global climate change. We design our learning analytics in accordance with a socioconstructivism-based pedagogical framework, called Knowledge Integration, and the principles of learning analytics Implementation Design. Our methodology for the design process draws upon the principle of the Orchestrating for Learning Analytics framework to engage stakeholders (i.e. teachers, researchers, and developers). The resulting learning analytics were aligned to unit activities that engaged students in key aspects of the knowledge integration process. They provided teachers with actionable insight into their students' understanding at critical junctures in the learning process. We demonstrate the efficacy of the learning analytics in supporting the optimization of the unit's learning design. We conclude by synthesizing the principles that guided our design process into a framework for developing and evaluating learning analytics for learning design.

CCS CONCEPTS

• **General and reference** → **Design**; *Empirical studies.*

KEYWORDS

learning analytics, learning design, design-based research, theory, TEL environment

ACM Reference Format:
Korah J. Wiley, Yannis Dimitriadis, Allison Bradford, and Marica C. Linn. 2020. From Theory to Action: Developing and Evaluating Learning Analytics for Learning Design. In *Proceedings of the 10th International Conference on LA and Knowledge (LAK '20), March 23–27, 2020, Frankfurt, Germany.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3375462.3375540

1 INTRODUCTION

The central purpose of learning analytics (LA) for learning design can be found in the LAK definition of LA: to support the understanding of learning and the optimization of the learning context [5]. While many design efforts have attended to either aspect of this purpose, few have been successful at attending to both [13]. In their recent systematic literature review of LA for learning design studies, Mangaroska and Giannakos [13] identify numerous factors that limit the efficacy of LA for learning design, including the lack of grounding in a theory of learning and the lack of alignment with a theory-grounded learning design. Not grounding LA in a theory of learning can lead to a haphazard selection of data for LA development, biasing towards data that is simply proximal to rather than consequential for learning [17]. Not aligning LA to the learning design makes them inactionable, regardless of how interpretable they may be. We argue that in order for LA for learning design to accomplish their dual purpose of supporting the understanding of learning and optimizing the learning context, they must be developed using a strategy that attends to both of these aspects.

In this paper, we describe the development and evaluation of LA for learning design for which we used a design-based research methodology focused on inter-stakeholder dialogue. We draw upon principles of LA Implementation Design (LAID) and the Knowledge Integration (KI) pedagogical framework to develop teacher-facing LA to improve the learning design of an online middle school inquiry science unit on global climate change. We collaborated with five teachers, who had previously taught the unit, to identify, in accordance with our chosen theory of learning, several assessment items for which to develop LA. These LA leveraged data from a simple platform feature to provide teachers with an idea-focused analysis of their students' understanding. Our evaluation focuses on four issues: 1.) the efficacy of the LA data to identify students' learning needs, 2.) whether the analysis supported teachers in understanding their students' learning needs, 3.) whether the analytics informed the optimization of the learning design, and 4.) whether the LA report informed teachers' pedagogical action.

LAK '20, March 23–27, 2020, Frankfurt, Germany

ACM ISBN 978-1-4503-7712-6/20/03…$15.00
https://doi.org/10.1145/3375462.3375540

2 RELATED WORK

2.1 Theory-grounded LA

In recent years, researchers have identified a need to ground LA in theory [6, 17]. Gašević et al. [6] identify theory—along with design and data science—as a key dimension of LA. While many studies attend to the dimensions of design and data science, greater attention needs to be given to theory since it is theory that differentiates LA from data analytics [17]. Reimann [17] argues that grounding LA in a theory of learning allows them to function as a methodology for learning science research and thereby fulfill its purpose of supporting better understanding of learning. If LA are to be used as a methodological tool, greater focus must be placed on their development. Bergner et al. [2] note that methodologies are positioned at the intersection of the "what" and the "how". Consequently, using LA as a methodology helps to interrogate what is being studied and how it is being studied. Reimann [17] argues that the theory dimension of learning theory is essential for its function as a methodology. He contends that theory informs the decision of which data are most appropriate for measuring a particular aspect of learning as well as facilitating the explanation of analytics-identified student outcomes and lighting the path for responsive action.

2.2 Aligning to theory-grounded learning design

With its focus on paving the path for learning, learning design is an intuitive partner for LA. Numerous studies highlight the importance and value of the alignment between LA and learning design [13, 15]. When properly aligned, LA can support the optimization of the learning design [8]. The grounding of LA and learning design in a common theory of learning [6, 12, 21] is critical to their ability to support learning design optimization. Lockyer et al. [12] argue that theory-grounded learning design documents the pedagogical intent and provides an interpretative lens for making sense of process analytics (i.e. LA that provide insight into the development and application of knowledge). They further contend that theory-grounded learning design supports pedagogical action by conveying expected student outcomes against which actual outcomes, as represented in the LA, can be compared. It is the theory component of LA that creates these affordances [17]. By virtue of their common theoretical grounding, discrepancies between expected learning outcomes and actual outcomes provide a clear signal for where and how to modify the learning design. [12, 27]. However, in their review of LA for learning design, Mangaroska and Giannakos [13] only found three studies in which LA were explicitly aligned to a theory-grounded learning design [7, 18, 26].

Several researchers argue that to make LA actionable for teachers, not only does it need to be aligned to a theory-grounded learning design but teachers need to understand the alignment [4, 19]. Rodriguez-Triana et al. [19] propose that teachers be brought "in the loop" during the design process, a notion that resonates with the emphasis of the Orchestrating Learning Analytics (OrLA) framework on inter-stakeholder dialogue [16]. Involving teachers in the design process provides them with a greater understanding of the utility and limitations of the LA for optimizing the learning design [19].

3 PERSPECTIVES AND FRAMEWORKS

In light of these identified needs, we argue for and employ a two-phase strategy for developing and evaluating LA. Here, we describe the perspectives and frameworks that guided this strategy.

3.1 Design-based Research (DBR) and OrLA

To develop the LA for this study, we used DBR in conjunction with the OrLA practice of inter-stakeholder dialogue [16, 20]. DBR entails the rigorous application of theoretical insight and principles of engineering design to solve complex, practice-based problems [20]. Like Reimann [17],we identify the centrality of theory in DBR as its primary affordance for developing LA.

Starting from theoretical conjectures, the DBR process engaged a multidisciplinary design team in iterative design-test cycles to create a design solution. To guide the team through each cycle, we implement the OrLA practice of inter-stakeholder dialogue. The OrLA framework posits that each stakeholder has specialized knowledge relevant to LA development [16]. For example, researchers have knowledge about learning theories and evidence-based pedagogical structures, teachers have knowledge about the specific classroom constraints and diverse learning needs of their students, and system developers have knowledge about technical constraints and possibilities. During each design iteration, the members of this multidisciplinary, multi-stakeholder design team contribute their ideas and perspectives. By sharing their specialized knowledge throughout the development and evaluation process, the stakeholders can collaboratively develop LA that can effectively support teachers in the classroom context. The OrLA framework contends that while this design process is fraught with negotiations around differences in values, constraints, and priorities, the outcome of such complex negotiations is a design solution that can be sustainably implemented. Starting from theory, the design process generates a design solution that translates those theoretical conjectures into embodied conjectures. We discuss our theoretical conjectures in the following section and describe our embodied conjectures in the Materials and Methods section. For this study, we draw upon the KI pedagogical framework and LAID to form our theoretical conjectures.

3.2 Knowledge Integration (KI) Framework

We used the KI pedagogical framework to develop the theoretical conjectures for how the LA in this study would achieve the goal of supporting teachers to understand their students' learning.

The KI framework is grounded in a socioconstructivist perspective of learning and provides guidance for supporting students to develop integrated science knowledge. It holds that students enter any learning environment with preformed ideas that they developed through interactions with their physical and social environments and which inform their understanding of new ideas [11]. In a KI-based learning design, students' ideas are elicited and made available for exploration and further development. Once elicited, normative ideas can be added, at which point students need opportunities to distinguish between these ideas. This distinguishing step is critical for students to develop integrated knowledge, because this step is when students determine which ideas are most productive for understanding the phenomena under study. The distinguishing step is often the step in which students and teachers need the greatest support [24, 25]. The final step of the KI process is reflection, during which students synthesize their new understanding. There is well-established research demonstrating the power of curriculum based on the KI framework to promote robust and coherent scientific knowledge, as measured by the integration of complex, normative science ideas [10, 23, 24]. We, therefore, conjecture that linking LA to curriculum components in a KI-based learning design will provide teachers with insight into students' progress in the knowledge integration process (i.e. learning).

3.3 LA Implementation Design

To develop the theoretical conjectures for how the LA in this study would support teachers in optimizing the learning design, we use the LAID principles of coordination, comparison, and customization [27].

Coordination. There are two aspects of the coordination principle, conceptual coordination and logistical coordination. Conceptual coordination calls for an alignment of the LA to the educational objectives of the learning design. It further calls for researchers to communicate the logic of this alignment to users (e.g. teachers). The conceptual coordination principle also encourages researchers to give forethought about how desirable and undesirable outcomes would manifest in the analytics. By adhering to this principle, we aim to support teachers in knowing how the LA can provide them with insight into their students' understanding. The logistical coordination principle encourages the design team to attend to teachers' needs regarding when and how they access and use the analytics. How this principle is implemented depends largely on the context, constraints, and preferences relevant to the teachers and system developers.

Comparison. The comparison principle has two aspects, absolute comparison and relative comparison. These principles relate to the measures used to create the analytics. LA that are based on an external, fixed standard (e.g. a scoring rubric) support absolute comparison. However, LA that juxtapose similar entities (e.g. class periods) support relative comparison. The selection of which data to use and with which comparison frame to analyze it is informed by the theory of learning that grounds them. We use the comparison principle to engage teachers in noticing discrepancies between expected versus actual student outcomes, so as to motivate and inform their pedagogical action.

Customization. The customization principle supports the recognition that each teacher is uniquely constrained and motivated and will, consequently, need the freedom and support to act accordingly. Implementing this principle requires researchers to be in close communication with teachers and system developers.

We conjecture that using the principles of coordination, comparison, and customization to guide inter-stakeholder dialogue and the implementation of design ideas will generate LA that support teachers to understand their students' learning needs and guide their subsequent pedagogical action.

Research Questions. With these perspectives, we ask the following research questions:

(1) Are the data used to generate the LA useful for:
  (a) understanding student learning?
  (b) optimizing the learning design?
(2) Is the resulting LA solution useful for informing pedagogical action?

4 MATERIALS AND METHODS

To investigate our research questions, we used a mixed methods approach following the principles of DBR and OrLA. Here, we describe how we developed a LA solution that embodies our theoretical conjectures.

4.1 Learning Design Description

The learning design used in this study consisted of a multi-lesson, online middle school inquiry science unit about global climate change (GCC). (GCC Unit for Cycle 1 - https://wise.berkeley.edu/project/24105#/vle/group1; GCC Unit for Cycle 2 - https://wise.berkeley.edu/project/24852#/vle/group1) The unit was developed according to the KI framework and embedded in the Web-based Inquiry Science Environment (WISE). Since states across the U.S. are rapidly adopting the Next Generation Science Standards (NGSS), we saw value in aligning the GCC unit to them. The NGSS call for students to develop a coherent understanding of science concepts across discipline and grade level. To help students meet this goal, the unit was designed in accordance with evidence-based, inquiry science learning design principles [3], including direct instruction and numerous animations and models to support students in understanding the various processes that result in the warming of the Earth. The unit had a pre/post-test and embedded assessments of varying formats, which were completely individually and in groups, respectively.

4.1.1 Assessments and Rubrics. We used two types of assessment items for this study: multiple-choice (M-C) items and open-response items. All the assessment items used in this study supported students in meeting the Next Generation Science Standards (NGSS) performance expectation, MS-ESS3-5: "Ask questions to clarify evidence of the factors that have caused the rise in global temperatures over the past century" [1].

The M-C items engaged students in the KI process of distinguishing amongst their own ideas, which they generated from previous experience and engagement with the unit resources, and those offered in the answer choices. Each M-C item had a single correct answer and thus generated data that teachers could use to make absolute comparisons (i.e. comparing their students' answer choice and the correct choice). Additionally, each M-C item had automatic feedback guidance associated with each incorrect answer choice that prompted students to revisit relevant learning resources in the unit.

The open-response items engaged students in the KI process of integrating their ideas to develop a coherent scientific explanation. The pretest item, "Car on a Cold Day" calls for student to explain the affect that the sun has on a parked that has been parking outside over the course of two weeks of cold temperatures. The unit-embedded item, "What do you think now?", calls for students to explain how the sun warms the earth. To score the open-response item, we used a crater-ML scoring algorithm [9]. The algorithm

Table 1: The Knowledge Integration rubric, with sample responses, used to score the pretest and open-response assessment items.

Score	Level Description: Sample Response
1	Off-task: I don't know.
2	Irrelevant/Incorrect: The car is an artificial habitat by all means and retains the lowest tempurature
3	Partial - Normative isolated ideas, no valid link: The inside of the car would be warmer then the outside environment because the sun is directly shining on the car warming it up via radiation, and not taking into account the cold air on the outside of the car.
4	Simple - Elaborate a scientifically valid link: Heat has been trapped inside the car from previous days, or from the sun. As the sunlight is able to pass through the windows of the car, but cant escape, allowing a buildup of heat, despite the cold temperature outside.
5	Complex - Elaborate 2+ scientifically valid links: The car works similar to the solar oven. Even though the car was parked on a cold day, it was under the sun. The glass from the windows allow radiation from the heat inside and can trap the heat inside. Therefore, making the car have a greenhouse effect that allows the car to become warmer than the outside air.

was trained on 1,000 student responses scored by human coders using the content-specific KI rubric in Table 1. Consistent with the KI framework, the KI rubric prioritizes the scientifically-sound links that students use to connect their ideas while not penalizing non-scientifically normative ideas.

4.2 Participants

A total of three researchers, three system developers, and five local middle school science teachers and their 885 students participated in this study. To preserve anonymity, the teachers were given the following pseudonyms: Ms. Huey, Ms. Joyner, Ms. Kerrington, Mr. Lewis, and Mr. Scofield. The study took place at one school across two consecutive academic years, corresponding to the two design cycles. All teachers participated in each design cycles, except Ms. Joyner, who only participated in Cycle 1. The teachers reported that they shared the content of all the interviews they participated in with each other during their regular department planning meetings.

4.3 Data Collection and Analysis

The chronology of this DBR study is reflected in the sequence of the following sections.

4.3.1 Cycle 1 - Identifying and Validating the Data (RQ 1).

Classroom Study. During Cycle 1, we activated the "maximum number of attempts" (MNA) feature on every M-C embedded assessment item in the GCC unit. The MNA feature is only available for items that have automated feedback guidance. This decision was based on what previous teacher-partners and researchers saw as its potential utility for helping to evaluate students' engagement with the unit. Our design conjecture was that LA developed from MNA-feature data could: 1.) support teachers to engage in both aspects of the LAID comparison principle (e.g. comparing students' response against the correct answer and trends across their class in the number of attempts needed to correctly respond),and 2.) provide teachers and researchers with insight into how students understood the target concepts.

For this classroom study, the maximum number of attempts was set to the total number of options. While engaging the unit, students could see, directly under the answer choices, the ratio of the number of attempts they used versus the maximum number of attempts available.

After participant teachers used the MNA-activated version of the GCC unit, we conducted classroom observations of Ms. Joyner's and Ms. Kerrington's class, one period for each teacher. We took field notes of the teachers' and students' responses to and use of the MNA feature.

Stakeholder Interviews. In keeping with the OrLA framework, we conducted semi-structured interviews with all five teachers, who all had previously used the GGC unit, and three WISE system developers. The interviews with each stakeholder group were conducted separately and, we, the researchers, functioned as liaisons between the groups. The teacher interviews were audio recorded and the transcripts served as our primary data source. Handwritten notes were the primary data source for the system developer interviews. We analyzed the transcripts and written notes for data relevant to Research Question 1a.

Teachers. The teacher interviews primarily focused on issues related to the LAID principles of coordination, comparison, and customization, and more generally on what information they needed and wanted to receive when their students engaged the GCC unit. We also solicited their feedback about the potential utility of the MNA feature for supporting teaching and learning.

System Developers. The system developer interviews primarily focused on issues related to the coordination and customization principles of LAID (e.g. the available GCC log data and how to access and add to it).

Statistical Analysis. To determine whether log data associated with the MNA-feature correlated with student learning outcomes, we conducted an ordinal logistic regression analysis. We combined data from the students in our participant teachers' classes during Year 1 and 2. We limited our analysis to two M-C items and the open-response item. Although the other M-c items were only excluded from the statistical analysis, they were used in all other parts of the study. We selected the M-C items, "What Happens to Solar Radiation on Earth?" and "When does the energy from the Sun warm Earth?, based on their conceptual alignment with the open-response items, "Car on a Cold Day" and "What do you think now?". For the analysis, we used the following variables: outcome variable = students' KI rubric score on the "What do you think now?" item (categorical: range 1-3, corresponding to low (rubric scores = 1 or 2), middle (rubric score = 3), and high (rubric score = 4 or 5); explanatory variables = the total number of attempts need to answer the two

M-C items (categorical: range 2-7), and teacher (categorical: value 1-5).

4.3.2 Cycle 2 - Developing and Evaluating the LA (RQ 2).

Stakeholder Interviews. In Cycle 2, we conducted both informal and semi-structured stakeholder interviews with the system developers and teachers.

System Developers. The informal interview with the system developers focused primarily on issues related to the LAID principle of logistical coordination (i.e. the logistics and limitations of providing teachers with the requested information.)

Teachers. One week before teachers used the GCC unit, we conducted a semi-structured interview with Ms. Kerrington in her role as the science department chair to address issues related to the LAID principles of coordination and comparison. We presented Ms. Kerrington with the plan for where and how the MNA feature would be used in the unit and solicited her feedback on the plan.

LA Report. The LA report represents the embodiment of our theoretical conjectures implemented in conjunction with the information and negotiations from the stakeholder interviews (see Figure 1). To generate the LA for this study, we developed a Python script to automate the analysis of logged student data associated with the two M-C assessment items. The specific data used in the script was:

- Revision count (i.e. number of changes made to selection irrespective of submission status)
- Binary submission status (i.e. submitted [1] or not submitted [0])
- Response (i.e text of selected answer choice)
- Binary correctness status (i.e. correct [1] or incorrect [0])
- ID code (6-digit integer code for student/group)
- Class Period (integer value)

The analytics in the report were designed to support the following activities: monitor student engagement in the GCC unit, facilitate the identification of students needing targeted support, and identify potential opportunities for intervention.

Classroom Study. As part of the classroom study, we conducted informal interviews with individual teachers after they received the LA report but while they were still using the GCC unit. The teachers who used the GCC unit in Cycle 2 were: Ms Huey, Ms. Kerrington, Mr. Lewis, and Mr. Scofield. The interviews were conducted either in-person or virtually, depending on the teachers' availability. In the interviews, we asked teachers to discuss whether and how they used the LA report. We also solicited their feedback regarding how the report was provided to them and its content features.

5 RESULTS

The study results are present in terms of the research question they address. The subsections are mapped to the design structure using the following abbreviation system: Cycle Number-Cycle Activity.

5.1 RQ1a: Understanding Student Learning

5.1.1 Cycle 1-System Developer Interview. In this interview, the system developers shared with us the log data associated with the MNA feature. The developers explained that this feature constrains the number of times a student can submit an answer to a M-C assessment embedded in the unit and is only available on items that have automated feedback guidance. Rather than being punitive, the MNA feature along with the automated feedback are designed to encourage students to utilize relevant unit resources and to promote a habit of learning from their mistakes. They shared that the log data for the MNA feature includes the attempt count and timestamp associated with each answer submission. Also associated with the MNA feature is the specific automated feedback for the answer selection as well as whether or not the answer selection is correct. The system developers said that information regarding students' use of the feedback guidance could be found in the event log files. These files contained timestamp and action data (e.g. page entrance, exit, item submission) for each page in the unit.

Sample Analysis. Taken together, the information shared by the system developers provided insight for developing a plan for converting the log data into LA. Prior to our interview with teachers, but while they were using the GCC unit, we created a sample analysis of student learning using MNA-feature data. The sample analysis highlighted an emerging pattern in students' answer selection sequence. It read, "We noticed that in response to the question, "When does energy from the Sun (SR) warm Earth?" most students who needed more than one attempt chose "When SR reaches the Earth". Upon getting feedback that said, "Go back to the model and watch more sunrays!" they correctly selected "Only when SR is absorbed" even without revisiting the model (which was on a different step in the unit). From this analysis, we hypothesized that students need support distinguishing the difference between when solar radiation "reaches the Earth" and when it is absorbed."

5.1.2 Cycle 1-Classroom Study. To assess the value to teaching and learning, we activated the MNA feature on all the M-C assessment items in the GCC unit and informed the teachers of this feature addition. We conducted classroom observations in Ms. Kerrington and Ms. Joyner's class to investigate the impact of the feature on teacher and student behavior. During the class observations, most students seemed unaffected by the MNA feature, despite getting on-screen feedback regarding the number of attempts remaining. A few students noticed the feature and asked their teacher about its significance.

5.1.3 Cycle 1-Teacher Interview. After the unit was completed, we conducted semi-structured interviews with all five participant teachers. The teachers corroborated the classroom observations that few students seemed to notice the feature. They shared that those who noticed were highly concerned about the unknown implications of using the attempts. The teachers also stated that even though we told them about the feature they, like their students, did not notice it. When asked more generally about the potential benefit and use of the MNA feature, the teachers all expressed that they perceived great value for both teaching and learning. In regards to learning, they commented that the feature could encourage students to more seriously engage the units, a challenge expressed by other teachers who have used the units on the platform [22]. They hypothesized that students would perceive the limited number of attempts as an indicator of importance and take their time to provide more thoughtful responses. This hypothesis was based on their previous

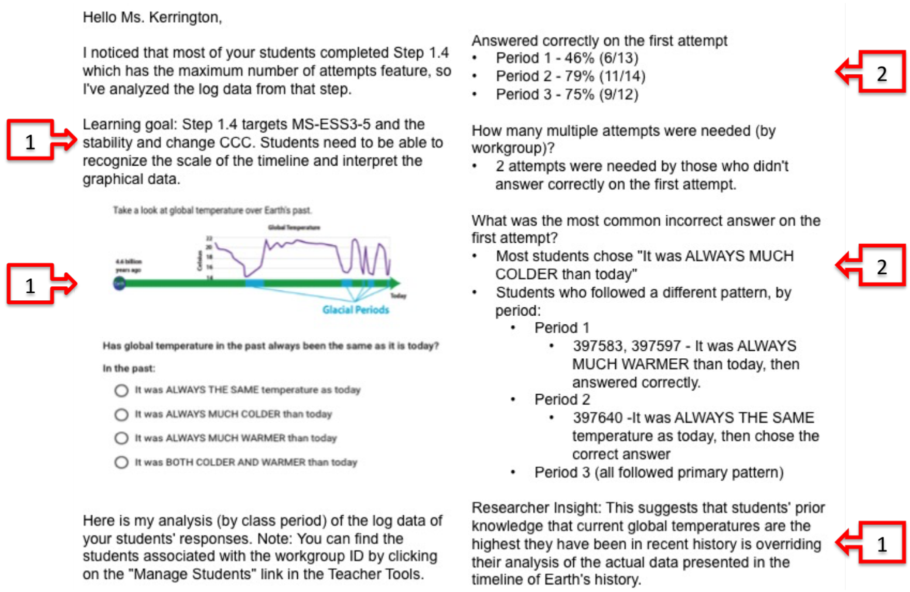

Figure 1: Reconstructed Step 1.4 LA Report Email to Ms. Kerrington. Red BoxArrows: 1 = LAID-Coordination, 2 = LAID-Comparison

observations of how their students responded to graded versus ungraded work.

Teachers also perceived value in the feature as a means to provide insight into student thinking. However, their vision for what information would be valuable varied, as evidenced in the following exchange: "I'm not looking for did they get it right or wrong, I want to know that explanation (Ms. Kerrington)". "I [want to] see right away who missed it,...instead of me having to look through (Ms. Joyner)".

When we shared our sample analysis with teachers they all expressed a desire to have such analysis. In particular, the analysis resonated with Ms. Kerrington who commented that she helped a small group of students make the distinction between what happens to solar radiation with it "reaches" the Earth versus when it is "absorbed" by the Earth. She commented that had she known that more students were struggling with the same issue, she would have brought the discussion up to the whole-class level. Based on this feedback, we decided to provide teachers with LA that would give them insight into why students might have selected a particular incorrect answer as well as data regarding general correctness. Ms. Kerrington's response to the sample analysis also raised the question of whether individual or aggregated data should be presented. There was no consensus on the issue so we used it as an opportunity to implement the LAID customization principle and therefore decided to provide analytics at both the aggregate and individual level.

During the interview, we also solicited teachers for feedback on the number of assessment items that utilize the MNA feature (i.e. all the M-C items or a select few). Teachers had differing opinions, some believed that having the MNA feature on all M-C items would cause students to no longer recognize it as an indicator of item importance. However, other teachers believed that the effect would not be lost and the potential benefit of students engaging more seriously on all the items would be "worth the risk". Given the lack of consensus, we proposed to teachers that they make the determination of which and how many assessment items to use the MNA feature on. Although this proposal would support the LAID principle of customization, they stated that they did not feel confident enough in their understanding of the unit's learning design to make those decisions. Using our intimate knowledge of the learning design, we, therefore, decided to restrict the MNA feature to only the M-C items that aligned with the "distinguishing ideas" step in the KI process. In this way, we implemented the LAID conceptual coordination principle, and the MNA-based LA would, thus, provide teachers with insight into student's progress in knowledge integration at a critical and often challenging point of the process.

5.1.4 Cycle 1-Statistical Analysis. Although we expected that developing MNA-based LA for M-C assessment items targeted to the distinguishing ideas step of the KI process would provide insight into their learning process, we wanted to evaluate this conjecture. We wanted to determine whether the number of attempts needed to correctly answer the M-C assessment items could function as a proxy for student learning. With students scores on a subsequent open-response assessment item (i.e. "What do you think now?") as the outcome variable, and total number of attempts used and teacher as explanatory variables, we conducted ordinal logistic regression analysis (see Table 2). After controlling for the other covariates, students who use more than 5 total attempts to answer the two, M-C items (possible range: 2-7 attempts) are 76% less likely to express integrated ideas in the explanation of how the sun warms to the Earth (i.e. rubric score level 4 or 5) than students who use no more than 3 attempts. Providing an integrated scientific explanation (i.e. rubric score level 4 or 5) on the "Car on a Cold Day" pretest item significantly affects the odds of doing the same on the open-response item in the unit (4.06 times as great as low performance, score level 1 or 2). No other model variables were significant at the 5% level.

5.2 RQ1b: Optimizing the Learning Design

When we reviewed the event log files, namely the sequence of timestamps associated with page entrance and exit, we noticed that none of the students revisited the recommended resources after being prompted to do so by the automated feedback guidance. The sequence of answer selections also showed that numerous students were choosing answers in a top-down pattern, suggesting a "click through till correct" selection strategy. This revelation led us to redesign the unit so as to place the M-C assessment items on the same page as their related resources. This redesign was based on the logic that students might be more inclined to reference the model if it appeared on the screen alongside the assessment item. We also modified the setting on the MNA feature to limit the maximum number of attempts from the total number of answer choices to one less than this value. We made this adjustment to motivate students to attend to the feedback in order to get the correct answer rather than just clicking through the responses till they made the correct selection.

5.3 RQ2: Supporting Pedagogical Action

5.3.1 Cycle 1-System Developer Interview. We interviewed the system developers to discuss the logistics of providing teachers with real-time LA within the WISE platform. The system developers stated that while all the necessary log data was available, generating platform-embedded analytics would require major changes to the system infrastructure. Specifically, they would need to develop the graphical displays and interface for the analyzed data. They explained that while the re-build was possible, they would not be able to complete the task in advance of the teachers' next use. We asked for their suggestions on how to meet the deadline and provide teachers with a LA report. They suggested that we develop an off-platform LA report that could be emailed to teachers. They stated that doing so would economize time and resources, which could later be used toward implementing an evidence-based LA strategy. Emailing the report to teachers would allow us to apply the LAID logistical coordination principle, since teachers would be able to access and act on the analytics in accordance with their teaching practice.

The LA Report. Based on this interview, we planned a MNA-based LA report that would be generated after 50% of students completed the items and provide teachers with near-time (within 24 hours) analysis to answer the following questions:

- How many students/groups (absolute and relative frequency) chose the correct answer on their first attempt?
- Of those who needed multiple attempts, what was the most commonly selected incorrect answer?
- What were alternative incorrect answer patterns? (Note: The student/group ID codes were listed with the respective alternative pattern.)

The student/group ID codes were listed with the respective alternative pattern.

Since the assessment items had a correct answer, the teachers would be able to make absolute comparisons. Providing them with both class-level aggregated and individual data also supported teachers in making relative comparisons. With this design approach, we implemented both aspects of the LAID comparison principle.

To contextualize the LA in the learning design and implement the LAID coordination principle, we included the specific learning goal and a screenshot of the assessment item in the email, just above the analytics. We sent the LA report to each teacher as an email using their school email address.

Based on teachers' stated desire to have information similar to the sample analysis that we showed them in Cycle 1, we included a section called "Researchers' Insight". The "Researchers' Insight" reflected our conjectures about student thinking based on specific response patterns and automated feedback guidance associated with the incorrect submission. As an implementation of the coordination principle, we included, at the top of the report, the specific learning goal and a screenshot of the assessment item.

5.3.2 Cycle 2-Classroom Study. In preparation for the classroom studies, we conducted a semi-structured interview with Ms. Kerrington, the science department chair, to finalize the plan for generating and sharing the LA report. We shared and solicited her feedback about the specific assessment items for which the MNA feature would be activated and the planned content for emailed report. We also asked her about the foreseeable actions teachers might take in response to the analytics in the report. She approved of the selected assessment items and commented that they would function perfectly as formative assessments given their position in the unit and their focus on supporting students in understanding the key concepts associated with global climate change. Ms. Kerrington said that she would instruct the other teachers to use the items as formative assessments as well. For the classroom study, we conducted classroom observations of Ms. Kerrington's class and conducted informal interviews with all Cycle 2 teachers after they received the report to determine how they used it and the pedagogical actions they took in response.

The pedagogical actions that teachers took in response to the analysis in the emailed MNA-based LA varied. For example, Ms. Kerrington received a LA report for which included a "Researcher's Insight" stating, [the results] suggest that students' prior knowledge that current global temperatures are the highest they have been in recent history is overriding there analysis of the actual data presented in the timeline of Earth's history." After implementation, Ms. Kerrington reported the following, "I started each class period this morning by showing [the step] and addressing the need to use evidence in our answer choices (claims). I then talked about prior knowledge and how prior knowledge can lead us into misconceptions. For my part, I recognize how last week's graphing of temperature data for 2006-2016 (done to match the Mauna Loa CO2 data from the Exploratorium activity) helped students develop this strong misconception so I'll definitely be making changes next year." Ms. Kerrington went on to comment, "[The LA report] was very valuable for my students and for me! Looking forward to more feedback!"

Table 2: Ordinal Logistic Regression for Student Performance on the Open-Response Item (N = 479)

Variable	Odds Ratio	Std. Err.	z	P>z	[95% Confidence Interval]	
Total Attempts Used						
3 to 4	0.98	0.21	-0.10	0.920	0.64	1.49
5 to 7	0.24	0.09	-3.79	<0.001***	0.121	0.50
Pretest Score						
Rubric Level 3	1.55	0.35	1.93	0.054	0.99	2.42
Rubric Level 4 or 5	4.06	2.22	2.56	0.011*	1.39	11.87
Teacher	0.92	0.07	-1.20	0.230	0.80	1.06

Note: Reference Variables - Total Attempts Used = 0; Pretest Score = Rubric Level 1 or 2; Teacher = Ms. Kerrington
*p<0.05; ***p<0.001

In regards to his use of the LA, Mr. Lewis stated, "The way I've used this data so far is by first noting which students used all attempts without answering the question correctly. For each report you've sent me, I've made a list of those students and while the class works independently on [the GCC unit], I have quick, one-on-one conferences with those aforementioned students. In addition, I review the most common incorrect answer and have table talks and then classroom discussions about why students might have that as a misconception, why it's a misconception, and why the correct answer is correct. For a couple of the questions, I have supplemented the classroom discussions with various simulations and videos to try and change the students' understanding of the misconception."

Mr. Scofield said the following in response to receiving the LA report: "My plan is to use this to teach students how to read graphs and to better understand their own prior knowledge. I created a page with the question on it and a prompt for the students to discuss in groups why they or others chose the main incorrect answer. Also, this data makes me want to implement more pre-activities that help students understand their background knowledge before beginning the next unit."

In contrast to the other teachers, Ms. Huey did not fully utilize the MNA-based LA, despite offering positive feedback during the Cycle 1 interviews. While Ms. Huey received a report for the first M-C item, she did not express a desire to receive more, and therefore no additional reports were sent to her. During our informal interview, Ms. Huey stated that she did not take any action because she interpreted the report analysis to say that the automated feedback guidance provided students with sufficient support and no further action was warranted. She further commented that she did not yet feel comfortable using the unit and that this was the primary reason for not wanting to receive additional LA reports.

6 DISCUSSION

Using a theory-grounded development and evaluation strategy allowed us to use LA as a tool for: identifying, understanding, and supporting student learning; optimizing the learning design; and supporting pedagogical action. We take the empirical evidence from this study as validation of our strategy and, therefore, offer the Theory-Grounded LA Development and Evaluation (T-GLADE) framework. In the following sections, we elaborate on the T-GLADE framework through a discussion of our findings.

6.1 T-GLADE Framework

Theory is central to the T-GLADE framework. Every aspect of developing and evaluating LA for learning design is grounded in and guided by theory.

6.1.1 T-GLADE Principle 1: Use a learning theory to guide data selection, analysis, and evaluation. Designing the GCC unit in accordance with learning theory (e.g. the socioconstructivism-based KI framework) allowed us to identify the assessment items that would be ideal for providing students with learning support. Prior research on KI-based learning designs indicate that the distinguishing ideas step of the KI process is a critical aspect of the learning process for which students need support [24]. This knowledge informed our decision to develop LA for this step type and to use the MNA-feature as a means to evaluate students' learning process. On its face, the number of attempts that a student needs to correctly answer an M-C question yields little insight to guide learning design decisions. However, by using our learning theory to guide our data analysis strategy for MNA-associated data, we were able to gain insight into how students were distinguishing amongst the available ideas.

We used regression modeling to evaluate our MNA-based LA. This decision to use students' rubric score on the open-response item as the response variable and use the total number of attempts they need to correctly answer the two M-C items as an explanatory variable was informed by our theory-grounded framework. Based on the KI framework, we expected that students' performance on M-C items that engaged them in distinguishing amongst relevant ideas would be a predictor for their performance on the open-response item. Our modeling results aligned with this expectation, thus substantiating our conjecture that developing LA for M-C items that support students in distinguishing their ideas could provide teachers with insight into student thinking.

6.1.2 T-GLADE Principle 2: Use an implementation theory to guide the feature development and evaluation of the LA solution. Using the principles of LAID as our other theoretical grounding allowed us to develop a LA solution that supported responsive pedagogical action. We used this theoretical grounding to govern our data collection during the stakeholder interviews. In this way, we were able to leverage the expertise of our stakeholders to develop a LA solution that instantiated each of the three LAID principles (see Figure 1). During Cycle 1 interviews, we drew upon the teachers' and researchers' expertise to decide which data analysis to present as well as when and how to present it. Ms. Joyner's comment stressing the importance of being able to quickly determine which students needed support led us to not only provide aggregate class-level data but also include a list, by attempt, of the students who selected each incorrect answer. Ms. Joyner's expertise was held in balance with the researchers' and Ms. Kerrington's expertise, which dictated an exploration into why students might have chosen a particular incorrect answer. The inclusion of the Researchers' Insight reflected this value for exploring the student thinking and was recognized by teachers as a valuable feature for guiding their pedagogical actions.

While the teachers expressed that the LA helped them understand their students' learning needs, we found, like others, that having deep knowledge of the learning design is essential for knowing how to take actions in response to the analytics. [14]; [19]. Ms. Kerrington's deep knowledge of the learning design, derived from her role as the science department chair, allowed her to readily identified from the LA report how she could improve the unit learning design for future classes. In contrast, Ms. Huey stated that her limited understanding of the learning design inhibited her ability to take anayltics-informed pedagogical actions. This finding suggests that our instantiation of the LAID coordination principle (see Figure 1) were insufficient in providing teachers with requisite knowledge to act on the LA. Providing a complete representation of the learning design, as recommended and supported by research literature [8, 15] may be an effective way to support teachers in customizing both the learning design and LA solution.

6.1.3 T-GLADE Principle 3: Use theory as an arbiter during the development process. Our LAID-focused interviews surfaced differences across and amongst our stakeholder groups in terms of the values and priorities for implementing the LAID principles. Take for example the topic of which M-C items to activate the MNA feature: the system developers were agnostic; some teacher provided justified reasons for having the feature on all M-C items while other teachers gave justification for only having it on some M-C items, irrespective of step type; the researchers saw value in applying it to all M-C items of a certain step type, namely the steps that supported students in distinguishing their ideas. The arbiter of these differences was the theory. It is not likely that a single theory can inform all aspects of LA development and evaluation. Therefore, reliance on multiple theories may be necessary. In the above example, the LAID principle of customization was insufficient to resolve the issue since the teachers felt ill-equipped to make decisions about the unit's learning design. This situation could have stymied the development process or led to an ineffective LA solution. Since the unit's learning design was grounded in socioconstructivism, it was able to guide our decision. Having a theory to guide each aspect of the development process ensures that during the inevitable event of conflicting values and perspectives decisions can be made that safeguard the integrity of the resulting LA solution. The stakeholder that is most knowledgeable of the theor(y/ies), in this study this was the researchers, should be positioned as liaisons for the stakeholder groups. With their deep knowledge of the theory, the liaison stakeholder would know when an impasse in the development and evaluation process has been reached and determine which theory or aspect of the theory can successfully guide the design team through it.

7 CONCLUSION

This study provides empirical evidence for the value and importance of grounding all aspects of developing and evaluating LA for learning design in theory[17]. Leveraging our learning theory allowed us to develop LA for M-C items rather than open-response items and, therefore, use simpler data analysis techniques (i.e. descriptive statistics vs. NLP or predictive modeling) while maintaining (and perhaps enhancing) pedagogical value. Leveraging our implementation theory allowed us to develop features for our LA solution that helped with understanding student learning, optimizing the learning design, and supporting analytics-informed action.

The DBR and OrLA principles provide practical scaffolds for employing our theory-grounded strategy. Although the context of this study was a middle school lesson on climate change, which happened to be covered by all 7th grade teachers at one time during an academic term (hence the two-year study time frame), we imagine other contexts with shorter turnaround times, such as higher education or MOOC courses. In different contexts, the T-GLADE framework could guide a process LA development and evaluation to completion within a few months. Having a core team or individual to usher the stakeholders through the process would likely increase efficiency and the likelihood of implementing an effective LA for learning design solution.

With the T-GLADE framework and stakeholders willing to engage the process, LA can be developed that enrich the teaching and learning context on multiple levels. Teachers will have course specific LA that complement their teaching practices. Students will receive timely and targeted support as they engage the learning process. Researchers will have data closely aligned to the learning process with which they can further explore and expand theories of learning. System developers will have opportunities to apply and extend analysis techniques to generate analytics that are easily understood and implemented in the classroom context.

7.1 Future Research

Our immediate next step is to extend the evaluation of our LA solution to investigate the impact of teachers' pedagogical actions on student learning (paper forthcoming). Additionally, we are developing a system-integrated dashboard for our LA, using the Educational Data Storytelling approach as a theoretical guide [4].

To expand the impact of this study, we invite other researchers to test the robustness and versatility of the T-GLADE framework to guide the development and evaluation of LA for learning design in other contexts and withe different theories and stakeholder groups. The T-GLADE framework and these study findings present two promising areas of research for the next 10 years: 1.) using LA as a methodological tool to explore teaching and learning, 2.) designing LA for learning design that can be an educative resource for teachers, thereby bridging the gap between theory and classroom action.

ACKNOWLEDGMENTS

This material is based upon work supported in part by the National Science Foundation under Grant DRL-1813713: Supporting Teachers in Responsive Instruction for Developing Expertise in Science

(STRIDES). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The research of the second author is partially funded by the European Regional Development Fund and the National Research Agency of the Spanish Ministry of Science, Innovations and Universities under project grants TIN2017-85179-C3-2-R, by the Ministry of Science and Education (PRX17/00410), by the European Regional Development Fund and the Regional Council of Education of Castile and Leon under project grant VA257P18, and by the European Commission under project grant 588438-EPP-1-2017-1-EL-EPPKA2-KA.

REFERENCES

[1] 2013. *Next Generation Science Standards: For States, By States.* National Academies Press, Washington, D.C. https://doi.org/10.17226/18290

[2] Yoav Bergner, Geraldine Gray, and Charles Lang. 2018. What Does Methodology Mean for Learning Analytics? *Journal of Learning Analytics* 5, 2 (Aug. 2018), 1–8. https://doi.org/10.18608/jla.2018.52.1

[3] Ton de Jong. 2019. Moving towards engaged learning in STEM domains; there is no simple answer, but clearly a road ahead. *Journal of Computer Assisted Learning* 35, 2 (April 2019), 153–167. https://doi.org/10.1111/jcal.12337

[4] Vanessa Echeverria, Roberto Martinez-Maldonado, Roger Granda, Katherine Chiluiza, Cristina Conati, and Simon Buckingham Shum. 2018. Driving data storytelling from learning design. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*. ACM Press, Sydney, New South Wales, Australia, 131–140. https://doi.org/10.1145/3170358.3170380

[5] Rebecca Ferguson. 2012. Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning* 4, 5/6 (2012), 304. https://doi.org/10.1504/IJTEL.2012.051816

[6] Dragan Gašević, Vitomir Kovanović, and Srećko Joksimović. 2017. Piecing the learning analytics puzzle: a consolidated model of a field of research and practice. *Learning: Research and Practice* 3, 1 (Jan. 2017), 63–78. https://doi.org/10.1080/23735082.2017.1286142

[7] Pablo A. Haya, Oliver Daems, Nils Malzahn, Jorge Castellanos, and Heinz Ulrich Hoppe. 2015. Analysing content and patterns of interaction for improving the learning design of networked learning environments: Analysing content and patterns of interaction. *British Journal of Educational Technology* 46, 2 (March 2015), 300–316. https://doi.org/10.1111/bjet.12264

[8] Davinia Hernández-Leo, Roberto Martinez-Maldonado, Abelardo Pardo, Juan A. Muñoz-Cristóbal, and María J. Rodríguez-Triana. 2019. Analytics for learning design: A layered framework and tools: Analytics layers for learning design. *British Journal of Educational Technology* 50, 1 (Jan. 2019), 139–152. https://doi.org/10.1111/bjet.12645

[9] Claudia Leacock and Martin Chodorow. 2003. Crater: Automated Scoring of Short-Answer Questions. *Language Resources and Evaluation - LRE* 37 (2003), 389–405. https://doi.org/10.1023/A:1025779619903

[10] Hee-Sun Lee and Ou Lydia Liu. 2009. Assessing learning progression of energy concepts across middle school grades: The knowledge integration perspective. *Science Education* 94, 4 (Nov. 2009), 665–688. https://doi.org/10.1002/sce.20382

[11] Marcia C. Linn and Bat-Sheva Eylon. 2011. *Science learning and instruction: taking advantage of technology to promote knowledge integration.* Routledge, New York. OCLC: ocn176946367.

[12] Lori Lockyer, Elizabeth Heathcote, and Shane Dawson. 2013. Informing Pedagogical Action: Aligning Learning Analytics With Learning Design. *American Behavioral Scientist* 57, 10 (Oct. 2013), 1439–1459. https://doi.org/10.1177/0002764213479367

[13] Katerina Mangaroska and Michail N. Giannakos. 2018. Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies* (2018), 1–1. https://doi.org/10.1109/TLT.2018.2868673

[14] Yishay Mor, Rebecca Ferguson, and Barbara Wasson. 2015. Editorial: Learning design, teacher inquiry into student learning and learning analytics: A call for action: Learning design, TISL and learning analytics. *British Journal of Educational Technology* 46, 2 (March 2015), 221–229. https://doi.org/10.1111/bjet.12273

[15] Donatella Persico and Francesca Pozzi. 2015. Informing learning design with learning analytics to improve teacher inquiry: Informing LD with LA to improve teacher inquiry. *British Journal of Educational Technology* 46, 2 (March 2015), 230–248. https://doi.org/10.1111/bjet.12207

[16] Luis P. Prieto, María Jesús Rodríguez-Triana, Roberto Martínez-Maldonado, Yannis Dimitriadis, and Dragan Gašević. 2018. Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level. *Australasian Journal of Educational Technology* (Nov. 2018). https://doi.org/10.14742/ajet.4314

[17] Peter Reimann. 2016. Connecting learning analytics with learning research: the role of design-based research. *Learning: Research and Practice* 2, 2 (July 2016), 130–142. https://doi.org/10.1080/23735082.2016.1210198

[18] María Jesús Rodríguez-Triana, Alejandra Martínez-Monés, Juan I. Asensio-Pérez, and Yannis Dimitriadis. 2015. Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations: Scripting and monitoring meet each other. *British Journal of Educational Technology* 46, 2 (March 2015), 330–343. https://doi.org/10.1111/bjet.12198

[19] María Jesüs Rodríguez-Triana, Luis P. Prieto, Alejandra Martínez-Monés, Juan I. Asensio-Pérez, and Yannis Dimitriadis. 2018. The teacher in the loop: customizing multimodal learning analytics for blended learning. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*. ACM Press, Sydney, New South Wales, Australia, 417–426. https://doi.org/10.1145/3170358.3170364

[20] William A. Sandoval and Philip Bell. 2004. Design-Based Research Methods for Studying Learning in Context: Introduction. *Educational Psychologist* 39, 4 (Dec. 2004), 199–201. https://doi.org/10.1207/s15326985ep3904_1

[21] Anouschka van Leeuwen. 2015. Learning analytics to support teachers during synchronous CSCL: Balancing between overview and overload. *Journal of Learning Analytics* 2, 2 (Dec. 2015), 138–162. https://doi.org/10.18608/jla.2015.22.11

[22] Keisha Varma, Freda Husic, and Marcia C. Linn. 2008. Targeted Support for Using Technology-Enhanced Science Inquiry Modules. *Journal of Science Education and Technology* 17, 4 (Aug. 2008), 341–356. https://doi.org/10.1007/s10956-008-9104-8

[23] Tammie Visintainer and Marcia Linn. 2015. Sixth-Grade Students' Progress in Understanding the Mechanisms of Global Climate Change. *Journal of Science Education and Technology* 24, 2-3 (April 2015), 287–310. https://doi.org/10.1007/s10956-014-9538-0

[24] Jonathan M. Vitale, Elizabeth McBride, and Marcia C. Linn. 2016. Distinguishing complex ideas about climate change: knowledge integration vs. specific guidance. *International Journal of Science Education* 38, 9 (June 2016), 1548–1569. https://doi.org/10.1080/09500693.2016.1198969

[25] Korah J Wiley, Allison Bradford, and Marcia C. Linn. 2019. Supporting Collaborative Curriculum Customizations Using the Knowledge Integration Framework. In *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning*, Kristine Lund, Gerald P. Niccolai, Elise Lavoue, Cindy Hmelo-Silver, Gahgene Gweon, and Michael Baker (Eds.), Vol. 1. International Society of the Learning Sciences (ISLS), Lyon, France, 480–487.

[26] Alyssa Wise, Yuting Zhao, and Simone Hausknecht. 2014. Learning Analytics for Online Discussions: Embedded and Extracted Approaches. *Journal of Learning Analytics* 1, 2 (2014), 48–71. https://doi.org/10.18608/jla.2014.12.4

[27] Alyssa Friend Wise and Jovita Vytasek. 2017. Learning Analytics Implementation Design. In *Handbook of Learning Analytics* (first ed.), Charles Lang, George Siemens, Alyssa Wise, and Dragan Gasevic (Eds.). Society for Learning Analytics Research (SoLAR), 151–160. https://doi.org/10.18608/hla17.013