Learning Analytics for Online Discussions: A Pedagogical Model for Intervention with Embedded and Extracted Analytics

Alyssa Friend Wise
Simon Fraser University
250 - 13450 102nd Avenue
Surrey, BC, V3T0A3 Canada
1-778-782-8046
alyssa_wise@sfu.ca

Yuting Zhao
Simon Fraser University
250 - 13450 102nd Avenue
Surrey, BC, V3T0A3 Canada
1-778-782-8046
yza174@sfu.ca

Simone Nicole Hausknecht
Simon Fraser University
250 - 13450 102nd Avenue
Surrey, BC, V3T0A3 Canada
1-778-782-8046
shauskne@sfu.ca

ABSTRACT
This paper describes an application of learning analytics that builds on an existing research program investigating how students contribute and attend to the messages of others in online discussions. A pedagogical model that translates the concepts and findings of the research program into guidelines for practice and analytics with which students and instructors can assess their discussion participation are presented. The analytics are both e*mbedded* in the learning environment and *extracted* from it, allowing for integrated and reflective metacognitive activity. The pedagogical intervention is based on the principles of (1) Integration (2) Diversity (of Metrics) (3) Agency (4) Reflection (5) Parity and (6) Dialogue. Details of an initial implementation of this approach and preliminary findings are described. Initial results strongly support the value of student-teacher dialogue around the analytics. In contrast, instructor parity in analytics use did not seem as important to students as was expected. Analytics were reported as useful in validating invisible discussion activity, but at times triggered emotionally-charged responses.

Categories and Subject Descriptors
K.3.1 Computer uses in education

General Terms
Measurement, Design, Human Factors.

Keywords
Online learning, Computer mediated communication, Learning analytics, Asynchronous discussion groups, Student participation.

1. INTRODUCTION
This paper presents a learning analytics application that builds on an existing research program investigating how students contribute and attend to the messages of others in online discussions. Each of these activities is important in realizing the theoretical potential of online discussions to support group knowledge construction and individual development of understanding [36]. However a substantial research base shows that in actual discussions learners often pay limited attention to others' posts and make disconnected comments [32,14], resulting in conversation patterns that can be characterized as shallow and incoherent rather than dialogic [12, 33]. Early research into these problems reported disturbingly low overall statistics, suggesting that the problems were global and thus in part systemic products of the online discussion environment [14, 19, 31]. These findings spurred efforts to improve discussion environments to support productive engagement and discussion [e.g. 22, 28]. However, more recent work disaggregating data across individuals has revealed that students in fact engage in very different kinds of behaviors in online discussions [35, 36, 37]. This suggests that discussion participation can also be improved with more targeted efforts to provide guidance to students individually. This work unites these two kinds of efforts to present a pedagogical model using both embedded and extracted analytics to support online discussion participation.

LAK '13, April 08-12 2013, Leuven, Belgium

1.1 Situating the Present Effort in the Field
The field of learning analytics is concerned with the collection and analysis of data traces related to learning in order to inform and improve the process and/or its outcomes [29]. Within this space, a distinction can be made between classes of analytics based on the types of data collected and the kinds of decision making targeted [8]. At a macro level, administrators and policy makers have the opportunity to use learning analytics to make programmatic or legislative decisions. In such situations data on past learning events is used to make decisions about future ones; these choices tend to be based on relatively long data time-cycles [5], affect large numbers of people, and involve outcome-type data, for example summative assessments, performance indicators and the like [3]. For these purposes data may also be aggregated at various levels (i.e. student, class, department, institution etc.).

In contrast, at a micro level learners and teachers have the opportunity to use learning analytics to make more local decisions about the current learning event they are involved in [5]. In this case, the relevant data relates to tracking learning processes and an important element of interpretation is having a model of learning for the particular environment - i.e. a research-based framework for understanding what productive activity in the specific learning context looks like. In some cases the model may be specified such that analytics data is processed and interpreted according to some system of rules leading to automatic changes in the learning system [e.g. 26]. In other cases, data may be processed into categories according to the model, but then

presented to stakeholders to support decision making [e.g. 17]. In this latter situation, two core challenges that are currently receiving a great deal of research attention are how to determine meaningful traces of learning, and how to present them in a form that will be useful to decision makers [7]. Another important, but less studied component is the model for intervention [5]; that is the framing of the activity (for teachers and/or students) of interpreting and making decisions based on the analytics.

This work contributes to all three of these areas of interest: capturing meaningful traces of learners' activity in online discussions based on a specific model of learning through discussions, presenting these traces to learners in a useful form, and framing this presentation as part of pedagogical activity to guide the use of the analytics in decision-making by learners and teachers. While the traces and analytics presentation are specific to online discussions, the pedagogical model for framing reflective activity is described in terms of general principles that can be applied to a variety of learning contexts. This pedagogical and analytics model has recently been piloted in a semi-automated implementation with a small group of learners in an authentic class setting. It is argued that such lightweight testing of the efficacy of theoretically-grounded analytics implemented with the minimally necessary toolset is a valuable validation step prior to the development of full-blown learning analytics platforms and dashboard.

2. LEARNING ANALYTICS FOR ONLINE DISCUSSIONS

The increasing amount of information automatically captured by online learning environments is attractive as a source of data with which to better understand and support learning processes. However, there is a wide gap between the kinds of data that are easily capturable in social learning environments such as online discussions and the kinds of constructs that have been established as pedagogically valuable [3].

To address this challenge, we begin by laying out our theoretical framework for learning in online discussions with the relevant research base. We then describe two classes of analytics for online discussions that we have developed based on this model: *embedded* analytics and *extracted* analytics. The first class refers to a set of analytics that are embedded in the discussion interface and can be used by learners in real-time to guide their participation; here interpretation of the analytics and participation in the learning activity are unified as a single endeavor. In contrast, the second class refers to analytics that are extracted from the traces of learning activity and presented back to learners for interpretation as a separate exercise from participating in the learning activity itself. As an example, analytics presented to learners via dashboard-like systems fall into the second category of this taxonomy. That is while the presentation of the analytics may be integrated into the overall learning environment, the activity of interpreting them is separate from that of engaging in the learning activity itself. This distinction will be clarified through the description of the specific instantiation of each type of analytic that follows in the subsequent sub-sections.

In addition to the development of these analytics as a source of informative data, we describe their intended use in terms of pedagogical design. That is the intervention consists not simply of providing these analytics to learners but also framing their interpretation as an integral course activity tied to goals and expectations.

2.1 Theoretical Framework and Research Base for Learning through Discussions

In this work online discussions are conceptualized from a social constructivist perspective as a venue in which learners can interact to build both collective and individual understanding through dialogue [2, 30]. Scholars differ in their theoretical accounts of the mechanisms underlying the learning process, referring variously to the importance of being exposed to multiple viewpoints, articulating one's own ideas, experiencing socio-cognitive conflict, and the negotiation and internalization of group understandings [36, 21, 30]. However, at a fundamental level all explanations depend on two basic processes that learners must engage in: "speaking" (externalizing one's ideas by contributing posts to the discussion); and "listening" (taking in the externalizations of others by accessing existing posts) [36]. Speaking in online discussions is clearly visible to others; this may explain why the bulk of research on and guidance for participation in online discussions is focused on posting activity [13]. However, while listening is largely invisible, it is also critical for discussions that build understanding in the ways described above [36]. Recent work has begun to document the different kinds of listening behaviors in which students engage [35, 36, 37] and has shown empirical connections between students' listening and speaking behaviors [34].

While the language of speaking and listening draws on a metaphor based in face-to-face conversations, online discussions offer different affordances and constraints for these activities [34, 35]. Specifically, in online discussions learners have greater control over the timeline and pace of their engagement [16]. This creates opportunities for thoughtful listening and reflective speaking [21], but also additional challenges of time management, especially for prolific discussions [25]. For this reason, helping learners to actively monitor and regulate how they speak and listen in online discussions is an important tool for supporting productive engagement in discussions.

Given the above-described goal of using dialogue to build individual and collective understandings and the existing research base on online discussion, particular speaking and listening behaviors can be characterized as more or less productive. For example, in terms of *speaking quantity*, multiple posts are needed to respond to others' ideas and questions, elaborate on the points made, and negotiate with others [24]. These posts should be distributed throughout the discussion (rather than concentrated at the start or end), relating to the *temporal distribution* of participation. In addition posts of moderate length best support rich dialogue since very short posts tend to be shallow in their presentation of ideas [6] but long posts are often perceived as overwhelming and thus not read [25]. Precise specifications of moderate length may differ by context and age-level; in higher education this is often specified around 100 to 200 words. In terms of *speaking quality,* posts that are clear, critical and connected to the existing conversation support the generation of insight and understanding of the topic [27]. Posts whose arguments are based on evidence and/or theory can also trigger others to productively build on or contest the points [4], and responses that clarify points, elaborate or question existing ideas, or synthesize different ideas together help deep the exploration of ideas and move the discussion forward [24]. In terms of listening activity there are also multiple dimensions to attend to. Considering *breadth of listening*, viewing a greater proportion of others' posts exposures students to more diversity of ideas and is

associated with richer responses [37]. *Depth of listening* is also important as an indication of the degree to which learners' are considering others' ideas and greater listening depth is predictive of richer argumentation in posts made [34]. *Listening reflectivity* (revisiting one' own and others' posts from earlier in the discussion) is also important to provide context for interpreting recent posts and examine how thinking has changed. Revisiting others posts has also been shown to be predictive of more substantive responses to others' ideas [34]. Finally *temporal distribution* of listening is important since engaging in multiple sessions during a discussion and integrating listening and posting in the same session can support making connections between ideas and posts which productively build on previous ones [36].

2.2 Analytics Embedded Real-Time in the Discussion Interface

In this work, we have chosen a specific discussion forum tool to use because of its inherent affordances for providing embedded analytics. The Visual Discussion Forum environment [22] was developed to present discussion threads as a hyperbolic (radial) tree structure, allowing students to easily see the structure of the discussion and the location of their comments within it (see Figure 1). Posts are represented as scaled colored spheres connected by lines indicating their reply relations. When a post is selected, it enlarges to the maximum size and moves to the center of the diagram while the other posts are rearranged around it; sphere size decreases with distance from the currently selected post. For each student, new posts are shown in red and previously viewed posts are shown in blue. The initial (seed) post always remains yellow. The design rationale and general benefits over traditional linear text-based forums have been described previously [22] and include more purposeful reading and replying behaviors by students, the ability to focus on a part of the discussion in the context of the whole, and increased reading of posts in a connected fashion.

In addition to these general benefits for interaction, the Visual Discussion Forum also provides an embedded visual analytic of listening and speaking behaviors in terms of how the group's discussion is proceeding and how each individual is attending and contributing to it. At the group level, the graphical representation of the tree's branches allows for easy inspection of the structure of the discussion. Learners can see how many different threads have been created thus far and how deep each is, using this information to inform their decision about where to read and contribute [22]. They can also examine which threads and posts are receiving the most attention (responses) and if any are being neglected. For example, in Figure 1. the post labeled "Transfer"…? has not yet been taken up by the group. This kind of analytic can be useful in addressing the problem of inadvertent thread abandonment or death [14].

At the individual level, the red/blue coloring of the posts helps each student easily track which parts of the discussion they have listened to already and which parts they have not yet attended to. For example in Figure 1 the student has been heavily attending to one thread, moderately attending to a second, and very minimally attending to a third. In addition, for the current application, we made an adaptation from the previous version of the forum to color the active learner's posts differently (light blue). In in this way, we provide an analytic to each student of how they have contributed to the discussion thus far in term of quantity (are their contributions high or low volume compared to overall quantity of discussion), breadth (to what extent are their posts distributed throughout the discussion), and intensity (if they have contributed multiple times to a specific thread). Students can also easily see which of their posts has been taken up by others and which have not. In the example shown in Figure 1.the learner has made five posts in two of the three active threads. Several of their posts have stimulated further discussion, in particular the one labeled "Synthesizing." Here they can also see their round-trip interaction [1, 11] with the student responding with "A question," and also that no one has yet addressed their post entitled "A concern."

This section has described a set of analytics that are naturally embedded in the design of the Visual Discussion Forum tool. An

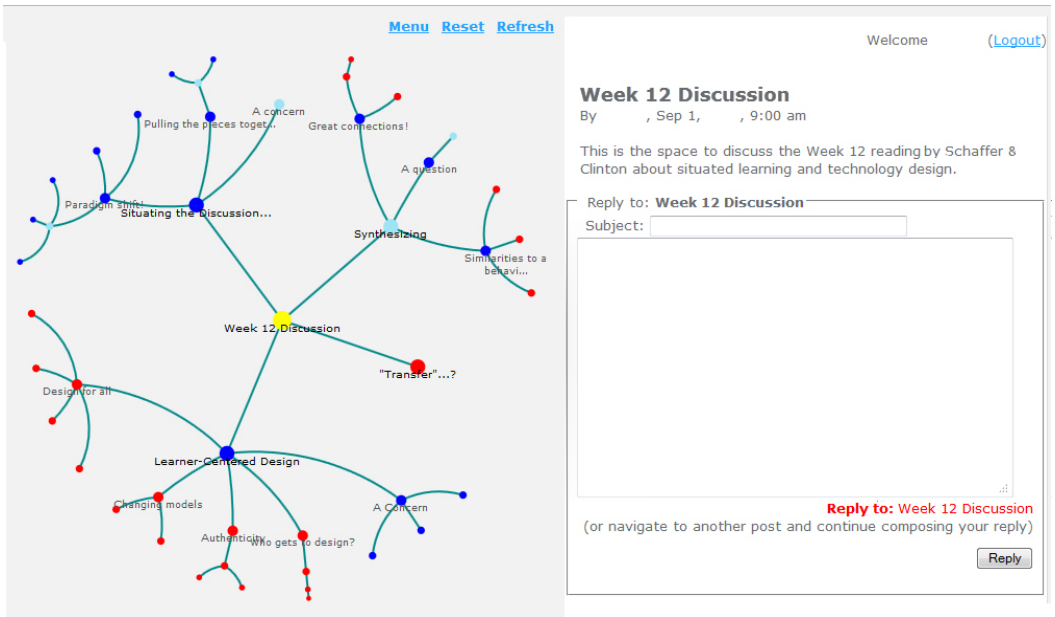

Figure 1. Visual Discussion Forum environment adapted for analytics

advantage to such embedded analytics is that they can be used seamlessly to support metacognitive aspects of the discussion activity itself. However, a weakness of their being embedded is that there is also the possibility that they may be ignored; i.e. there is no reason to assume that students will naturally use these affordances of the tool to support their participation individually or as a group. Thus a key aspect of effective use is support for such activity through pedagogical framing. In other words, such use needs to be specifically encouraged by structuring it in to the discussion activity parameters. The details of our pedagogical model for doing so and a description of its initial implementation are explained in Sections 3 and 4.

2.3 Analytics Extracted Periodically from the Discussion

In contrast to the embedded analytics described above, there is other useful information about students' online discussion activity that does not easily lend itself to presentation through the graphical interface (e.g. temporal distribution of participation). Thus, in our work with extracted analytics, we sought to make log-file trace data of speaking and listening activity visible to learners. The metrics used were developed based on our prior research investigating how students attend to the messages of others in online discussions described earlier [34, 35, 36, 37] and are summarized in Table 1. Data processing was implemented using a toolkit consisting of a combination of mySQL queries and Excel VBA macros as described below.

Table 1. Summary of discussion participation metrics

Metric	Definition	Participation Criteria
Range	Span of days a student logged in to the discussion.	*Temporal distribution*
Sessions	Number of times a student logged in to the discussion.	*Temporal distribution*
Percent of sessions with posts	Number of sessions in which a student made a post, divided by his/her total number of sessions.	*Temporal distribution*
Average session length	Total length of time spent in the discussion divided by his / her number of sessions.	*Temporal distribution*
Posts	Total number of posts a student contributed to the discussion.	*Speaking quantity*
Average post length	Total number of words posted by a student divided by the number of posts he/she made.	*Speaking quantity*
Percent of posts read	Number of unique posts that a student read divided by the total number of posts made by others to the discussion.	*Listening breadth*
Number of reviews of own posts	Number of times a student reread posts that he/she had made previously.	*Listening reflectivity*
Number of reviews of others posts	Number of times a student reread others' posts that he/she had viewed previously.	*Listening reflectivity*

Data was initially extracted from the log-file and posts tables in the discussion forum database and merged into a single spreadsheet file. This file lists each action taken by a student in the system in a row with the following information: action type (view-post, create-post, edit-post, delete-post), a time-date stamp, ID of user performing the action, ID of post being acted on, length of post being acted on, ID of user who created post being acted on. Macros were then used to clean the data, separate data by user, calculate action duration (through subtraction of sequential time stamps), divide actions into sessions-of-use (based on a 60-min abandonment threshold, see [36]), and make adjusted estimates for duration of session-ending actions (based on the relevant post's length and the average speed of the user in conducting the indicated action). View actions made on a user's own posts were re-coded as self-review actions and all view and review actions were sub-categorized as reads or scans based on a maximum reading speed of 6.5 words per second (wps) see [15]. Finally, nine variables were calculated based on the definitions shown in Table 1. Averages for the group were also calculated and a summary table was created for each learner (see Figure 2) as a straightforward way of presenting the data.

Please review the analytics about your discussion participation below, noting anything interesting you see in the data in the observations box and using this information to help you as you answer the reflection questions.

Metric	Your Data (Week X)	Class Average (Week X)	Observations
Range of participation	***4 days***	5 days	
# of sessions	***6***	13	
Average session length	***33 min***	48 min	
% of sessions with posts	***67%***	49%	
# of posts made	***8***	12	
Average post length	***149 words***	125 words	
% of posts read	***82%***	87%	
#of reviews of own posts	***22***	13	
#of reviews of others' posts	***8***	112	

Figure 2. Sample learner analytics summary

The metrics chosen for the extracted analytics were designed to be complementary to the embedded analytics; however two overlaps occur. First, the metric *Number of Posts Made* is viewable through the embedded analytics. However in the embedded analytic, the total number of posts made is less salient than their distribution. Thus providing this sum and the average for the group (which is not easily determinable from the interface) was deemed useful additional information. Second, the metric *Percent of Posts Read* is similar to the embedded red/blue color tracking of posts viewed in the interface. However, while the interface tracks all posts *viewed*, this metric is only based on posts actually *read* (not scanned) and thus provides complementary (and often quite different) information.

3. PEDAGOGICAL DESIGN OF THE LEARNING ANALYTICS INTERVENTION

When working with learning analytics a number of concerns have been raised about the dangers of rigidity of interpretation, lack of transparency with regards to data capture and access, the hegemony of optimizing to only that which can be measured, and possible impediment of learner's development of metacognitive and self-regulative learning skills [3, 5, 7]. Addressing such concerns requires attention not only to the how the analytics are developed and presented, but the pedagogical framework of activity that surrounds their use [5]. Thus while the analytics described above have the potential to be useful to instructors and students in monitoring and improving discussion participation, doing so requires active interpretation in the larger context of a shared understanding about the qualities of productive participation described earlier. Our pedagogical design for the use of the online discussion analytics carefully considered these challenges and attempted to address them through a number of core guiding principles. While the principles are described below in the context of our work with online discussion analytics, we believe that at the conceptual level they have the potential for applicability in a broad variety of learning analytics contexts.

3.1 Guiding Principles and their Instantiation

3.1.1 Integration

In order for analytics to be used by students in meaningful ways, there is a need for the metrics to be connected to the larger framework of purpose and expectations for the activity. For this reason it is important to *integrate the analytics metrics with the goals of the learning activity*. In other words, students need to understand (1) the purpose of the learning activity, (2) the characteristics of what the instructor views as productive engagement in the activity, and (3) how the learning analytics provided serve as a representation of this. In our work, we instantiate the first element by introducing the online discussions at the start of a course with a conversation about the goals of the activity as a vehicle for clarifying and building understandings through dialogue. This is particularly important given that students may view discussions quite differently—for example as a social space, a place to receive information, or a chance to show off what they know [20]. Depending on the scale and context of the learning experience, this first element might alternatively be enacted through a simple presentation or by including the students in the determination of the activity's goal. For the second element, we provide the students with clear guidelines for what is expected (and will be evaluated) for their discussion participation in terms of quantity, quality and timing of posting as well as broad, deep, integrated and reflective attention to the posts of others (see Figure 3).

These guidelines are reflective of the learning framework and research described earlier and students are made privy to this rationale. Finally, the analytics are introduced in the context of this framework. For the embedded analytics, mention is made in the participation guidelines in the appropriate section of how the interface can support this element of their participation (see Figure 3); for the extracted analytics, a separate guideline sheet is given with a chart describing each metric and how it relates to the participation criteria (see Figure 4). In this way, the metrics are not presented simply as a set of numbers, but ones which have clear meaning in the context of the learning activity.

3.1.2 Diversity & Agency

One important concern in using learning analytics is that the analytics alone will dictate how people engage in the learning activity and thus we "become what we measure", even though the metrics only capture some aspects of the overall activity [5, 7]. To address this concern it is important to *include multiple diverse measures* (so no one metric becomes the sole focus) and *support students in actively interpreting their meaning* (in the context of the larger framework of the activity goals and criteria described earlier). Of course the drive to provide multiple metrics needs to be balanced with care not to overload or overwhelm students unproductively. In our work, a selection of metrics are presented to students in a simple table format (see Figure 2). Importantly the guidelines present them as a starting point for consideration, not as absolute arbiters of one's engagement in the activity. This is done to help students develop an awareness of how they are participating in the discussion and take responsibility for the discussion and their actions in it. We also provide students with class averages for each metric to give them a context in which to consider the numbers.

Discussion Participation Guidelines

Attending to Others Posts

Broad Listening: Try to read as many posts as possible to consider everyone's ideas in the discussion. This can help you examine and support your own ideas more deeply. However, when time is limited it is better to view a portion in depth, then everything superficially.

**The visual interface shows posts that you have viewed in blue and new ones in red to help you track this.*

Purposeful Participation

Group Responsibility: As a group, we have a collective responsibility to tend to our discussion and make sure there aren't parts being neglected.

**The visual interface allows you to see which branches of discussion have received more attention than others.*

Figure 3. Excerpts from discussion participation guidelines

Learning Analytics Guidelines

Attending to Others' Posts	
% of posts read	The proportion of posts you read (not scanned) at least once. *It is good to read as many posts as possible to consider everyone's ideas in the discussion However, when time is limited it is better to view a portion in depth, then everything superficially.*
# of reviews of others' posts	The number of times you revisited others posts. *It is good to review others' posts to help you develop a higher level of response by relating back to others ideas. This number may be inflated by click-through, so should be evaluated relative to the group average.*

Figure 4. Excerpts from learning analytics guidelines

Finally, as described in the following sub-section, students are encouraged to set personal goals for their participation and to use the analytics to help monitor these. This supports individual student agency in using the analytics and sets up a situation of multiple possible profiles of productive participations, rather than a single goal to which all students must aspire. By making visible previously hidden listening activity, this approach can also highlight different students' various strengths in discussion participation; for example some students may need to work on their listening, while others may discover they are posting far less than others. In these ways we aim to create an environment of analytics use that is active, dynamic and personalized, with the goal of empowering students [3].

3.1.3 Reflection

One of the key attractions of learning analytics is the possibility to support the learner in actively reflecting on and taking action to manage their learning process [10]. From a constructivist perspective, reflection is an essential part of constructing one's understanding; in turn, as one's understanding develops, reflection can also be used more effectively [23]. However, online activities that can happen at anyplace and anytime often happen nowhere and never [18]; conversely attention to constantly available analytics can draw away from engagement in the activity itself. To support productive reflective activity, we *provide explicit time and space for reflection on the analytics*. In our work this is operationalized in the form of an online reflective journal shared between each student and the course instructor. The technology we employ for this purpose is a series of private wikis, however a variety of other technological solutions could serve the same function. At the start of the term, after students are given the guidelines, they are asked to set concrete goals for their participation in the discussion in the journal. They are then periodically provided with their analytics (as well as class averages as a benchmark) and given a series of reflective questions to respond to in the online journal (See Figure 5). Because the journal is shared, the instructor can review students' analytics and reflections as needed and respond in the same space. In this way interpretation of the analytics is not owned solely by the teacher or student but becomes negotiated between them. Storing the analytics, prompts, reflections and instructor comments in a digital journal also facilitates longitudinal review of changes and progress over the course of the term by both the student and instructor. The frequency with which the analytics are provided and reflective activity engaged in will vary depending on the context, but the goal of setting up specific timing is to avoid overwhelming students or making them overly reliant on the analytics [3]. As described in Section 4, in our current implementation we have found it most meaningful to provide analytics at the end of each week-long discussion since this is the unit of activity that students experience.

Below you are provided with some detailed information about your discussion participation for the past week generated from the system clickstream data. Please refer to the analytics guideline sheet to aid your interpretation and remember that more isn't always better. Note anything interesting you see in the data in the observations box and then use this information to help you answer the reflection questions.

1. What do you think went well in the group's discussion this week? What could be improved?
2. What do you think went well in your individual discussion participation in this week's discussion? What could be improved?
3. How does your participation in this week's discussion compare to previous weeks?
4. How well did you do in meeting the goal(s) you set for your participation in this week? How do you know?
5. Please set at least one goal for your participation in the coming week and write it below.

Figure 5. Sample reflective journal question prompts.

3.1.4 Parity & Dialogue

Another set of important issues in implementing learning analytics relate to questions of power, access and transparency [7, 9]. We address these issues through the principle of *establishing parity between the students and instructor in analytics use* and *creating a space of dialogue around the analytics and their interpretation*. To implement parity within our pedagogical design both the students and instructor keep a reflective journal based on their own analytics as described above. The instructor has access to and the ability to comment on students' reflection, creating a space of dialogue around the interpretation of the metrics in the context of each student's current goals. In turn, the students are free to read the instructor's reflections allowing him or her to model the reflective process and creating a sense of openness and equity around the use of the data. In addition, by having the instructor go through the same experience as the students he or she will have a better sense of how the metrics relate to actual activity, helping them to work with students to interpret meaning. The instructor is also thus exposing themselves to the same vulnerability as the students, can experience the same kinds of reactions to seeing their own analytics and thus have greater empathy in working with students. These activities alter the power balance from one in which the instructor collects data *on* the students into one in which data is used as a reflective and dialogic tool *between* the instructor and students.

It is important to note, however, that there are also possible negative implications of setting up parity with the instructor. Specifically, instructors may be unaccustomed to having their activity scrutinized by students and thus this principle may make them more hesitant to use analytics. It may also add additional self-imposed pressure on their discussion participation activities as they feel the need to set a good (or even ideal) model for students. In this sense the experiences of teachers and students are unbalanced as the teacher faces a situation in which the entire class may be focused on his or her activity traces, while each student knows that only the teacher (and possibly a few assistants) will be looking at theirs. Finally, the need for and expectations of teacher participation in a discussion may be different from that of students; if this is not communicated clearly then the teacher's analytics may create an inappropriate reference point for students.

4. IMPLEMENTATION AND INITIAL FINDINGS

The pilot implementation of this learning analytics approach was conducted in a semester-long blended graduate seminar on educational technology. This setting afforded willing students and a manageable class size with which to roll out the approach. This was done as a reasonably light-weight way to test our theoretical notions of how to provide useful analytics prior to the large investment needed to build a fully-automated system. After the model has been evaluated (and likely revised) in this "best case" scenario, it will be developed into a more automated and robust system and sequentially rolled out with larger classes, to the undergraduate population, and in fully online courses.

4.1 Implementation Context and Approach

The graduate seminar met once a week in a face-to-face session with a series of ten week-long online discussions interspersed between meetings. While discussions remained open for continuing the conversation after their designated week, in practice students focused their posting activity almost exclusively on the current-week discussion. The first discussion week was facilitated by the instructor to model good practice and give students a chance to acclimate to the tool; each subsequent discussion week was facilitated by one of the course's nine students. The assigning of student facilitators can be both a negative and positive in that by increasing learner responsibility for the discussion in one week, we potentially induce some level of abdication of responsibility for it in the others [39]. However, this is a tradeoff that we think is worthwhile since in past implementations of the course, the opportunity to facilitate was something that students found very valuable about the discussion activity and also something that they reported helped them understand the purpose of the discussions more deeply.

The parameters of the assignment were designed to support students in taking both individual and collective responsibility for the discussion. At the start of the term, students were introduced to the Visual Discussion Forum and engaged in a conversation about the goal and purpose of online discussions in the course, effective discussion participation strategies, effective discussion facilitation strategies and use of the ongoing embedded analytics as objects of reflection to understand and effect change on their discussion participation. Students were also provided with discussion participation guidelines and an online wiki-based reflective journal as described above. At the beginning of each four-hour class session students were given 10-15 min to write in their reflective journal based on the prompts described earlier. In between classes the instructor was invited to read students' comments and respond as needed.

For the first half of the term (which included five week-long discussions) the reflections were based solely on students' perceptions of the discussions and the embedded analytics. This was done to separate out the effects of the embedded analytics and provide a baseline for comparison once the extracted analytics were introduced. In the second half of the term (which also included five week-long discussions) students were provided with extracted analytics as described above. Because of the structure of the course, analytics were calculated using the discussion week as the unit of analysis. Using the semi-automated toolkit described, the weekly extraction, processing and preparing of the data took approximately 45 minutes. Both the embedded and extracted analytics were presented explicitly as objects of reflection to understand and effect change on one's discussion participation and not a tool of evaluation.

4.2 Initial Findings

In this section we report initial findings on the experience of using the analytics from the perspective of the course instructor and discuss their implications for the future revision of the analytics and pedagogical intervention model. In the future we will enhance this understanding by reporting on the experience from the student perspective and presenting a detailed analysis of the text of students' reflective journals and their log-file data (both efforts are currently in progress). Through this work we aim to assess both the degree to which the participants found the analytics useful for monitoring and reflecting on discussion participation, and to what extent this resulted in actual impact on students' discussion participation across the term.

4.2.1 Power of Dialogue

The reflective journal was created a space to encourage dialogue between the students and instructor around the interpretation of the analytics; however we were unsure to what extent this use would be taken up, especially in the first half of the course when only the embedded analytics were available. Surprisingly, the instructor reported that the reflective journaling was eagerly engaged in by students, even before the extracted analytics were provided. She further reported that this journaling provided a useful window into students' thinking around their discussion participation; both explaining external circumstances affecting activity and also showing that in many cases students were aware of the areas in which they needed to improve. While she had not planned to comment on every student's reflection every week, she found herself spending the time each week to do so because she felt it was useful in connecting with students individually, especially for those less vocal in the face-to-face setting. In contrast, the instructor's reflective journal, set up to enact the parity principle, was not taken up as a site for dialogue and in fact seems to be viewed rarely, if at all, by students.

Our reaction to the usefulness of the reflective journal before the extracted analytics were provided is mixed: on one hand this element seems to have contributed to a productive class environment and metacognitive attention by students to their participation in the online discussions; on the other hand there is a concern both with the evidentiary base for these reflections and that the pattern of reflection without the extracted analytics may have reduced attention to them when they were introduced.

4.2.2 Diverse Reactions to Extracted Analytics

The instructor reported that students had diverse reactions to the extracted analytics when they were introduced. Some students found them useful in providing "hard" numbers; however many pointed out that there is much they don't capture. While the analytics were referred to in the reflective journals, many learners still based much of their reflections on general perceptions. This may be because they found the analytics only moderately useful, they were continuing patterns from the initial reflections without data, or because the reflection prompts were not explicit enough in referring to the analytics. These questions will be addressed through the analysis of student interview data.

In general students' reflections on the analytics seemed to fall into two classes: validation of things students were already aware of and metrics that were surprising. Some of the surprises were positive; for example the instructor had felt she wasn't contributing enough to the discussions but the metrics showed her she was far above the class average. Other surprises were negative; for example, for many people the extracted metric of percent of posts read was substantially lower than what they expected based on the embedded red/blue posts-viewed analytic. The difference was due to scanning of posts; thus showing certain individuals that while they were attending to all posts at a minimal level, they were not listening to all of their peers' comments deeply. While this is an important piece of feedback for students to receive, there seemed to be an emotionally-charged element to some of the reactions to these results, with students variously feeling pleased, upset, or ashamed by their metrics.

4.2.3 Validating Invisible Activity

Finally, one of the most valuable outcomes of using the analytics that the instructor reported was that it honored and validated discussion forum activity occurring under the surface. Specifically with respect to the metrics capturing listening data it made her aware of the intense involvement of certain students who were very engaged with the discussion but did not always post a lot of comments. It also highlighted a lack of listening by some of the vociferous speakers. This led to reflective journal dialogues in which some students were able to point out their listening efforts while others realized that they needed to listen more deeply.

5. LIMITATIONS

The current efforts have several limitations; these are primarily related to conducting the pilot in a small context with advanced and highly motivated students. First, the model of one-on-one dialogue between the instructor and each student is not sustainable at scale. To replicate this interaction in more populous contexts, instructors can consider using peer commenters or consolidating the reflections in larger units as a formal assignment. We have implemented the latter approach to reflection (without analytics) in a large undergraduate class, suggesting it would also be viable for this purpose. Second, findings in this context may not be generalizable to students who are earlier in their post-secondary studies or less keen on learning; thus further testing will be needed. Finally, as measures of speaking quality are not directly assessable from log-file data, they are not currently included in the analytics; future metrics using computational linguistic approaches [e.g. see 9] are needed.

6. CONCLUSIONS

In conclusion, this paper has presented a theoretical framework for considering students' speaking and listening in online discussions, used this to develop analytics both embedded in and extracted from the learning environment to inform and improve these activities, and explicated a pedagogical model for the analytics intervention based on the principles of Integration, Diversity, Agency, Reflection, Parity, and Dialogue. Together these elements address the challenges of establishing traces of learning that are meaningful and presenting them in a format that is useful and transparent to learners while avoiding rigidity of interpretation, a myopic focus on only that which can be measured, and impediment of learner's development of metacognitive skills [3, 5, 7]. By framing the use of the analytics in a carefully designed pedagogical model of intervention, we seek to present them as a guide for sense-making that can empower students to take responsibility for regulating their own learning processes.

7. REFERENCES

[1] Anderson, T. 2008. Towards a theory of online learning. In T. Anderson (Ed.) *The theory and practice of online learning.* (pp. 45-74). Edmonton, Canada: Athabasca University Press.

[2] Boulos, M. N., & Wheeler, S. 2007. The emerging web 2.0 social software: An enabling suite of sociable technologies in health and health care education. *Health Information and Libraries Journal, 24*, 2-23.

[3] Buckingham Shum, S.& Ferguson, R. 2012. Social learning analytics. *Educational Technology & Society, 15*(3),3-26.

[4] Clark, D., Sampson, V., Weinberger, A., & Erkens, G. 2007. Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review, 19*, 343-374.

[5] Clow, D. 2012.The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd International Conference on Learning Analytics & Knowledge*, (Vancouver, Canada, 2012).

[6] Dennen, V.P. 2001. *The design and facilitation of asynchronous discussion activities in web-based courses: Implications for instructional design theory.* Doctoral thesis, IU-Bloomington

[7] Duval, E. & Verbert, K. 2012. Learning analytics. *E-Learning and Education, 1*(8).

[8] Ferguson, R. 2012. *The state of learning analytics in 2012: A review and future challenges.* Technical Report KMI-12-01, Knowledge Media Institute, The Open University

[9] Ferguson, R. & Buckingham Shum, S. 2011. Learning analytics to identify exploratory dialogue within synchronous text chat. In *Proceedings of the 1st International Conference on Learning Analytics & Knowledge*, (Banff, Canada, 2011).

[10] Govaerts, S., Verbert, K., Klerkx, J., Duval, E., 2010. December). Visualizing activities for self-reflection and awareness. In *Proceedings of 9th International Conference on Web-based Learning*, (Shanghai, China, 2010).

[11] Henri, F. 1992. Computer conferencing and content analysis. In A. R. Kaye (Ed.), *Collaborative learning through computer conferencing: the Najaden papers* (pp. 117– 36). Berlin: Springer-Verlag.

[12] Herring, S. 1999. Interactional coherence in CMC. *Journal of Computer-Mediated Communication, 4*(4).

[13] Hew, K.F., Cheung, W.S., & Ng, C.S.L. 2010. Student contribution in asynchronous online discussion: A review of the research and empirical exploration. *Instructional Science, 38*(6), 571-606.

[14] Hewitt, J. 2003. How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computer Research, 28*, 31-45.

[15] Hewitt, J., Brett, C., & Peters, V. 2007. Scan rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *American Journal of Distance Education, 21*(4), 215-231.

[16] Jonassen, D. H., & Kwon, H. I. 2001. Communication patterns in computer mediated versus face-to-face group problem solving. *Educational Technology Research and Development, 49*(1), 35-51.

[17] Jovanović, J., Gaševic, D., Brooks, C., Devedžic, V., Hatala, M., Eap, T., et al. 2008. LOCO-Analyst: Semantic web technologies in learning content usage analysis. *International Journal Of Continuing Engineering Education And Life Long Learning 18*(1), 54-76.).

[18] Jun, J. 2005. Understanding e-dropout. *International Journal on E-Learning, 4*(2), 229-240.

[19] Kear, K. 2001. Following the thread in computer conferences. *Computers & Education, 37*, 81–99.

[20] Knowlton, D. S. 2005. A taxonomy of learning through asynchronous discussion. *Journal of Interactive Learning Research, 16*(2), 155-177.

[21] Lipponen, L. 2002. Exploring foundations for computer-supported collaborative learning. In Stahl, G. (Ed.) In *Proceedings of Computer Supported Collaborative Learning* (Boulder, USA, 2002)

[22] Marbouti, F. 2012. Design, implementation and testing of a visual discussion forum to address new post bias. Masters thesis, Simon Fraser University.

[23] McAlpine, L., & Weston, C. 2000. Reflection: Issues related to improving professors' teaching and students' learning. *Instructional Science, 28*(5), 363-385.

[24] Pena-Shaff, J.B. & Nicholls, C. 2004. Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education, 42*, 243-65.

[25] Peters, V. & Hewitt, J. 2010. An investigation of student practices in asynchronous computer conferencing courses. *Computers & Education, 54*(4), 951-961.

[26] Roll, I., Aleven, V., & Koedinger, K. R. 2010. The invention lab: Using a hybrid of model tracing and constraint- based modeling to offer intelligent support in inquiry environments. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*. (Berlin, Germany, 2010)

[27] Rovai, A. P. 2007. Facilitating online discussions effectively. *The Internet and Higher Education, 10*, 77-88.

[28] Scardamalia, M. 2004. CSILE/Knowledge Forum. In *Education and technology: An encyclopedia* (pp.183-192). Santa Barbara: ABC-CLIO.

[29] Siemens, G.,Gašević, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson,R., et al. 2011. *Open Learning Analytics: An Integrated and Modularized Platform* (Concept Paper): SOLAR.

[30] Stahl, G. 2005. Group cognition in computer-assisted collaborative learning. *Journal of Computer Assisted Learning, 21*(2), 79-90.

[31] Swan, K. 2003. Learning effectiveness: What the research tells us. In J. Bourne & J. C. Moore (Eds), *Elements of quality online education, practice and direction* (pp.13--45). Needham,MA: Sloan Center for Online Education.

[32] Thomas, M. 2002. Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning, 18*, 351-366.

[33] Webb, E., Jones, A., Barker, P., & van Schaik, P. 2004. Using e-learning dialogues in higher education. *Innovations in Education and Teaching International, 41*(1), 93-103.

[34] Wise, A. F. Hausknecht, S., & Zhao, Y. In review. Relationships between speaking and listening in online discussions: An empirical examination. In *Proceedings of Computer Supported Collaborative Learning,* (Wisconson, USA, 2013)

[35] Wise, A. F., Perera, N., Hsiao, Y., Speer, J. & Marbouti, F. 2012. Microanalytic case studies of individual participation patterns in an asynchronous online discussion in an undergraduate blended course. *The Internet and Higher Education, 15,* 108–117.

[36] Wise, A. F., Speer, J., Marbouti, F. & Hsiao, Y. 2012. Broadening the notion of participation in online discussions: Examining patterns in learners' online listening behaviors. Advance online publication. *Instructional Science.*

[37] Wise, A., Hsiao, Y., Marbouti, F., Speer, J. & Perera, N. 2012. Initial validation of "listening" behavior typologies for online discussions using microanalytic case studies. In *Proceedings of International Conference of Learning Sciences*, (Sydney, Australia, 2012).

[38] Wise, A. F., Marbouti, F., Hsiao, Y. & Hausknecht, S. In press. A survey of factors contributing to learners' "listening" behaviors in asynchronous discussions. Accepted for publication in the *Journal of Educational Computing Research.*

[39] Wise, A. F. & Chiu, M. M. 2013. The effects of summarizing roles on learner's listening behaviors in online discussions. Paper presented at the *2013 Annual Meeting of the American Educational Research Association*, San Francisco, CA.

8. Acknowledgements

We would like to thank the Social Sciences and Humanities Council of Canada for their generous support of this work and Farshid Marbouti for the use of the Visual Discussion Forum and making the adaptations to use it in this work.