

A Process Mining Approach to Linking the Study of Aptitude and Event Facets of Self-regulated Learning

Sanam Shirazi Beheshitha
School of Interactive Arts and
Technology
Simon Fraser University
Burnaby, Canada
sshirazi@sfu.ca

Dragan Gašević
School of Computing Science
Athabasca University
Athabasca, Canada
dgasevic@acm.org

Marek Hatala
School of Interactive Arts and
Technology
Simon Fraser University
Burnaby, Canada
mhatala@sfu.ca

ABSTRACT

Research on self-regulated learning has taken main two paths: self-regulated learning as aptitudes and more recently, self-regulated learning as events. This paper proposes the use of the Fuzzy miner process mining technique to examine the relationship between students' self-reported aptitudes (i.e., achievement goal orientation and approaches to learning) and strategies followed in self-regulated learning. A pilot study is conducted to probe the method and the preliminary results are reported.

Categories and Subject Descriptors

K.3.1[Computers and Education] Distance Learning

General Terms

Human Factors, Measurement

Keywords

Self-Regulated Learning, Clustering, Learning Patterns, Process Mining

1. INTRODUCTION

Existing research on self-regulated learning has taken two main paths: self-regulated learning as aptitudes and more recently, self-regulated learning as events [1]. Aptitudes are theoretical constructs underlying the observed differences between individual learners in specific contexts such as motivational factors, learning styles and epistemic beliefs [2]. They provide as the basis of standards that students use to regulate their learning. More specifically, they influence how learners metacognitively monitor and then control their learning by enacting certain cognitive tactics and strategies in action. Self-reported questionnaire is a common method of measuring aptitudes.

Self-regulated learning can also be conceived in terms of events, the actual actions learners execute rather than underlying mental states that is caused by them [3]. Recent advances in technology-enhanced learning enables for capturing fine-grained trace data of learners' actions in online learning environments. Investigating occurrence, temporal sequence or regular patterns of events that are products of learners' cognitive activities and strategies allows for tracing aptitudes in practice [1].

Several studies looked at occurrence and sequence of self-regulated learning actions using the concordance analysis [4], activity transition graph [5] and sequence mining [6]. It is argued that self-regulated learning behaviors, although weakly sequenced, can be assumed to be driven by an underlying holistic model [1]. A recent study introduced adoption of process mining techniques to uncover regular patterns governed by the holistic process [7]. The reported study examined differences in the self-regulated learning strategies and activities and students that have different aptitudes measured by two different constructs.

1.1 Aptitude Measures

In this paper, we specifically focus on the two aptitude measures: achievement goal orientation and approaches to learning. Achievement goal orientation is a common motivational construct which describes "the purpose of engagement in an achievement behavior" [8]. Mastery goal orientations focus on the development of task competence, whereas performance goal orientations emphasize on the illustration of performance competence [9] (with respect to self or others [10]). In terms of valence, these achievement goals may be further distinguished by approaching success and avoiding failure in a certain competence [11]. Prior research has examined association of achievement goal orientations with self-regulated learning behaviors and strategies mostly through self-report instruments measuring constructs such as strategy use [12] or learning outcome [13].

Another aptitude construct describes students' preferred approaches to learning for a specific task, in a particular teaching context with respect to two dimensions – motives and strategies [11]. Surface study approaches have a preference towards rote learning which is motivated by failure avoidance. Whereas, deep approaches have an orientation towards comprehending and sense making with intrinsic motivation [14]. Relevant literature, has investigated influence of personal and context factors on study approaches [15], as well as how study approaches affect learning outcome and processes through self-report questionnaires [16].

1.2 Research Questions

To our best knowledge, in the context of self-regulated learning, process mining techniques have not yet been applied to identify learning patterns with respect to different aptitudes (e.g., achievement goal orientations and study approaches). This study uses process mining to examine the relationship between students' self-reported aptitudes and how they self-regulate their learning in action. More specifically, we aim to address the following questions:

- Can we identify groups of students with similar self-reported study approaches/achievement goal orientations?

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference'10, Month 1–2, 2010, City, State, Country.
Copyright 2010 ACM 1-58113-000-0/00/0010 ...\$15.00.

- Do groups of students with different self-reported study approaches/achievement goal orientations follow different learning strategies and perform different study activities?

To probe a research method that can address the above research questions, we carried out a pilot study with undergraduate students who were asked to conduct a research project and write an essay in an online tool for self-regulated learning.

2. METHOD

2.1 Participants

Participants were third year undergraduate students of an interdisciplinary Interactive Arts and Technology program (n=22) at a research-intensive university in Canada (Female:27%, Male:73%; Age 18 to 24: 14%, 25 to 34: 86%).

2.2 Course and Learning Task

The study was conducted in the context of a course dedicated to internet computing technologies. The course was mainly about the software development with XML and Web 2.0 technologies. Students' learning was assessed through several assignments, quizzes, final exam and a research project. As a learning task within this study, the students were asked to perform an independent research project on the topic of their choice and write a report in the length of 1200-1500 words. The outcome was evaluated in terms of writing, structure, quality of sources and argumentation and was worth 5% of their final grade. Participation in the study was voluntary and a bonus of 3% counted towards the final grade was offered in return for participation. The participants were asked to fill out several questionnaires and submit those electronically prior to the learning task. Then, they were asked to use the nStudy tool [17] to perform their research and write a report. The students were introduced to nStudy during a lecture session and were advised to use it on their own and seek help if needed.

2.3 Instruments

The R-SPQ-2F questionnaire was adopted to evaluate students approaches to study [14]. It is composed of 20 items, grouped into 4 sub-scales corresponding to study approaches dimensions (surface motive, surface strategy, deep motive, and deep strategy). The students gave responses on a Likert-type scale, from 1 (never or only rarely true of me) to 5 (always or almost always true of me). Out of all students, 20 completed this questionnaire.

The 3x2 AGQ questionnaire was also used to investigate students achievement goals [10]. It consists of 18 items, grouped into 6 categories corresponding to achievement goals (task-approach, task-avoidance, self-approach, self-avoidance, other-approach, and other-avoidance), whereby self and task represent mastery goals and other represents performance goals. The responses were recorded on a Likert-type scale, from 1 (not at all true of me) to 7 (very true of me). Nineteen students completed this questionnaire.

2.4 nStudy Tool

nStudy is a learning tool offered as a Firefox add-on. It is primarily designed to help learners organize and process information as they self-regulate their study of online materials (Figure 1) [17]. It also captures log data of learners' activities within the environment. A summary of all learning activities is provided in Table 1. In this study, the participants were asked to use nStudy within their research project to bookmark and organize online resources, highlight and quote key points, take notes, define terms and write the final report document. Time-stamped trace data of participants' interaction with nStudy was collected

for analysis. Twenty students used nStudy for their research project.

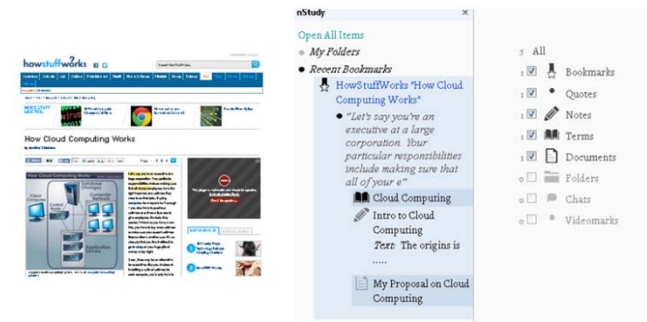


Figure 1: nStudy Tool

2.5 Analysis

To address the research questions, the following analysis were conducted. First, clustering was adopted on questionnaires' results to identify groups of participants based on similarity of scores on goal orientation and study approaches questionnaires' scales. Given the small size of dataset, agglomerative hierarchical clustering algorithm was used with Euclidian distance measure and Ward's criterion.

Further, the trace data logged by nStudy was analyzed on two levels:

- *Learning Strategy*: Coarse-grained coding based on cognitive self-regulated learning strategies – Rehearsal, Organization, and Elaboration – was done based on the framework introduced in [18]. Rehearsal strategies that involve repetition and revisit of studied resources have lowest cognitive demand (e.g., Create Quote, Expand Note or Switch to Bookmark). However, Organization strategies which emphasize on changing the structure of information and studied resources (e.g., Include Note or Edit Folder Title), require higher cognitive processing. Finally, Elaboration strategies, the most cognitive demanding strategies, focus on activities that modify or transform studied information (e.g., Create Term or Edit Document Content). The complete list of nStudy learning activities and corresponding learning strategy codes is presented in Table 2.
- *Learning activity*: Fine-grained coding based on nStudy learning activities (actions performed on a certain nStudy object).

Given the characteristics of the dataset, the Fuzzy miner algorithm was used for process mining. Inspired by the metaphor of maps, Fuzzy miner was proposed for exploratory process mining of complex, less-structured data based on four principles [12]: 1) Aggregation of low-level details and displaying it as one cluster, 2) Abstraction of low-level information which is not significant in a particular context, 3) Emphasis on significant information, 4) Customization of level of details based on particular context and purpose. Once the Fuzzy miner algorithm is performed on a sequence of coded events, the output process model was a simplified descriptive transition diagram. Within this diagram, nodes illustrated event classes and directed edges displayed the relationships between the event classes.

Table 1. nStudy Learning Activities

Action	nStudy Object	Learning Activity Description
Create	All Artifacts*, Chat	Generate a new instance
Delete	All Artifacts, Chat	Delete an existing instance
Permanently Delete	All Artifacts, Chat	Delete an existing instance permanently
Undelete	All Artifacts, Chat	Undo deletion of an existing instance
Edit Content	Document, Note, Quote, Term, Chat	Modify content text of an object
Edit Title	All Artifacts	Modify title text of an object
Expand	All Artifacts, All System Components ⁺	Open on the sidebar tree structure
Collapse	All Artifacts, All System Components	Close on the sidebar tree structure
Include	All Artifacts	Include as child of other components on the sidebar tree structure
Exclude	All Artifacts	Exclude child from parent components on the sidebar tree structure
Switch To	Bookmark, Webpage, nStudy Page	Switch the active page

Note. A learning activity is referred to an action applied on an nStudy object.

* (User-generated) Artifacts: Bookmark, Document, Folder, Note, Quote and Term.

⁺System Objects: Chat, (root) Collection (of folders), History (of Bookmarks) and Trash

In order to implement above principles, two fundamental metrics were used:

- *Significance*: It measures relative importance of an event class (*Unary Significance*) or a relationship (*Binary Significance*).
- *Correlation*: It is used for relationships (*Binary Correlation*) and computes "how closely related two events following one another are" [12, p.333].

Table 2: Mapping nStudy learning activities to self-regulated learning strategies

Action	Learning Activity nStudy Object	Learning Strategy Code
Create	Bookmark/Quote	Rehearsal
Create	Document/Folder	Organization
Create	Note/Term	Elaboration
Delete	(any applicable) object	Organization
Undelete	(any applicable) object	Organization
Permanently Delete	(any applicable) object	Organization
Edit Content	Note/Quote/Term/Document	Elaboration
Edit Title	Bookmark/Document/Folder/Note	Organization
Edit Tile	Term	Elaboration
Expand	(any applicable) object	Rehearsal
Collapse	(any applicable) object	Rehearsal
Include	(any applicable) object	Organization
Exclude	(any applicable) object	Organization
Switch to	Bookmark/Webpage	Rehearsal
Switch to	nStudy Page	Remove

Then the following steps were applied to simplify process model according to the principles:

- *Edge Filtering*: If utility of an edge is below a configured threshold then the edge is removed from the output model. The utility of an edge is computed as the weighted sum of *Binary Significance* and *Binary Correlation*. The weight is called *Utility Ratio*:

$$Util = UtilRatio * (BinSig) + (1 - UtilRatio) * (BinCorr)$$

- *Node Filtering and Aggregation*: If utility of a node is below a configured threshold then the node is either removed or clustered with other nodes as follows: 1) Low utility nodes with *low Binary Correlation* are removed. 2) Low utility nodes with *high Binary Correlation* are clustered together. The utility of a node is equal to *Unary Significance*.

The following parameter settings were used in our process mining analysis: *Significance* and *Correlation* were computed based on the frequency and routing measures for *Unary*, the frequency and distance measures for *Binary*, and the temporal proximity for *Correlation* [12]. The significance cut off for the node filter was set to 0.25. The edge *Utility Ratio* was fixed at 0.75. The edge cut off was set to 0 for the Learning Strategy level and 0.2 for Learning Activity level.

3. RESULTS

Trace files included 6758 interactions of 20 participants generated over the period of 3-9 study sessions each lasting 20-120 minutes. In total, 718 artifacts were produced using nStudy (Table 3).

Table 3: Descriptive statistics of nStudy artifacts produced

nStudy Artifact	M	SD	Total
Quote	15.5	14.28	310
Bookmark	13.2	7.22	264
Folder	2.15	2.6	43
Term	2.05	2.95	41
Document	1.7	1.26	34
Note	1.25	1.71	25

3.1 Cluster Analysis

The result of cluster analysis revealed two clusters based on the study approaches. As presented in Table 4, these two clusters (i.e., *Surface Learners* and *Deep Learners*) had significantly different scores on the surface and deep approach scales. However, no significant difference was found in the nStudy variables describing the use of different functionalities.

Clustering based on the achievement goal orientation scale scores did not yield satisfactory results as the difference between valences of same competence were not salient within a cluster.

Thus, no further analysis of the learning strategy and learning activities was possible using process mining.

Table 4: Cluster analysis result based on study approach scale scores

Variable	Cluster 1: Surface Learners		Cluster 2: Deep Learners		t-test (t-value)
	M	SD	M	SD	
Surface Ap	0.67	0.70	-0.86	0.54	-5.32***
Deep Ap	-0.51	0.99	0.63	0.58	3.02**

N=18; Surface Learners (n=10); Deep Learners (n=8);

p<.01, *p<.001

3.2 Process Mining

Figure 2 shows the resulting model with events coded on the *Learning Strategy level*. Rehearsal strategies had the highest utility in both models. *Deep Learners* (Figure 2, b) manifested a slightly higher utility of Elaboration strategies and a lower utility of Organization strategies than *Surface Learners* (Figure 2, a). In addition, *Surface Learners* demonstrated more transitions between strategies. Both models show a one-way loop from Rehearsal to Elaboration, then to Organization and at last back to Rehearsal. However, for *Surface Learners* there is an additional strong transition from Organization to Elaboration. On the other hand, process mining at the Learning Activity level did not uncover substantial differences between the two groups of learners.

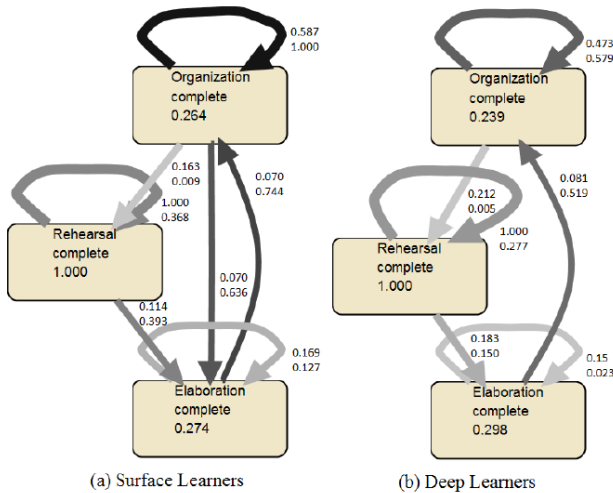


Figure 2: Process model of Surface Learners (n=10) and Deep Learners (n=8) at learning strategy level. The numbers in the nodes indicate unary significance and the numbers on the edges represent binary significance (upper) and binary correlation (lower).

4. DISCUSSION AND CONCLUSIONS

Cluster analysis on aptitude questionnaire subscales identified two clusters based on study approaches (i.e., *Deep Learners* and *Surface Learners*). This is aligned with findings of several previous studies that had similar findings. For example, Bliuc et al. identified two groups of students based on approaches to online and face-to-face learning, as well as conceptions of

learning and academic performance [19]. According to their findings, one group of students aimed for deeper understanding, whereas the other group was concerned with reproduction to complete the learning task.

Process mining on the interactions of each group with nStudy revealed a slightly higher utility of Elaboration strategies for *Deep Learners* and a slightly higher utility of Organization strategies for *Surface Learners*. Generally, *Deep Learners* have orientation towards deeper understanding and less memorization. Hence, their strategy around self-study involves higher cognitive processing of new information [14]. This is assumed to be a probable reason for the higher utility of Elaboration strategies, which demand the highest amounts of cognitive processing. In addition, the higher number of edges in the process model of *Surface Learners* is indicative of a more frequent use of learning strategies. It seems that *Deep Learners* were more selective in their use of strategy than *Surface Learners*, which aligns with prior research showing that effective learners with on-track solutions to the task adopt strategies more effectively rather than more frequently. Hence, they exhibit less actions and complex patterns [20].

Higher temporal proximity correlation observed in the process of *Surface Learners* is an indication of shorter time interval between subsequent switches in strategy use than *Deep Learners*. This could mean that *Deep Learners* spend more time applying a strategy of a certain type before switching to another, particularly Elaboration (e.g., editing content of a document, note or definition of a term) and Rehearsal strategies (e.g., highlighting important materials or reviewing them). Orientation of *Deep Learners* towards meaning seeking can serve as the underlying reason for this additional time [14]. An alternative explanation according to the Winne and Hadwin model of self-regulated learning is that *Deep Learners* deliberate after using a cognitive strategy to metacognitively monitor the learning progress against learning goals and objectives and control their learning by planning adjustments to future cognitive operations and strategies [21].

Our findings suggest that process mining techniques are useful for exploratory analysis of self-regulated learning data. They can uncover sequential and temporal relationships that are not otherwise observable through conventional methods of analysis. However, they do not provide causal explanations. In the case of fuzzy miner, the output is a descriptive transition diagram and therefore, cannot be subjected to a statistical significance test. Hence, significance of any observed difference between output models of the two groups of students cannot be statistically confirmed [7].

Several limitations might have affected both external and internal validity of the findings. First, the small sample size might have affected the statistical power. Moreover, with small number of cases, minor changes in parameter configuration of process mining result in considerable variations in the output process model.

Second, the study did not cover a separate learning session where students can familiarize themselves with the tool at hand before the actual experiment. Therefore, trace data includes both the learning phase and actual testing phase which might have influenced the accuracy of findings since students did not have sufficient time to regulate their tool use [22].

Third, post-study surveys about students' experience with the tool uncovers moderately low tool acceptance. Technical issues and low usability of the tool had negatively affected the perceived ease of use (M= 2.05 of 5). Two students withdrew from using the tool halfway through the study. Further interviews with one of the

participant revealed low perceived usefulness in comparison to similar available tools [23].

Fourth, students' self-reports of their aptitudes through close-ended questionnaires are not necessarily accurate. Very often responses are influenced by cognitive illusions, such as prior experience, and are not tuned to the learning task at hand [24]. In addition, students' aptitudes can change over the course of an experiment. Surveys fail to capture temporal shifts [1], particularly with study approaches where additional factors such as teaching and task context have high influence [14].

Despite these limitations, our methodology suggests a new venue for analysis of learning strategies and processes followed in self-regulated learning. Investigating a temporal sequence of events, which is assumed to be governed by a learning process, is in agreement with theoretical models for self-regulated learning [25]. The proposed method needs to be validated with larger sample sizes than the one in the present study and with alternative measures of aptitudes such as the trace data-based measures of achievement goal orientation as suggested by Zhou and Winne [26].

5. REFERENCES

- [1] P. H. Winne, "Improving measurements of self-regulated learning," *Educ. Psychol.*, vol. 45, no. 4, pp. 267–276, 2010.
- [2] R. E. Snow, "Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy," *J. Consult. Clin. Psychol.*, vol. 59, no. 2, p. 205, 1991.
- [3] P. H. Winne, "Issues in researching self-regulated learning as patterns of events," *Metacognition Learn.*, pp. 1–9, 2014.
- [4] F. De Jong, "Task and student dependency in using self-regulation activities: Consequences for process-oriented instruction," *Process-Oriented Instr. Verbal Pict. Aid Comprehension Strateg.*, pp. 87–99, 1994.
- [5] A. F. Hadwin, J. C. Nesbit, D. Jamieson-Noel, J. Code, and P. H. Winne, "Examining trace data to explore self-regulated learning," *Metacognition Learn.*, vol. 2, no. 2–3, pp. 107–124, 2007.
- [6] M. Kapur, J. Voiklis, and C. K. Kinzer, "Sensitivities to early exchange in synchronous computer-supported collaborative learning (CSCL) groups," *Comput. Educ.*, vol. 51, no. 1, pp. 54–66, 2008.
- [7] M. Bannert, P. Reimann, and C. Sonnenberg, "Process mining techniques for analysing patterns and strategies in students' self-regulated learning," *Metacognition Learn.*, pp. 1–25, 2014.
- [8] A. J. Elliot, A. J. Elliot, and C. S. Dweck, "A conceptual history of the achievement goal construct," *Handb. Competence Motiv.*, vol. 16, pp. 52–72, 2005.
- [9] C. Ames, "Classrooms: Goals, structures, and student motivation," *J. Educ. Psychol.*, vol. 84, no. 3, p. 261, 1992.
- [10] A. J. Elliot, K. Murayama, and R. Pekrun, "A 3 \times 2 achievement goal model," *J. Educ. Psychol.*, vol. 103, no. 3, p. 632, 2011.
- [11] J. B. Biggs, *Student Approaches to Learning and Studying. Research Monograph*. ERIC, 1987.
- [12] K.-L. Lau and J. Lee, "Examining Hong Kong students' achievement goals and their relations with students' perceived classroom environment and strategy use," *Educ. Psychol.*, vol. 28, no. 4, pp. 357–372, 2008.
- [13] L. Linnenbrink-Garcia, D. F. Tyson, and E. A. Patall, "When are achievement goal orientations beneficial for academic achievement? A closer look at main effects and moderating factors," *Rev. Int. Psychol. Soc.*, no. 1, pp. 19–70, 2008.
- [14] J. Biggs, D. Kember, and D. Y. Leung, "The revised two-factor study process questionnaire: R-SPQ-2F," *Br. J. Educ. Psychol.*, vol. 71, no. 1, pp. 133–149, 2001.
- [15] M. Baeten, E. Kyndt, K. Struyven, and F. Dochy, "Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness," *Educ. Res. Rev.*, vol. 5, no. 3, pp. 243–260, 2010.
- [16] Z. Akyol and D. R. Garrison, "Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning," *Br. J. Educ. Technol.*, vol. 42, no. 2, pp. 233–250, 2011.
- [17] P. H. Winne and A. F. Hadwin, "nStudy: Tracing and supporting self-regulated learning in the Internet," in *International handbook of metacognition and learning technologies*, Springer, 2013, pp. 293–308.
- [18] C. E. Weinstein, T. W. Acee, and J. Jung, "Self-regulation and learning strategies," *New Dir. Teach. Learn.*, vol. 2011, no. 126, pp. 45–53, 2011.
- [19] A.-M. Bliuc, R. Ellis, P. Goodyear, and L. Piggott, "Learning through face-to-face and online discussions: Associations between students' conceptions, approaches and academic performance in political science," *Br. J. Educ. Technol.*, vol. 41, no. 3, pp. 512–524, May 2010.
- [20] J. Malmberg, S. Järvelä, and P. A. Kirschner, "Elementary school students' strategic learning: does task-type matter?," *Metacognition Learn.*, pp. 1–24, 2014.
- [21] P. H. Winne and A. F. Hadwin, "Studying as self-regulated learning," *Metacognition Educ. Theory Pract.*, vol. 93, pp. 27–30, 1998.
- [22] G. Lust, J. Elen, and G. Clarebout, "Regulation of tool-use within a blended course: Student differences and performance effects," *Comput. Educ.*, vol. 60, no. 1, pp. 385–395, 2013.
- [23] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: four longitudinal field studies," *Manag. Sci.*, vol. 46, no. 2, pp. 186–204, 2000.
- [24] R. A. Bjork, J. Dunlosky, and N. Kornell, "Self-regulated learning: Beliefs, techniques, and illusions," *Annu. Rev. Psychol.*, vol. 64, pp. 417–444, 2013.
- [25] P. H. Winne and J. C. Nesbit, "Graph theoretic techniques for examining patterns and strategies in students' studying: An application of LogMill," in *Annual Meeting of the American Educational Research Association, San Francisco*, 1995.
- [26] M. Zhou and P. H. Winne, "Modeling academic achievement by self-reported versus traced goal orientation," *Learn. Instr.*, vol. 22, no. 6, pp. 413–419, 2012.