Investigating SMART Models of Self-Regulation and their Impact on Learning

Stephen Hutt[1], Jaclyn Ocumpaugh[1], Juliana Ma. Alexandra L. Andres[1], Nigel Bosch[2], Luc Paquette[2], Gautam Biswas[3], and Ryan S. Baker[1]

[1] University of Pennsylvania
[2] University of Illinois at Urbana-Champaign
[3] Vanderbilt University
hutts@upenn.edu

ABSTRACT

Self-regulated learning (SRL) is a critical 21st-century skill. In this paper, we examine SRL through the lens of the searching, monitoring, assessing, rehearsing, and translating (SMART) schema for learning operations. We use microanalysis to measure SRL behaviors as students interact with a computer-based learning environment, Betty's Brain. We leverage interaction data, survey data, *in situ* student interviews, and supervised machine learning techniques to predict the proportion of time spent on each of the SMART schema facets, developing models with prediction accuracy ranging from *rho* = .19 for translating to *rho* = .66 for assembling. We examine key interactions between variables in our models and discuss the implications for future SRL research. Finally, we show that both ground truth and predicted values can be used to predict future learning in the system. In fact, the inferred models of SRL outperform the ground truth versions, demonstrating both their generalizability and their potential for using these models to improve adaptive scaffolding for students who are still developing SRL skills.

Keywords

Self Regulation, SMART, Self Regulated Learning, Machine Learning, Student Interviews

1. INTRODUCTION

In traditional classrooms, most support for acquiring self-regulated learning (SRL) strategies comes from teachers, who might check in on projects and/or provide advice about next steps [33] in order to keep students focused on their end goals. However, teachers' external regulation alone is insufficient to encourage educational success [24]; the learner must also develop internal regulation schemas. SRL demands may increase when the student is completing a project in a computer-based learning environment that is no longer teacher-led. The software might scaffold learning activities, but identifying the complex behaviors involved with SRL is still not a typical function of most computer-based learning systems.

In most computer-based learning environments, learners must control, manage, plan, and monitor their learning [12], i.e., implement the definitional components of SRL. SRL has consistently been shown to facilitate knowledge acquisition and retention among learners in a structured and systematic way [12]. As such, work has called for a deeper understanding of SRL impacts in online learning [1, 8, 37].

A range of techniques have been used to better understand SRL both in computer-based learning environments (e.g., [1, 5, 12, 34]) and in other contexts (see [17, 27] for meta-analyses). Research in computer-based learning can be split into two groups: supporting SRL and detecting SRL behaviors [46]. Supporting SRL has taken a number of forms, but in general, these approaches typically scaffold students in either their goal-setting, self-evaluation, help-seeking, self-efficacy, or some combination of these [29]. This might be through verbal prompts (e.g. "Take time to read everything,") [7, 22] or more intricate support systems [25], such as progress bars [14], or tools such as notebooks, that better facilitate student reflection [2, 35].

In terms of detecting SRL in computer-based learning environments, Azevedo and colleagues have (using MetaTutor) considered the role that emotion plays in regulation, posing that affect should be considered as we scaffold SRL behaviors [4]. Segedy et al. [36] used interaction data and coherence analysis to measure self-regulation. Learner behaviors were tracked using log files to assess action coherence (i.e., did a student's actions present a coherent strategy relevant to the current tasks), which was shown to predict learning. Winne et al. [45] also leveraged log data in a scalable system that traces student actions, classifying each learning event into SRL categories in order to better understand student cognition, motivation, and metacognition. We build upon this approach in this work.

While interaction data has been successfully used to detect SRL, a number of researchers argue that this data should not be considered in isolation [3, 37, 40]. Instead, we must also consider contextual factors and individual differences not easily inferred from logs. This work combines interaction data with data from targeted in-situ student interviews and student survey data to predict SRL as characterized by the COPES and subsequent SMART models of SRL [42] (discussed in detail below). We examine the impact of SRL on learning, analyzing contextual and student-level factors that may influence SRL behavior and demonstrating the potential of the latent encoding of SRL for identifying students who need further support.

1.1 Related Works

At a high level, SRL is a process in which learners take initiative to identify their learning goals and then adjust their learning strategies, cognitive resources, motivation, and behavior to optimize their learning outcomes [11, 42]. First characterized in

1989 [47], SRL is now widely acknowledged as an essential skill for learning in the modern knowledge-driven society [23]. In learning technologies specifically, recent work has called for a deeper understanding of SRL and for learning technology that supports the development of SRL strategies [1, 3, 8, 20, 37].

In order to provide insight into how SRL works, researchers have proposed a number of theoretical models (e.g., [30, 47]). Winne & Hadwin's model [43], grounded in information processing theory, characterizes SRL as a series of events that happen over four recursive stages: (1) task definition, (2) goal setting and planning, (3) studying tactics, and (4) metacognitive adaption of studying techniques. Each stage is then characterized by Conditions, Operations, Products, Evaluations, and Standards (COPES). In later work, Winne subcategorized the COPES model further by detailing five kinds of operations—searching, monitoring, assembling, rehearsing, and translating—known as the SMART model [42].

In the context of educational data mining, we can study SRL by measuring these theoretical constructs and studying their relationships to each other and to external measures (such as achievement). SRL constructs can be measured either online (while an activity is happening) or offline (before or after an activity) [34]. Offline assessments typically rely on self-report questionnaires, but student interviews have also been used. These can be implemented either online and offline and can offer advantages over questionnaires that may limit students to pre-defined answers [16, 40].

Trace analysis is perhaps the main approach used (and endorsed [37]) to measure SRL online. Traces (such as log data) capture learning actions along with additional contextual and timing information, providing a detailed window into a learner's processes and behaviors [40]. This data can support microanalytic approaches, as sequences of actions can be aligned with different facets of a self-regulation model [21, 45]. Models that conceptualize SRL in terms of events or student actions (such as the COPES model [43]) lend themselves more to a trace-based analysis [42] than to offline measurement. However, many researchers argue that trace data should be supplemented with additional measurements (e.g., self-reports or think-alouds) when measuring SRL [3, 37, 40].

1.2 Current Study

The current study was conducted within the context of Betty's Brain, a computer-based learning environment for middle school science. We combine multiple data sources (interaction, surveys, and interview data) to analyze SRL patterns through the lens of Winne's COPES and SMART models [42].

We first demonstrate that combining features from different data sources yields the most successful models of the SMART facets. We present a feature analysis to investigate the key interactions in each model. We next examine how the different facets of SRL influence student learning. We consider not only the ground truth calculations of SMART facets but also our predicted models of these facets, showing that the latter better predicts future student outcomes than the original variables.

To our knowledge, this work presents the first exploration of how student interviews, surveys, and interaction data may be used in concert to predict SRL and learning. This approach provides detailed insight into how we may best support students in an environment where external regulation may be harder to provide.

2. DATA
2.1 The Learning Environment

In this project, we used the learning environment Betty's Brain. This system implements a learning-by-teaching model [9], where students teach a virtual agent named "Betty" by creating a causal map of scientific processes (e.g., thermoregulation or climate change). Betty demonstrates her "learning" by taking quizzes, graded by a mentor agent, Mr. Davis. In this open-ended system, students choose how to navigate a variety of learning sources, how to build their maps, and how often to quiz Betty. They may also interact with Mr. Davis, who can support their learning and teaching endeavors [10].

Betty's Brain is a suitable environment for examining SRL behaviors for two reasons. Firstly, students choose when and how to perform each step of the learning process (both their own and Betty's) [20, 33]. Indeed, the pedagogical agents in Betty's Brain are designed to facilitate the development of SRL behaviors by providing a framework for the gradual internalization of effective learning strategies. Secondly, students' interactions with Betty's Brain are logged to an online database with detailed timing information, enabling the microanalysis of student actions [37] for the measurement of SRL behaviors and strategies.
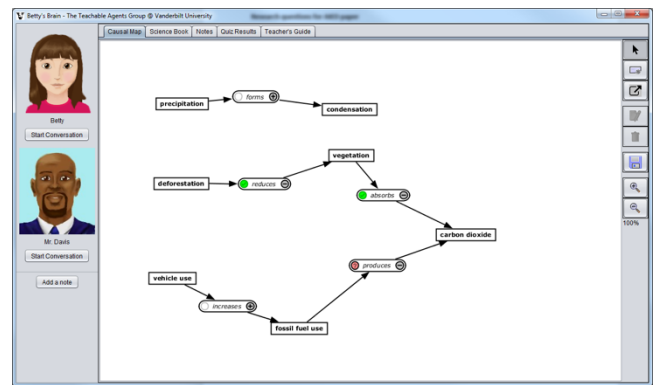

Figure 1. Screenshot of Betty's Brain showing a partial causal map constructed by a student.

2.2 Data Collection

This study examines data from 93 sixth graders who used Betty's Brain during their 2016–2017 science classes in an urban public school in Tennessee. The first data collection occurred over seven school days. On day 1, students completed a 30–45-minute paper-based pre-test that measured knowledge of scientific concepts and causal relationships. On day 2, students participated in a 30-minute training session about the learning goals and user interface. Afterwards (days 2–6), students used the Betty's Brain software for approximately 45–50 minutes each session, using concept maps to teach Betty about the causal relationships involved in the process of climate change. On day 7, students completed a post-test with the same questions as the pre-test. In addition to the data described, we also surveyed students on self-efficacy [31] and the task value [31].

A second data collection period occurred two months later, during which students were asked to model the causal relationships involved in thermoregulation. This was otherwise identical to the first session, but we consider only the learning data (pre – post test) from this second scenario (see section 4.2).

2.3 In-Situ Interviews

As students interacted with Betty's Brain, automatic detectors of educationally relevant affective states [19] and behaviors [26], already embedded in the software, identified key moments in the students' learning processes, either from specific affective patterns or theoretically aligned behavioral sequences. This detection was then used to prompt student interviews through Quick Red Fox (QRF), an app which integrates interview data with Betty's Brain events. Interviewers sought to take a helpful but non-authoritative role when speaking with students. Interviews were open-ended and occurred without a set script; however, students were often asked what their strategies were (if any) for getting through the system. As new information emerged in these open-ended interviews, questions were designed to elicit information about intrinsic interest (e.g., "What kinds of books do you like to read and why?"). Overall, however, students were encouraged to provide feedback about their experience with the software and talk about their choices as they used the software

2.4 Interview Transcription and Coding

A total of 358 interviews were conducted during this study and stored on a secure file management system. Interviews were manually transcribed by three members of the research team, preserving all metadata but scrubbing any identifying information.

The code development process followed [38]'s 7-stage recursive, iterative process: conceptualization, generation, refinement, codebook generation, revision and feedback, implementation, and continued revision. The conceptualization of codes involved a literature review to capture experiences relevant to affect and SRL. Using grounded theory [13], we worked with the lead interviewer (2nd author) to identify categories that were (1) theoretically valid and pertinent to the conditions in the COPES model and (2) likely to saliently emerge in the interviews.

We iteratively refined the coding scheme until the entire research team reached a shared understanding. Following the coding manual's production, external coders reached acceptable inter-rater reliability with the 3rd author before coding all of the transcripts. All codes had Cohen's kappa > .6, and the average Cohen's kappa across codes was .83. See Table 1 for details.

2.5 SMART Encoding

We operationalized SRL behavior within the log data using the COPES and SMART SRL frameworks [42]. In this work, we categorize all student actions recorded in the log files as "operations" within the COPES model (defined as "cognitive and behavioral actions applied to perform the task"). We then evaluate these operations using the SMART model, which subcategorizes actions by the information taken as input and product generated [39]. Specifically, the SMART model presents five primitive cognitive operation subcategories: Searching, Monitoring, Assembling, Rehearsing, and Translating [39]. Each category is briefly described below; for more details, see [39, 41, 42, 45]. Examples specific to Betty's Brain are shown in Table 2.

Searching is the operation where a learner focuses their attention on a knowledge base or resource to update their working memory.

Monitoring considers two types of information: (1) learner perceptions (current understanding, quiz answers, etc.), (2) standards for performance. In monitoring activities, the learner evaluates their perceptions compared to the standards.

Assembling involves building a network of internal links between acquired information to understand relationships (X precedes Y,

Table 1. Interview codes

Code	N	Description
Helpfulness	51	Utility of system resources for learning, and positive evaluations of the resources. κ=.643
Interestingness	11	Interestingness of system resources and continued desire to use the platform. κ=.726
Strategic Use	205	Indicates plan for interacting with the platform, or changes in strategy or interaction based on experiences. κ=.911
Positive Mr. Davis Attribution	8	Explicitly mentions interactions with Mr. Davis as positive experiences. κ=.838
Positive Science Attribution	26	Explicitly (positively) mentions science in relation to books, future careers, school subjects, and overall evaluations. κ=.837
Positive Persistence	105	Expression of a desire for challenge and that the current task is a challenge; there is active pursuit of a goal, and repeated attempts to complete a step/problem. κ=.911
Procedural Strategy	225	Step by step approach to the learning activity, active use of within-platform tools, reference to previous or upcoming step. κ=.862
Motivational Strategy	151	Explicit indication of expected outcome from behaviors/actions, explicitly mentions a pursuit for mastery, contains positive attribution/emotion for completion, and/or mentions desire to meet task demands. κ=.870
Self-Confidence	174	Positive description of own progress or ability, self-assessments of learning progress, willingness to encounter learning challenges/, recognition of helpful resources. κ=.877

Y causes Z, etc.). Assembling activities help students to connect individual items of knowledge in working memory.

Rehearsing operations repeatedly direct attention to information that the learner is currently working on. These actions reinforce the same information and prevent decay in working memory.

Translating operations reformat information into a new representation, providing the potential for alternate interpretations and understanding. Examples include converting a diagram to plain text or answering a question about a diagram.

To enable a trace analysis of student SRL patterns [37] we first assigned each of the possible student operations within Betty's Brain to one SMART category. We categorized operations that added new items to the concept map within Betty's Brain as *assembling*, and operations that edited existing items as *monitoring*. In ambiguous cases, such as between translation and monitoring tasks, we considered student agency. Specifically, actions initiated by the system were classified as *translating* even if they had an evaluative component. In our operationalization of the SMART model, we found that Betty's Brain logged no rehearsing actions; thus, this category was not analyzed.

3. MODEL TRAINING METHODS

We built supervised machine learning models to detect each facet of the SMART model. We leveraged a combination of activity, survey, and interview data (described further below).

Table 2. Example Betty's Brain actions by SMART facet.

SMART Facet	N	Example
Searching	8	Searching the virtual textbook (initiated by the student)
Monitoring	22	Reviewing and updating the label of a causal link (initiated by the student)
Assembling	2	Adding a causal link to the map (initiated by the student)
Rehearsing	0	-
Translating	3	Responding to a system-initiated multiple-choice questions (vs. those initiated by the student)

3.1 Features

We split features into three groups based on their origin. Each group is described in detail below. Due to differences in scale, we Z-scored each feature prior to model training.

(Other) Student Activity Features (N = 4). These features provide a high-level description of student actions: the raw number of student actions, the proportion of links made that were ineffective, time spent off-task/idle (as characterized in [36]), and number of successful quizzes. These features were designed to be more coarse-grained than the log data used to derive the SMART variables. None of the fine-grained features used to calculate the SMART encoding are included in this feature set.

Student Interview Codes (N = 9). These were derived from the transcribed student interviews (described in section 2.3). In cases where students had multiple interviews, codes were averaged to provide one feature per code per student.

Survey Features (N = 2). Survey features come from the two survey measures described in section 2.2: self-efficacy and task value. While each measure consisted of multiple survey questions, both were summarized down to one variable, respectively.

3.2 Dependent Variables

We initially considered four dependent variables, the proportion of the time a student spent on each of the SMART variables discussed in section 2.5. We considered time spent rather than raw action counts for a more standardized comparison and to avoid misinterpretation. For example, there are more monitoring actions than searching actions; however, it is common for students to spend considerably more time searching than monitoring. Due to time spent idle (at least 30 seconds of inactivity [36]), the sum of these four variables for any given student may not be 1. The most common category was searching (M=0.65, SD=0.07), followed by monitoring (M=0.16, SD=0.06), translating (M=0.10, SD=0.02), and assembling (M=0.09, SD=0.04).

We also considered a second set of dependent variables related to student learning. We derived two variables, one for the current scenario from which the rest of the data was collected, and one for the future scenario. In both cases, learning was characterized by post test – pre test. We consider both scenarios to examine how well our approach generalizes to future interactions and understand how immediate context may influence prediction.

3.3 Regression Models

We used scikit-learn [28] to implement Bayesian ridge regression, linear regression, Huber regression, and random forest regression, and also implemented XGBoost with a separate library [15]. Hyperparameters were tuned on the training set using scikit-learn's cross-validated grid search [28] where appropriate.

All models were trained using 4-fold student-level cross-validation and repeated for ten iterations, each with a new random seed. For evaluation, predictions were pooled across folds, and averaged across iterations. These models then underwent a decision tree based secondary analysis, discussed below.

4. RESULTS

We compare model accuracy by computing the correlation between the model predictions and the ground truth values derived from student logs. We measured the Spearman *rho* correlation coefficient in the test folds to evaluate models. In the majority of cases, random forest regressors yielded the best results. As such, results from these models are reported below.

4.1 Predicting SMART Operations

We first consider results predicting the proportion of time a student spent on each of the four SMART operations. For each operation (i.e., searching, monitoring, assembling, and translating), we developed models drawn from various combinations of our feature types (actions, surveys, and interview codes). Thus, we were able to test the modeling potential of seven different combinations of features for each SMART operation (see Table 3). To provide a point of comparison, we generated a chance baseline for each variable by shuffling the ground truth values. This allowed us to estimate a random baseline that still preserved the original distribution.

Table 3. Spearman correlations predicting ground truth labels of self-regulated learning operations

Features	Searching	Monitoring	Assembling	Translating
Chance Baseline	0.01	0	0.01	0
Individual Feature Sets				
Student Surveys (Surveys)	0.28	0.29	0.28	0.08
Student Interviews (Int)	0.31	0.37	0.35	0.09
Student Actions (Act)	0.27	0.47	0.59	0.11
Combined Feature Sets				
Int + Surveys	0.35	0.42	0.62	0.13
Act + Surveys	0.29	0.47	0.63	0.12
Act + Int	0.34	0.51	0.64	0.1
Act + Int + Surveys	0.39	0.55	0.66	0.19

We note that all models outperformed baseline, and that models consistently performed worst at predicting Translating. This may be due to the low variance between students as noted above. We note that the best model performance was achieved by combining the three feature sets (Actions + Interviews + Surveys). This suggests that even though these operations are derived from student log data, additional context from interviews and surveys can improve SRL predictions.

4.1.1 Feature Interaction Analysis

Our most successful models were tree-based, meaning that they may contain nonlinear relationships that would be unsuitable for linear feature analysis. Therefore, we trained one decision tree regressor per outcome and examined each tree's top two levels to observe the most important interactions, each of which was classified as "High" or "Low."

As Table 4 shows, Self-Confidence and Self-Efficacy frequently occur in these interactions, implying students' self-regulation

Table 4. Top 8 interactions for predicting SMART facets

Feature 1				Feature 2		Predicted Value	
Low	Self-Confidence	+	Low	Successful Quizzes	=	High	Searching
High	Self-Confidence	+	Low	Off Task Time	=	Low	Searching
Low	Off Task Time	+	Low	Self-Efficacy	=	High	Monitoring
High	Off Task Time	+	High	Self-Efficacy	=	High	Monitoring
Low	Action Count	+	Low	Self-Efficacy	=	High	Assembling
High	Action Count	+	Low	Ineffective Links	=	Low	Assembling
Low	Procedural Strategy	+	Low	Self Confidence	=	Low	Translating
High	Procedural Strategy	+	Low	Motivational Strategy	=	High	Translating

hinged on their perception of themselves. For example, students with high Self-Confidence who spent less time off task were still likely to have lower searching values, ostensibly because they may not feel the need to consult external resources.

4.2 Predicting Student Learning

Next, we explored how the four SMART facets predicted student learning (operationalized as post-test – pre-test) in both the *current* scenario (from which all the data used in the models was collected) and then the *future* scenario (collected in a second round of data collection with the same students; see section 2). In this future scenario, the content was different (climate change vs. thermoregulation), but the software remained the same.

We consider three feature sets: 1) the three feature sets used in section 4.1 combined; 2) the ground truth values for the SMART encodings (dependent variables in section 4.1); 3) predicted values for each of the SMART operations generated using the best models from section 4.1.

For both learning outcomes, we tested both the Ground Truth values collected from the first scenario (i.e., the actual searching or monitoring behaviors from that scenario) and Predicted SMART values (as predicted by the Act + Int + Survey models from the current scenario). This allowed us to examine how data collected in the current scenario generalizes to a future learning session.

As Table 5 shows, each learning model outperformed chance, demonstrating both predictive validity and generalizability. These results also present two findings of note. Firstly, learning models constructed from Predicted SMART values outperformed those constructed from the Ground Truth SMART values for both scenarios. It is possible that our models in fact, smooth over some of the noise that is present in the ground truth, thus presenting a more robust measure than the raw encodings [6].

Second, we note that for the future scenario, the predicted SMART values outperform model constructed directly from the Act + Int + Survey variables, despite this being the values from which the SMART predictions are made. The SMART values may provide a latent encoding of this data, which is more generalizable than the raw values to future occurrences, however further study would be required to confirm this hypothesis.

Table 5. Spearman rho for models predicting learning gains. All features are derived from the current scenario

Features	Current Scenario	Future Scenario
Chance Baseline	.01	.01
Act + Int + Survey	.45	.37
Ground Truth SMART	.21	.29
Predicted SMART	.32	.43

4.2.1 Feature Interaction

Using the same feature analysis methods described in section 4.1.1 we again examined the interactions involved when predicting learning gains. These results are shown in Table 6.

We note the need for the balance between SMART operations. For example, *high* monitoring and low translating resulted in lower learning on the current scenario, but so did high searching with *low* monitoring, suggesting it would be insufficient to simply increase monitoring activities; we must encourage more effective combinations of operations. Similarly, these results imply the need for a careful structure approach to assembling.

The results shown for the future scenario focus on more transferrable features than results for the current scenario. This makes sense given that we are no longer considering the immediate context. We found that students who had low off task time and high persistence in the first scenario were more likely to perform well in the second. Students with lower monitoring but high translating were likely to have lower learning, indicating it is not enough to simply test your knowledge, it is also important to review feedback and compare work to standards.

5. DISCUSSION

Adaptive learning technology that responds to students' learning patterns can improve both immediate and long-term goals by supporting the internalization of appropriate self-regulated learning behaviors. In this paper, we infer SRL using a combination of data mining and interviews/surveys.

5.1 Main Findings

Automated detection of SRL behaviors poses several challenges, as many of the processes it entails are highly internal [42]. In this work, we demonstrate that a combination of activity data, data from surveys, and student interviews provides a more robust prediction of SRL than any individual data stream. We find that predicted SRL behaviors (from students' first system interactions) predict future performance. In fact, models based on our inferred SRL measures outperform models constructed from the original features used to train them (action, interview, and survey data) and the SMART ground truth values. This finding is important for environments where detailed trace analysis may not be possible, but coarser-grained activity can be distilled.

Further, we show that a balanced combination of SRL behaviors is required for successful learning. For example, students with low learning are likely not spending enough time monitoring, but simply requiring them to check their work more often may not create improvement if they have not yet fully assembled the knowledge necessary to effectively examine their previous efforts. Future work should design scaffolds to create this balance.

Table 6. Top interactions for predicting learning. * indicates a predicted value

Scenario	Feature 1			Feature 2			Predicted Value	
Current Scenario	High	Monitoring*	+	Low	Translating*	=	Low	Learning
	High	Searching	+	Low	Monitoring	=	Low	Learning
	High	Successful Quizzes	+	High	Self-Efficacy	=	High	Learning
	Low	Successful Quizzes	+	High	Ineffective Links	=	Low	Learning
Future Scenario	High	Monitoring*	+	Low	Assembling*	=	High	Learning
	Low	Monitoring*	+	Low	Assembling*	=	Low	Learning
	Low	Off Task Time	+	High	Positive Persistence	=	High	Learning
	Low	Monitoring	+	High	Translating	=	Low	Learning

These results demonstrate the importance of considering log data in the context of other measures when understanding student SRL. This, in turn, underscores the need for more automated measures of complex noncognitive measures such as self-efficacy and persistence. Our work shows that these codes collected from interview data boost SRL detection. In order to scale SRL detection, we must first consider how we might automate the detection of some of the constructs discussed here (see future work below). These results offer the potential for designing pre-emptive interventions, providing a more informed, asset-based intervention as opposed to responding to a negative event.

5.2 Applications
The key application of this work is to develop adaptive online learning environments that respond to student SRL. As SRL detection continues to improve, systems like Betty's Brain might choose from wide range of intervention strategies that have already been shown to improve SRL (e.g., discussion in section 1). For example, once students who are not employing optimal strategies have been identified, additional scaffolding tasks might be used to encourage new behaviors. Similarly, the software could deliver interventions to increase motivation or interest.

It is important to note that the proposed intervention strategies rely on SRL detection, which is likely always to be imperfect. Self-regulation is highly internal [32], and as such, it is unlikely that we will ever be able to infer SRL perfectly. Any interventions should be designed to be "fail-soft" in that there are no damaging effects to student learning or future SRL if delivered incorrectly.

In situations where computer-based learning is being used to augment classroom instruction, a further application of this work would be in providing feedback to teachers. Such feedback could help them dynamically adapt their instruction, as outlined in [18] for example, providing real-time feedback or an early warning system, etc.

5.3 Limitations and Future Work
This work has limitations that should be addressed going forwards. Firstly, the SMART features only characterize student operations, and they do not give a complete SRL picture. Future work should look to combine the SMART framework with the broader COPES model [43]. The interview and survey measures used in this work may also capture aspects of the cognitive and task conditions referred to in the COPES model, but additional study would be required to confirm this hypothesis.

A further limitation is the slightly cyclic nature of using student activity features derived from log data, to predict SRL, also derived from log data. While we made every effort to ensure that our models were not confounded in some way, future work should

consider an external measure of SRL for additional validation [44].

Finally, interview data is time-consuming to collect, limiting scalability. In the future, we will employ alternate measures for some of the interview codes measured in this work, such as student surveys. It is possible that voice recognition and natural language processing could be used in the future to support this type of data collection.

5.4 Conclusions
This paper investigates predicting student SRL behavior in a computer-based learning environment from a complex dataset of coarse-grained activity data, in-situ student interviews, and student surveys. Our analyses indicated that SRL was best predicted from a combination of the three feature sets. We found our predicted SRL operations were better at predicting future learning than their ground truth equivalents, suggesting the potential for a smoother latent encoding and better supporting students in future endeavors. We envision this paper contributing to future technologies that will track and respond to student SRL behaviors and create more positive learning experiences.

6. ACKNOWLEDGMENTS

7. REFERENCES
[1] Azevedo, R., Johnson, A., Chauncey, A. and Burkett, C. 2010. Self-Regulated Learning with MetaTutor: Advancing the Science of Learning with MetaCognitive Tools BT - New Science of Learning: Cognition, Computers and Collaboration in Education. (2010), 225–247.https://doi.org/10.1007/978-1-4419-5716-0_11.

[2] Azevedo, R., Johnson, A., Chauncey, A. and Graesser, A. 2015. Use of Hypermedia to Assess and Convey Self-Regulated Learning. *Handbook of Self-Regulation of Learning and Performance*. 10622 (2015).https://doi.org/10.4324/9780203839010.ch7.

[3] Azevedo, R., Moos, D.C., Johnson, A.M. and Chauncey, A.D. 2010. Measuring Cognitive and Metacognitive Regulatory Processes during Hypermedia Learning: Issues and Challenges. *Educational Psychologist*. (2010).https://doi.org/10.1080/00461520.2010.515934.

[4] Azevedo, R. and Strain, A.C. 2011. Integrating Cognitive, Metacognitive, and Affective Regulatory Processes with MetaTutor. *New Perspectives on Affect and Learning Technologies*. (2011).https://doi.org/10.1007/978-1-4419-9625-1_11.

[5] Azevedo, R., Witherspoon, A., Chauncey, A., Burkett, C.

and Fike, A. 2009. MetaTutor: A MetaCognitive Tool for Enhancing Self-Regulated Learning. *AAAI Fall Symposium - Technical Report*. FS-09-02, (2009), 14–19.

[6] Baker, R.S.J. d., Corbett, A.T. and Aleven, V. 2008. Improving Contextual Models of Guessing and Slipping with a Truncated Training Set. (2008).

[7] Bannert, M. and Reimann, P. 2012. Supporting Self-Regulated Hypermedia Learning through Prompts. *Instructional Science*. 40, 1 (2012), 193–211.

[8] Biswas, G., Baker, R.S. and Paquette, L. 2017. Data Mining Methods for Assessing Self-Regulated Learning. *Handbook of Self-Regulation of Learning and Performance*. (2017), 388–403.https://doi.org/10.4324/9781315697048-25.

[9] Biswas, G., Leelawong, K., Schwartz, D., Vye, N., Davis, J., Belynne, K., Viswanath, K., Brandsford, J. and Katzlberger, T. 2005. Learning by Teaching: A New Agent Paradigm for Educational Software. *Applied Artificial Intelligence*. (2005).https://doi.org/10.1080/08839510590910200.

[10] Biswas, G., Segedy, J.R. and Bunchongchit, K. 2016. From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *International Journal of Artificial Intelligence in Education*. (2016).https://doi.org/10.1007/s40593-015-0057-9.

[11] Boekaerts, M. and Pekrun, R. 2016. Emotions and Emotion Regulation in Academic Settings. *Handbook of Educational Psychology*. (2016), 76–90.https://doi.org/10.1017/S0954579400006301.

[12] Broadbent, J. and Poon, W.L. 2015. Self-Regulated Learning Strategies & Academic Achievement in Online Higher Education Learning Environments: A Systematic Review. *Internet and Higher Education*. 27, (2015), 1–13.https://doi.org/10.1016/j.iheduc.2015.04.007.

[13] Charmaz, K. 1983. The Grounded Theory Method: An Explication and Interpretation. *Contemporary Field Research*. (1983), 109–126.

[14] Chen, C.-M. and Huang, S.-H. 2014. Web-Based Reading Annotation System with an Attention-Based Self-Regulated Learning Mechanism for Promoting Reading Performance. *British Journal of Educational Technology*. 45, 5 (2014), 959–980.

[15] Chen, T. and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2016), 785–794.https://doi.org/10.1145/2939672.2939785.

[16] Cleary, T.J. and Zimmerman, B.J. 2012. A Cyclical Self-Regulatory Account of Student Engagement: Theoretical Foundations and Applications. *Handbook of Research on Student Engagement*. (2012).https://doi.org/10.1007/978-1-4614-2018-7_11.

[17] Dignath, C., Buettner, G. and Langfeldt, H.-P. 2008. How Can Primary School Students Learn Self-Regulated Learning Strategies Most Effectively?: A Meta-Analysis on Self-Regulation Training Programmes. *Educational Research Review*. 3, 2 (2008), 101–129.https://doi.org/https://doi.org/10.1016/j.edurev.2008.02.003.

[18] Holstein, K., McLaren, B.M. and Aleven, V. 2017. Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (2017), 257–266.

[19] Jiang, Y., Bosch, N., Baker, R.S., Paquette, L., Ocumpaugh, J., Andres, J.M.A.L., Moore, A.L. and Biswas, G. 2018. Expert feature-engineering vs. Deep neural networks: Which is better for sensor-free affect detection? *Artificial Intelligence in Education* (2018), 198–211.https://doi.org/10.1007/978-3-319-93843-1_15.

[20] Kinnebrew, J.S., Biswas, G., Sulcer, B. and Taylor, R.S. 2013. Investigating Self-Regulated Learning in Teachable Agent Environments. *International Handbook of Metacognition and Learning Technologies*. (2013), 451–470.https://doi.org/10.1007/978-1-4419-5546-3_29.

[21] Kinnebrew, J.S., Segedy, J.R. and Biswas, G. 2014. Analyzing the Temporal Evolution of Students' Behaviors in Open-Ended Learning Environments. *Metacognition and Learning*. (2014).https://doi.org/10.1007/s11409-014-9112-4.

[22] Kizilcec, R.F., Pérez-Sanagustin, M. and Maldonado, J.J. 2017. Self-Regulated Learning Strategies Predict Learner Behavior and Goal Attainment in Massive Open Online Courses. *Computers & Education*. 104, (2017), 18–33.

[23] Klug, J., Ogrin, S. and Keller, S. 2011. A Plea for Self-Regulated Learning as a Process : Modelling , Measuring and Intervening. *Psychological Test and Assessment Modeling*. 53, 1 (2011), 51.

[24] de la Fuente, J., Sander, P., Kauffman, D.F. and Yilmaz Soylu, M. 2020. Differential Effects of Self- vs. External-Regulation on Learning Approaches, Academic Achievement, and Satisfaction in Undergraduate Students. *Frontiers in Psychology*. 11, (2020), 2678.https://doi.org/10.3389/fpsyg.2020.543884.

[25] Lee, H.W., Lim, K.Y. and Grabowski, B.L. 2010. Improving Self-Regulation, Learning Strategy Use, and Achievement with Metacognitive Feedback. *Educational Technology Research and Development*. 58, 6 (2010), 629–648.

[26] Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R.S. and Paquette, L. 2018. Modeling learners' cognitive and affective states to scaffold srl in open-ended learning environments. *UMAP 2018 - Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (2018), 131–138.https://doi.org/10.1145/3209219.3209241.

[27] Pandey, A., Hale, D., Das, S., Goddings, A.-L., Blakemore, S.-J. and Viner, R.M. 2018. Effectiveness of Universal Self-Regulation–Based Interventions in Children and Adolescents: A Systematic Review and Meta-Analysis. *JAMA Pediatrics*. 172, 6 (Jun. 2018), 566–575.https://doi.org/10.1001/jamapediatrics.2018.0232.

[28] Pedregosa, F. et al. 2011. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*. 12, (2011), 2825–2830.https://doi.org/10.1007/s13398-014-0173-7.2.

[29] Pérez-Álvarez, R., Maldonado-Mahauad, J. and Pérez-Sanagust\'\in, M. 2018. Tools to support self-regulated learning in online environments: literature review. *European conference on technology enhanced learning* (2018), 16–30.

[30] Pintrich, P.R. 2000. The Role of Goal Orientation in Self-Regulated Learning. *Handbook of Self-Regulation*. (2000), 451–502.https://doi.org/10.1016/b978-012109890-2/50043-

3.

[31] Pintrich, P.R.R., Smith, D., Garcia, T. and McKeachie, W. 1991. A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ). *Ann Arbor. Michigan*. (1991).https://doi.org/ED338122.

[32] Puustinen, M. and Pulkkinen, L. 2001. Models of Self-regulated Learning: A review. *Scandinavian Journal of Educational Research*.

[33] Roscoe, R.D., Segedy, J.R., Sulcer, B., Jeong, H. and Biswas, G. 2013. Shallow Strategy Development in a Teachable Agent Environment Designed to Support Self-Regulated Learning. *Computers and Education*. (2013).https://doi.org/10.1016/j.compedu.2012.11.008.

[34] Schraw, G. 2010. Measuring Self-Regulation in Computer-Based Learning Environments. *Educational Psychologist*. (2010).https://doi.org/10.1080/00461520.2010.515936.

[35] Schwartz, D.L., Chase, C., Chin, D.B., Oppezzo, M., Kwong, H., Okita, S., Biswas, G., Roscoe, R., Jeong, H. and Wagster, J.D. 2009. Interactive Metacognition: Monitoring and Regulating a Teachable Agent. *Handbook of Metacognition in Education*. (2009), 340–358.

[36] Segedy, J.R., Kinnebrew, J.S. and Biswas, G. 2015. Using Coherence Analysis to Characterize Self-Regulated Learning Behaviours in Open-Ended Learning Environments. *Journal of Learning Analytics*. (2015).https://doi.org/10.18608/jla.2015.21.3.

[37] Siadaty, M., Gasevic, D. and Hatala, M. 2016. Trace-Based Microanalytic Measurement of Self-Regulated Learning Processes. *Journal of Learning Analytics*. 3, 1 (2016), 183–214.https://doi.org/https://doi.org/10.18608/jla.2016.31.11.

[38] Weston, C., Gandell, T., Beauchamp, J., McAlpine, L., Wiseman, C. and Beauchamp, C. 2001. Analyzing Interview Data: The Development and Evolution of a Coding System. *Qualitative Sociology*. 24, 3 (2001), 381–400.https://doi.org/10.1023/A:1010690908200.

[39] Winne, P.H. 2011. A Cognitive and Metacognitive Analysis of Self-Regulated Learning. *Handbook of Self-Regulation of Learning and Performance*. (2011), 15–32.https://doi.org/10.4324/9780203839010.ch2.

[40] Winne, P.H. 2010. Improving Measurements of Self-Regulated Learning. *Educational Psychologist*. 45, 4 (2010), 267–276.https://doi.org/10.1080/00461520.2010.517150.

[41] Winne, P.H. 2014. Issues in Researching Self-Regulated Learning as Patterns of Events. *Metacognition and Learning*. 9, 2 (2014), 229–237.https://doi.org/10.1007/s11409-014-9113-3.

[42] Winne, P.H. 2017. Learning Analytics for Self-Regulated Learning. *Handbook of Learning Analytics*. (2017), 241–249.https://doi.org/10.18608/hla17.021.

[43] Winne, P.H. and Hadwin, A.F. 1998. Studying as Self-Regulated Learning. *Metacognition in Educational Theory and Practice*. (1998), 277–304.

[44] Winne, P.H. and Perry, N.E. 2000. Measuring Self-Regulated Learning. *Handbook of Self-Regulation*. (2000).https://doi.org/10.1016/b978-012109890-2/50045-7.

[45] Winne, P.H., Teng, K., Chang, D., Lin, M.P.C., Marzouk, Z., Nesbit, J.C., Patzak, A., Rakovic, M., Samadi, D. and Vytasek, J. 2019. NStudy: Software for Learning Analytics about Learning Processes and Self-Regulated Learning. *Journal of Learning Analytics*. (2019).https://doi.org/10.18608/jla.2019.62.7.

[46] Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G.J. and Paas, F. 2019. Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review. *International Journal of Human-Computer Interaction*. 35, 4–5 (2019), 356–373.https://doi.org/10.1080/10447318.2018.1543084.

[47] Zimmerman, B.J. 1989. Models of Self-Regulated Learning and Academic Achievement. *Self-Regulated Learning and Academic Achievement*. (1989), 1–25.https://doi.org/10.1007/978-1-4612-3618-4_1.