LINK-REPORT: Outcome Analysis of Informal Learning at Scale

Xiang Fu
Dept. of Computer Science
Hofstra University
Hempstead, NY 11549
Xiang.Fu@hofstra.edu

Tyler Befferman
Dept. of Computer Science
Hofstra University
Hempstead, NY 11549
tbeff1@pride.hofsra.edu

M.D. Burghardt
Dept. of Engineering
Hofstra University
Hempstead, NY 11549
M.D.Burghardt@hofstra.edu

L@S 2016, April 25–26, 2016, Edinburgh, UK.
ACM 978-1-4503-3726-7/16/04.
http://dx.doi.org/10.1145/2876034.2893407

Abstract
We present LINK-REPORT, a distributed learning outcome analysis module that is integrated with the WISEngineering platform for supporting informal learning in engineering. LINK-REPORT provides a coherent workflow of outcome analysis: starting from development of learning outcome goals, to learner behavior collection, to automated grading of open ended short answer questions, and to report generation and aggregation. It generates learning data for research opportunities in modeling of learner traits.

Author Keywords
outcome assessment; automated grading

Introduction
As children spend over 80% of their time out-of-school [2], informal learning has been a crucial tool for engaging pre-college students and broadening their participation in science, engineering, and mathematics (STEM). The outcome assessment, however, is challenging because typical evaluation approaches in classroom teaching are not applicable. Therefore, collecting *more* and *better* data on learning outcomes is essential [1].

The evolving online education platforms such as Blackboard and edX have shed a light on this problem.

Figure 1: WISEngineering Tablet Portal. It delivers step by step instructions of an engineering project (e.g., creating a prosthetic leg for disabled people). Tools such as a design journal and design wall allow reflection and discussion. Distributed clusters at backend serve large volume of video and picture uploads from students.

For example, Yang et al. studied the learning data of Scratch [6], an informal learning platform for computer programming. In this study, learning processes are modeled as the cumulative trajectory of the programming blocks mastered by learners [10].

It is interesting to see how the data mining in [10] can be generalized to the analysis of a variety of learning outcome goals which will need to draw upon many different types of data. For example, in [1] a collection of six strands of learning goals are proposed for informal learning: O1: interests, O2: knowledge, O3: scientific reasoning, O4: reflection, O5: practice, and O6: identifying with scientific enterprise. It is a great challenge to collect and analyze the evidence data.

Extending the Web-based Inquiry Science Environment (WISE) from UC Berkeley [8], the WISEngineering system [4, 5] provides a tablet portal (shown in Figure 1) that offers hands-on engineering project activities and content for informal science settings. In our project, learners use a 7 inch tablet to collect data, plot and analyze information, reflect upon their designs, and share products with peers.

LINK-REPORT draws upon the existing WISEngineering system. It provides a coherent workflow to collect, tag, analyze, synthesize, and aggregate reports, leveraging auto-grading and machine learning algorithms. In this paper, We first present the learning dataset that is available in the WISEngineering system. Then we briefly introduce the workflow of LINK-REPORT and its software architecture. Finally, we present how learning indicators are modeled and assessed for the set of six learning goals in [1].

Raw Dataset

WISEngineering inherits the user management, curriculum design, and visual rendering module from Berkeley WISE [8]. All learner related data are stored in several MySQL databases. The data include the learner profile, session information (such as login/out time-stamp, duration of stay on each page), student responses (to multiple choice, match sequence questions, and open-response short answer questions).

LINK-REPORT also produces raw data. It injects an additional client-end script into the curriculum pages of WISE, and collects learner behavioral data such as keyboard and mouse events, which could be potentially used for user traits analysis in future.

Data Synthesis

Raw data is further processed for data mining and report aggregation in the next stage. LINK-REPORT provides a learning goal editor for defining learning outcome goals. Then a collection of indicators (called "grading criteria") can be defined for each learning goal. A learning indicator can be defined as the student performance over a specific question. The system has the capability to evaluate open-ended questions using multiple criteria (e.g., whether a learner is familiar with terminology used in the project, and the writing proficiency of the learner). As an example, Figure 2 shows that a learning goal "understanding engineering specification" is evaluated using three indicators in a project called "Balance Challenge".

One benefit of LINK-REPORT is that it leverages the use of machine learning for processing open-ended short answer questions and datasets that usually need labor intensive human evaluation. As shown in [3], a stacked generalization algorithm [9] is used for combining multiple

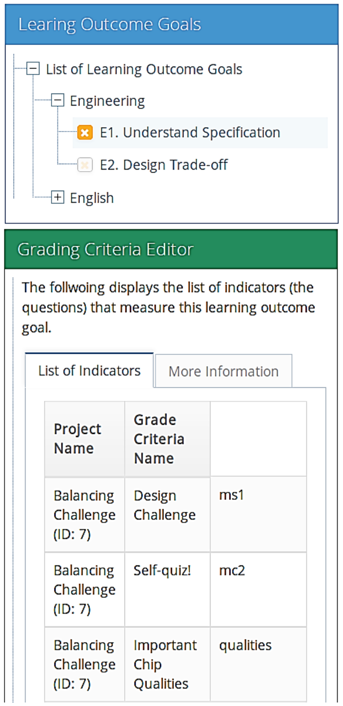
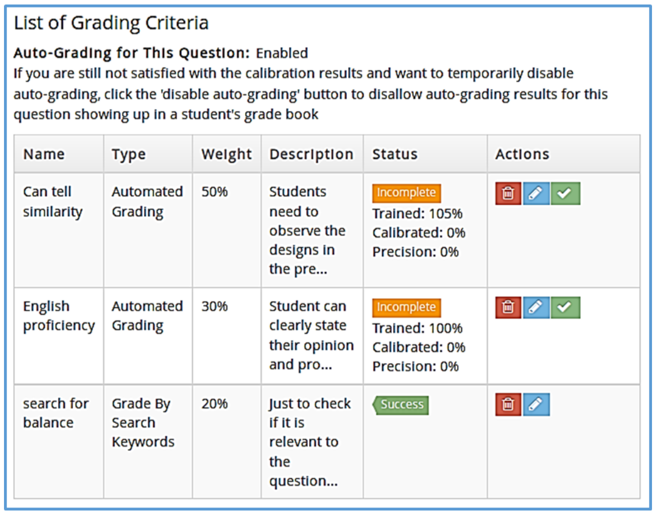


Figure 2: Indicator of Learning Goals. Three grading criteria (using multiple choice, match sequence, and auto-grading on short answer, respectively) are defined for a learning goal "understanding engineering specification".

machine learning/auto-grading algorithms (including the edX EASE engine [7]), for generating an optimized classifier as the auto-grader. We deployed these auto-grading classifiers in a distributed environment and can provide instant grading services for short answer questions [3]. Figure 3 shows an example of the summary page of a project and the statistics of its auto-graders.


List of Grading Criteria

Auto-Grading for This Question: Enabled

If you are still not satisfied with the calibration results and want to temporarily disable auto-grading, click the 'disable auto-grading' button to disallow auto-grading results for this question showing up in a student's grade book

Name	Type	Weight	Description	Status	Actions
Can tell similarity	Automated Grading	50%	Students need to observe the designs in the pre...	Incomplete Trained: 105% Calibrated: 0% Precision: 0%	
English proficiency	Automated Grading	30%	Student can clearly state their opinion and pro...	Incomplete Trained: 100% Calibrated: 0% Precision: 0%	
search for balance	Grade By Search Keywords	20%	Just to check if it is relevant to the question...	Success	

Figure 3: Auto-grading Training Statistics

Report Aggregation

WISEngineering [3] was designed to support nationally distributed informal science projects by learning communities. For example, the Wise Guys and Gals (WGG) project [4] leverages the national Boys and Girls Clubs for broadening the participation of middle school ages in engineering. These projects have multiple sites of implementation across the nation. One question that LINK-REPORT can address is how implementation and learning outcomes are similar and different across the various sites. Analysis can be broken down by site and compared.

To allow scalability WISEngineering adopts a loose confederation of servers. When a report is ordered, a template is sent to satellite servers, then these reports are aggregated.

On each satellite server, each learning outcome goal is assessed by aggregating the performance data on all the performance indicators associated with it (as shown in Figure 2). For example, the goal "understanding engineering specification" can be evaluated by a multiple choice question in one project, and another essay question which is graded by auto-grader.

```
$dtPost35 = array("type"=>"dataItem",
  "cols"=>array("Learning Outcome", "Club"),
  "presentation"=>array("approach"=>"histogram"),
  "function"=>"getOverallPerformance",
  "person_selector"=>array("type"=>"PERSONSEL_NONE"),
  "aggregate_op"=>array("type"=>"AGGREGATEOP_AVG"),
  "groupby_op"=>array("type"=>"GROUPBYOP_CLUBID"),
  "daterange_op"=> array("type"=>"GENBY_REPORT"),
  "histogram_op"=> array("frequency"=>"WEEKLY", "count"=>"10")
);
```

Figure 4: Code Snippet of Weekly Report Template

Data can be grouped by individual students, clubs and geographical districts. Figure 4 shows a snippet of a weekly report template. The report generation tool allows an educator to search data in a flexible way using different aggregation and grouping options.

Figure 5 shows a sample report generated by aggregating the reports from satellite servers. Using a simple weighted sum approach, the performance of all clubs on each learning outcome goal can be visually displayed and compared.

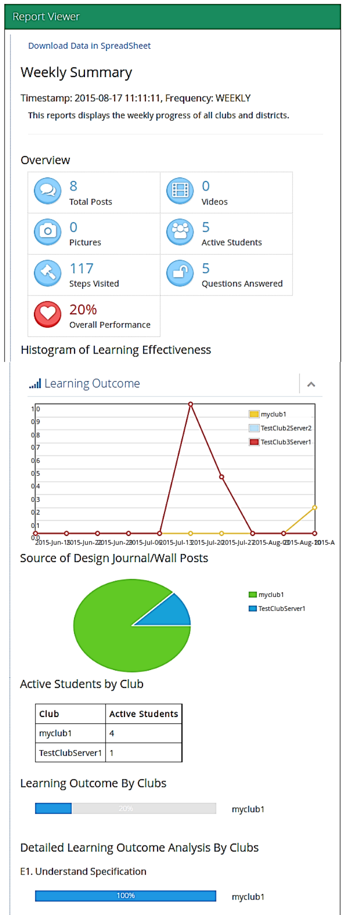


Figure 5: Sample Weekly Report. The clubs shown in the report are actually located on two separate physical servers. The bottom of the report shows the progress of each learning outcome goal at each club.

Work In Progress: Learning Model

LINK-REPORT is expandable. New learning indicators can be added into the system and be correlated with existing learning outcomes. We now briefly describe our plan to establish a collection of indicators for the six learning goals outlined in [1].

Most of the attitude outcomes (such as O1: interests in science and O6: identifying with scientific enterprise) can be assessed by online surveys built into the system. O2 (knowledge), O3 (scientific reasoning) and O4 (reflection) can be evaluated by standardized questions and short answer questions (graded by machine learning algorithms automatically). O5 (engagement in scientific practice) are related to the learner behavior data collected, such as the student tablet usage and activity records on design journals. We suspect that other usage data such as keystroke frequency and mouse events can also be used as indicators.

A learning indicator based on technology vocabulary can be established, similar to [10]. The "relevance" of terms can be automatically synthesized from semantic dictionary such as Wordnet, Wikipedia, and sample essays provided by human graders. Thus, by analyzing the bag of words, and their distance to the engineering discipline, auto-grader can assess the depth and effectiveness of a learner's use of scientific reasoning. Another direction is to assess the relevance and correlation between all indicators, so that the assessment model can be automatically adjusted.

Acknowledgments This material is based on the work supported by the National Science Foundation under grant DRL 1422436.

References

[1] Bell, P., Lewenstein, B., Shouse, A. W., and Feder, M. A. *Learning Science in Informal Environments: People, Places, and Pursuits*. National Academic Press, 2009.

[2] Cardella, M., Wolsky, M., Paulsen, C., and Jones, T. Informal pathways to engineering. In *ASEE Annual Conference* (2013).

[3] Fu, X., Befferman, T., Chiu, J., and Burghardt, M. Wisengineering: Achieving scalability and extensibility in massive online learning. In *Proc. Web Information System Engineering (WISE)* (2015), 323–337.

[4] Hofstra STEM Research Center. Research and Development: Advances in Wise Guys & Gals – Boys and Girls as WISEngineering STEM Learners. http://www.hofstra.edu/Academics/Colleges/SEAS/CTL/wise/research-development.html.

[5] Hofstra STEM Research Center. WISEngineering Web Portal. http://wgg1.hofstra.edu.

[6] MIT Media Lab. Scratch Online. http://scratch.mit.edu/.

[7] Paruchuri, V., Huang, D., Jarvis, J., Tauber, J., Aune, N., and Kern, J. EASE Auto-Grading Module. https://github.com/edx/ease.

[8] UC Berkeley. Berkeley WISE System. https://wise.berkeley.edu/.

[9] Wolpert, D. H. Stacked generalization. *Neural Networks 5* (1992), 241–259.

[10] Yang, S., Domeniconi, C., Revelle, M., Sweeny, M., Gelman, B. U., Beckley, C., and Johri, A. Uncovering trajectories of informal learning in large online communities of creators. In *Proceedings of 2nd ACM Conference on Learn at Scale* (2015), 131–140.