A Quality Data Set for Data Challenge: Featuring 160 Students' Learning Behaviors and Learning Strategies in a Programming Course

Owen H.T. LU[a*], Anna Y.Q. HUANG[b], Brendan FLANAGAn[c], Hiroaki OGATA[c], Stephen J.H. YANG[b]
[a]*International College of Innovation, National Chengchi University, Taiwan*
[b]*Computer Science & Information Engineering, National Central University, Taiwan*
[c]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*owen.lu.academic@gmail.com

Abstract:

Emerging science requires data collection to support the research and development of advanced methodologies. In the educational field, conceptual frameworks such as Learning Analytics (LA) or Intelligent Tutoring System (ITS) also require data. Prior studies demonstrated the efficiency of academic data, for example, risk student prediction and learning strategies unveiling. However, a publicly available data set was lacking for benchmarking these experiments. To contribute to educational science and technology research and development, we conducted a programming course series two years ago and collected 160 students' learning data. The data set includes two well-designed learning systems and measurements of two welldefined learning strategies: Self-regulated Learning (SRL) and Strategy Inventory for Language Learning (SILL). Then we summarized this data set as a Learning Behavior and Learning Strategies data set (LBLS-160) in this study; here, 160 indicates a total of 160 students. Compared to the prior studies, the LBLS data set is focused on students' book reading behaviors, code programming behaviors, and measurement results on students' learning strategies. Additionally, to demonstrate the usability and availability of the LBLS data set, we conducted a simple risk student prediction task, which is in line with the challenge of cross-course testing accuracy. Furthermore, to facilitate the development of educational science, this study summarized three data challenges for the LBLS data set.

Keywords: Educational data, learning analytics, AIED

1. Introduction

Artificial Intelligence (AI) can be implemented as a software agent to handle routine human tasks (Russell, 2010). AI has rapidly grown due to the rise of big data and machine learning technologies. Several AI applications constructed by data have already been implemented in the real world. For example, Devlin, Chang, Lee, and Toutanova (2018) designed a bi-directional model to let machines construct the capability to understand human language. The model supported four tasks: Next Sentence Prediction (NSP), Masked Language Model (MLM), Single Sentence Tagging (SST), and QuestionAnswering (QA). The performance evaluation results demonstrated that the work is the state-of-the-art machine reading comprehension model after training the parameters using 2,500M+800M words from Wikipedia articles and BooksCorpus. Many researchers involved in foundation work on data to support AI research and development. For example, Deng et al. (2009) hosted a project named ImageNet, which collected 14M images till May 11, 2021. Follow-up researchers benchmarked their AI model based on ImageNet, such as Krizhevsky, Sutskever, and Hinton (2012) evaluated the Convolutional Neural Network (CNN) performance on the data set. The above works indicated training a machine as an agent for supporting human's routine tasks, data playing a significant role in model research and development, and an opened data set is also essential for model performance benchmarking.

AI in education can be referred to the Intelligent Tutoring System (ITS), which goal is providing near real time personalized feedbacks to learner (Hwang, Xie, Wah, & Gašević, 2020). One kind of implementation is video recommendation system, which push a video with missing context to a student based on their learning pathway or concept proficiency, priori studies also demonstrated its efficiency on students' learning performance (L. Leite et al., 2022). Meanwhile, to implement ITS, machine learning techniques, algorithms, datasets are also required components (Khanal, Prasad, Alsadoon, & Maag, 2020). When it comes to data applications in education, learning analytics is another implementation. Learning analytics is defined as a kind of data-driven application in educational field (Cristobal Romero & Ventura, 2020). The goal is improving students' learning performance based on data analysis results (Clow, 2013). One of the popular implementations is predicting at-risk students. For example, Cristóbal Romero, López, Luna, and Ventura (2013) demonstrated the potential to identify risk population by using students' level of discussion participation in early semester. Choi, Lam, Li, and Wong (2018) demonstrated a 7% improvement on learning performance after intervening risk population that identified by prediction model that trained by students' clicker data. On the other hand, to offer correct intervention to risk population, researchers applied learning strategy approaches. For example, Jovanović, Gašević, Dawson, Pardo, and Mirriahi (2017) unveiled students' self-regular learning strategy by using sequence clustering on students' book reading behaviors.

The currently opened data sets has several characteristics. The earliest opened educational data set can be traced back to 2008 (Cortez & Silva, 2008), containing 30 variables, such as gender and school, from the Mathematics and Portuguese language course. The latest opened educational data set can be found in 2022 (Flanagan, Majumdar, & Ogata, 2022), which contains 120 students with 17 variables. Perhaps due to the advancement of technology, the data collected has become more diverse. From the beginning, most data set only collected demographics such as gender and age (Amrieh, Hamtini, & Aljarah, 2016; Cortez & Silva, 2008; MITx). In recent years, logs from the learning system have appeared, such as video viewing behavior on MOOCs (Kellogg & Edelmann, 2015), discussions on LMS (Amrieh et al., 2016), and book reading behavior on e-Book (Flanagan et al., 2022). Algebra (Stamper, NiculescuMizil, Ritter, Gordon, & Koedinger, 2010) and Mathematics (Cortez & Silva, 2008) were the most popular subjects for the data collection task. But on the other side, some data sets do not specify subjects or contain almost all subjects (MITx).

To sum up, we can understand the importance of open data sets to contemporary education technology development, but the currently available data sets focused on learning behaviors but overlooked learning strategies. Therefore, this study aims to define and publish a data set that measures learning behaviors and strategies and then attempts to facilitate learning analytics research and development through several scientific challenges.

1. We conduct a programming course for continuing to open a new educational data set for the development of AIED research.
2. In addition to reading behavior, this data set will disclose students' learning strategies and coding behavior, which will be more extensive than the previous data sets.
3. To confirm the data quality, we define an evaluation process and demonstrate the usability of the proposed data set. Meanwhile, we explained a few potential data challenges for future works.

2. Literature Reviews

2.1 How to evaluate the data quality?

Data would not achieve the research goal if we released it after collection directly. A suitable evaluation process can make subsequent research and development more effective. Therefore, our idea is to challenge a topic before opening, review its effectiveness and use it as a baseline for future challenges.

Among the functionality of education data, Cristobal Romero and Ventura (2020) defines (1) Educational Data Mining (EDM), which explores methods to recognize learning patterns or critical factors in education data, and (2) Learning Analytics, which uses educational data to optimize the learning environment or to improve learning performance. This study will focus on the early stage of LA because EDM requires more data sets to benchmark the methods. In practice, the LA presented by Cristóbal Romero et al. (2013) used students' participation in the discussion forum to predict students' learning performance, and they tested the Accuracy of various classification algorithms.

In another review, Conijn, Snijders, Kleingeld, and Matzat (2016) conducted a large-scale analysis on the risk student prediction task. They collected 17 courses from LMS and adopted the same methodologies mentioned above to understand the characteristics of the learning analytics research. The results demonstrated rich classification Accuracy of the risk prediction model on verification of the same course but poor on cross-course validation. In summary, to evaluate the data quality, this study will implement a learning analytics task of predicting student risk while cross-validating individual courses and courses by using a convention classification algorithm and indicator. The result is expected to be consistent with Conijn et al. (2016), and some problems encountered could be defined as future challenges.

2.2 What else data could be collected in classroom?

Learning strategy indicates to a student how to construct knowledge in the classroom; therefore, students' strategy measurement becomes another critical factor to be recorded. The self-regulated learning proposed by Zimmerman and Schunk (2001) aims to explore the process of students' acquisition of knowledge in the three stages of forethought, monitoring, and self-reflection. During the programming process, students learn to solve problems and develop programming skills during the phases of understanding the problem, planning and implementing solutions, and evaluating potential solutions. It can be seen that the stages of the programming process are similar to the self-adjustment behaviors of planning, goal setting, organization, self-monitoring, and self-evaluation in self-regulated learning. In view of this, Shin and Song (2022) have proposed that students' self-regulated learning ability need to guide to promote students' programming skills in the problem-solving process. Therefore, Shin and Song (2022) further defined the learning tasks corresponding to the three stages of selfregulated learning in the programming process. In addition, previous research literatures found that students' self-regulated learning ability is closely related to programming learning performance (Cigdem, 2015; Echeverry, Rosales-Castro, Restrepo-Calle, & González, 2018).

On the other hand, researchers tried to adopt learning strategies for natural languages in programming courses. They demonstrated high potential because of (1) the similar aspects of syntax, lexicon, and semantics (Ernst, 2017) and (2) the efficiency of students' learning performance. In the efficiency aspect, for instance, Sun and Frederick (2015) proposed a framework named: SLA-aBLe (Second language acquisition to facilitate a blended learning); they demonstrated the improvement of student's learning motivation and performance by constructing connections between vocabulary and programming syntax. Many researchers adopted Strategy Inventory for Language Learning (SILL)(Oxford & Burry-Stock, 1995) to measure students' natural language learning strategies. Essentially, it is because it provides an explicit intervention action if students' measurement results are below the average; for example, teachers should add more images in learning material to increase their memory retention rate. Therefore, this study will use SILL to measure students' capability of language learning strategy.

3. Methods

3.1 Participants

As shown in Figure 1, LBLS-160 data sets were collected from three programming language classes which Class A, B, and C have 63, 56, and 41 students, totaling 160 students. The teacher, learning materials, content, syllabus, homework, grading policies, and learning durations were all the same in this course. The only difference is that Class A and B are in the same semester, and Class C is in another semester. The 160 participants in the course were all from university, non-computer science-related departments, and they were all learning programming languages for the first time. LBLS-160 collects two kinds of data: learning behaviors and learning strategies, and it will be explained in detail at next section.

		Learning Behaviors		Learning Strategies		
	Participants	BookRoll	VisCode	SRL Motivation	SRL Strategy	SILL
Class A	63	✓	✓	✓	✓	✓
Class B	56	✓	✓	✓	✓	✓
Class C	41	✓	✓	✓	✓	

Total: 160 Non-CS Students

Figure 1. The number of students and features in LBLS-160.

3.2 Data Collection

LBLS-160 includes two parts. The first is individuals' learning behaviors collected from two online learning environments. The second is individuals' learning strategies which are measured using two different questionnaires.

Two learning environments include BookRoll(Ogata et al., 2015) and VisCode(Lu, Huang, Huang, Huang, & Yang, 2016) as shown in Figure 2; both software was designed for teachings and have a complete learning logs recording function for learning analytics research. BookRoll is an e-book software for teachers to upload and manage teaching materials; for learners to read through the Internet. BookRoll provides functions such as Maker and Memo and saves learners' reading records as a Log file. VisCode is an integrated development environment for teachers to upload sample codes; learners to develop scripting programming languages, execute and test. VisCode stores the programs and development records created and interacted with learners.

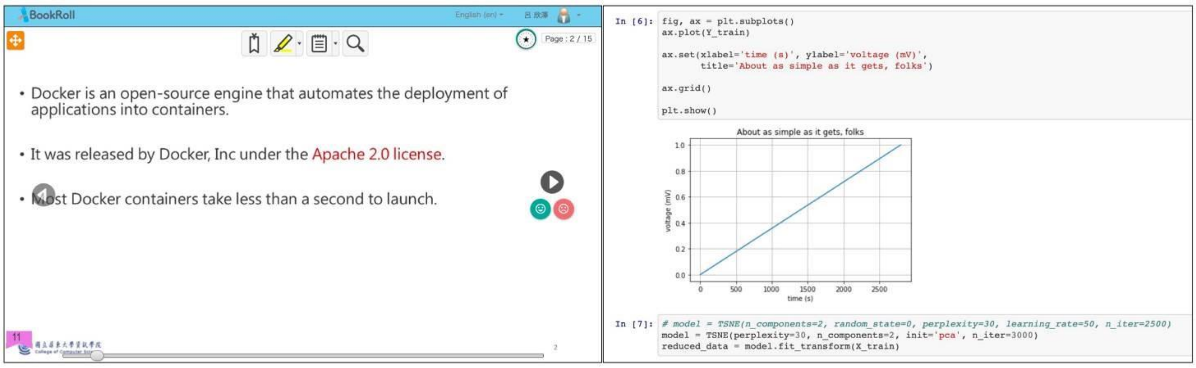

Figure 2. Learning environment, BookRoll (L) and VisCode (R).

Learning strategies include Self-regulated Learning (SRL)(Zimmerman & Schunk, 2001) and Strategy Inventory of Language Learning (SILL)(Oxford & Burry-Stock, 1995). According to the survey from prior studies in the literature review section, we can consider learning strategies an essential factor in supporting learners in improving their learning performance. In SRL, the Learning Motivation Strategies Questionnaire (MSLQ) with a 5-point Likert scale proposed by Pintrich (1991) has gradually been used to measure learners' self-regulation ability, it mainly includes two scales of learning motivation and learning strategy. The learning motivation scale has 31 questions, including six phases of intrinsic goal orientation, extrinsic goal orientation, task value, control belief, learning and performance self-efficacy, and test anxiety. The learning strategy has 50 questions, including nine phases of rehearsal, elaboration, organization, critical thinking, metacognitive self-regulation, time and study environment, effort regulation, peer learning, help seeking. In SILL, we measured six phases by

48 questions with 5 points Likert scale: the cognitive phase, compensation phase, social phase, affective phase, meta-cognitive phase, and memory phase.

3.3 Learning Activities

The course aims to teach the basic Python programming knowledge. There are eight basic Python programming concepts: (1) *input and output*, (2) *variables*, (3) *lists*, (4) *conditions*, (5) *for-loop*, (6) *while-loop*, (7) *dictionary*, and (8) *functions*. This course contained three stages: before-class, in-class, and after-class. In the stage of before-class, each student required to preview learning materials on BookRoll. There are three actions has to be done in stage of in-class: teacher will give an instruction based on learning materials, students have to take a program challenge assigned by teacher, and teacher will explain the solution of the challenge before the end of class. In the stage of after-class, each student has to practice three to five Python assignments on VisCode.

3.4 Data Preprocessing

We took a few preprocesses before releasing LBLS-160. The first is de-identification by removing direct recognition fields such as students' first and last names and encoding students' identification as a unique code: "userid" in each data set. Students will obtain the same code in different subsets but unique code in the same subsets to fulfill the subset merge subsets requirement. The second is to fill the ignored responses to be 0 in the questionnaire measurement process. The last one is to encode headers to make data more readable. For example, the first question in the SRL motivation questionnaire will be encoded into "srl_m_1". On the other hand, although it is important to use features that are critical to learning outcomes, but in this study, because the data has not been validated, we used data that can be collected in two learning environments for validation, as listed in Table 1 and Table 2:

Table 1. Features of book reading behaviors (BookRoll).

Features	**Description**
userid	Anonymized student userid, eg: b1dfc5c6ec04d46d1823c5fa972ad320
ADD BOOKMARK	Added a bookmark to current page.
ADD MARKER	Added a marker to current page.
ADD MEMO	Added a memo to current page.
ADD_HW_MEMO	Added a handwrite memo to current page.
BOOKMARK_JUMP	Jump to a specific page with a bookmark.
CHANGE MEMO	Modify the content of an existing memo on current page.
CLEAR_HW_MEMO	Clear the content of an existing handwrite memo on current page.
CLOSE	Closed the book.
CLOSE_RECOMMENDATION	Deleted an exist bookmark in the e-book.
DELETE BOOKMARK	Deleted a bookmark on current page.
DELETE MARKER	Deleted a marker on current page.
DELETE_MEMO	Deleted a memo on current page.
GETIT	Press the smiley face icon to indicate the understanding on current page.
MEMO_JUMP	Select a note to jump to the specific page.
NEXT	Went to the next page.
NOTGETIT	Press the crying face icon to indicate the misunderstanding on current page.
OPEN	Opened the book.
PAGE_JUMP	Jumped to a particular page.
PREV	Went to the previous page.
SEARCH	Searched for something within the e-book.
SEARCH_JUMP	Jumped to a page from the search results.
UNDO_HW_MEMO	Undo the last action of handwriting.

Table 2. Features of programming coding behaviors (VisCode).

Features	Description
id	Anonymized student userid, eg: b1dfc5c6ec04d46d1823c5fa972ad320
code_length	Nunber of lines of code (LOC) coded in this semester.
code_copy	Number of times a student copy codes.
code_execution	Number of times a student execute codes.
code_paste	Number of times a student paste codes.
code_speed	Average input digits per minutes.
notebook_open	Number of times a student open coding environment.
Features	**Description**
tree_open	Number of times a student open a folder looking for a code.
AttributeError	Raised when attribute reference or assignment fails.
ConversionError	Failed to convert value(s) to axis units.
FileExistsError	Raised when trying to create a file or directory which already exists.
FileNotFoundError	Raised when a file or directory is requested but doesn't exist.
IndentationError	Base class for syntax errors related to incorrect indentation.
IndexError	Raised when a sequence subscript is out of range.
JSONDecodeError	Raised if the given JSON document is not valid.
KeyError	Raised when a mapping (dictionary) key is not found in the set of existing keys.
KeyboardInterrupt	Raised when the user hits the interrupt key (normally Control-C or Delete).
LookupError	The base class for the exceptions that are raised when a key or index used on a mapping or sequence is invalid: IndexError, KeyError.
ModuleNotFoundError	A subclass of ImportError which is raised by import when a module could not be located.
NameError	Raised when a local or global name is not found.
OperationalError	Exception raised for errors that are related to the database's operation, and not necessarily under the control of the programmer.
SyntaxError	Raised when the parser encounters a syntax error.
TabError	Raised when indentation contains an inconsistent use of tabs and spaces.
TypeError	Raised when an operation or function is applied to an object of inappropriate type.
UnboundLocalError	Raised when a reference is made to a local variable in a function or method, but no value has been bound to that variable.
UnicodeDecodeError	Raised when a Unicode-related error occurs during decoding.
ValueError	Raised when an operation or function receives an argument that has the right type but an inappropriate value, and the situation is not described by a more precise exception such as IndexError.
ZeroDivisionError	Raised when the second argument of a division or modulo operation is zero.

3.5 Evaluation

To evaluate LBLS-160 quality and establish the baseline performance for the follow-up challenges, we conducted a risk prediction experiment to assess if the data set achieved self-predictable goals. Due to the data set being merged from learning logs and questionnaire measurement results, we referred to prior studies to define the evaluation process, as shown in the Figure 3.

1. **Normalization:** Learning behaviors and learning strategies were specified on different scales. Learning logs were the behavior count, and questionnaire measurement results were the level of

learning strategy capabilities determined by Likert Scales. Therefore, a normalizer is necessary to consistent both scales.

2. **Principal Component Analysis (PCA):** The proposed LBLS-160 data set has 56 features merged from two learning logs and one questionnaire measurement; not every feature was critical for the evaluation task. Therefore, we adopted PCA to identify critical features.
3. **Support vector machine (SVM):** In this experiment, we divided students into risk and non-risk, then used the proposed LBLS-160 data set to teach a computer how to classify each other. Therefore, the classification algorithm was picked-up for this sub-process.

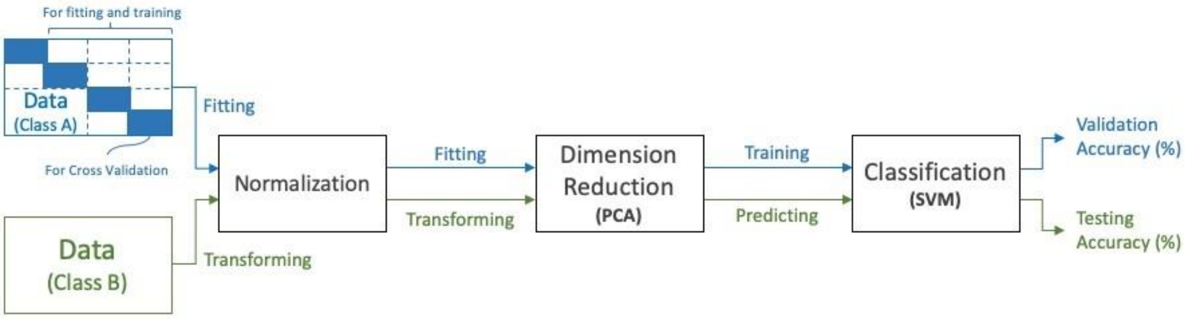

Figure 3. Cross-class risk prediction model evaluation process.

A cross-class evaluation has been conducted in this experiment. We will target one subset from a specific class to fit the normalizer, PCA, and to train the model. Here "class" specified a course in a particular duration of time instead of categories in the machine learning dictionary. Validation accuracy was obtained when we input a split training set into the model, and the accuracy should present outstanding performance to prove the self-predictable concerns. On the other hand, testing accuracy was obtained when we input subsets from different classes.

4. Results and Discussions

This experiment proves that Class A and Class B's feature has similar distributions. Still, Class C has fewer generalization characteristics than Class A and B. This observation is in line with the survey results on different data sets by Conijn et al. (2016) on different data sets. In the evaluation process, first, since the data set includes a total of 131 features, as mentioned in the literature review section, we need to perform PCA to reduce the dimensions. Using Class A as an example, the PCA results are shown in Figure 4. When the Cumulative Explanation Variance (CEV) reaches 80%, 10 components are found, and the 10 vectors along with components are used to reduce the dimension of Class B and C.

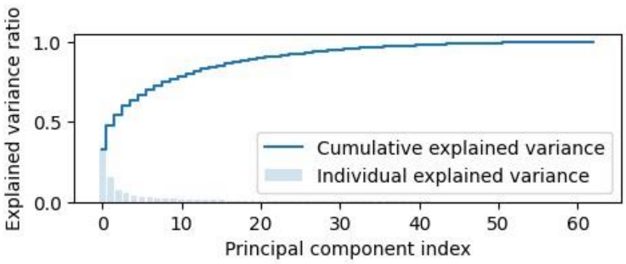

Figure 4. PCA results of Class A, CEV=80%, then Components=10

After normalization, the overall accuracy evaluation results of the training model are shown in the Figure 3. The horizontal axis represents the final grade of the course. First of all, taking the final grade of 80 points as an example, it means we labeled students with grades lower than 80 as True(T) and others as False(F). Then we calculated a T/F balance ratio by using the number of T divided by the number of F. The ideal T/F balance ratio is one, which represents the number of T and F were equaled (Lu, Huang, & Yang, 2021). At this time, the blue bar on points of 80 in this figure means the T/F balance ratio of Class A is around 0.5, which might encounter unbalanced issues, making the Accuracy of the model non-referenceable. In this case, the ideal final grade for training the model is between 84 and 88, marked in red in this figure. In addition, the figure's vertical axis represents Model A's Accuracy. It shows the validation accuracy of Model A for Class A is 0.81, and the cross-course test accuracy for Class B and Class C is 0.77 and 0.67, respectively.

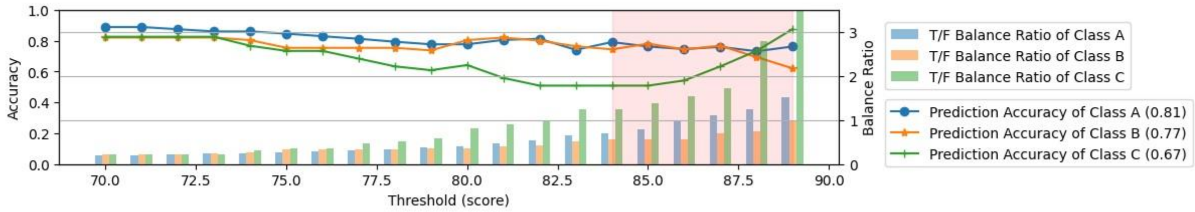

Figure 5. Prediction accuracy of Model A and True(T)/False(F) balance ratios for each class.

Table 3 lists the model's validation and testing accuracy, which was trained by components with CEV larger than 80% and individually classified the training set by final grades of 84 to 87. In the table, Model A indicates that the model is trained with the data of Class A and validates the classification accuracy of Class A, testing the classification of Class B and C, respectively. Then got average Accuracy on Class A: 0.78, Class B: 0.71, and Class C: 0.57. Model B and Model C, and so on. The results of all three experiments show that the validation accuracy is higher than the test accuracy.

The above evaluation results can be discussed that LBLS-160 has the potential to be the data challenge target. There are two significant reasons. First, Class A and Class B are in the same semester, under the condition that the syllabus, homework, and progress are almost consistent; therefore, the Accuracy of Model A and Model B are similar. However, the current evaluation used the most common SVM for the model training; although the validation accuracies are all around 0.8, it still has the possibility to be improved by applying other algorithms. Second, although the syllabus and homework of Class C are consistent with Class A and Class B, but there may be some biases in the implementation of the curriculum for teachers. Therefore, the accuracy of Class C is low compared with the other two classes. This result shows that the model does not have generality. This illustrates the issue where risk prediction models are not available in practical scenarios so that it will become one of the challenges in the future.

Table 3. Cross-class risk prediction accuracy (CEV>80%, final grade: 84~89).

	Model A	**Model B**	**Model C**
Class A	**0.78**	0.61	0.56
Class B	0.71	**0.80**	0.67
Class C	0.57	0.60	**0.71**

5. Challenges

Finally, to achieve the goal of facilitating learning analytics research and development, we have sorted out three learning analytics applications raised in recent years and considered to be potentially achieved through LBLS-160 as follows:

1. **Educational data visualization:** In recent years, educational data visualization has become increasingly popular to support learners' monitoring and tracking of their learning status. Researchers summarized a few meaningful research questions on this topic, for example: "Who are the learners?", "What do they do while learning?"(Schwendimann et al., 2016).
2. **Learning strategies unveiling:** This is also a young topic since 2017(Jovanović et al., 2017). Researchers demonstrated learners' book reading behaviors were a piece of evidence of their SRL strategy (Akçapinar, Chen, Majumdar, Flanagan, & Ogata, 2020). Measuring learners' learning strategies using logs instead of questionnaires could be in more real-time and reliable. Therefore, unveiling the correlation between students' learning behaviors and learning strategies will be a reasonable research question in the proposed data set.
3. **Cross-class risk prediction:** Prior studies proved learning logs were valuable materials to identify risk students in the classroom (Conijn et al., 2016; Lu et al., 2016). However, the prediction model in prior studies didn't confirm the generalizable in the cross-class scenario. Model performance benchmark on one opened data set has also not been considered.

Conclusions

This study aims to release LBLS-160, which is an educational data set to facilitate the development of AI in education. We collect data sets from three programming courses containing students' reading behavior, coding behavior, self-regulated learning ability, and language learning strategy. To evaluate the quality of LBLS-160, we conducted an experiment that predicted students' learning performance and whether the indicator: Accuracy, was consistent with prior studies or not. The results indicate the model has acceptable Accuracy if we train and evaluate it in the same classes and unacceptable Accuracy in the cross-class scenario. This issue is consistent with prior studies and establishes a benchmark baseline for future challenges. Based on the characteristics of LBLS-160, several possible contributions could be conducted in the future; the first one is using logs to unveil students' learning strategies instead of using a questionnaire. And a more generalized risk prediction model for diverse curriculum design.

Acknowledgments

This work was supported by National Science and Technology Council, Taiwan under grants MOST 109-2511-H-008-007-MY3, 108-2511-H-008-009-MY3 and 110-2511-H-153 -001 -.
.

References

Akçapinar, G., Chen, M.-R. A., Majumdar, R., Flanagan, B., & Ogata, H. (2020). *Exploring student approaches to learning through sequence analysis of reading logs.* Paper presented at the Proceedings of the tenth international conference on learning analytics & knowledge.

Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining educational data to predict student’s academic performance using ensemble methods. *International Journal of Database Theory and Application, 9*(8), 119-136.

Choi, S. P., Lam, S. S., Li, K. C., & Wong, B. T. (2018). Learning analytics at low cost: Atrisk student prediction with clicker data and systematic proactive interventions. *Journal of Educational Technology & Society, 21*(2), 273-290.

Cigdem, H. (2015). How does self-regulation affect computer-programming achievement in a blended context? *Contemporary Educational Technology, 6*(1), 19-37.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683-695.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies, 10*(1), 17-29.

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *Imagenet: A large-scale hierarchical image database.* Paper presented at the 2009 IEEE conference on computer vision and pattern recognition.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Echeverry, J. J. R., Rosales-Castro, L. F., Restrepo-Calle, F., & González, F. A. (2018). Selfregulated learning in a computer programming course. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 13*(2), 75-83.

Ernst, M. D. (2017). *Natural language is a programming language: Applying natural language processing to software development.* Paper presented at the 2nd Summit on Advances in Programming Languages (SNAPL 2017).

Flanagan, B., Majumdar, R., & Ogata, H. (2022). Fine Grain Synthetic Educational Data: Challenges and Limitations of Collaborative Learning Analytics. *IEEE Access, 10*, 26230-26241.

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. In (Vol. 1, pp. 100001): Elsevier.

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education, 33*(4), 74-85.

Kellogg, S., & Edelmann, A. (2015). Massively open online course for educators (MOOC-E d) network dataset. *British Journal of Educational Technology, 46*(5), 977-983.
Khanal, S. S., Prasad, P., Alsadoon, A., & Maag, A. (2020). A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies, 25*(4), 2635-2664.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems, 25*.
L. Leite, W., Roy, S., Chakraborty, N., Michailidis, G., Huggins-Manley, A. C., D'Mello, S., . . . Jing, Z. (2022). *A novel video recommendation system for algebra: An effectiveness evaluation study.* Paper presented at the LAK22: 12th International Learning Analytics and Knowledge Conference.
Lu, O. H., Huang, A. Y., Huang, J. C., Huang, C. S., & Yang, S. J. (2016). *Early-Stage Engagement: Applying Big Data Analytics on Collaborative Learning Environment for Measuring Learners' Engagement Rate.* Paper presented at the 2016 International Conference on Educational Innovation through Technology (EITT).
Lu, O. H., Huang, A. Y., & Yang, S. J. (2021). Impact of teachers' grading policy on the identification of at-risk students in learning analytics. *Computers & Education, 163*, 104109.
MITx, H. HarvardX-MITx Person-Course Academic Year 2013 De-Identified dataset, version 2.0. Harvard Dataverse (2014). In: ed.
Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). *EBook-based learning analytics in university education.* Paper presented at the International conference on computer in education (ICCE 2015).
Oxford, R. L., & Burry-Stock, J. A. (1995). Assessing the use of language learning strategies worldwide with the ESL/EFL version of the Strategy Inventory for Language Learning (SILL). *System, 23*(1), 1-23.
Pintrich, P. R. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).
Romero, C., López, M.-I., Luna, J.-M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education, 68*, 458-472.
Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(3), e1355.
Russell, S. J. (2010). *Artificial intelligence a modern approach*: Pearson Education, Inc.
Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., . . . Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies, 10*(1), 30-41.
Shin, Y., & Song, D. (2022). The Effects of Self-Regulated Learning Support on Learners' Task Performance and Cognitive Load in Computer Programing. *Journal of Educational Computing Research*, 07356331211052632.
Stamper, J., Niculescu-Mizil, A., Ritter, S., Gordon, G., & Koedinger, K. (2010). Challenge data set from KDD Cup 2010 Educational Data Mining Challenge.
Sun, L., & Frederick, C. (2015). Applying Second Language Acquisition to Facilitate a Blended Learning of Programming Languages.
Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*: Routledge.