Chang, M. et al. (Eds.) (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education

Measuring Analysis Skill in Data-informed Self-directed Activities

Yuanyuan YANG[a*], Rwitajit MAJUMDAR[b], Huiyong Li[a], Gökhan AKÇAPINAR[b], Brendan FLANAGAN[b] & Hiroaki OGATA[b]
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*44.yangoo@gmail.com

Abstract: Current technology enables tracking of various learning and physical activities. User can use the data to analyze issues in the execution of those activities. Current work focuses on this analysis phase of data-informed self-directed activity cycle and proposes a measurement framework of the skill while learners work in a data-rich context. It is a paradigm shift to support and measure analysis skill from previous approaches which mostly rely on questionnaire-based measurements. In our approach, we emphasize the monitoring of learner's analytical process and the automatic evaluation of the analysis results through system. Based on that, an automated measurement is carried out in the system to depict learner's analysis skill and changes of skill. Additionally, we elaborate the framework in the context of the GOAL system which provides affordances of analysis based on physical and reading activity data.

Keywords: Data analysis, Quantified-self, Self-directed learning, Self-direction skills, Automated skill measurement, GOAL

1. Introduction and Background

Various frameworks for the 21st century learning consider self-direction skill (SDS) is a necessary skill for learners (P21 framework, 2015; enGauge 21st century skills 2003; SCANS report 1991). According to Knowles (1975), Self-Directed Learning (SDL) is described as "a process in which individuals take the initiative, with or without the help from others, in diagnosing their learning needs, formulating goals, identifying human and material resources, choosing and implementing appropriate learning strategies, and evaluating learning outcomes". Researchers also propose different models to understand the process of SDL, such as Candy's model (1991) comprises personal autonomy, self-management, learner-control, and autodidaxy. Garrison's model (1997) contains self-management, self-monitoring, and motivation. Loyens, Magda, and Rikers (2008) point out that SDL generally entails goal setting and task analysis, implementation of the constructed plan and self-evaluation of the learning process.

We notice that both learning and physical activities play an important role in students' life and being self-directed is important in those contexts. For example, developing a habit of running one hour every day and a habit of reading one hour every day, both of them require individuals to choose appropriate strategies, to monitor process, to evaluate outcomes and so on. In our research, we pay our attention to learner's daily life and focus on self-directed activities in learning and physical activity contexts.

Moreover, with the current e-learning tools and wearable devices, tracking and logging learning behaviors and physical activities has become more affordable. It provides us a chance to search for data-driven way to support execution and acquisition of learner's SDS. We proposed a process model called DAPER (Data Collection – Data Analyze – Setting Goal and Plan – Execution and Monitoring - Reflect) for execution of data-informed self-directed activity and acquisition of the sub skills (Majumdar et.al., 2018). The five phases in the model relates to the five subskills of being self-directed. We developed GOAL (Goal Oriented Active Learner) system based on DAPER model, as a technology support for promoting learner's SDS.

This paper focuses on the Data Analysis phase of the DAPER model showing in figure 1. Analysis phase requires learners to get aware of their own situation, trend and current skill level by checking and comparing the collected data of an activity context. It is a crucial precondition of setting meaningful goals and feasible plans. In past researches, scholars have proposed some useful strategies to help improve students' SDS. For example, Diclemente and Proschaska (1998) designed guideline for it, which contains various items to help people recognize their possible problematic behavior. However, such strategies don't take the affordances of the current e-learning tools and often can't be directly adopted for implementing in an e-learning system. In our work, we provide strategies within digital environment, which asks the learner to analyze their own activity data synthesized in the GOAL system. Figure 1 highlights data analysis phase in DAPER model and learner's analysis task in this phase.

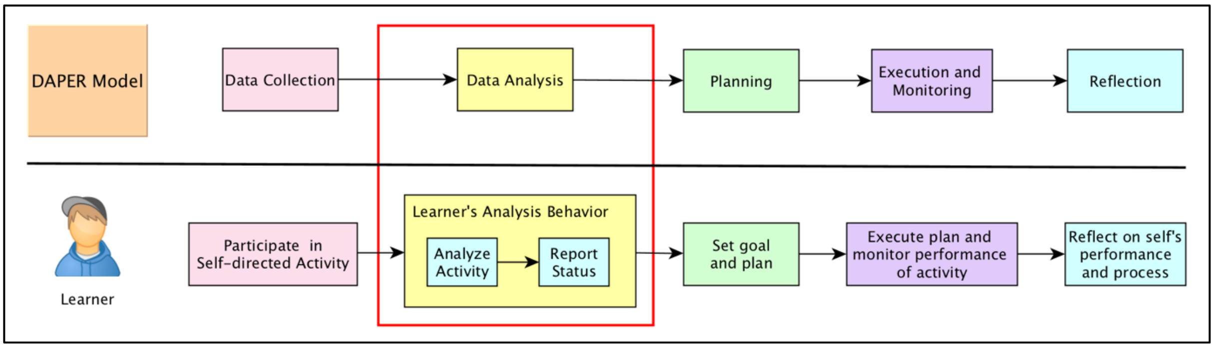

Figure 1. Data Analysis Phase and Analysis task of learner

Based on previous literature, we elaborate our notion of Data Analysis as a sub-skill of Self-directedness. Then we propose our framework to answer the following research question: *How to measure data analysis skill of learners in a data-rich system?* We implement this framework in the GOAL system and demonstrate on simulated data.

2. Literature Review

2.1 Analysis Skill and Strategies of Analysis

Analyzing own activities is crucial for being self-directed in learning. Loyens, Magda, and Rikers (2008) consider that analysis happens in the starting point of SDL. They state that analysis in the practice of SDL is to analyze the task at hand and to determine the task (e.g., what is the task about?) and personal features (e.g., what knowledge can I apply? Do I find the task interesting?). In Thornton's research (2010), he mentioned analyzing needs and analyzing current skill in planning phase which is one phase in four phases of a self-directed learning cycle. Noguchi, J., and McCarthy, T. (2010) states that analytical skill is the ability to examine what happened in their learning process in detail and discern the cause and effect relationship among various elements involved in the process.

In order to achieve this objective, most common strategies are either suggesting learner following questions or hints to think and self-report. For example, Diclemente and Proschaska (1998) designed guideline that contains various items for helping learner to analyze. Curry, Wagner and Grothaus (1990) suggest learners think about their personal reason in analysis. M.E. Gredler and L. S. Schwartz (1997) designed questionnaire for helping learner to self-report. However, none of them is able to observe learners' analysis behavior itself but only relies on their self-report.

In DAPER model, we treat analysis as the second phase of any self-directed activity. The learner is required to analyze their activity data for understanding their own status. By using the system affordances, the learners can complete specific analysis task with their activity data and identify their own status.

2.2 Measuring Data-analysis skill as a Sub-skill of Self-directedness

To help learner to acquire skill, one needs to measure and estimate current skill level. To measure SDS, interview and questionnaire were widely used in past research. In Williamson's research (2007), SRSSDL (the self-rating scale of self-directed learning) is used as the instrument to measure the level of self-directedness in learner's learning process. It covers five constructs, which are awareness, learning strategies, learning activities, evaluation and interpersonal skills. Stockdale and Brockett (2010) designed PRO-SDLS (Personal Responsibility Orientation to Self-Direction in Learning Scale). Noguchi, J., and McCarthy, T. (2010) asked advisors of learners to speak their thinking process out while they are evaluating and deciding final grades for the submitted module work of their students and he recorded the think-aloud sessions. A list of the criteria of grading was made on that record in this research. These survey- or questionnaire-based evaluations give score on learners' answers for rating learners' skill. Measuring analysis is covered by some items in the scales. For example, in Williamson's scale (2007), the item in 3.10 asks students to rate their ability to analyzing new ideas, information or any learning experiences. Too many items in the scale would overload the learners while rating the skill. And it significantly relies on learners' memory but not their actual analysis behavior.

In our research, we explore the possible way to automatically measure learners' analysis skill from their actual performance in analysis task as a part of the self-directed activity and performing a real-time track on their skill's changing. Automated measurement based on performance would eventually be a more seamless experience for learners comparing with past researches.

3. Framework for Skill Measurement of Data Analysis Skill

Based on the previous works and contextualizing it for a data-rich environment, we set the objectives of analysis skill of learners. In the analysis phase of DAPER model, the individual needs to check their collected data in a specific context, then analyze issues in their activities which can lead to planning phase to set appropriate plans to overcome those problems or pursue challenges. In this section, we describe our framework for analysis skill measurement.

3.1 Framework and Data Flow

Figure 2 shows the overview of the components of our framework for analysis skill measurement. Before analysis phase, learner should participate in self-directed activities in learning or physical context.

The framework contains 4 components. The first component provides analysis function that allows learners to analyze their activity with the help of system. The second component is to gather data while learner is using analysis functions in the system. Two kinds of data are gathered, learner's interaction log and their analysis report. The third component is the system analysis. In automatic analysis, activity data compares with criterion to report status of learner's activity. Based on that, the system analysis component classifies whether an individual has problems. The last component rates skill from learner's interactions, learner's report, system report, and the scale of analysis skill.

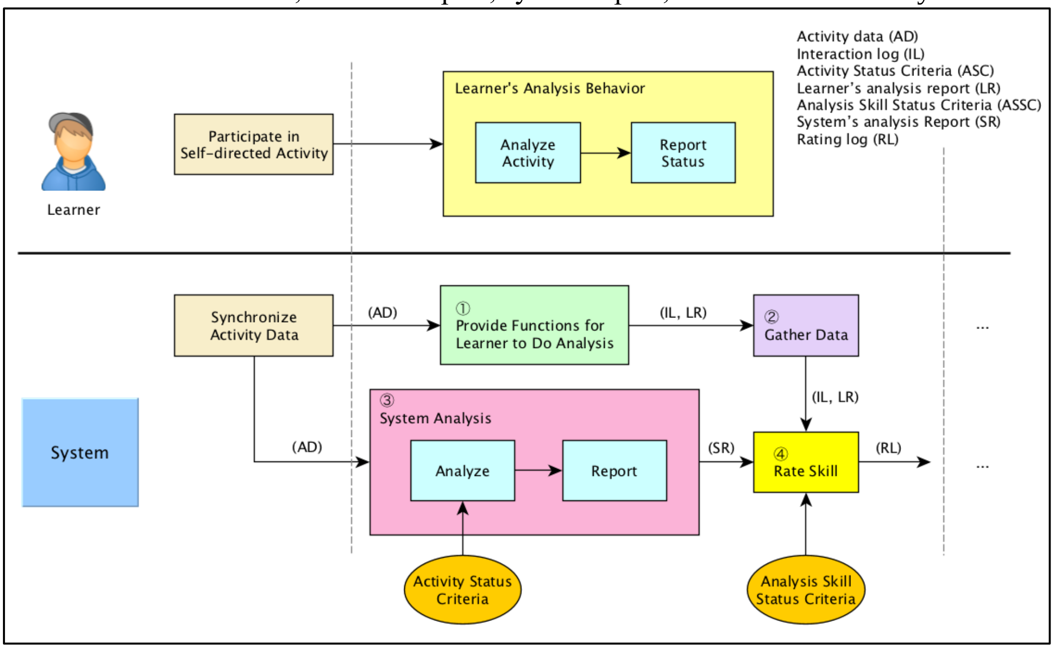

Figure 2. Framework for Analysis Skill Measurement and Relation to Learner's Behavior

3.2 Data Model and Criteria

Our method advocates automatic data gathering and calculating for skill measurement by the system. This section describes what kind of data the system will gather and use for calculation.

Table 1 describes the five kinds of data and the two types of criteria in our measurement. Activity Data (AD) is collected automatically from other data-logging platforms, such as, physical activity data from iOS Healthkit or Androids Google Fit, learning data from digital learning system. While the learners use the system to analyze their Activity Data, Interaction Log (IL) and the Learner's analysis Report (LR) are recorded. System's analysis Report (SR) are generated automatically by comparing Activity Data based on the Activity Status Criteria (ASC). Considering the interaction log, the learner's report and the system's report and comparing to the Analysis Skill Status Criteria (ASSC), the Rating Log (RL) is generated by the system.

Table 1 *Data Model and Criteria*

	Description	Sample	Relation to Analysis
Activity data (AD)	A specific activity, which might be a physical activity or a learning activity.	One walked 4000 steps in 2019/5/15.	Object of analysis
Interaction log (IL)	Interactions while learner do analysis in the system.	One checked his activity data for last 7 day in 2019/5/12 on the system.	Process of doing analysis by using system's features
Activity Status Criteria (ASC)	Criteria for analyzing learners' activity	Tudoe-Locke C.'s (2008) zone-based hierarchy on daily steps	Help to get result from analyzing activity data
Learner's analysis report (LR)	Contents that learner reports their analysis result.	One found himself lacking physical exercise.	Result of analysis as noted by learners
Analysis Skill Status Criteria (ASSC)	Criteria for rating learners' analysis skill	Criteria of SDS proposed by Majumdar. et.al. (2019)	Help to rate learners' analysis skill
System's analysis Report (SR)	Result of automatically analyzing activity data by the system	System distinguished one lacking physical exercise	Result of analysis as computed by system
Rating log (RL)	Related data and result on rating analysis skill	Comparing results from the learner and system, rate the learner's skill based on the scale of analysis skill	Related to get analysis skill level

3.3 Implementation of framework in GOAL System

In the GOAL system, we implement this framework to automate the measurement of analysis skill. This section details how we implement each component in GOAL system.

3.4 Component 1: Functions for Analysis

Based on DAPER model, we developed GOAL (Goal Oriented Active Learner) system, which contributes to technology solutions of promoting learner's SDS (Self-Direction Skill). Learner's analysis behavior becomes observable and traceable through the GOAL system. It supports data logging of physical learning activities.

We provide visualized graphs for learners which allows them to check their own activity status. We consider from Williamson's (2007) and Noguchi, J., and McCarthy, T. 's (2010) research. In Williamson's (2007) research, strategy is considered as an important element of analysis. In order to understand self's status, we suggest learner comparing their own data with other related data as one of strategies in analysis phase. Hence, various options are designed in the graph, such as showing recommend value, or maximum, minimum, average value of all users' activity data, as shown in figure 3a. Noguchi, J., and McCarthy, T. (2010) mentioned that the key word "details" reflects learner's analysis skills. Details means how deep learner understand data. Checking detail data helps learner know the exact gap which relates to setting an effective plan. The graph showing in figure 3b provides function for exhibiting detail data of every related value.

Additionally, analysis report function shown in figure 3c, requires learners to report the result of their analysis and note the issue they have found. And learner can see system analysis result to compare their own analysis.

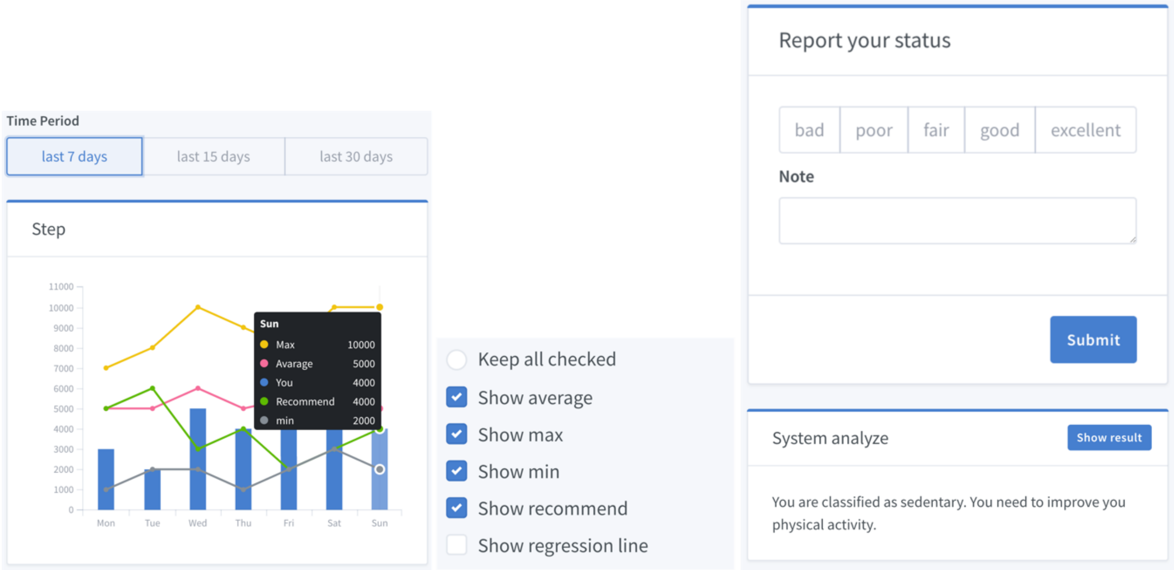

Fig 3a. Visualized Data to analyze

Fig 3b. Details on Demand for Analysis

Fig 3c. Record report and verify with system

Figure 3. the Analysis Page in GOAL System

3.5 Component 2: Data Gathering in Analysis phase

Two kinds of data are captured in the data gathering component. The first one is interaction logs when user using system's features, such as choosing an activity to analyze, clicking "show result" button to check system's analysis result. The second one is the content of report from using feature in Fig 3c.

3.6 Component 3: Analysis of Activity by System

To measure learner's analysis skill, our approach is to compare the student's analysis report with the system's analysis report regarding an activity. One can potentially implement various methods in this component by using statistical techniques or different machine learning approaches. To demonstrate we use a linear regression method for system analysis. Linear regression is a method used for explaining the relationships among variables. In our context, we hope to find out daily changes in physical activities and learning activities. We use following linear regression formula to express the relationship between activity and day.

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Where i : the number of the data; α: the constant coefficient. β: the slope of the line, which reflect the tendency of the variation. X: the number of days. Y: the value of an activity in one day. ε: random disturbances and cannot be directly observed.

System does analysis on collected activity data in recent time period and calculate the coefficients of expression from it. Based on that we predict the value of activity in one day later than that time period. Next, we compare it with the criteria of the activity. Depending to different type of activity, we use different criteria. The criteria of the activity can come from other researches, for example Tudor-Locke, C.'s (2008) zone-based hierarchy for computing status of daily steps activities.

3.7 Component 4: Rating Analysis Skill

We aim to measure their analysis skill automatically by using technologies. The GOAL system captures data and calculates their skill level from that data. This section gives our data structure of rating log and approach for rating data analysis skill from the rating log.

Rating log keeps the unified record of the related computed values in component 2 for rating analysis skill and result of skill level. The related computed values contains count of interaction logs, learner's analysis result and system' analysis result. These data will be used for classifying the learner's analysis skill. Table 2 gives all of attributes of the log.

Table 2 *Attributes of Analysis Rating log*

Attributes	Description
id	Unique identifier of report
uuid	Whose activity belongs to
activity_type	Report for which activity
count_interaction	The count of interaction logs while the learner is analyzing
check_sys_result	Whether system analysis result has been shown before learner report the result
self_result	Result from learner doing the report 0: bad, 1: poor, 2: fair, 3: good, 4: excellent
system_result	Result from system analyzing learner's activity data 0: bad, 1: poor, 2: fair, 3: good, 4: excellent
skill_level	The result of rating analysis skill
date	When learner did report

On the other hand, we introduce criteria of analysis skill shown in the first two columns of table 3 (Majumdar. et.al. 2019), which contains five levels to describe different analysis skill. In addition, we give more specific logic expression on it shown in the third column of table 3.

Table 3 *Scoring Rubric for Analysis Skill*

Score	Analysis behavior	Logic expression
4	Check data - Successfully identify status WITHOUT system support	$N_1 > 0 \ \&\& \ N_2 > 0 \ \&\& \ (R_1 == R_2) \ \&\& \ S$
3	Check data - Successfully identify status WITH system support	$N_1 > 0 \ \&\& \ N_2 > 0 \ \&\& \ (R_1 == R_2) \ \&\& \ !S$
2	Check data - PARTIALLY identify status	$N_1 > 0 \ \&\& \ N_2 > 0 \ \&\& \ (R_1 != R_2)$
1	Check data - DID NOT identify status	$N_1 > 0 \ \&\& \ (N_2 == 0)$
0	DID NOT Check data	$N_1 == 0$

Where N_1: Count of interaction logs; N_2: Count of learner's analysis reports; R_1: Result of learner's analysis report; R_2: Result of system report; S: Boolean data, whether system analysis result has been shown before learner report his result.

4. Demonstration with Pilot Data

In this section we present an illustration of our framework in a data-informed self-directed activity by picking one kind of activity to explain how the framework works and what the result would be.

4.1 Sample of Activity Data and System Analysis

The GOAL system collected physical and learning activity data of 15 learners. We select a set of daily steps data from a male master student which is most complete. It had 183 records of the daily steps from April 1, 2018 to September 30, 2018. The daily steps data ranges from 1490 to 55903. Mean of sample is 12265.85, and the sample standard deviation is 8467.808. The data set in each month and the regression line is shown in figure 4.

On the other hand, we introduce the criteria of adults' daily steps proposed by Tudor-Locke, C. (2008), which classified adults into 5 status: 1) < 5000 steps/day (sedentary); 2) 5000–7499 steps/day (low active); 3) 7500–9999 steps/day (somewhat active); 4) ≥10,000–12,499 steps/day (active); and 5) ≥12,500 steps/day (highly active). At the end of each month, system executed a linear regression on the data for that month and got the regression coefficients shown in table 4. According to these coefficients, the system predicted the value of the next day of the data set. Then comparing it with the criteria of adults' daily steps, system computed the status of his physical activity in each month.

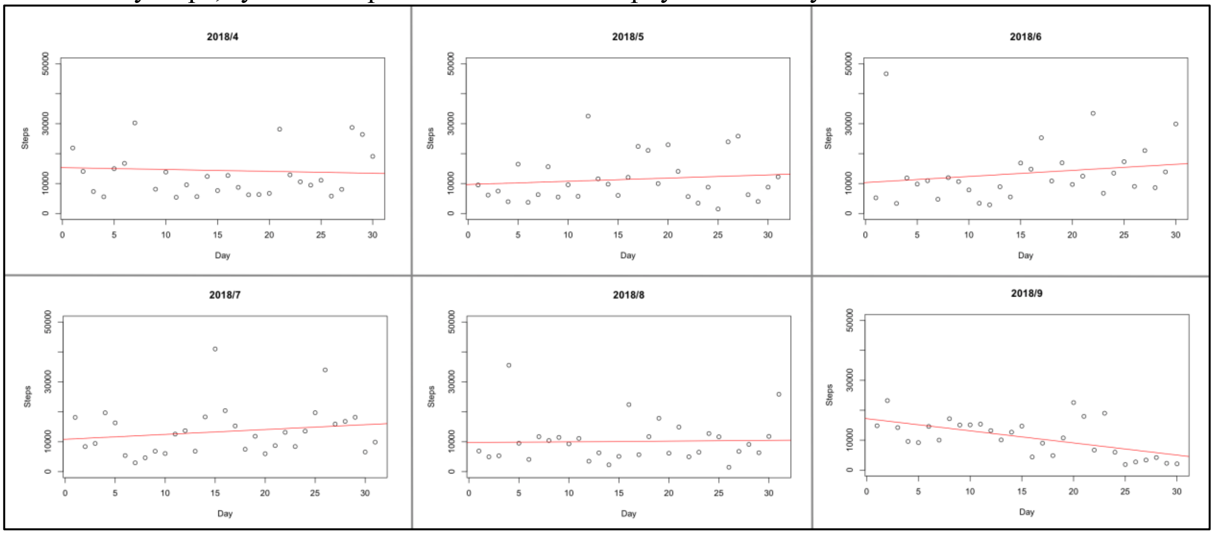

Figure 4. Daily Steps over 6 Months along with Regression Line

Table 4 *Coefficients and Result Analyzed by System*

month	α	β	predicted value	system_result
2018/04	15327.01	-62.93	13376.18	5
2018/05	9670.4	107.7	13116.8	5
2018/06	10327.7	203.8	16645.5	5
2018/07	10816.7	163	16032.7	5
2018/08	9723.11	24.59	10509.99	4
2018/09	17172.9	-405.7	4596.2	1

4.2 Simulated Interaction Logs and Rating Report

At the time of writing this paper the interaction logging module was under development. Hence, we simulated analysis interaction data for the same student and the same activity shown in table 5. By comparing with system analysis result and the criteria of analysis skill level, system got this student's skill level in each month (see table 5).

Table 5 *Rating Report*

id	date	count_ interaction	check_system_ result	self_result	system_result	skill_level
1	2018/4/30	0	0	0	5	0
2	2018/5/31	15	0	0	5	1
3	2018/6/30	15	0	3	5	2
4	2018/7/31	14	0	3	5	2
5	2018/8/31	10	1	4	4	3
6	2018/9/30	12	1	1	1	3

From the results we can see, this student didn't have awareness of analyzing his own status at first. Then, his analysis skill keeps improving over the next five months.

5. Conclusion

In this article, we focus on Analysis phase in SDS which examines learners' propensity and actual situation which leads to successfully pursue challenges. As strategy of doing analyze, we suggest learner to get aware of their own situation, trend and current skill level by checking and comparing the collected activity data. We proposed a framework for analysis skill measurement in data-informed self-directed activities, which considers the object, process, and outcome of the analysis. Then we implemented this framework in the GOAL system. With the pilot activity data that we collected, we demonstrate a proof of concept of the framework to measure and track the changes in skill level in the context of physical activity.

We contribute by proposing a new method to automatically measure analysis skill in data-rich context. The automated measurement aims to benefits the learners to understand their skill better. With this approach and understanding of the current level of students' skill would enable the system to give a further adaptive support for the learners in the future.

In this paper, we use regression to support component 3, system analysis. In this future, it is worthwhile to explore and compare other statistical methods and machine learning approaches, to build the system analysis component. In our demonstration with pilot data, we let the system measure the analysis skills monthly. However, choosing the appropriate time granularity for different activity context and developing flexible algorithm is also one of our future work.

Lastly, this research considers both learning and physical contexts. It potentially gives us more insights for supporting SDS in multiple contexts.

Acknowledgments

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304 and JSPS KAKENHI Research Activity Start-up Grant Number 18H05746 (F.Y.2018), 19K20942 (F.Y.2019). This research is also supported by the Leading Graduates Schools Program, “Collaborative Graduate Program in Design” by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

Candy, P. C. (1991). Self-Direction for Lifelong Learning. A Comprehensive Guide to Theory and Practice. Jossey-Bass, 350 Sansome Street, San Francisco, CA 94104-1310.

Curry, S., Wagner, E. H., & Grothaus, L. C. (1990). Intrinsic and extrinsic motivation for smoking cessation. *Journal of Consulting and Clinical Psychology, 58(3)*, 310-316.

DiClemente, C. C., & Prochaska, J. O. (1998). Toward a comprehensive, transtheoretical model of change: Stages of change and addictive behaviors. *In W. R. Miller & N. Heather (Eds.), Applied clinical psychology. Treating addictive behaviors (pp. 3-24)*. New York, NY, US: Plenum Press.

Framework for 21st century learning http://www.p21.org/our-work/p21-framework.

Garrison, D. R. (1997). Self-directed learning: Toward a comprehensive model. *Adult education quarterly, 48(1)*, 18-33.

Knowles, M. S. (1975). Self-directed learning: A guide for learners and teachers. New York: Association Press.

Loyens, S. M. M., Magda, J., & Rikers, M. J. P. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational Psychology Review*.

M.E. Gredler, & L. S. Schwartz (1997). Factorial Structure of the Self-Efficacy for Self-Regulated Learning Scale. *Psychological Reports*, 1997,81, pp 51-57.

Majumdar R., Yang Y.Y., Li H., Akçapınar G., Flanagan B., & Ogata H. (2018) GOAL: A System to Support Learner's Acquisition of Self Direction Skills. *26th ICCE,* Manila, Philippines, Nov 2018.

Majumdar R., Li H., Yang Y.Y., Akçapınar G., Flanagan B. and Ogata H. (2019). Towards Adaptive Support for Acquisition of Self-direction Skills in GOAL system, 19th IEEE ICALT, Maceió, Brazil July 2019.

NCREL and Learning Point associates (2003) http://www.cwasd.k12.wi.us/highschl/newsfile1062_1.pdf

Noguchi, J., & McCarthy, T. (2010). Reflective Self-study: Fostering Learner Autonomy. In A. M. Stoke (Ed.), JALT2009 Conference Proceedings. Tokyo, Japan: JALT.

Stockdale, S. L., & Brockett, R. G. (2011). Development of the PRO-SDLS: A measure of self-direction in learning based on the personal responsibility orientation model. *Adult Education Quarterly, 61(2)*, 161-180.

Thornton K. (2010), Supporting Self-Directed Learning: A Framework for Teachers. *Language Education in Asia, 1(1),* 158-170.

Tudor-Locke, C., Hatano, Y., Pangrazi, R. P., & Kang, M. (2008). Revisiting “how many steps are enough?” Medicine and Science in Sports and Exercise, 40(7 SUPPL.1).
What work requires of schools. A SCANS report for America 2000 (1991), https://wdr.doleta.gov/scans/whatwork/whatwork.pdf
Williamson, S. N. (2007). Development of a self-rating scale of self-directed learning. Nurse researcher, 14(2).