Enabling Precision Education by Learning Analytics Applying Trace, Survey and Assessment Data

Enabling Precision Education by Learning Analytics Applying Trace, Survey and Assessment Data

Dirk Tempelaar
School of Business & Economics
Maastricht University
Maastricht, the Netherlands
d.tempelaar@maastrichtuniversity.nl

Bart Rienties
Institute of Educational Technology
The Open University UK
Milton Keynes, UK
bart.rienties@open.ac.uk

Quan Nguyen
School of Information
University of Michigan
Ann Arbor, MI, US
quanngu@umich.edu

***Abstract*— Accurate and timely measurement of learning engagement is crucial for the application of precision education. At the same time, it is still a central research theme, both in the learning analytics community as in the broader area of educational research. 'Engagement is one of the hottest research topics in the field of educational psychology' is for a good reason the opening sentence of a recent special issue. In our contribution, we propose a holistic approach to the measurement of engagement by integrating data of behavioral type through traces of learning processes captured from log files into affective, behavioral, and cognitive measures of engagement collected with surveys and cognitive measures from assessments for and as learning. We apply this holistic approach in an empirical analysis of dispositional learning analytics. Starting from four different engagement profiles created by two-step clustering, we find that these profiles primarily differ in their timing of engagement with learning. Next, we develop regression-based prediction models that make clear that trace, survey, and assessment data have complementary roles in signaling students at risk for failure and are all three crucial constituents of prediction equations that differ in the timing of learning feedback.**

Keywords—dispositional learning analytics, precision education, trace data, survey data

I. Introduction

Precision education [1] aims to identify at-risk students as early as possible and provide timely intervention through diagnosis, prediction, treatment, and prevention. The topic of student engagement and the measurement of engagement is crucial because of its close connection to self-regulated learning, the condition sine qua non for all learning, and learning in technology-enhanced environments in specific [2, 3]. The above-mentioned special issue [4] contains several contributions. Each provides an alternative solution for the measurement of learning engagement: self-report surveys, log data from technology-enhanced learning systems, think alouds' and tests. Rather than creating such antithesis, we propose in this contribution a synthesis of information from different data sources, much in line with contemporary approaches to Educational Big Data [5, 6]. Our proposal for such synthesis is based on ideas derived from dispositional learning analytics (DLA, [7, 8]). The DLA infrastructure combines learning data (generated in learning activities through technology-enhanced systems) with learner data (i.e., self-reported student dispositions, values, and attitudes.

Beyond a general agreement on the importance of the construct, 'engagement could be described as the holy grail of learning,' [9, p. 1], research literature demonstrates a lack of agreement on how to operationalize learning engagement. Traditional educational research applies survey instruments to investigate the role that engagement plays in the learning process. One of the instruments broadly validated in empirical research is the Motivation and Engagement Scale (MES), based on the 'motivation and engagement wheel' framework [9]. Of more recent times is the data analytics-inspired research tradition of investigating traces in digital learning environments to operationalize learning engagement (see, e.g. [3]). In general, empirical studies in learning engagement are either based on survey data or trace data but nearly never attempting to integrate both approaches [2].

The aim of this article is to provide such 'multi-modal data' based contribution to the research of student engagement in precision education steered learning. In this study, we investigate quantitative aspects of engagement as the intensity of learning activities and add to that qualitative aspects of engagement. We do so by connecting to the research of the role of tutored and untutored problem-solving and using worked examples [11]. This line of research investigates learning behaviors and students' preferences for feedback formats in their learning. Traditionally, research on the use of worked examples and other instructional formats of problem-solving took place in the non-authentic settings of labs. The introduction of learning analytics and, more in general, the use of technology-enhanced instruction created new opportunities for the research of students' preferences for different formats of learning feedback. This development led to a convergence of learning analytics-based studies in the use of feedback by students and instructional design-based research, such as [11]. Our current study is aligned with this development, adding an extra dimension to the research of students' preferences: the temporal dimension. Our study aims to operationalize precision education by classifying students in different clusters [11] and builds on previous research by the authors [6, 8, 13-16], which focused on the early prediction of drop-out or low performance.

II. This study

The integration of two approaches of operationalizing learning engagement, the survey approach, and the data analytics approach, is the primary goal of this empirical study.

978-1-6654-4106-3/21/$31.00 ©2021 IEEE
DOI 10.1109/ICALT52272.2021.00114

The integration of both approaches is enabled by the dispositional learning analytics context of the course we investigate. The instructional format is blended or hybrid learning, which generates a rich set of trace variables that are indicators of learning engagement. Examples of such indicators are overall student activity in the digital learning tool, measured by the number of attempts to solve problems and time-on-task, next to more specific indicators as the number of worked examples studied, the number of hints called for, and the number of finished packages: successfully solving a set of related problems. Measurements of these indicators are dynamic in nature: they are measured in each of the eight sequential, weekly learning cycles. The dispositional aspect of our research refers to the administration of several self-report surveys.

A. Context

This study takes place in large-scale introductory course mathematics and statistics for first-year students of a business administration and economics program in the Netherlands. The educational system is best described as 'blended' or 'hybrid'. The essential component is face-to-face: Problem-Based Learning (PBL), in small groups (14 students), coached by expert tutors (in parallel tutor groups). Participation in the tutor group meetings is required. The online component of the blend is optional: the use of the two e-tutorial platforms SOWISO (https://sowiso.nl/) and MyStatLab (MSL). This design is based on the philosophy of student-centered education, in which the responsibility for making educational choices lies primarily with the student. Since most of the learning takes place outside the classroom during self-study through e-tutorials or other learning materials, the class time is used to solve advanced problems. The educational format, therefore, has most of the characteristics of the flipped-classroom design in common.

In the use of e-tutorials, three different learning phases can be distinguished. In Phase 1, students prepare for the next tutorial session. To be able to take part in solving 'advanced' problems, students are expected to prepare by self-study outside class, e.g., by studying the literature together with some peers or practicing in the e-tutorials. Phase 1 was not formally assessed. Phase 2 was the preparation of the quiz session, one or two weeks after the respective tutorial. Quizzes were taken every two weeks in "controlled" computer labs and consisted of items drawn from the same item pools applied in the practicing mode. Assessment through quizzes was primarily for formative purposes, but students can score a bonus point in each quiz that adds to their exam score. Phase 3 consisted of preparing the final exam at the end of the course. The final exam is summative in nature and has the largest share in the course score (86%). Students' timing decisions, therefore, related to the amount of preparation in each of the three consecutive phases. The timing of data measurements corresponds to the consecutive weeks, where T0 refers to measurements taking place before the course starts, and T1 to T7 refer to measurements taking place in weeks one to seven.

The subject of this study, the entire cohort of students 2018/2019 (1072 students), is of substantial diversity: 79% are international students. It is, therefore, crucial that this introductory module is flexible and allows for individual learning paths. On average, students spend 27 hours connect-time in SOWISO and 32 hours in MSL, which is 30% to 40% of the 80 hours available to learn both subjects. Although students work in two e-tutorial platforms, this analysis will focus on student activity in one of them, SOWISO, because of the availability of fine-grained and time-stamped trace data.

B. Instrument and procedure

E-tutorial systems follow a test-driven learning and practice approach. Each step in the learning process is initiated by a problem and students are encouraged trying to solve each. If a student has not entirely mastered a problem, he or she can ask for hints to solve the problem step by step or ask for a fully worked out example. Upon receipt of feedback, a new version of the problem is loaded (parameter-based) to enable the student to demonstrate his or her newly acquired mastery.

Our study combines trace data from the SOWISO e-tutorial with self-report data measuring learning dispositions and course performance data. In this study, we focus on process data, such as the clicks to initiate the learning support, since those represent students' engagement with learning in the e-tutorial. Dynamic trace data were assigned to the three learning phases, next aggregated over time, to arrive at static, full course period accounts of trace data. A total of six trace variables were selected: #Attempts (total number of attempts at individual exercises), #Examples: (number of worked examples called), #Hints (number of hints called), #Views (number of theory pages called), #Packages (number of sets of related exercises a student finished), and TimeOnTask (total time on task in problem-solving).

Survey-based engagement indicators are taken from three instruments with two measurement times. The first is the MES-instrument, derived from 'Motivation and engagement wheel' framework by [10], measured at the start of the course (T0) and in the fifth week of the course (T5). Martin breaks down learning cognitions and learning behaviors into four categories: adaptive versus maladaptive types and cognitive versus behavioral types. The classification is based on the theory that thoughts and behaviors can either enable learning and act as boosters, or hinder learning by acting as mufflers and guzzlers. The MES-instrument [10] provides an operationalization of the four higher-order factors into eleven lower-order factors. Self-belief, Value of School, and Learning Focus shape the adaptive, cognitive factors, as cognitive boosters. Planning, Task Management, and Persistence shape behavioral boosters. The mufflers, maladaptive cognitive factors are Anxiety, Failure Avoidance, and Uncertain Control, while Self-Sabotage and Disengagement are the maladaptive, behavioral factors or guzzlers. Cognitive factors are best interpreted as learning motivations, whereas the behavioral factors represent facets of learning engagement. In this study, we apply student scores administered in the first week of the course so that these survey-based engagement scores can be taken as antecedents of the trace-based engagement indicators.

Learning attitudes of our students were administered before the start of the course (T0). The attitude towards learning mathematics and statistics was assessed using the SATS instrument [17]. The instrument contained six quantitative methods-related learning attitudes: Affect, CognComp (cognitive competence), Value, Difficulty, Interest, and Effort.

A third survey measurement taking place halfway through the course (T5) refers to learning activity emotions. From the Achievement Emotions Questionnaire [18] measuring learning emotions, we selected four scales: Enjoyment, Anxiety, Boredom, and Hopelessness.

The last survey administered at the start of the course (T0) relates epistemic learning emotions. While achievement emotions arise from doing learning activities, like doing homework, epistemic emotions are related to cognitive aspects of the task itself. Prototypical epistemic emotions are curiosity and confusion. In this study, epistemic emotions were measured with the Epistemic Emotion Scales (EES: [19]), which was distributed at the start of the course. That instrument included the scales: Surprise, Curiosity, Confusion, Anxiety, Frustration, Enjoyment, and Boredom.

C. *Methods*

The primary statistical method we apply is of classifying approach type [11], creating learning profiles with k-means cluster analysis. Profiles were estimated on a broad range of educational measurements: trace variables Attempts, Solutions, and Hints, next to dispositional variables and prior knowledge indicators. Eight dispositional variables were selected, four of adaptive type and four of maladaptive type, known to be predictive of academic success from previous research [13-16]: Persistence and Task Management as adaptive behaviors of the MES instrument, Disengagement and Self-sabotage as two maladaptive behaviors of the MES, Anxious and Frustrated as two maladaptive epistemic emotions (EES) and Cognitive competence and Interest as two adaptive learning attitudes (SATS). This person-oriented modeling approach allowed us to profile students based on the combined trace, disposition, and prior knowledge data. The number of clusters was chosen based on several practical arguments: to have maximum variability in profiles without going into very small clusters and maintaining the interpretability of cluster solutions. We opted for a six-cluster solution, as solutions with higher dimensions did not strongly change the characteristics of the clusters but tended to split the smaller clusters into even smaller ones. As a next step in the analysis, differences between profiles were investigated with ANOVA, and prediction equations were generated with hierarchical regression models. All analyses were done using IBM SPSS statistical package.

III. Results

A. *Cluster-based learning profiles*

The interpretation of the final six-cluster solution is primarily based on differences in overall activity in the e-tutorial. Six variables describe that overall activity: the number of Attempts in each of the three learning phases and the number of Examples called in these three learning phases. The first cluster is labeled as the profile of 'Inactive' students: a relatively large group of students opt to study mainly outside the digital learning environment (or study at a minimal level). The largest cluster is labeled 'Low activity' profile, followed by two large 'High activity' profiles. These profiles differ in the timing of their learning activities: either concentrated in the second learning phase, preparing the quiz session ('High activity Quiz' profile), or more or less equally spread out over first and second learning phase ('High activity TutGr' profile). The small fifth cluster of students champions in activity levels, both in the first and the second learning phases: 'Extreme activity' profile. The last cluster is the only cluster not described by activity level but by prior knowledge/schooling. Students in the 'High prior knowledge' profile score highest on the diagnostic entry test and are the single profile with a majority of students educated at an advanced level in high school.

B. *Profile differences in overall activity levels*

Comparing overall levels (counts referring to the three learning phases are summed) of Attempts, Examples, and Hints for the six profiles is best made with Figure 1. Appreciating the large differences in total counts between Attempts, Examples, and Hints, visible from the vertical axes, the outstanding position of the Extreme activity profile is clear from the first two panels. However, not in the third panel: where Hints are concerned, the High prior knowledge profile takes the lead, and the profile of most active students is in the one-but-last position.

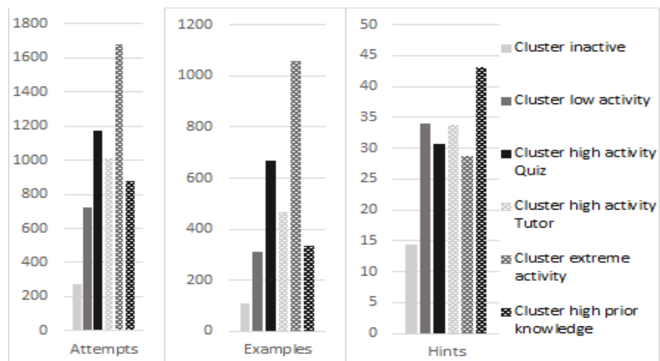

Fig. 1. Profile differences in the total number of Attempts, first panel, total number of Examples, second panel, total number of Hints, third panel.

C. *Profile differences in activity levels per learning phase*

Further differences between the several profiles are found when we disaggregate overall activity levels into levels of the three consecutive learning phases: preparing for the tutorial group session, quiz session, and exam. The timing of Attempts distinguishes the first three profiles from the last three. Students in the first three profiles concentrate on the second learning phase, preparing the quizzes: more than 70% of their Attempts fall in that phase. Their preparation in the first learning phase, directed at the tutorial session, occurs outside the e-tutorial or is absent at all: about 10% of the Attempts occur in that phase. The fourth and fifth profiles spread out preparation over the first and second phase: about 40% of Attempts in preparing the tutor session, about 50% in preparing the quiz session. The last profile of the students with high prior knowledge meets the learning pattern that best fits problem-based learning. Most of the preparation occurs in the first phase, so these students enter the tutorial session well prepared. The e-tutorial plays a minor role in preparing the final exam by practicing problem-solving: on average, less than 10% of all Attempts are falling in this phase for all profiles except the first.

The temporal pattern in Example calls is very different. In all profiles but the last one, most Example calls are positioned in the last learning phase, the preparation of the exam. Only students from the last three profiles use substantial amounts of Examples to prepare their quizzes. The profile of students having high prior knowledge is again that of the ideal students: they make extensive use of examples to prepare the first assessment they need to write, which frees them from further preparation for the final exam.

The call for Hints temporal pattern demonstrates that the second learning phase is the crucial one here. Except, once again, students of the profile of high prior knowledge: they use most of their Hints calls already in the first learning phase. Students of fourth and fifth profiles spread the use of Hints over the first two learning phases.

D. Profiles and survey measurements

Student dispositions, as measured by the several survey instruments, demonstrate differences between the several profiles. For most of these differences, statistical and practical significance have unequal sizes: nearly all cluster differences are statistically significant at levels beyond .01, but effect sizes are typically limited and mainly in the range between 2% and 5%.

The two profiles characterized by low activity in the e-tutorials start with very high levels of self-belief, much higher than the other profiles, even the profile of students with very high levels of prior knowledge, who might be expected to have the highest self-belief levels at the start of the course. The self-belief of the first two profiles drops strongly over the course, as does students' self-belief in the profile of late preparing students: see Figure 2. It suggests these three profiles have been overconfident at the start of the course, in the self-belief levels are becoming more realistic during the course. The single profile that shows a rise in self-belief, be it non-significant in size, is that of the highly active and timely preparing students: they start with modest self-belief levels and manage to leave them intact.

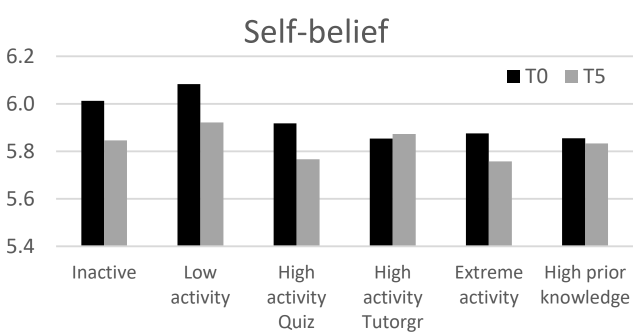

Fig. 2. Change in Self-belief from T0 to T5, by profile.

E. Profiles and course performance

Strong profile differences are found in course performance measures. Differences are not only statistically significant beyond .001, also effect sizes are modest in size: 9.3% for the score in the exam, 19.2% for the quiz scores. Prior knowledge is a crucial determinant of performance score, as is clear from the domination of the student profile, counting most students with high levels of prior schooling. Next in line is the profile of active and timely preparation. Low activity levels have a detrimental effect on quiz scores but much less on exam scores. The explanation of this is undoubtedly in the partial observation of the learning process: students with low activity levels in the digital learning environments might be very active in the other learning modes.

Course performance prediction equations can be classified along two dimensions: source of data used as predictor set, and time as the last moment included in the data set. Concerning the source of data, we distinguish trace data derived from the e-tutorial, survey data measuring student dispositions, assessment data of assessment for learning and learning type, and the full data set comprising all three types. Concerning the timing, we distinguish the seven weeks of education labeled as T1 to T7, together with T0 indicating the data available at the moment the course starts. We analyze cumulative data sets: for any time period, data measured in previous time periods are included. Figure 3 provides the longitudinal development of prediction power, expressed as a proportion of explained variation, of these four data constellations of the primary performance variable: score in the final exam.

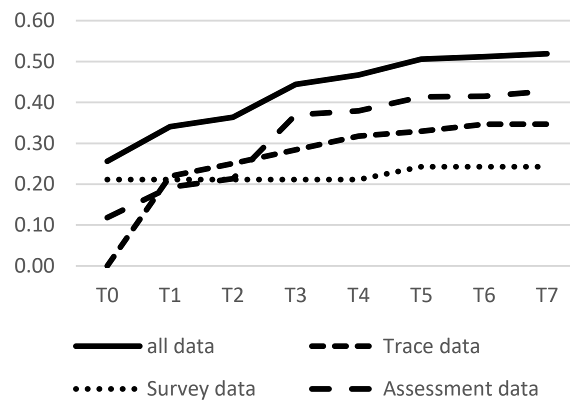

Fig. 3. Explained variation in final exam score, by three types of data, as a function of time.

The most stable factor is the prediction of performance by survey-based student dispositions. Learning attitudes, engagement, and motivation, together with epistemic emotions, all measured at the start of the course, explain 21% of the variation in performance. That percentage stays the same in the first four weeks, since no further survey is collected, to make a small jump to 24% explained variation in the fifth week when the second measurement of engagement and motivation and activity emotions occur. The collection of trace data only starts in the first week of the course, explaining that predictive power starts from zero. It jumps to 22% at the end of the first educational week, and grows gradually, adding new trace data every week, toward 35% at the end of the last educational week. A similar growth path is visible for predictions based on assessment data. There are four moments where assessment data become available: the diagnostic entry test (T0), and the three intermediate quizzes (T3, T5, T7). Added to that is the weekly assessment as learning data: the mastery data of the several educational weeks. The growth path of assessment data has two clear discontinuities: at the very start, where the diagnostic test taken at the start of the course explains 12% of the variation, and between the second and third week of education, where the first quiz data become available, leading to a jump of 16% in predictive power. The predictive power of all data sources together is distinctive from the predictive power of the three individual data sources. It tells that the different data sources are at least partially complementary.

IV. Discussion and Conclusions

An essential finding of this study is that a click is not a click. There are different indicators of engagement and what they tell tends to be different. We collected several kinds of click data, telling different messages: their relationships with course performance are different, giving rise to other learning profiles. Next to types of clicks having different impacts, the timing of clicks is of crucial importance. Using three demarcated learning phases, we demonstrated that the interpretation and impact of learning engagement indicators differ per learning phase. Learning activities undertaken in the first learning phase tend to have a much stronger positive effect on course performance than learning activities undertaken in later phases. This finding has major repercussions for learning feedback and interventions. If the measurement of learning engagement has the purpose of signaling inactivity to intervene, the question is if such intervention can ever be in time.

Timing does play a crucial role too in the predictive power of learning analytics applications. Figure 3 provides a concise summary of the dilemma of any such application: the longer you wait, the more powerful one's prediction, but the less time left for intervention. That dilemma is to be combined with the characteristics of the data sources. Student disposition data measured with surveys represent stable student characteristics. These won't change a lot throughout a course and miss the fluidity in learning activity data. For that reason, these survey data are critical at the very start of a course, when the other sources of data, trace and assessment data, are not very informative yet. In our data-rich environment, survey data shape the leading predictor at the very start of the course, shortly followed by trace data measuring learning activity. The dominant predictor is, however, assessment data collected for and as learning. But that dominant position is only acquired halfway to the course; maybe too late for effective intervention in an application of precision education.

Those interventions constitute another reason to continue applying survey data in precision education. Our profiling signals that students with low activity levels are at risk for failing the course. But they may be at risk for different reasons. Being too optimistic about one's proficiency level suggests being one reason for low activity: these students judge that they do not need to practice to pass the course. The very high levels of self-belief at the start of the course, and the substantial drop in self-belief in the two clusters of low activity, is the narrative of that mechanism. At the same time, some other students opt out of the digital learning mode and focus on different learning modes, entirely in line with the principles of student-centered learning within a blended learning context. We cannot tell from activity data alone what type of student is showing low activity levels. Yet, it is a crucial difference in terms of learning feedback and intervention. The only way to find out in a successful application of precision education is by using other types of data, such as survey data containing self-belief measures.

References

[1] S. J. H. Yang 2021. Precision Education - A New Challenge for AI in Education. Educational Technology & Society, 24 (1), 105–108.

[2] S. L. Christenson, A. L., Reschly, and C. Wylie (eds) 2006. Handbook of Research on Student Engagement. NY, NY: Springer.

[3] L.-C. Lin, I.-C.Hung, N.-S. Chen (2019). The impact of student engagement on learning outcomes in a cyber-fipped course. Education Tech Research Dev 67:1573–1591 DOI: 10.1007/s11423-019-09698-9

[4] Educational Psychologist, 2015: 1.

[5] Nian-Shing Chen, Chengjiu Yin, Pedro Isaias & Joseph Psotka (2020). Educational big data: extracting meaning from data for smart education, Interactive Learning Environments, 28:2, 142-147, DOI: 10.1080/10494820.2019.1635395

[6] D. Tempelaar, B. Rienties, & Nguyen, Q. 2021. The Contribution of Dispositional Learning Analytics to Precision Education. Educational Technology & Society, 24 (1), 109-122.

[7] S. Buckingham Shum and R. Deakin Crick. 2012. Learning Dispositions and Transferable Competencies: Pedagogy, Modelling and Learning Analytics. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, S. Buckingham Shum, D. Gasevic, and R. Ferguson (Eds.). ACM, New York, NY, USA. DOI: 10.1145/2330601.233062992-101

[8] D. Tempelaar, B. Rienties, and Q. Nguyen 2017. Towards Actionable Learning Analytics Using Dispositions. IEEE Transactions on Learning Technologies, 10(1), 6-16.

[9] B. Rienties, S. Cross, and Z. Zdrahal 2017. Implementing a Learning Analytics Intervention and Evaluation Framework: What Works? In Big data and learning analytics: Current theory and practice in higher education, B. Kei Daniel (Ed.). Springer, Cham, 147–166.

[10] A. J. Martin 2007. Examining a multidimensional model of student motivation and engagement using a construct validation approach. British Journal of Educational Psychology, 77(2), 413-440.

[11] B. M. McLaren, T. van Gog, C. Ganoe, M. Karabinos, and D. Yaron 2016. The efficiency of worked examples compared to erroneous examples, tutored problem solving, and problem solving in computer-based learning environments. Computers in Human Behavior, 55, 87–99.

[12] Anna Y. Q. Huang, Owen H. T. Lu, Jeff C. H. Huang, C. J. Yin & Stephen J. H. Yang (2020) Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs, Interactive Learning Environments, 28:2, 206-230, DOI: 10.1080/10494820.2019.1636086

[13] Q. Nguyen, D. T. Tempelaar, B. Rienties, and B. Giesbers 2016. What learning analytics based prediction models tell us about feedback preferences of students. Quarterly Review of Distance Education. 17(3).

[14] B. Rienties, D. Tempelaar, Q. Nguyen, and A. Littlejohn 2019. Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. Computers in Human Behavior, DOI: 10.1016/j.chb.2019.07.007

[15] D. T. Tempelaar, B. Rienties, and B. Giesbers 2015. In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. Computers in Human Behavior, 47, 157-167.

[16] D. T. Tempelaar, B. Rienties, J. Mittelmeier, and Q. Nguyen 2018. Student profiling in a dispositional learning analytics application using formative assessment. Computers in Human Behavior, 78, 408-420.

[17] D. T. Tempelaar, W.H. Gijselaers, S. Schim van der Loeff, and J.F.H. Nijhuis 2007. A structural equation model analyzing the relationship of student achievement motivations and personality factors in a range of academic subject-matter areas, Contemporary Educational Psychology, 32(1), 105-131. DOI: 10.1016/j.cedpsych.2006.10.004.

[18] Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. 2011. Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). Contemporary educational psychology, 36(1), 36-48. DOI: 10.1016/j.cedpsych.2010.10.002

[19] Pekrun, R., Vogl, E., Muis, K.R. and Sinatra, G.M. 2017. Measuring emotions during epistemic activities: the Epistemically-Related Emotion Scales, Cognition and Emotion, 31(6), 1268-1276, DOI: 10.1080/02699931.2016.120