

# Chapter 9

## Learning Analytics and the Measurement of Learning Engagement



Dirk Tempelaar, Quan Nguyen, and Bart Rienties

### 9.1 Introduction

The topic of student engagement is of crucial importance because of its close connection to self-regulated learning, the condition sine qua non for all learning, and learning in technology-enhanced learning environments in specific (Ifenthaler, Gibson, & Zheng, 2018a, 2018b). Nevertheless, beyond a general agreement on the importance of the construct, *engagement could be described as the holy grail of learning*, (Sinatra, Heddy, & Lombardi, 2015, 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 (Martin, 2007). This instrument distinguishes cognitive or motivational and behavioural or engagement facets, and within each category, adaptive and maladaptive facets. 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., Azevedo, 2015; Ifenthaler et al., 2018a, 2018b). Some proponents of the data analytics tradition base the choice for engagement measures generated by logs on a total denial of the validity of survey type of data. However, more in general, one can observe that empirical studies in

---

D. Tempelaar (✉)

Maastricht University School of Business and Economics, Maastricht, Netherlands

e-mail: [d.tempelaar@maastrichtuniversity.nl](mailto:d.tempelaar@maastrichtuniversity.nl)

Q. Nguyen

School of Information University of Michigan, Ann Arbor, MI, USA

B. Rienties

Open University UK, Institute of Educational Technology, Milton Keynes, UK

© Springer Nature Switzerland AG 2020

D. Ifenthaler, D. Gibson (eds.), *Adoption of Data Analytics in Higher Education Learning and Teaching*, Advances in Analytics for Learning and Teaching, [https://doi.org/10.1007/978-3-030-47392-1\\_9](https://doi.org/10.1007/978-3-030-47392-1_9)

learning engagement are typically based on survey data, or log data, but nearly never attempting to integrate both approaches (Christenson, Reschly, & Wylie, 2006).

The aim of this chapter is to provide such ‘multi-modal data’-based contribution to the research of student engagement in learning. In this study, not only quantitative aspects of engagement are investigated in terms of measured or self-reported intensity of learning activities but also qualitative aspects of engagement. For example, learners make conscious choices of what type of learning activity to engage, such as using un-tutored and tutored problem-solving as well as worked examples (Aleven, McLaren, & Koedinger, 2006; Aleven, McLaren, Roll, & Koedinger, 2004; Aleven, Roll, McLaren, & Koedinger, 2016; McLaren, van Gog, Ganoë, Karabinos, & Yaron, 2016). This line of research investigates learning behaviours and student’s 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, along the lines of an experimental design with the different instructional formats as different treatments, in search for differences in efficiency and effectivity of learning. The introduction of learning analytics (Ifenthaler, 2015; Ifenthaler, Yau, & Mah, 2019), 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. It made it possible to move from the lab to authentic educational setting, to move from the experimental design to observational settings, investigating individual differences in preferences for feedback formats rather than the efficiency or effectivity of them. This development led to a convergence of learning analytics-based studies in the use of feedback by students, such as Ifenthaler (2012), and instructional design-based research, such as Aleven et al. (2004, 2006, 2016) and McLaren et al. (2016). Our current study is aligned with this development, adding an extra dimension to the research of student’s preferences: the temporal dimension (Rienties, Cross, & Zdrahal, 2017). Our study builds on previous research by the authors (Nguyen, Tempelaar, Rienties, & Giesbers, 2016; Rienties, Tempelaar, Nguyen, & Littlejohn, 2019; Tempelaar, Rienties, & Giesbers, 2015; Tempelaar, Rienties, Mittelmeier, & Nguyen, 2018; Tempelaar, Rienties, & Nguyen, 2017; Tempelaar, Rienties, & Nguyen, 2018) that focused on the issue of early prediction of drop-out or low performance.

## 9.2 This Study

The integration of the two approaches of operationalizing learning engagement, the survey approach and the data analytics approach, is the primary goal of this empirical study. The integration of both approaches is enabled by the dispositional learning analytics context of the course we investigate. The instructional format is that of blended or hybrid learning, which generates a rich set of log variables that are indicators of learning engagement. Examples of such indicators are overall student activity in the digital learning tool as 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 and the number of hints called for or, very specific to this context, the number of finished packages. Problems are offered to students in the format of small sets of related problems, called a package. A finished package is when a student studies all problems of such a set in one run. All of the 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 that measure learning dispositions of students, both at the start of the course and during the course.

### 9.2.1 Context

This study takes place in a large-scale introduction course in mathematics and statistics for first-year students of a business administration and economics program in the Netherlands. The educational system can best be described as ‘blended’ or ‘hybrid’. The most important 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-centred 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 the e-tutorials or other learning materials, the class time is used to discuss how to solve advanced problems. The educational format, therefore, has most of the characteristics of the flipped-classroom design in common. The intensive use of the e-tutorials and achievement of good scores in the e-tutorial practice modes is encouraged by giving performance bonus points in quizzes that are taken every 2 weeks and consist of items drawn from the same item pools that are used in the practice mode. This approach was chosen to encourage students with limited prior knowledge to make intensive use of the e-tutorials.

In the use of the e-tutorials, three different learning phases can be distinguished. In Phase 1, students prepare for the next tutorial session. Knowing that they will face the discussion of ‘advanced’ maths problems in that tutorial session, students are expected to prepare by self-study outside class, e.g., by studying the literature together with some peers, or practising in the e-tutorials. Phase 1 was not formally assessed, other than that such preparation allowed students to actively participate in the discussion of the problem tasks in the tutorial session. Phase 2 was the preparation of the quiz session, one or 2 weeks after the respective tutorial. The three quizzes were taken every 2 weeks in ‘controlled’ computer labs and consisted of test items that were drawn from the same item pools applied in the practising mode. Although the assessment through quizzes was primarily for formative purposes, students can score a bonus point in each quiz that is added to their written exam score. Phase 3 consisted of the preparation of the final exam, at the end of the course. The written exam was a multiple-choice test of 20 questions on mathematics,

**Table 9.1** The three learning phases: preparing the tutorial session as Phase 1 (light grey), preparing the quiz session as Phase 2 (grey), and preparing the exam as Phase 3 (dark grey)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Topic Week 1	Phase 1	Phase 2	Phase 2	Phase 3	Phase 3	Phase 3	Phase 3	Phase 3
Topic Week 2		Phase 1	Phase 2	Phase 3	Phase 3	Phase 3	Phase 3	Phase 3
Topic Week 3			Phase 1	Phase 2	Phase 2	Phase 3	Phase 3	Phase 3
Topic Week 4				Phase 1	Phase 2	Phase 3	Phase 3	Phase 3
Topic Week 5					Phase 1	Phase 2	Phase 2	Phase 3
Topic Week 6						Phase 1	Phase 2	Phase 3
Topic Week 7							Phase 1	Phase 3

as well as 20 questions on statistics. These questions could be practised using textbook materials and e-tutorial modes. The final exam is mostly summative of nature and has by far the largest share in the course score (86%). Students' timing decisions, therefore, are related to the amount of preparation in each of the three consecutive phases and are summarized in Table 9.1.

The subject of this study is the full cohort of students 2018/2019 (1072 students). The diversity of the student population was large: only 21% of the student population was educated in the Dutch secondary school system, compared to 79% educated in foreign systems, with 50 nationalities. A large part of the students had European nationality, with only 4.0% of the students from outside Europe. Secondary education systems in Europe differ widely, particularly in the fields of mathematics and statistics. 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 log data.

## 9.2.2 Instrument and Procedure

Both e-tutorial systems SOWISO and MSL follow a test-driven learning and practice approach. Each step in the learning process is initiated by a problem and students are encouraged to (try to) solve each problem. If a student has not (fully) 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. The alternative feedback strategies that students can choose for are:

- Check: the unstructured problem-solving approach, which only provides correctness feedback after solving a problem

- Hint: the tutored problem-solving approach, with feedback and tips to help the student with the different problem-solving steps
- Solution: the worked examples approach
- Theory: asking for a short explanation of the mathematical principle

Our study combines log data from the SOWISO e-tutorial with self-report data that measure learning dispositions, and course performance data. Azevedo (2015) distinguishes between log data of product type and process type, where click data is part of the process data category. In this study, we will focus on process data only, such as the clicks to initiate the learning support mentioned above of Check, Hint, Solution and Theory, since those represent the engagement of students with learning in the e-tutorial. SOWISO reporting options for log data are very broad, which requires making selections from the data. All dynamic log data were assigned to the three consecutive learning phases in line with the scheme depicted in Table 9.1, next aggregated over time, to arrive at static, full course period accounts of log data. For all three learning phases, six log variables were selected:

- #Attempts: the total number of attempts at individual exercises
- #Examples: the number of worked examples called
- #Hints: the number of hints called
- #Views: the number of theory pages in which a mathematical principle is explained, called
- #Packages: the number of sets of related exercises that all correspond to one mathematical principle a student finishes
- TimeOnTask: total time on task in problem-solving

Survey-based engagement indicators are taken from the MES-instrument, derived from ‘Motivation and Engagement Wheel’ framework by Martin (2007). Martin breaks down learning cognitions and learning behaviours into four categories of adaptive versus maladaptive types and cognitive versus behavioural types. The classification is based on the theory that thoughts and behaviours can either enable learning and act as boosters or hinder learning by acting as mufflers and guzzlers. The instrument Motivation and Engagement Wheel (Martin, 2007) provides an operationalization of the four higher-order factors into 11 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 the behavioural boosters. The mufflers, maladaptive cognitive factors are Anxiety, Failure Avoidance, and Uncertain Control, while Self-Sabotage and Disengagement are the maladaptive, behavioural factors or guzzlers. Cognitive factors are best interpreted as learning motivations, whereas the behavioural 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 log-based engagement indicators.

### 9.2.3 *Data Analysis*

Given the purpose of connecting the data analysis with student feedback and interventions, we opt for person-centred methods rather than variable-centred methods in the data analysis phase. Person-centred methods result in profiles of students demonstrating similar learning behaviours. These profiles are constructed by two-step clustering. The subsequent step in the analysis is to investigate profile differences with regard to the antecedents of these profiles, the student learning dispositions, and with regard to the consequences of these profiles, the learning outcomes. Inputs for the clustering step are all learning engagement indicators of log type: the number of Attempts, Examples, Hints, Views, and Packages plus TimeOnTask to prepare the tutorial sessions, to prepare the quiz sessions, and to prepare the final exam, in total 18 engagement indicators. As a next step in the analysis, differences between profiles were investigated with ANOVA, and prediction equations were estimated with hierarchical regression models. In the derivation of these prediction models, special attention was given to the issue of collinearity, also coined as multicollinearity. Collinearity arises when predictors in a regression model are correlated, what is typically the case in many learning analytics applications, where prediction models are estimated with learning logs as predictor variables. As a result of collinearity, regression coefficients are not stable but can take surprising values, with large standard errors. When collinearity is strong, a rule of thumb being the variance inflation factor exceeding the value of five, the model needs to be adapted, e.g., by eliminating one of the highly correlated predictor variables. Ethics approval for this study was achieved by the Ethical Review Committee Inner City faculties (ERCIC) of the Maastricht University, as file ERCIC\_044\_14\_07\_2017.

## 9.3 Results

### 9.3.1 *Descriptive Statistics of Survey-Based Measures*

Survey-based measures of engagement that follow the ‘Motivation and Engagement Wheel’ framework are administered with a Likert 1...7 scale having the value four as the neutral anchor. Descriptive statistics are provided in Table 9.2.

Mean scores of adaptive cognitions and behaviours are all beyond the neutral score. Most scores are quite high, with the exception of Planning: students perceive their proficiency in planning their study at a rather modest level. Maladaptive cognitions score, with one exception, below the neutral score. That exception is Anxiety: students express high levels of anxiety, relative to the other maladaptive constructs.

Standard deviations are low for variables with extreme scores, both in the high end of the scale (the adaptive constructs) and the low end of the scale (Disengagement),

**Table 9.2** Descriptive statistics of engagement measures from the ‘Motivation and Engagement Wheel’ framework

Variable	Scale	Mean	Standard deviation	Cronbach alpha
Self-belief	Adaptive cognitions	5.98	0.74	0.78
Valuing school		6.03	0.63	0.64
Learning focus		6.34	0.60	0.74
Planning	Adaptive behaviours	4.84	1.06	0.78
Task management		5.62	0.94	0.76
Persistence		5.58	0.79	0.77
Anxiety	Maladaptive cognitions	4.63	1.29	0.84
Failure avoidance		2.49	1.27	0.84
Uncertain control		3.42	1.17	0.80
Self-sabotage	Maladaptive behaviours	2.18	1.02	0.80
Disengagement		1.73	0.73	0.65

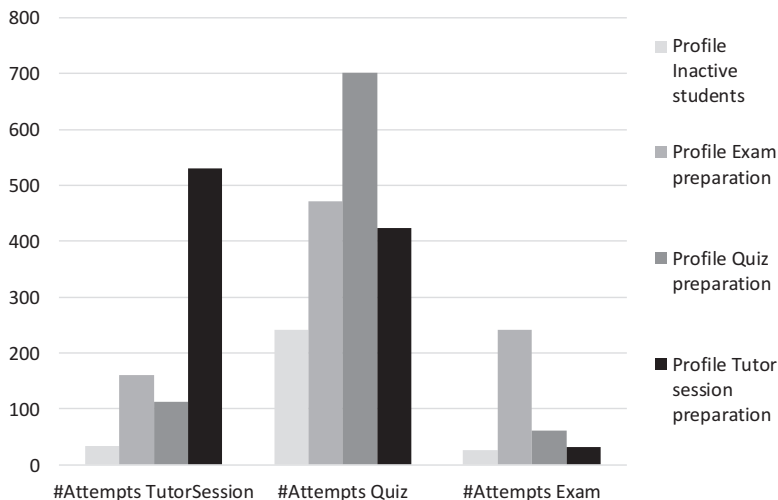
with higher standard deviations found in variables ending up in the middle of the scale.

Reliability scores range from satisfactory to good, with two exceptions: those for Valuing School and Disengagement are weaker.

### 9.3.2 Cluster-Based Learning Profiles

The cluster analysis results in four different learning profiles, similar to previous research when applying cluster analysis to longitudinal log data (Rienties et al., 2019). The temporal aspect of the log data contributes strongly to distinguishing the four profiles, much stronger than the aspect of different instructional formats. That is students of different profiles first and for all concentrate on different learning phases. The labelling of the clusters we have opted for is based on these temporal aspects of learning processes:

- Profile Inactive students: The 257 students in this cluster demonstrate low engagement levels in the e-tutorial. These students ‘opt-out’ with regard to the digital learning environment and prepare themselves in different ways, or not at all. The few learning activities in the digital mode are mostly in the second learning phase, the preparation of the quizzes.
- Profile Exam preparation: This smallest cluster counting 69 students prepares in both the second and third learning phases. As the next cluster, their preparations in the digital mode are primarily assessment based.
- Profile Quiz preparation: The largest cluster with 468 students shares with the previous profile that preparations are directed at assessments but differs in timing: they focus completely on learning in the second phase, preparing the quiz sessions.



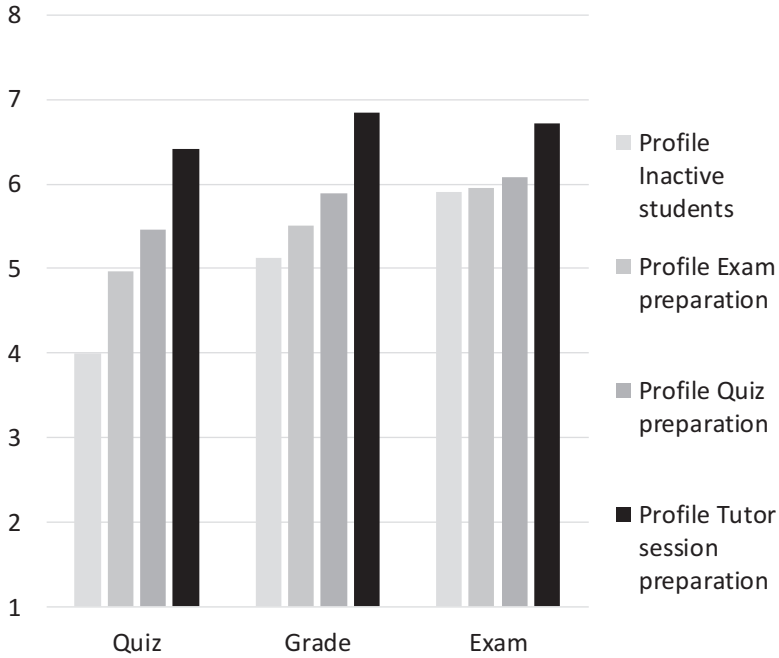
**Fig. 9.1** Number of Attempts, for each of the three learning phases, and all four learning profiles

- **Profile Tutor session preparation:** These 315 students are the ‘ideal’ students in a PBL-based curriculum: they seriously prepare the tutorial sessions by learning and practising in the e-tutorial and finish their preparations in the second learning phase, rehearsing to prepare the quizzes. They seem not to need any further preparation in the third learning phase.

Figure 9.1 describes the differences between these four learning profiles graphically by means of the distribution of the number of Attempts over the three learning phases, for each cluster. Other engagement indicators, as #Examples or TimeOnTask, generate very similar patterns, due to the collinearity of engagement indicators.

Figure 9.1 makes clear that most students postpone learning until after the tutorial session. It is only the approach of an assessment, first the quiz and later the final examination, that creates sufficient stimulus to do most of the learning for students in the first three clusters. Most of their learning takes place in the second learning phase and is finished in the third learning phase. The exception to this pattern of postponing the learning process is found in the last cluster, labelled as the profile directed at the preparation of the tutorial session. Most of their learning takes place in the first phase, and learning is finished in the second phase, leaving little to study in the preparation of the final examination.





**Fig. 9.2** Means of course performance indicators Quiz, Grade and Exam, standardized to the grading range 1...10, for the four different cluster-based learning profiles

### 9.3.3 Learning Profiles and Course Performance

The relevance of engagement indicators and student profiles based on these engagement indicators is in the relationship with course performance variables. Figure 9.2 provides an impression of that relationship. There is indeed a consistent relationship between profiles, ordered from less to more adaptive learning behaviours, and course performance, where all course performance variables are re-expressed as school grades (1...10). Differences between profiles are even larger when performance is expressed in a pass or fail, because the typical passing benchmark is at 5.5. Effect sizes of profile differences calculated by ANOVA analyses are 18.7%, 9.5%, and 4.0% for Quiz, Grade, and Exam, respectively.

All three ANOVA analyses are statistically significant with significance levels below 0.001. Post-hoc analyses indicate that differences in mean quiz scores are statistically significant for all four clusters, whereas statistical significant differences in grades and exam scores refer to the differences between the fourth cluster of students with the profile of preparing the tutor session, and the three other clusters.

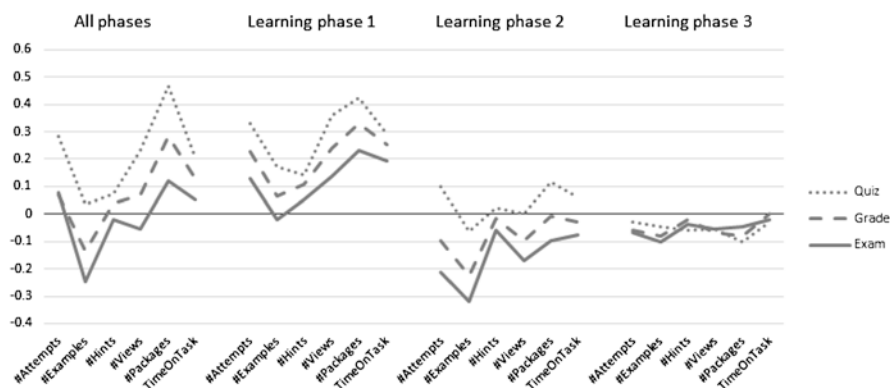
### 9.3.4 Bivariate Relationships Between Engagement Indicators and Course Performance

Although the several engagement indicators are collinear, bivariate relationships with course performance variables demonstrate characteristic differences in patterns (see Fig. 9.3). All correlations in Fig. 9.3 equal to 0.1 or larger are statistically significant at significance levels of 0.001; correlations of absolute size of 0.075 or larger are statistically significant at significance levels of .01 and correlations of absolute size of 0.060 or larger are statistically significant at significance levels of 0.05. Taking the strict benchmark of the 0.001 significance level implies that correlations in the first three panels are mostly significant, but not those in the last panel of Fig. 9.3.

First: The timing plays a crucial role in those relationships. Engagement indicators referring to learning in the first phase are all positive, indicating that higher levels of engagement correspond on average with higher performance levels. However, bivariate relationships referring to the second learning phase become negative, or approximately zero, for performance categories Grade and Exam: only Quiz performance is positively related to some of the engagement indicators. That trend continues into the third learning phase: all bivariate correlations are negative and small in size.

Second: Quiz performance is more positively related to performance indicators than the other performance categories, and final Grade is more positively related to performance indicators than Exam score for mathematics.

Third: Highest correlations are found for the engagement indicator of finished Packages, much higher than the indicators based on the number of clicks (such as problem-solving Attempts started, the number of Examples studied) or Time on task.



**Fig. 9.3** Correlations of engagement indicators #Attempts, #Examples, #Hints, #Views, #Packages, and TimeOnTask with performance indicators Quiz, Grade, and Exam, for the full course and separate learning phases

### 9.3.5 *Multivariate Relationships Between Engagement Indicators and Course Performance*

In the multivariate relationships explaining the two course performance measures from the set of traced engagement indicators, we find strong collinearity caused by #Attempts and #Examples being collinear. To diminish collinearity and arrive at variance-inflation-factors below five for all predictor variables, #Examples is removed from all hierarchical regression relationships. What remains is weak collinearity, visible from the negative signs of several of the regression coefficients, knowing that most bivariate relationships between engagement indicators and course performance variables are positive (as discussed in the previous section). See Table 9.3 for the regressions predicting Exam score and Table 9.4 for the regressions predicting Quiz score. In Table 9.3, the only predictor variable with a consistent positive regression coefficient is the number of Packages finished by the student. The higher the number of finished packages, the higher the expected exam scores. The other main predictor is the number of Attempts, always with a negative beta.

Negative betas are caused by collinearity of #Attempts and #Packages and need to be interpreted as: for a given number of finished packages, students who need more attempts to finish those packages are expected to score less well in exam, on average. A similar relationship regards the number of Views: students who use more views to finish a certain number of packages are expected to score less well in the exam, on average.

The pattern of Table 9.3 is repeated in Table 9.4. Again, the number of Packages students finish is the dominant predictor in explaining Quiz score. Collinearity amongst the five log-based engagement constructs (more Attempts go with more time-on-task with more Hints and more Views and lead to more finished Packages) together with the dominant role of #Packages variable makes the other engagement variables become non-significant, or significant but with a negative beta: if you need more Attempts to reach a certain level of #Packages, it decreases the expected Quiz score.

**Table 9.3** Hierarchical regression equations explaining Exam score from log-type engagement indicators, for the full sample and each of the four cluster-based profiles: betas (standardized regression coefficients) and explained variation

Regression betas exam score	Full sample	Profile inactive students	Profile exam preparation	Profile quiz preparation	Profile tutor session preparation
#Attempts	-0.537***	-0.582***	-0.166	-0.529***	-0.462***
#Hints	-0.037	-0.103	-0.007	-0.060	-0.043
#Views	-0.109**	0.023	-0.014	-0.119**	-0.110
#Packages	0.637***	0.715***	0.186	0.375***	0.274***
TimeOnTask	-0.002	0.038	-0.089	0.021	-0.061
R <sup>2</sup>	0.126	0.099	0.025	0.163	0.195

Note: \*\*\*:  $p < 0.001$ ; \*\*:  $p < 0.01$

**Table 9.4** Hierarchical regression equations explaining Quiz score from log-type engagement indicators, for the full sample and each of the four cluster-based profiles: betas (standardized regression coefficients) and explained variation

Regression betas quiz score	Full sample	Profile inactive students	Profile exam preparation	Profile quiz preparation	Profile tutor session preparation
#Attempts	-0.345***	-0.335***	0.090	-0.418***	-0.386***
#Hints	-0.015	-0.048	-0.010	-0.037	-0.035
#Views	-0.027	0.006	0.006	-0.038	0.019
#Packages	0.768***	0.701***	0.376*	0.536***	0.299***
TimeOnTask	0.015	0.070	-0.001	0.057	-0.095
R <sup>2</sup>	0.258	0.193	0.196	0.145	0.121

Note: \*\*\*:  $p < 0.001$ ; \*\*:  $p < 0.01$

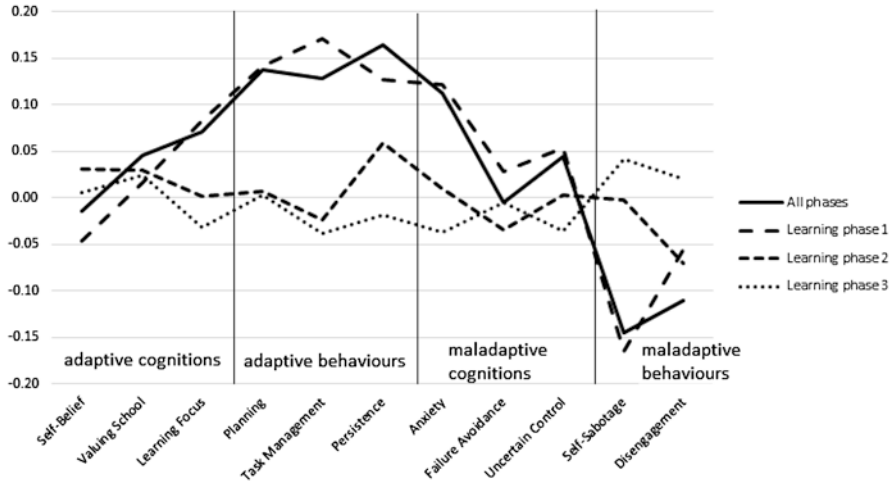
When we compare the two tables, we find that performance in the Quiz is better predicted than performance in the Exam. Since quizzes are administered in the e-tutorials and quiz questions are similar to problems students practice with, this is no coincidence. However, there is an exception to this rule, what can be seen by comparing columns in the two tables. That exception regards the profile of students who focus on the first learning phase, preparing the tutorial sessions. In this profile, engagement indicators predict exam performance better than they do for quiz performance. The relationships in the profile of the student who focuses on exam preparation may differ from the relationships in other profiles, but the evaluation is slightly more difficult, due to the small sample size of this cluster.

In all clusters, both #Views and TimeOnTask are statistically insignificant in the prediction of exam and quiz performance. In all cases, we investigated the change in the prediction equations would the main predictor, #Packages, not be incorporated in the regression equations. Without reporting these outcomes, the pattern that emerges is that #Attempts becomes the main predictor, with positive betas in the several regressions, and that TimeOnTask is the secondary predictor with negative betas. Giving rise to the interpretation that for a given number of attempts, students who need more time-on-task to do these attempts are expected to score less well in exam and quiz.

### 9.3.6 *Bivariate Relationships Between Survey-Based Engagement Scores and Log-Based Engagement Indicator*

As a last step in the analysis, the relationships between the main log-based engagement indicator, #Packages, and the survey-based engagement scores were investigated. We express these relationships again as bivariate correlations (see Fig. 9.4).

The first observation from Fig. 9.4 is the dominant role of learning engagement factors: both the adaptive (Planning, Task Management, and Persistence) and mal-



**Fig. 9.4** Correlations of engagement indicator #Packages with motivation & engagement survey scores, for the full course and separate learning phases

adaptive behaviours (Self-Sabotage and Disengagement) are all statistically significant related to #Packages (all correlations larger than 0.10 in absolute size are statistically significant at 0.001 significance level), whereas the motivational variables are not, with one exception: Anxiety.

The second observation is that the maladaptive cognitions are not maladaptive in the sense that they are positively related to #Packages as a measure of learning engagement. Failure avoidance, Uncertain control and especially Anxiety, although acting as mufflers to learning in general, tend to increase learning activity in the digital learning environment.

The third observation is that the pattern of correlations is very different for the first learning phase and the second and third learning phases. In fact, measured learning activity in second and third learning phases is unrelated to any of the engagement and motivation scores (with the single exception of activity in the second learning phase being marginally significantly related to Disengagement).

### 9.4 Findings and Discussion

From a methodological perspective, our main finding emphasized the issue heterogeneity in engagement measures, in many different respects. There are different indicators of engagement and the story they tell tends to be different. In this study, we collected several kinds of click data, next to time on task data and engagement data rather unique to this study: the number of finished packages, or complete runs through a problem set. One of the main findings of this study is that basic measures

of engagement as clicks and time are dominated in predictive power by this more complex measure of engagement. And in a multivariate context, these clicks and time-related engagement indicators get a reversed interpretation: relative to the number of packages finished by a student, taking more time, or making more attempts, has a negative impact on expected performance levels. The lesson we learned from this is that learning engagement does not have a unique and straightforward operationalization. Different contexts may demand different operationalizations and require investigations to find out what suits best.

Another source of heterogeneity is the timing of learning efforts. Profiting from the existence of three clearly 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, that of preparing the tutorial session, 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 to signal inactivity in order to intervene, the question is if such intervention can ever be in time. In our context, the first moment to find out if a student fell short in the preparation of the tutorial session is at the start of the second learning phase. That in itself leaves the student ample time to catch up, unless learning activities in phases two and three appear to be consistently less effective than those in learning phase one. (Note: it is dangerous to extrapolate data from this study, since the effectivity of learning in later phases may be impacted by doing an intervention that is not in place when we collected the current data set).

Differentiating the timing of learning over these three learning phases appeared being a crucial facet of the four learning profiles: different types of learners prepare in different ways with different temporal patterns. Moreover, and of crucial importance in the context of this study, different engagement indicators are relevant to these different profiles.

Profiles are predictive for course performance, with profiles of more and more timely engagement achieving higher levels of performance. Largest effects are for quiz scores, due to the circumstance that quiz questions are generated from the same item pools students work with in the practice mode of the e-tutorials, and in line with general findings that engagement better predicts low-level tests than high-level tests (Sinatra et al., 2015). Highly engaged students who practised many problem sets have a cognitive advantage over less well-prepared students. Remarkably, the next highest effect size is found in the course grade, rather than the mathematics exam score. The course grade is a weighted mean of quiz and exam scores of both mathematics and statistics. Where engagement indicators summarizing learning activities of mathematical content will represent both cognitive and behavioural aspects, those same indicators will not signal the knowledge of statistical concepts. The effect on course grade being stronger than the effect on exam score thus indicates that the behavioural aspect is not limited to learning mathematics only but extends to the learning of other topics.

Diversity by learning phases is not restricted to the consequences of learning in different phases, as addressed above. Diversity also refers to the antecedents of

learning activity in the several learning phases. All of the engagement factors from the motivation and engagement wheel framework are related to measured engagement in phase one, all along expected directions: the booster behaviours are positively related to the number of finished packages; the guzzlers or maladaptive behaviours are negatively related to the number of finished packages. However, learning in phases two and three is, with one exception, unrelated to any of the dispositional measures of engagement.

Next to heterogeneity, another crucial concept in the analysis of engagement data is collinearity. We found strong collinearity in our set of traced engagement scores and corrected for that by leaving out one of the engagement variables from multivariate modelling. The resulting data is still containing weak collinearity, visible from the differences between multivariate and bivariate relationships. In our context, we find that the number of attempts and time on task are negatively related or unrelated to performance indicators, rather than positively related.

From a theoretical perspective, our findings highlighted the relationship of behavioural trace data with the antecedents of measured engagement: the engagement dispositions. If the outcomes of predictive modelling suggest that some at-risk students would profit from becoming more engaged, it is a poor intervention to tell those students to spend more time-on-task, try more attempts or finish more packages. Such interventions are tackling the symptoms rather than the causes of low engagement. The causes of low engagement might be found in the learning dispositions students bring to class based on previous learning experiences. Examples of learning dispositions associated with engagement as measured in the learning platform are low levels of booster behaviours, such as Planning, Task Management and Persistence, and high levels of guzzlers, the maladaptive behaviours as Self-Sabotage and Disengagement. One can imagine designing learning interventions that address these dispositions. But even if these interventions turn out to be productive in changing learning behaviours in the adaptive direction, they will not be very helpful if, as in this study, learning dispositions have little effect on learning engagement in later phases than phase one.

From a practical perspective, the ultimate aim of all learning analytics applications is intervention. We collect data in order to make predictions, e.g., about which students are at risk and why. However, these predictions are not the aim in themselves. We make these predictions in order to intervene: provide learning feedback to the student at risk, hoping that the student will be able to adapt the learning, or to change the instructional context with the purpose to improve learning. But these interventions cannot be any better than the quality of the prediction models they are based on. Traditionally, many learning analytics applications apply the number of clicks and/or time on task as measures of learning engagement to predict course performance or risk of dropout. Clicks and time on task are easy to generate, and in many digital learning environments still the only types of log data available, but may not be the best predictors of course performance. It is only in a data-rich context as provided in our context, or in Ifenthaler et al. (2018a, 2018b), that one can sort out the relative importance of log-based engagement indicators and find out if some may even have a reversed effect on performance indicators. Stimulating stu-

dents to try more attempts or to spend more time on task would constitute an inferior intervention when it is the number of finished packages rather than the number of problems attempted being the main predictor of course performance (this study), or when not time on task but the number of launched tasks is the dominant predictor (Ifenthaler et al. studies).

However, even in the case of a rich set of traced engagement indicators, allowing selecting the dominant predictors of course performance and estimating the multivariate relationships between course performance indicators and measured engagement factors as their antecedents, as in this study or the Ifenthaler et al. studies, there is no guarantee for arriving at adequate prediction models. The first issue at stake is that of collinearity: rich sets of measured engagement indicators demonstrate collinearity by default and very few empirical studies in the learning analytics area investigate the presence of collinearity. Collinearity expresses itself in regression coefficients taking surprising values, both in sign and size, and in large standard errors of the coefficients. Since the choice of intervention is typically based on what variables act as dominant predictors in course performance prediction equations, collinearity may be one cause of choosing a suboptimal format of intervention. In order to prevent collinearity dimension reduction can be utilized or obtain an interaction score from this variable or metrics.

The other obstacle to successful intervention investigated in this study has been labelled as heterogeneity or diversity. Having access to time-stamped engagement data in a learning context where three different learning phases can be distinguished, we were able to investigate both the consequences of learning engagement, in terms of course performance, and the antecedents of learning engagement, in terms of learning dispositions from the motivation and engagement wheel framework. In short, we concluded that learning engagement in the early phase of learning is predictive of course performance, but not learning that takes place in later phases. And we concluded that learning in that first phase is related to engagement dispositions, but not the learning in later phases. These conclusions have a major impact on the perspectives of learning interventions based on learning analytics generated feedback. Our early learning phase lasts for only 1 week; after that week, students enter the second learning phase. But it takes time to find out that a student lacks engagement in this first learning phase and to design an intervention in order to stimulate the student to become more engaged. In our context, that intervention would impact learning in the second phase, at the earliest, and learning in the third learning phase. However, the relationships between engagement and performance in later learning phases than the first one differ substantially and are in fact absent. So if the intervention is not that powerful that it also changes the relationship between engagement in later learning phases and course performance, there is little perspective in pushing students to become more engaged learners.



## 9.5 Conclusion

In conclusion, this study investigates how behavioural traces of engagement at three different learning phases (i.e. before tutorial, before quiz, and before exams) aligned with self-report measures and their impact on academic performance. Our findings demonstrated strong effects of early engagement pattern on dispositional measures of engagement as well as performances in formative and summative assessments. The issue of temporal heterogeneity and collinearity in behavioural measurements of engagement as well as its implications for learning analytics interventions were discussed. Looking forward, we propose that learning analytics studies combining measured engagement indicators of sufficient fine-grained type, such as time-stamped log data, with survey-based disposition data, can have a great potential to bring empirical research on student engagement to a next level. At the same time, this suggests being a necessary but not a sufficient condition to design effective educational interventions based on learning feedback generated by predictive modelling.

## References

- Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2004). Toward tutoring help seeking. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Intelligent tutoring systems. ITS 2004* (Lecture notes in computer science) (Vol. 3220). Berlin, Heidelberg: Springer.
- Aleven, V., McLaren, B. M., & Koedinger, K. R. (2006). Towards computer-based tutoring of help-seeking skills. In S. Karabenick & R. Newman (Eds.), *Help seeking in academic settings: Goals, groups, and contexts* (pp. 259–296). Mahwah, NJ: Erlbaum.
- Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, 26(1), 205–223.
- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, 50(1), 84–94.
- Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2006). *Handbook of research on student engagement*. New York: Springer.
- Ifenthaler, D. (2012). Intelligent model-based feedback: Helping students to monitor their individual learning progress. In S. Graf, F. Lin, Kinshuk, & R. McGreal (Eds.), *Intelligent and adaptive learning systems: Technology enhanced support for learners and teachers* (pp. 88–100). Hershey, PA: IGI Global.
- Ifenthaler, D. (2015). Learning analytics. In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 447–451). Thousand Oaks, CA: Sage.
- Ifenthaler, D., Gibson, D. C., & Zheng, L. (2018a). Attributes of engagement in challenge-based digital learning environments. In *Proceedings of the 15th International Conference on Cognition and Exploratory Learning Technologies, CELDA 2018*, pp. 225–232. Lisbon, Portugal: IADIS Press.
- Ifenthaler, D., Gibson, D. C., & Zheng, L. (2018b). The Dynamics of Learning Engagement in Challenge-Based Online Learning. in *Proceedings 2018 IEEE 18th International Conference on Advanced Learning in the Digital Age, ICALT 2018*, pp. 178–182. Los Alamitos, California: IEEE.

- Ifenthaler, D., Yau, J. Y.-K., & Mah, D.-K. (Eds.). (2019). *Utilizing learning analytics to support study success*. New York: Springer.
- Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology*, 77(2), 413–440.
- McLaren, B. M., van Gog, T., Ganoë, C., Karabinos, M., & Yaron, D. (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.
- Nguyen, Q., Tempelaar, D.T., Rienties, B., & Giesbers, B. (2016). What learning analytics based prediction models tell us about feedback preferences of students. In R. Amirault, & Y. Visser (Eds.), *e-Learners and their data, Part 1: Conceptual, research, and exploratory perspectives. Quarterly Review of Distance Education*. 17(3).
- Rienties, B., Cross, S., & Zdrahal, Z. (2017). Implementing a learning analytics intervention and evaluation framework: What works? In B. Kei Daniel (Ed.), *Big data and learning analytics: Current theory and practice in higher education* (pp. 147–166). Cham, Switzerland: Springer International Publishing.
- Rienties, B., Tempelaar, D., Nguyen, Q., & Littlejohn, A. (2019). Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2019.07.007>
- Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist*, 50(1), 1–13.
- Tempelaar, D., Rienties, B., & Nguyen, Q. (2017). Towards actionable learning analytics using dispositions. *IEEE Transactions on Learning Technologies*, 10(1), 6–16.
- Tempelaar, D., Rienties, B., & Nguyen, Q. (2018). Investigating learning strategies in a dispositional learning analytics context: The case of worked examples. In *Proceedings of the International Conference on Learning Analytics and Knowledge*, Sydney, Australia, March 2018 (LAK'18), pp. 201–205.
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (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.
- Tempelaar, D. T., Rienties, B., Mittelmeier, J., & Nguyen, Q. (2018). Student profiling in a dispositional learning analytics application using formative assessment. *Computers in Human Behavior*, 78, 408–420.