KU LEUVEN

FACULTY OF SCIENCE

Learning Dashboard Activity as a 'Learning Trace'

Does dashboard activity predict first-year student success?

Margaux Delporte

Supervisor: Prof. Tinne De Laet
KU Leuven, Louvain, Belgium, Faculty of Engineering Science

Mentor: *Tom Broos*
KU Leuven, Louvain, Belgium, Department of Computer Science

Thesis presented in fulfillment of the requirements for the degree of Master of Science in Statistics

Academic year 2018-2019

Acknowledgements

I would like to use this opportunity to thank those who played an important role in the process of writing this thesis.

First and foremost I really felt blessed to be supervised by Prof. Tinne De Laet and Dr. Tom Broos. I would like to thank Prof. Tinne De Laet for her guidance and insight in the subject. Secondly I would like to thank my mentor Dr. Tom Broos for his counseling and providing the data. I also want to thank the jury for taking the time to read my thesis.

Furthermore, I would like to thank my colleagues of the Leuven Statistics Research Center for the agreeable work environment and the valuable statistical experience I gained. I would like to thank Prof. An Carbonez in particular for giving me this opportunity.

Next, I am very grateful to my friends who were always a helping hand and a pleasant company. Without these people, creating this thesis would have been even more challenging.

Most importantly, I express gratitude to my parents and sister for believing in me anytime I was doubting myself.

Margaux Delporte

Abstract

In this thesis, a case study is presented involving a learning dashboard sent out to first-year students at an open admission university. Since the access to the university is almost unconstrained, there is a lot of heterogeneity in the educational background and skills of incoming first-year students. Often this group of students faces difficulties meeting the academic expectations. As a result, the majority of first-year students do not pass all their courses. In contrast, the current labour market is in dire need of scientists and engineers. This makes early detection and remediation of students at high risk for failing an important task. Learning analytics can be used to facilitate detection and interventions for students with a high failure risk.

The present study applies learning analytics to predict first-year student success in science and engineering programs. In contrast to most other research in this field, the data is generated from a learning dashboard. The present study aims to investigate whether dashboard usage has incremental predictive validity on top of already available data. In addition, fine-grained measures of dashboard activity are examined.

A first finding is that dashboard usage has an incremental predictive validity for academic success on top of already available data. This is shown in the prediction of weighted average grade in September, category of cumulative study efficiency and in the prediction whether students are at risk. Next, fine-grained measures of dashboard activity are defined and related to the weighted percentage in September. The duration spend on the dashboard, viewing the learning skill tips and the amount of visitations are positively related to the weighted percentage. In contrast, the time between receiving the link to the dashboard and visiting the dashboard has a negative relation with weighted percentage in September for a lag up to 8.766 days. Lastly, students that visit the dashboard again after receiving the results of the first semester have a higher weighted percentage.

To conclude, dashboard usage provides a tool that has predictive value for academic success. The deployment of this dashboard is low cost and the results are available early in the academic year. Hence, it addresses one of the gaps in learning analytics: the lack of early predictors. It can therefore potentially help to detect students in need for intervention.
The relation with academic success is probably not causal; performing actions on the dashboard does not cause higher academic success by itself. More likely, it signals a certain disposition of the student. Further research is required to investigate this underlying disposition.

List of Abbreviations

math.hrs	Hours of mathematics in high school
math.score	Secondary mathematics grade
fys	Secondary physics grade
chem	Secondary chemistry grade
bio	Secondary biology grade
schooltype	School type
mot	Motivation
tmt	Time management
con	Concentration
tst	Test strategy
anx	Anxiety
skill	Learning skill
Pct_sept	Weighted percentage in September
wavg	Weighted average grade in September
atrisk	Weighted average grade below 8.5
LASSI	Learning and study strategies inventory
STEM	Science, Technology, Engineering and Math
LMS	Learning management system
VLE	Virtual learning environment
MOOC	Massive open online course
CSE	Cumulative study efficiency
MCAR	Missing completely at random
MAR	Missing at random
MNAR	Missing not at random
LOESS	Locally estimated scatterplot smoothing
Anova	Analysis of variance

List of Figures

List of Tables

Contents

Chapter 1

Introduction

The focus of this thesis is an open admissions university in Flanders (KU Leuven). Different from the Anglo-Saxon countries and some other countries in Europe, no central examination at the end of secondary education takes place. In addition, there is no entrance exam when entering higher education. The only requirement to enter this university in any field of study is a valid secondary education qualification (except for Medicine, Dentistry and Arts Education). Furthermore, tuition fees are low (below $1000 per year) in order to give economically disadvantaged groups the opportunity to participate in higher education. As a consequence, incoming first-year students are very heterogeneous in terms of their educational background. The high degree of heterogeneity imposes problems in the education of science and engineering students, where students that lack a sufficient background and skills have difficulties keeping up. This implies that the first year of university is a 'selection year'. The amount of the university students that pass all courses in the first year has dropped below 40 % (Fonteyne, Duyck & De Fruyt, 2017). KU Leuven reports a dropout rate of about 30 % in the STEM programs included in the present study. In contrast, the current labour market has a deficit of scientists and engineers (National Math and Science Initiative, 2015). Therefore, early detection and remediation of students at high risk for dropout in those programs is beneficial for everyone involved.

Learning analytics is a fairly new tool that can help with this detection and remediation. A popular definition of learning analytics is "*the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs*" (Long & Siemens, 2011). Often applications of learning analytics give students feedback about their learning behaviour with the goal to help the students become more strategic learners (Wolff, Zdrahal, Nikolov & Pantucek, 2013). This thesis applies learning analytics to predict first-year engineering and science student success. In contrast to other papers in this field, this thesis uses data generated by student-facing dashboards. This dashboard aims to provide feedback about learning skills. The first goal of this thesis is investigating if the dashboard usage has predictive capacity for first-year student success on top of already available data. The second goal is defining fine grained measures of dashboard activity that are related to first-year student success.

Chapter 2

Literature study

2.1 Definitions

Learning analytics has only been a separate field since 2010. It has emerged as a subfield from educational analytics. Educational analytics contains several subfields, which are not easily demarcated and share a lot of common characteristics. Still, defining them separately is important. Firstly, Mohamad and Tasir (2013) define educational datamining as a discipline that focuses on developing methods to explore data that is unique for an educational setting with the goal to better understand students and their learning context. This discipline mainly handles large volumes of data by the use of datamining methods. Secondly, academic analytics monitors the success of individual students for the purpose of management of the academic enterprise (Ferreira & Andrade, 2016). This field focuses on governmental, political and economical challenges.
Thirdly, student analytics lies on the verge of learning analytics and educational analytics. Just like learning analytics, this field aims to have a personalised approach. Data analysis is executed with the goal to discover predictive factors of study behaviour and study success. These factors can ultimately guide tailored and data-driven student counselling.

2.2 Previous research

2.2.1 Learning analytics in learning dashboards

A learning analytics dashboard is a dashboard that contains multiple visualisations of different characteristics of the learners, learning processes and/or learning contexts (Schwendimann, Rodrguez-Triana, Vozniuk, Prieto, Boroujeni, Holzer, Gillet & Dillenburg, 2016). The application of learning analytics in learning dashboards is relatively new. Not much research is conducted in this field, in contrast to the application of learning analytics in Massive open online courses (MOOC) and virtual learning environments (VLE). The following paragraphs summarise studies that did apply learning analytics in the context of a learning dashboard.

Broos, Verbert, Langie, Van Soom and De Laet (2018) investigated in the same context as this thesis the relationship between accessing of the learning analytics dashboard and the test positioning score. They divided the students in three groups: students who

(a) did not use the dashboard at all, (b) did visit the dashboard but did not view all feedback categories and (c) did visit the entire dashboard. The authors did not conduct a formal statistical test, but there is clearly a relationship present; the group that did not access the dashboard has a median positioning test score of 8.6 (SD=3.3). The middle group has a median score of 9.4 (SD=3.7) and the group that viewed the entire dashboard has a median score of 10.7 (SD=3.8).

A second study in this context is the study of Broos, Verbert, Van Soom, Langie and De Laet (2018). The authors conducted a study about a learning analytics dashboard for first-year students. This dashboard was developed to give feedback, facilitate reflection and give recommendations about the exam results. This paper reported statistics about the relationship between cumulative study efficiency (CSE) and click-through rate from an invitation link to the dashboard. The most successful students in the high CSE group have a click-through rate of 56, 3%, while the medium CSE group has a click-through rate of 45.9%. In contrast, the low CSE group has a click-through rate of only 34.8%.

A pilot project investigated the dashboard of the present study (Broos, Peeters, Verbert, Van Soom, Langie & De Laet 2017). The dashboard intends to give feedback and tips for learning skills based on the Learning and Study Strategies Inventory (LASSI). The skills in this dashboard are motivation, anxiety, concentration, time management and test strategies. The researchers conducted an in-depth analysis of the relationship between learner profile and use. Students with better learning skills have a higher probability to visit the dashboard. The difference was significant for each of the five LASSI skills in isolation. Next, the skills are integrated in a logistic regression model to predict whether the students visit the dashboard. Also gender and study program are in the full model and afterwards variable selection is performed. Only time management was significant in the reduced model. If the students did visit the dashboard, the lower the score on the skill, the more likely they viewed the corresponding tips. The latter relation was also significant in an integrated logistic regression model with all the skills, gender and study program. Further, motivation has a significant impact in the prediction whether the student viewed any of the learning skill tips.

2.2.2 Learning analytics in VLE's and MOOC's

In contrast of the application of learning analytics on learning dashboards, the application of learning analytics in Massive open online courses (MOOC) and virtual learning environments (VLE) has a long history and a lot of research is conducted in this field.

The very first study that used data from a Web-based learning environment to predict student performance is the study of Rafaeli and Ravid (1997). They responded to the criticism on the use of technologies in the classroom by designing a study on the role of internet-based education in learning. The reliability was suboptimal, because one third of the students reported they occasionally borrowed usernames and passwords from fellow students. In addition, the internet was not as widespread at that time, with less than half of the students that used internet prior to the course. Different linear regression models are fitted to predict the final grade by the use of measures of online usage in the different class groups. The amount of pages read and the grades for online questions could predict in each model more than 20% of the variation in the final grade. Furthermore, students seemed to accept the, at that time, fairly new use of online tools.

An important study that revealed some caveats of the research in learning analytics is the study of Conijn, Snijders, Kleingeld and Matzat (2017). The researchers used data from a learning management system (LMS) to predict student success. They state that the effects of LMS behaviour on student performance might be different in different institutions and that there even might be differences in different courses within the same institution. Therefore they raise questions about the portability of the prediction models. Another goal of their study was finding early predictors or other words, variables that can be measured within the first weeks. These can facilitate early interventions for students at risk. Learning analytics often uses aggregated variables over the whole learning process, which have limited value for intervention. Their results implied that the portability of prediction models is indeed limited. In addition, more online sessions, lower standard deviation of time between sessions and less time until the first session are associated with higher grades. Furthermore, LMS data has limited value for early interventions since the LMS data at an early stage have less predictive value for final grade than in-between assessments. If the dependent variable was coded as pass/fail, the prediction of LMS data turned out to be inaccurate.

A great diversity presents itself within the literature between the studies that predict student performance with LMS data. Especially in the predictor variables there is a lot of variety, because not all researchers have access to the same variables in the LMS (Conijn et al, 2017). In addition, different institutions and courses use different tools in different kinds of LMS. Therefore, it is not surprising that there are important differences between studies in the significance of predictor variables.

2.3 Important predictors of student success

Variables related to demographic characteristics, academic integration, social integration, psycho-emotional and social factors explain a big part of the variation in student success (Tempelaar, Rienties & Giesbers, 2015). For example, in the study of Tempelaar et al. (2015), the mathematics track in high school could explain 20% of the variation in mathematics related performance measures. Still, LMS user behaviour has incremental explanatory value on top of the traditional variables (Pinxten, Langie, Van Soom, Peeters, De Laet, 2017). In addition, this thesis focuses on changeable characteristics, where feedback is useful.
As noted above, a problem in learning analytics research is the lack of portability of the statistical models, which is also present on the level of predictive variables. While in some studies a particular variable can explain a large part of the variation in student success, in other studies the same variables yields a nonsignificant effect.

2.3.1 Demographic variables

Some demographic variables are able to explain student success very well. Trussel and Burke-Smalley (2018) conducted a study in order to discover important variables in the prediction of overall GPA. The sample consisted of 1919 undergraduate business students in a public institution located in the southeastern United States. The authors fitted a linear regression model with stepwise variable selection. This regression model could explain 28.7 % of the variation in cumulative GPA. Six factors have a significant impact on overall GPA: the female gender, household income, college admission score (ACT/ SAT), financial independence and high school GPA. Black race is negatively related to GPA, while other categories of race are not significant.

Van den Broeck, De Laet, Lacante, Pinxten, Van Soom and Langie (2018) conducted a study in the same context as this thesis about the role of academic background variables and diagnostic testing in bridging students. The overall GPA at the end of the professional bachelors program turned out to be the most predictive variable. Academic background variables have a higher predictive value compared to general characteristics (gender and SES) and the diagnostic test.

Further, Pinxten and Hockicko (2016) discovered predictive factors of study success of first-year science and engineering students at the university of Zilina. This study confirms the result of Trussel and Burke-Smalley (2018) that females perform significantly better than males. In addition, school type, math grades and effort expenditure in secondary school have a significant relation with students GPA and credits earned after the first semester.

Another important demographic variable is the education of the parents of the student. A first-generation student is defined as a student where none of the parents have a degree in higher education. A study of Spruyt, Kavadias and Roggemans (2014) conducted in the context of the Flemish entrance examination of students medicine and dentistry revealed that first-generation students have a disadvantage. When both parents of the student have a degree in higher education, the student has a two times higher probability to pass the exam compared to first-generation students. An exception is present for students that followed a track in secondary school combining Latin and science, where the difference is considerably smaller. Choy (2001) investigated the relation between first-generation students and dropping out during the first year or failing to return for the second year. First-generation students are twice as likely to have these outcomes compared to students whose parents have a bachelor's degree (resp. 23% and 10%).

2.3.2 Clicking behavior

Clicking behaviour in a LMS is a poor predictor for student success according to the study of Wollf et al. (2013). This study in the context of a virtual learning environment (VLE) found that some students never click and still pass the course, while other students clicked a lot and failed. A possible explanation is that some student print the online learning material, make notes or download it for offline use. Their main predictor of a performance drop was the relative difference between the clicking activity. In other words, the clicking activity of the student compared to the activity of the same student on a previous moment.

2.3.3 Interactions

A strong predictor in some learning analytics studies are interactions (Agudo-Peregrina, Iglesias-Pradas, Conde-Gonzlez & Hernndez-Garca, 2014). Moore (1989) partitioned the interactions in three groups: student-student interactions, student-teacher interactions and student-content interactions. Malikowski, Thompson and Theis (2007) propose an additional trichotomy based on the frequency of interactions. A first category are the most frequent interactions: the transmission of content. The second category are moderately used interactions: creation of class interactions (discussions between course members) and evaluation of students (quizzes and assignments). The last category are the rare interactions: evaluating courses/teacher and computer-based instruction (e.g. self-assessment quizzes, examining of prerequisites to get access to content).

Agudo-Peregrina et al. (2014) found that the different types of interactions within each classification are related to student academic performance only if the courses are entirely online. In this study student-student interactions are the most important predictor. Furthermore, student-teacher interactions and evaluating students are significant variables. Student-content interactions are nonsignificant, but this can be explained again because students can print or download the online material. If the courses have a face-to-face format with the support of a VLE, the variables have no significant effects.

Macfadyen and Dawson (2010) got similar results. This study was carried out in the context of the LMS data of a fully online course. The authors defined a fully online course as a course where all the communication, assessment and content transmission is done online. Firstly they investigated the correlations between LMS variables and final grade. They found significant correlations for the variables listed in Table 2.1. Hence, it seems that there exists an association between performing interactions and final grade. In addition, Macfayden and Dawson fitted a regression model for the student final grade which could explain 33% of the variation. The key predictors are the total number of discussion messages posted, total number of mail messages sent and total number of assessments completed. All these variables have a significant correlation with the final grade ($p < 0.05$).

Table 2.1: *Significant correlations in the study of Macfayden and Dawson (2010).*

Variable	Correlation	p-value
Total amount of discussion messages posted	.52	.00
Total number of online sessions	.40	.00
Total time online	.34	.00
Amount of files viewed	.33	.00
Amount of assessments finished	.31	.00
Amount of assessments started	.31	.00
Amount of replies to discussion messages	.30	.00
Amount of mail messages sent	.28	.00
Amount of assignments submitted	.26	.00
Amount of discussion messages read	.25	.00
Amount of web links viewed	.25	.00
Amount of new discussion messages posted	.24	.01
Amount of mail messages read	.22	.01

2.3.4 Time online

Time online is one of the predictors with many contradicting results in research. In this paragraph three studies are listed that each found different results.

The study of Macfayden and Dawson (2010) found that measures of time online showed a weak correlation with final grade ($r = 0.34$). In addition, it was not a significant predictor in the final regression model of final grade.
In contrast, Boulton, Kent and Williams (2018) got mixed results. The authors operationalized VLE usage as time online in a bricks-and-mortar learning setting. Hence, in this context the VLE has a hub for transferring lecture slides, worksheets and extra learning material. In this study VLE usage showed a Spearmans correlation ranging from 0.15 to 0.52, depending on the study program. The correlation was lower in BA programs compared to BSc. programs. Different usage of the VLE in the different programs causes the differences in correlations. For example, biology and medical science students have to log in each week for information about the practical sessions and to complete graded assessments. In addition, high VLE usage has an association with high grades, but low VLE usage not with low grades.

Yu and Jo (2014) provide support for a significant relationship between time online and student success in their study. The goal of this study was the prediction of students academic achievement based on LMS data in a South Korean female university. The subject of the study was a single course, where 20 % of the final grade was assigned for the participation in online discussion in a LMS. Yu and Jo fitted a multiple linear model with six covariates. The model could explain 33.5 % of the variation in the data ($R^2 = 33.5$). Two covariates are significant: total studying time in LMS and interactions with peers. In contrast, interactions with instructors, total login frequency in LMS, regularity of learning interval in LMS and the number of downloads are not significant in the multiple regression model. Note that this result also partially confirms the conclusion of Agudo-Peregrina et al. (2014), where student-student interactions are the most important predictor. But in the latter study student-teacher interactions are significant, which is not the case in the study of Yu and Jo (2014).

2.3.5 Academic skills

Five important academic skill variables in this thesis are concentration, motivation, time management, anxiety and test strategy. All these variables stem from the Learning and Study Strategies Inventory (LASSI). Pressure is defined as the preference for time scale.

Pinxten et al. (2017) examined the relation between LASSI and student success in the context of STEM first-year students of the KU Leuven. The LASSI variables are attitude, motivation/persistence, time management, anxiety, concentration, information processing, selecting main ideas, study aids, self-testing and test strategies. The authors investigated the correlations of ten LASSI variables with weighted GPA and the incremental predictive value of these variables over prior achievement to predict weighted GPA. Firstly, four out of ten variables correlate significantly with weighted GPA. These variables are motivation/persistence ($r = .26$), time management ($r = 0.24$), concentration ($r = 0.22$) and test strategies ($r = 0.21$). Thus the self-regulation skills related to effort have an association with weighted GPA. Next, the incremental predictive value are investigated in a stepwise regression model including the ten LASSI scales. Motivation, test strategies and time management have a significant impact on weighted GPA. In addition, the incremental predictive validity of the LASSI scales over prior achievement is investigated. The inclusion of motivation and time management on top of secondary school GPAs and math level results in an increase of explained variance of 2% in the total sample and of 3% in engineering science. Test strategies was not added since it did not result in a significant improvement of variance explained.

Mothilal, De Laet, Broos and Pinxten (2018) conducted a study to predict student success in the same context as this thesis. The authors performed principal component analysis with an oblique rotation because of collinearity issues. On the first principal component motivation, time management and concentration have high loadings ($loadings > .72$). Anxiety and test strategy have high loadings on the second principal component ($loadings > .76$). The first and second principal component are named respectively affective strategies and goal strategies. Next, these components are used to predict academic achievement. The coefficients of the regression model are jointly significant ($p< 0.001$), but the R^2 is only 0.06. Thus soft skills can explain only 6% of the variation in academic achievement. In addition, only the coefficient of affective strategies is significant and equivalently, this variable can explain most of the variance. Goal strategies and pressure are not significantly related to weighed GPA.
Mothilal et al. (2018) also fitted an additional sequential regression model where grades of the last year of high school (mathematics, physics and chemistry), number of hours of mathematics in the curriculum and the effort level are taken into account. It turned out that goal related strategies have additional explanatory value, while pressure preference and affective strategies have not.

Pinxten and Hockicko (2016) conducted a study to discover predictive factors of student success in first-year engineering and science students. Next to academic background variables, the value of the LASSI variables was examined. The authors investigated all ten LASSI variables. Three of those variables have a moderate correlation to the study results after the first semester: motivation, time management and test strategies. Hence, this study has similar results as the study of Pinxten et al. (2017) at KU Leuven. The only difference is that, in the latter study, there is also a significant relation between concentration and weighted average grade.

2.3.6 Student engagement

There is no single definition of student engagement. It is a multidimensional construct that contains multiple sub-dimenstions. Fredericks, Blumenfeld and Paris (2004) distinguish three types of engagement. First, behavioural engagement is participation in academic, social and extracurricular activities. This kind of participation is crucial for good results and prevents the student from dropping out. Second, emotional engagement involves affective responses to learning and the learning environment. Also identification with the school is part of this construct. Third, cognitive engagement relates to investment; the willingness to make the effort needed to understand complex ideas and master difficult skills. It entails self-regulation and being strategic. These factors are not isolated processes, but are intertwined in one dynamic process within the student.

Behavioural and cognitive engagement correlate with higher academic achievement. The correlation between emotional engagement and academic achievement is less documented, but there is some evidence of its presence (Fredericks et al., 2004). A meta analysis of Lei, Cui and Zhou (2018) reported significant correlations of academic achievement with overall engagement and the three sub-dimensions of engagement. More specifically, the correlation with overall engagement is 0.269 ($k=30$, $p<.001$). Behavioural engagement has a correlation of 0.350 ($k=55$, $p<.001$). The effect size of emotional engagement was 0.216 ($k= 47$, $p<.001$). Lastly, cognitive engagement has a correlation of 0.245 ($k=31$, $p <.001$).

Jung and Lee (2018) investigated the effect of student engagement on learning persistence in a MOOC. The authors measured student engagement via self-reports in questionnaire that followed the theory of Fredericks et al. (2004). There was a direct effect of student engagement on learning persistence. This result is supported by Pursel, Zhang, Jablokow, Choi and Velegol (2016) that operationalized student engagement as the watching of videos, making of quizzes and completion of a course project. In a logistic regression model, student engagement has incremental predictive value on top of other variables in the prediction of course completion.

Chapter 3

Methods

3.1 Linear Regression

Linear regression is applied to model the relationship between a continuous dependent variable and one or more independent variables. There are several assumptions underlying the linear regression model. The first assumptions to statisfy are the Gauss-Markov conditions:

$$E[\epsilon_i] = 0,$$
$$Var[\epsilon_i] = \sigma^2$$

and

$$E[\epsilon_i \epsilon_j] = 0 \text{ for all } i \neq j.$$

Secondly, independent and normally distributed errors are expected to make inferences about the regression parameters

$$\epsilon \sim N_n(0, \sigma^2 I_n).$$

Under this condition, the general linear model satisfies

$$y \sim N_n(X\beta, \sigma^2 I_n)$$

and

$$\hat{\beta}_{LS} \sim N_p(\beta, \sigma^2 (X^t X)^{-1}).$$

According to Kutner, Li, Nachtsheim and Neter (2005) departures from normality does not impose large problems. The sampling distribution of the intercept and slopes is still normal as long as the probability distribution of the errors does not depart seriously from the normal distribution. In addition, the confidence intervals and p-values will be approximately provided. If the departures are serious, the distribution of the slope and the intercept is asymptotically normal. Schmidt and Finan (2018) advocate that transformations in order to obtain a more normally distributed error are unnecessary. They argue that given a sufficient sample size ($\frac{p}{n} > 10$), linear regression models with non-normal errors are still valid, while transformations can bias model estimates.

3.2 Logistic regression model

A logistic regression model is used to model a binary outcome variable. It belongs to the family of generalized linear models. Several assumptions are underlying to the method. First, there should be no over- or underdispersion present. Over- or underdispersion is defined as the presence of respectively more or less variability in the data than expected under the logistic model. Second, the observations should be independent. Third, absence of perfect separation. Perfect separation occurs when one or a combination of predictor variables can perfectly predict a class. In addition, there should be no multicollinearity. Multicollinearity is defined as the existence of a linear relationship among two or more variables (Alin, 2010).

3.3 Multinomial logistic regression model

Multinomial logistic regression is applied to predict class membership of multiple non-overlapping classes. It is a special case of the generalized linear model. More specifically, it is a generalization of logistic regression to a setting with multiple classes. Garson (2014) states multiple assumptions of multinomial logistic regression:

- independence of observations
- absence of perfect separation
- no multicollinearity
- no over-or underdispersion
- independence of irrelevant alternatives (IIA)

The last assumptions entails that the odds ratio of any two categories is independent of the attributes or availability of a third category (McFadden, Tye & Train, 1976).

3.4 Two-sample inference

3.4.1 Unpaired t-test

An unpaired t-test is conducted when two means of independent groups are compared. The null hypothesis states that the population means are equal, the two-sided alternative hypothesis states that they are different. The test statistic is the following:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}}} \sim t_{df=n_1+n_2-2}.$$

The test has two assumptions. First, the variances of the groups of samples should be equal. Second, this test assumes that both groups follow a normal distribution. However, a sufficiently large sample size meets the latter condition. The reason is that from the central limit theorem it follows that the distribution of means calculated from repeated sampling will approach normality.

3.4.2 Welsch test

If the assumption of equal variances does not hold, the Welsch test can be conducted. The null and alternative hypothesis are equal to the null and alternative hypothesis of the unpaired t-test. Also the assumption of normality is maintained. The test statistic is the following:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_{x_1}^2}{n_1} + \frac{s_{x_2}^2}{n_2}}}.$$

Under the null hypothesis, the test statistic follows a Student's t-distribution with the following degrees of freedom:

$$df = (\frac{s_{x_1}^2}{n_1} + \frac{s_{x_2}^2}{n_2})^2 / (\frac{s_{x_1}^4}{n_1^2(n_2-1)} + \frac{s_{x_2}^4}{n_2^2(n_1-1)}).$$

3.5 K-sample inference

3.5.1 Anova

A first method is an one-way analysis of variance (Anova). The null hypothesis of an analysis of variance is that all group means are equal versus the alternative that at least one group mean is different from the others. The test statistic compares the variance between groups and the variance within groups.
The test makes several assumptions (Kutner et al., 2005). First, the test assumes normality of the dependent variable. The Shapiro-Wilk test can validate this assumption. Second, homogeneity of variances is assumed; each probability distribution should have the same variance. The Brown-Forsythe test assesses whether this assumption is valid. The Shapiro-Wilk and Brown-Forsythe tests are discussed in Appendix A. A third assumption is the independence of observations. Residual plots provide the means to visually check these assumptions.

Effects of deviations of normality and homogeneity of variance assumptions

Kutner et al. (2005) discusses the effects of departures of the assumptions. Deviations from normality does not impose a big issue, given that it is not extreme. In general, the estimates of group means are unbiased in case of non-normality. Still, it can affect the type I error of the F-test; the actual α can be larger than the nominal α. However, the effect of the type I error is quite contained.
The violation of the assumption of homogeneity of variance has only a large impact when the sample sizes of the groups are not similar. However, in the case of comparison of single group means the results becomes unreliable. Weighted least squares provides a valid solution in the case of heterogeneity of variance.

3.5.2 Kruskal-Wallis test

When the residual analysis shows departures from normality, the Kruskal-Wallis one-way Anova provides an alternative (Kutner et al., 2005). The test does not assume normality of the dependent variable. Still, the test assumes that the samples are drawn from distributions with the same general shape (Cytel inc., 2007). More specifically, the test assumes similar shapes of ranks of the groups (Vargha & Delaney, 1998). This entails homogeneity of variances of the ranks.

3.5.3 Median test

The median test is a non-parametric test to evaluate the equality of the medians of k distributions. The alternative hypothesis entails that at least one of the medians of the distributions is unequal to the others. The test is not as powerful as the Kruskal-wallis test, but has as advantage that no distributional assumptions are made (Cytel inc., 2007). The test is described in more detail in Appendix B.

3.5.4 Post-hoc tests: Holm's method

In Holm's procedure, hypotheses are sequentially rejected in order to maintain the nominal type I error rate. This means rejection of hypotheses until no further rejections are possible (Holm, 1979). The p-values are first ordered from small to large (P_1 until P_m). Next, for each level, k equals the minimal level such that:

$$P_k > \frac{\alpha}{m - 1 + k}.$$

The null hypotheses before H_1 until H_{k-1} are rejected, while the remaining null hypotheses are not rejected. If k equals 1, no hypotheses are rejected. If there is no k, all the null hypotheses are rejected.

Chapter 4

Dashboard usage as a binary variable

4.1 Description of the dataset

The research question of this chapter is whether dashboard usage has a substantial predictive value on top of other known predictive factors. At the beginning of the academic year, students had to complete a LASSI test and several additional questions. The additional questions included the grades of the students' in secondary school, the advice of their teachers about their study choice in higher education and the weekly number of hours of mathematics in high school. A couple of weeks later students received an invite by e-mail to access the online LASSI dashboard. When clicking on the link, students encountered their personalized dashboard. This dashboard usage was tracked. The dataset contains information about 3479 science and engineering students. The variables used for analysis are:

Weighted average grade (*wavg*): The weighted average score in September. The score is weighted by the amount of ECTS credits and ranges between 0 and 20.

Cumulative study efficiency (*cse.sept, cse.final*): the amount of credits that a student obtains in September from the total amount of credits he was subscribed for at the start of the academic year. The university has regulations for students with a low CSE. When a student obtains a CSE of less than 30%, he cannot continue with the same program. Students with a CSE of less than 50% receive binding study advice. *Cse.sept* is continuous, *cse.final* is binned according to the consequences ("$>$ 80%", "50%-80%", "30%-50%", "$<$ 30% or dropout").

School type (*schooltype*): The track the student followed in high school. In Flanders the high school system is organized in four general types: ASO, TSO, BSO and KSO. ASO contains a broad general curriculum that prepares the student for higher education. TSO also contains mathematics and science in their curriculum, but at a lower level than most ASO courses. The focus is more practical and technical focus instead of theoretical. The students are prepared to either enter the vocational market or to study a masters or bachelors degree. BSO students get a practical and job specific education and are not expected to pursue higher education. Lastly, KSO mixes a general and broad education with the practice of arts.

Math level (*math.hrs*): Students in Flemish high schools can choose between three levels of mathematics. The curriculum and the amunt of hours of mathematics per week depend on this choice. The low level corresponds to less than 6 hours of mathematics. The medium and high level correspond to respectively to 6 or 7 hours and more than 8 hours of mathematics.

High school grades (*math.score, fys, chem, bio*): The self-reported scores of respectively mathematics, physics, biology and chemistry in high school. Because there are no national-level school leaving exams in Flanders, the grades highly depend on the high school and teachers. To correct for this, the grades are binned into categories: "60%", "60-70%", "70-80%", "80-90%" and "> 90%".

Learning and study strategies: The learning and studying skills of students are assessed with an instrument constructed by Weinstein and Palmer (2002). The students answer 77 questions on a five-level Likert scale to operationalize and evaluate ten scales. Pinxten et al. (2017) showed that four scales have a substantial significant relation with academic achievement at the end of the first year: **motivation (*mot*), time management (*tmt*), test strategy (*tst*), concentration (*con*)**. Both the instrument and dashboard only incorporated the latter four and one extra scale in order to avoid survey fatigue and to keep the dashboard concise. **Performance anxiety (*anx*)** was also included as the focus lies on actionable feedback to the students. The final LASSI scores take values between 8 and 40, where higher scores correspond to better skills.

Advice of high school teachers (*advice*): The self-reported advice a student received from his board of high school teachers about their study choice in higher education. The categories are "positive", "partially positive", "negative" and "unknown".

Dashboard user (*dbuser*): Binary variable to indicate whether the student clicked on link of the dashboard.

First generation student (*pioneer*): Binary variable that indicates if the student is the first person of his family that pursues higher education.

4.2 Exploratory analysis

4.2.1 Descriptive statistics

The descriptive statistics for the continuous variables are listed in Table 4.1 and the Pearson correlations are shown in Figure 4.1. As can be expected, there is a very high correlation (r=.94) between the percent of the total credits a student gained and the weighted average grade. The Pearson correlations between the *wavg* and the LASSI variables are weak. The correlations between the LASSI variables are examined in more detail in the next section to asess multicollinearity. Multicollineariy occurs when there exists a linear relationship between two or more regressors (Alin, 2010). When this occurs, the parameter estimates become unreliable.

Table 4.1: *Descriptive statistics of the continuous variables.*

Variable	Mean	SD	Min	Max	Median	N missing
Wavg	9.992	3.753	0	18.750	10.775	99
Mot	28.493	4.194	12.000	40.000	29.000	52
Tmt	24.444	4.593	9.000	39.000	25.000	61
Anx	27.200	5.370	8.000	40.000	28.000	97
Tst	29.788	3.772	14.000	40.000	30.000	52
Con	27.624	4.749	9.000	40.000	31.000	71
cse.final	62.443	34.420	0.000	100.000	72.000	3

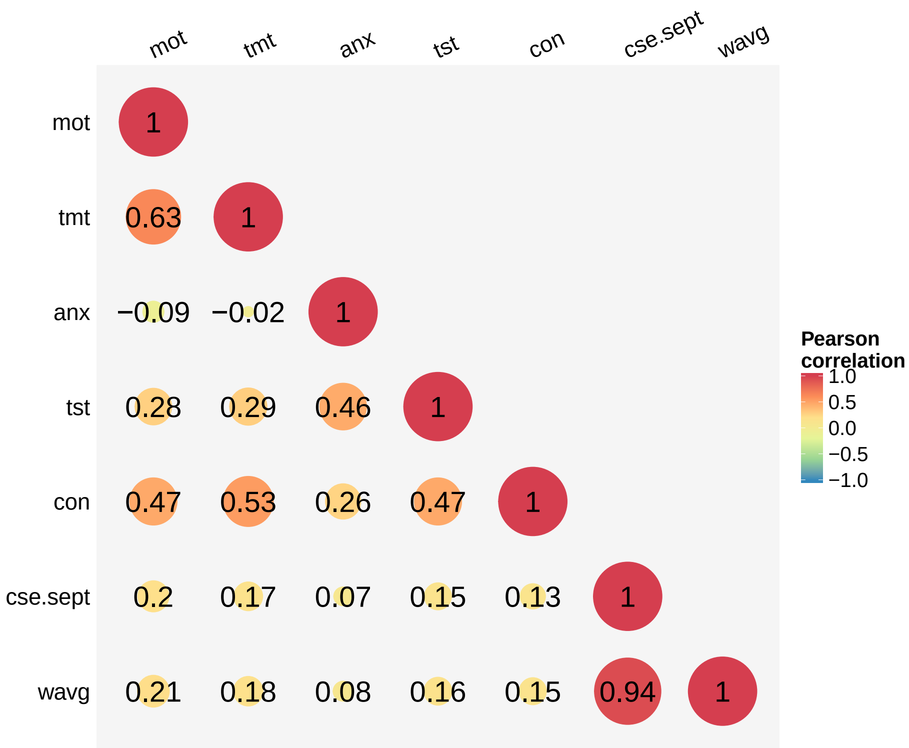


Figure 4.1: Correlations between the continuous variables.

Figure 4.2 displays the bar chart of the frequencies of the program groups. Engineering technology is by far the most frequently chosen university program. Engineering architecture is chosen the least frequent.

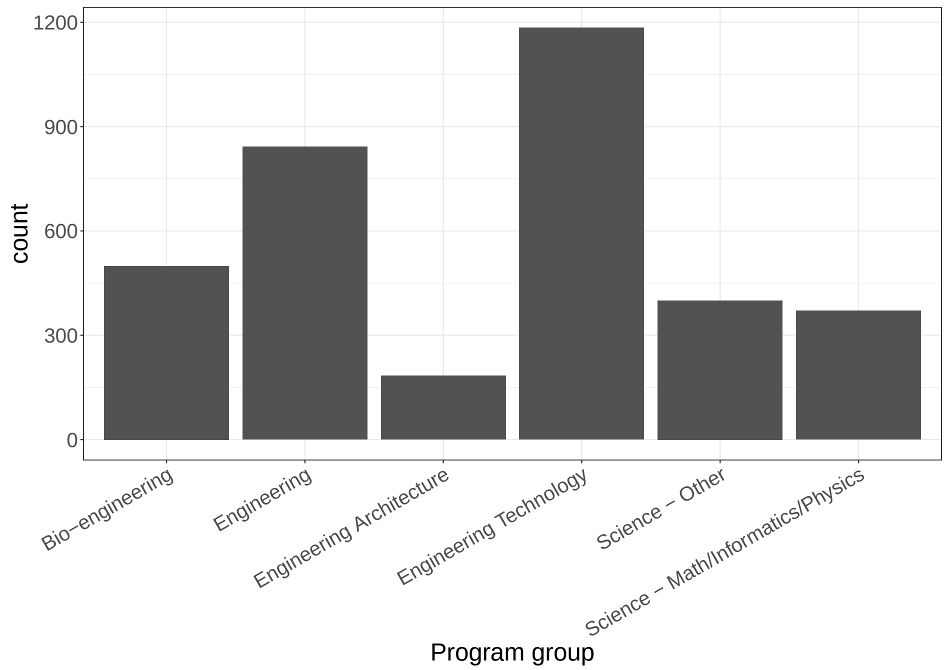


Figure 4.2: Bar chart of the school programs.

The proportions of the variables about the high school results are shown in Table 4.2. The mode of each of the high school results lies in the 70-80% interval. This expresses that students who opt for an engineering or science degree in university have in general good results for science courses in high school. Still, caution is advised when interpreting these scores. There are no national-level school leaving exams in Belgium and thus, these scores highly depend on the school and teachers.

Table 4.2: *Proportions of scores of high school courses.*

Variable	$< 60\%$	$60-70\%$	$70-80\%$	$80-90\%$	$> 90\%$	NA
Biology	0.070	0.211	0.326	0.237	0.047	0.109
Chemistry	0.111	0.283	0.332	0.200	0.050	0.025
Math.score	0.111	0.326	0.334	0.184	0.043	0.020
Physics	0.092	0.278	0.343	0.208	0.057	0.022

Next, the academic background of the students is examined. The dataset contains both information about the amount of hours mathematics, the school type and the advice the student received. As expected, the majority of the students followed the ASO type of secondary education (p=.795). From the students that followed a TSO type of secondary education (p=.165), most students opt for engineering technology (p=0.728). A negligible amount of students followed a BSO ($p<0.001$, n=1) and KSO school type (p=0.001, n=3). 3.9% of the students is not willing to disclose this information.
Only 11.5% of the students followed a program with less than 6 hours of mathematics a week. The majority (54.0%) followed a high school track with 6 or 7 hours of mathematics and the rest of the students (32.4%) had weekly 8 hours of mathematics scheduled in high school. Note that 2.1% of the students have missing values for this variable.

Most secondary teachers formulate an advice for the program choice of their students. Table 4.3 displays of each advice . The mode of the table is "Completely positive" and 71.7% of the students received at least partially positive advice. To conclude, in general students received positive advice of their board of high school teachers.

Table 4.3: *Proportions of the advice the students received from the secondary school.*

Advice	p
Completely positive	0.488
Partially positive	0.229
Negative	0.130
Unknown/ no advice	0.148
NA	0.004

The educational degree of the parents of the student is also measured. As noted in Section 2.3.1, students whom both parents do not have a higher education degree have a disadvantage. 86.6% of the students are not first-generation students, 10.5% of the students are the first of their generation to pursue a higher education degree and 2.4% of the students' status is unknown.

The dashboard variable signals whether or not the student clicked on the link with the dashboard that was send out at the start of the first year. It turns out that the majority of the students (87.3%) did. Only 12.7% of the students ignored the e-mail.

Next the final CSE in september is examined in Table 4.4. The mode is the category that has a CSE of more than 80% and hence a successful first year. Still, 25% of the students cannot further pursue their degree and 11% receive binding study advice. This proves that there is a large group of students that fail their first year.
A last variable is *atrisk*, which is defined as a weighted average score in September below 8.5. With this definition, 29.2% of the students are at risk, while 68.0% of the students are not at risk or at moderate risk ($wavg < 11.5$). In addition, 2.8% of the students have a missing value because they dropped out before the exams.

Table 4.4: *Proportion of each level of cumulative study efficiency.*

CSE	p
$< 30\%$ or dropout	0.246
$30 - 50\%$	0.112
$50 - 80\%$	0.187
$> 80\%$	0.454
NA	0

4.2.2 Multicollinearity

The Pearson correlations between the LASSI variables are shown in Figure 4.1. Since some correlations are of moderate magnitude, there exists a possibility that multicollinearity is present. Multicollinearity is defined as the existence of a linear relationship among two or more variables (Alin, 2010). According to Alin (2010) multicollinearity causes various problems in the regression analysis. For example, the regression coefficients are unreliable and have an inflated variance. Formal methods to detect multicollinearity are the conditioning numbers and variance inflation factors (Salmerón, García & García, 2018).

The variance inflation factors (VIF) are defined as

$$VIF_j = \frac{1}{1 - R^2}.$$

R^2 equals the coefficient of multiple determination when x_j is regressed on the remaining explanatory variables (Alin, 2010). It can be proved that this equitation equals $(R_{XX}^{-1})_{jj}$, the jth diagonal element of the inverted correlation matrix. The variance of the coefficients is more inflated when the VIF becomes larger. A common threshold to distinguish large and small VIF is 10 (Alin, 2010). The VIFs of the LASSI variables are listed in Table 4.5. The table proves that the variance inflation factors are small and therefore the variance is not severely inflated.

Table 4.5: *Variance inflation factors of the LASSI variables.*

Variable	*mot*	*tmt*	*anx*	*tst*	*con*
VIF	1.835	1.904	1.413	1.602	1.800

A second tool to detect multicollinearity is the conditioning number. The conditioning number is defined as

$$\eta = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}.$$

With λ_{max} and λ_{min} respectively the maximum and minimum eigenvalues of correlation matrix. A η larger than 30 implies evidence of multicollinearity (Belsley, 1982). In this data the conditioning number equals 2.616. To conclude, there is no evidence for collinearity in the LASSI variables.

4.2.3 Missing data

There is a considerable amount of missingness in the data. There are 99 students that have no observation for their weighted average score. The other variables in the dataset also contain missing values. Figure 4.3 displays the bar chart with the proportions of missing values. The variables with the most missing values are biology, school type and weighted average. The amount of missingness in the biology variable is striking; the score of biology in high school is missing in 11% of the cases. The reason behind this is that the biology score was not asked to certain program groups in a certain year. Due to the large amount of missingness, the variable is excluded from the analysis. Without the biology variable, 84.16 % of the cases are complete.

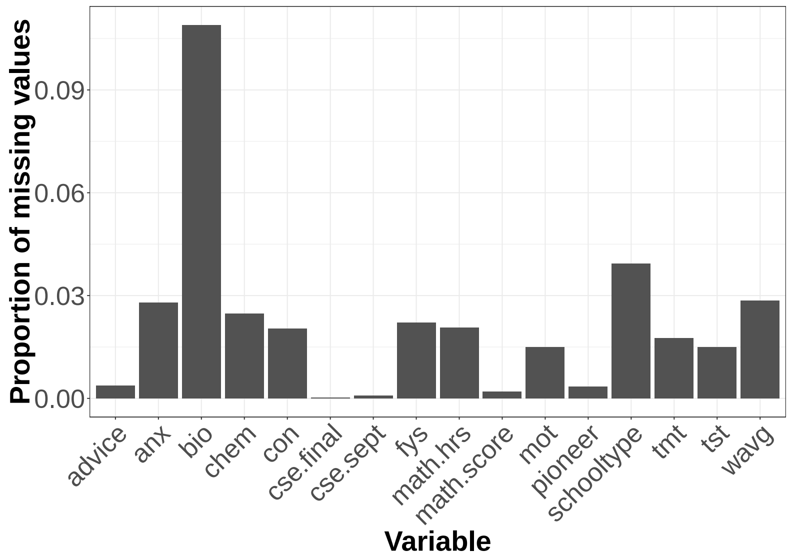


Figure 4.3: Proportion of the missing values in the different variables.

The variable *schooltype* is missing in 3.94% of the cases. The variable is not self-reported, but part of the database of KU Leuven. In this variable the reason behind the missingness is known; some students followed high-school in the Netherlands or in other countries abroad. Since these high school systems are organised in a different way, this variable is not applicable. This is also the case for students who did not go to regular high school, but only did exams in front of an examination selection board. The relation between the high school track and a binned version of *wavg* is illustrated in Figure 4.4. It is clear that the weighted average grade of the 'other' category is similar as TSO, but has a higher proportion of *wavg* between 15 and 18.8. In the subsequent analyses, sensitivity analysis is performed. The sensitivity analysis examines whether the results are sensitive to the creation of an 'other' category versus the exclusion of the incomplete cases.

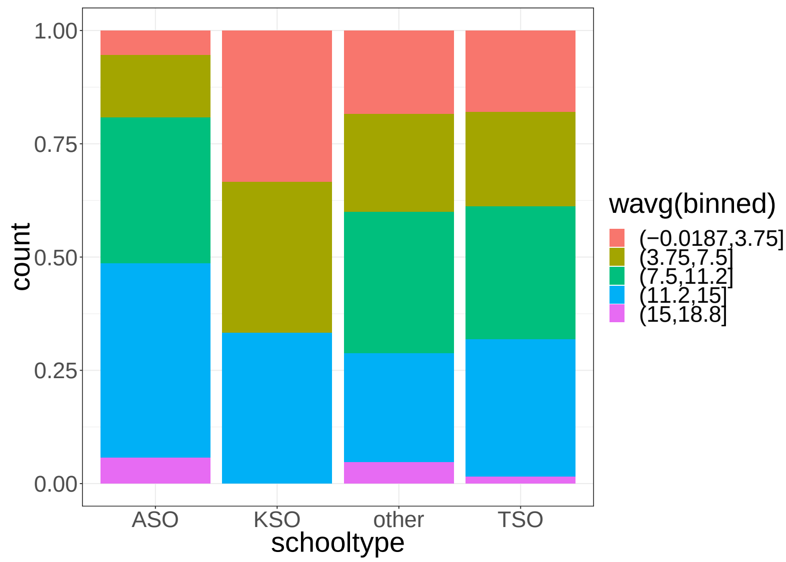


Figure 4.4: Relation between the high school track and a binned version of *wavg*.

A possible explanation of the missingness of the weighted average score is students dropping out early without participating in any exam. These students do not have a weighted average grade. As a consequence, the value of *wavg* is replaced with 0. In the subsequent analyses, the robustness of the conclusions to the replacement with 0 versus the exclusion of incomplete cases is assessed. After the creation of the 'other' category for *schooltype* and the replacement of missing values of *wavg* with 0, a test is conducted to assess the mechanism behind the missingness of the data.

According to Little and Rubin (2002) there are three mechanisms of missing data. Firstly, data can be missing completely at random (MCAR). In the first category missingness does not depend on any observed or unobserved values in the data. In the second mechanism, missing at random (MAR), missingness can depend on observed but not on unobserved values. Missing not at random (MNAR) means that the missingness can also depend on the unobserved values.

Complete case analysis is only valid under the MCAR assumption. Testing for MCAR can entail testing for homogeneity of covariances, means or other parameters between the different missing data patterns (Jamshidian & Jalal, 2010). One test to assess whether there is evidence against MCAR assumption is the Little MCAR Chi-square test (Little, 1988). This test assumes multivariate normality, which is not always valid. Jamishidian and Jalal (2010) propose to first conduct the Hawkins test. When the Hawkins test is significant, it signals either heteroscedasticity of covariances or absence of multivariate normality. As a consequence, when the Hawkins test is significant a non-parametric test is conducted to assess homoscedasticity of covariances.

The method of non-parametric testing is described by Jamishidian and Jalal (2010). Only independence of observations and continuity of the cumulative distribution function is assumed. In addition, it is implicitly assumed the variables are linearly related as imputation techniques are used. After imputation, the equality of the distribution of the test statistic amongst the groups of missing data patterns is tested. If this test is rejected, there is evidence against homogeneity of covariances. As a consequence, there is also evidence against missing completely at random.

The Hawkins test applied to the data leads to a significant result ($p<0.001$). The method of Jamshidian and Jalal (2010) proposes to ensure whether the data is missing completely at random with a non-parametric follow-up test. The null hypothesis of homogeneity of covariances is not rejected with this test ($p=0.293$). There is no evidence against homogeneity of covariances and as a consequence, missing completely at random.

Two options to handle the missing data arise. Firstly, a complete case analysis. In this analysis only the complete cases are used. A second popular option is single imputation. This method will result in valid point estimates, but an overestimation of precision (Molenberghs & Kenward, 2007). Hence, complete case analysis is preferred.

4.3 Linear regression model

A multivariate regression model is fitted to predict the weighted average score (*wavg*) of students. First, the full model is investigated. Next, backward variable selection is applied to reduce the number of variables. The full model is the following:

wavg = $\beta_0 + \beta_1$dbuser + β_2 schooltype + $\beta_3$math.hrs + $\beta_4$math.score + $\beta_5$physics + $\beta_6$https://www.overleaf.com/project/5c1212d5ebab477c3afafd85 chemistry+ $\beta_7$pioneer + $\beta_8$mot + $\beta_9$tmt + β_{10}anx + $\beta_{11}tst$ + β_{12}con + β_{13}advice

The reference categories of the categorical variables are less than 6 hours of mathematics a week for *math.hrs* and less than 60% for *math.score*, *fys* and *chem*. Partially positive advice is the reference category for advice. The reference category of *schooltype* is ASO. For pioneering status, no pioneering background is the reference category.

After fitting the model, the F-statistic denotes that there exists a relationship between the predictors and the response; the null hypothesis that all coefficients are jointly equal to 0 is rejected (F=44.71, p<0.001). In addition, dashboard usage has a significant impact on *wavg* in the full model (p<0.001). Keeping all other regressors fixed, using the dashboard increases the expected value of *wavg* with 1.292 points. The model has a R squared of 0.296, which means that the model can explain 29.6% of the variance of *wavg*. The parameter estimates of the full model are listed in the Appendix C.

Next, stepwise variable selection is applied to reduce the number of variables in order to have a more parsimonious model. Backward variable selection based on the F test is made with α=0.10 as threshold. The final reduced model is the following:

wavg = $\beta_0 + \beta_1$dbuser + $\beta_2$schooltype + $\beta_3$math.hrs + $\beta_4$math.score + $\beta_5$physics + $\beta_6$chemistry+ $\beta_7$pioneer + $\beta_8$mot + $\beta_9$tmt + β_{10}anx + β_{11}tst + β_{12}advice

The R squared of the final model is equal to 0.296. The coefficients, standard errors and p-values of the model are listed in Table 4.6. As expressed in Table 4.6, dashboard usage has a significant impact on *wavg*. Given that all the other variables stay fixed, using the dashboard results in 1.329 units increase in the expected value of *wavg*. In addition, a likelihood ratio test is applied to compare the final model with and without dashboard usage as a predictor. The test statistic is the following:

$$-2(\ell^{(0)}(\boldsymbol{\beta}|\boldsymbol{X}) - \ell^{(1)}(\boldsymbol{\beta}|\boldsymbol{X})) \sim \chi^2(1).$$

The test statistic has a value of 47.091 (p<0.001). It follows that dashboard usage significantly increases the likelihood of the final model.

Table 4.6: *Estimates, standard errors, t-values and p-values of the parameters of the linear regression model.*

	Estimate	SE	t-value	p-value
dbuser	1.329	0.194	t = 6.856	p < 0.001
schooltype BSO	−4.833	3.425	t = −1.411	p = 0.159
schooltype KSO	−4.461	1.959	t = −2.277	p = 0.023
schooltype Other	−1.206	0.327	t = −3.686	p < 0.001
schooltype TSO	−0.925	0.178	t = −5.194	p < 0.001
math.hrs 6-7u	2.231	0.216	t = 10.349	p < 0.001
math.hrs 8u	2.720	0.231	t = 11.786	p < 0.001
math.score>90%	1.966	0.390	t = 5.043	p < 0.001
math.score 60-70%	0.691	0.214	t = 3.219	p = 0.002
math.score 70-80%	1.543	0.225	t = 6.849	p < 0.001
math.score 80-90%	2.092	0.262	t = 7.995	p < 0.001
fys>90%	1.435	0.370	t = 3.873	p < 0.001
fys 60-70%	0.437	0.232	t = 1.885	p = 0.060
fys 70-80%	0.890	0.240	t = 3.702	p < 0.001
fys80-90%	0.928	0.273	t = 3.396	p = 0.001
chem>90%	2.088	0.374	t = 5.588	p < 0.001
chem60-70%	0.587	0.218	t = 2.691	p = 0.008
chem 70-80%	0.680	0.228	t = 2.987	p = 0.003
chem 80-90%	1.204	0.264	t = 4.567	p < 0.001
mot	0.038	0.020	t = 1.950	p = 0.052
tmt	0.085	0.017	t = 4.986	p < 0.001
anx	0.025	0.013	t = 1.861	p = 0.063
tst	−0.035	0.020	t = −1.760	p = 0.079
advice negative	−0.923	0.213	t = −4.329	p < 0.001
advice positive	0.519	0.162	t = 3.203	p = 0.002
advice unknown	−0.500	0.204	t = −2.449	p = 0.015
pioneer pionieer	−1.353	0.202	t = −6.692	p < 0.001
pioneer unknown	−1.183	0.408	t = −2.897	p = 0.004
Constant	1.260	0.684	t = 1.843	p = 0.066
Observations	3,150			
R^2	0.296			
Adjusted R^2	0.290			
Residual Std. Error	3.378 (df = 3121)			
F Statistic	46.825 (df = 28; 3121)			

Assumptions of the reduced model with dashboard usage

Figure 5.4 displays a diagnostic plot of the linear regression model. The picture on the upper left illustrates that the residuals are independent of $\hat{y}$ since there is no pattern present. In addition, the assumption of an expected value of 0 seems valid. The below left plot illustrates the Cook's distance of the observations. The Cook's distance is a measure to assess the amount of influence a certain observation has. More specifically, it measures the influence the i^{th} case has on all n fitted values (Kutner et al., 2005). The formula is

$$D_i = \frac{\sum_{j=1}^{n}(\hat{y}_j - \hat{y}_{j(i)})^2}{pMSE}.$$

with $\hat{y}_{j(i)}$ the estimate of y_j when the i^{th} case is deleted. There is no formal way to test when D_i is large, but Cook (2000) suggests to compare the values to a F-distribution with p and n-p degrees of freedom. In this case, when a D_i surpasses the critical value of 1.48 the observation is deemed influential. As a consequence, there are no highly influential observations present in this model according to this rule. The two spikes in the influence plot represent two students that followed the KSO school type. Since in total only three students followed KSO, they have a high impact on this dummy variable.

The plot below right suggests that the residuals have a constant variance and are thus homoskedastic. The upper right plot illustrates that the residuals deviate from normality. This is also formally tested with the Kolmogorov-Smirnov test and Shapiro-Wilk test. Both tests confirm the visual results of the QQ-plot, (resp. D=0.292, p< 0.001 and W=0.963, p< 0.001). As noted in Section 1.1, inferences are quite robust against deviations from normality (Kutner et al., 2005).

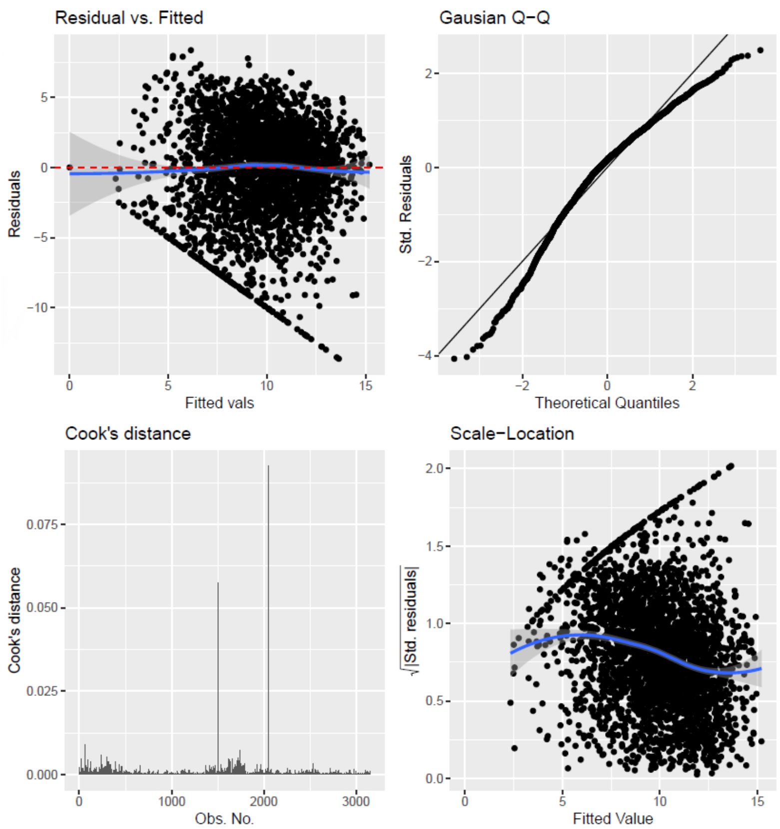


Figure 4.5: Diagnostic plot of the reduced linear regression model with dashboard usage.

The residual plots of the full model are displayed in the Appendix D. The conclusions are similar as the reduced model.

4.3.1 Sensitivity analysis

One can discuss two important decisions relating to the missing data. When *wavg* is missing, the value is replaced with 0. In addition, missing data of the *schooltype* variable is put into an 'other' category. This section examines whether these decisions affect our conclusion: dashboard usage has incremental predictive validity on top of other known predictors. More specifically, the conclusions are compared to conclusions in the case where no values are filled in or 'other' category is created. In addition, the conclusions are contrasted to the scenario where only the 'other' category is created or the missing values of *wavg* are set to 0.

In the case nothing is altered to the data, the same analysis is applied as described in the previous subsection. Hence, with a likelihood ratio test the reduced model with and without dashboard usage are compared. The null hypothesis of no significant increase in likelihood is again rejected ($\chi^2 = 45.53, p < 0.001$). Also when only missing values of *schooltype* are altered and not those of *wavg*, the null hypothesis is rejected ($\chi^2 = 50.948, p < 0.001$). Lastly, when only the missing values of *wavg* are set to 0, the same conclusion is drawn; the null hypothesis also rejected in this scenario ($\chi^2 = 45.913, p < 0.001$). To conclude, the conclusions are not altered by the decisions that have been made about the missing values.

4.4 Multinomial logistic regression model

Multinomial logistic regression is a special cases of a generalized linear model. The goal of this model is to predict class membership of multiple non-overlapping classes. The reference category is the category of students with > 80% CSE . The model is the following:

$$ln\frac{P(30\% < CSE)}{P(CSE > 80\%)} = \boldsymbol{X\beta},$$

$$ln\frac{P(30\% < CSE < 50\%)}{P(CSE > 80\%)} = \boldsymbol{X\beta},$$

$$ln\frac{P(50\% < CSE < 80\%)}{P(CSE > 80\%)} = \boldsymbol{X\beta}.$$

The reference categories are equal to those of the linear regression model in the previous section. Variable selection is made on the model and afterwards the multinomial logistic regression model with dashboard usage is compared to the multinomial logistic regression model without dashboard usage.

Firstly, models with and without dashboard usage are fitted. Before variable selection the Nagelkerkes R squared for the model with dashboard usage equals 0.302. The model without dashboard usage has a Nagelkerkes R squared of 0.297. This statistic does not measure the proportion of explained variance, but higher values imply a better model fit (Garson, 2014).

Next, variable selection is applied. Of course it is taken into account that several parameters reflect the effect of one covariate. Stepwise variable selection is applied on the models by use of the chi-square statistic. Each time the largest of these statistics is added, until no variables are significant at the 0.10 level. After entering each variable, all the variables in the model are candidate for removal with as a threshold a significance level of 0.15. Firstly, stepwise variable selection is applied on the model without the dashboard usage covariate. The Nagelkerke R squared of this model is 0.295. Out of the 12 possible predictor variables, 9 stay in the model: *schooltype*, *math.hrs*, *math.score*, *fys*, *chem*, pioneer, time management, concentration and the advice of the high school teachers.
Secondly, the model with dashboard usage is analysed. The final model after variable selection has a Nagelkerke R squared of .300. The final covariates are dashboard usage, *schooltype*, *math.hours*, *math.score*, *fys*, *chem*, pioneer, time management, concentration and the advice of the high school teachers. Notice that the predictors are not exactly the same as in the linear regression model. In this model concentration is included, while in the linear regression model motivation, test strategy and anxiety are present.

For the interpretation, the exponent of the coefficient is taken. The interpretation of the parameter estimates is that the odds for an one-unit increase in the variable is e^{β} times larger for being in the category of the numerator versus the 'CSE >80' category. Of course, this is on the condition that the other variables stay fixed. For the odds of being in the '30%>CSE' versus the 'CSE>80%' group, the exponent of the coefficient of dashboard usage is 0.538. This shows that for students that used the dashboard, the odds are multiplied a factor of 0.538. The coefficients of dashboard usage for the odds of being in the '30<CSE<50' and '50<CSE<80' categories versus the 'CSE>80' category are not significantly different from 0. As a consequence, there is no interpretation for the point estimates. The coefficients, standard errors, p-values and type III tests of fixed effects are listed in the Appendix E.

Lastly, a nested likelihood ratio test is applied to test whether the final model with dashboard usage has a significantly higher likelihood than the final model without dashboard usage. The test statistic is the following

$$-2(\ell^{(0)}(\boldsymbol{\beta}|\boldsymbol{X}) - \ell^{(1)}(\boldsymbol{\beta}|\boldsymbol{X})) \sim \chi^2(1).$$

The test statistic has a value of 18.596, which corresponds to a p-value smaller than 0.001. This denotes that the inclusion of dashboard usage yields in a significant increase in the likelihood of the model.

Alternatively, the ordinality of the response variable enables the application of proportional (cumulative) odds models. An advantage of this model is an increase of precision since there is taken advantage of the fact that the response is ordinal. But these models make a proportional odds assumption; the covariates have the same impact on the odds regardless of the split of the data (O'Connell, 2006). For example, it would imply that the coefficients that describe the relationship between the 'CSE <30' category and the category of '30<CSE' are the same coefficients that describe the relationship between the 'CSE< 50' and the '50<CSE' groups. In the multinomial logistic regression model, it is clear that this assumption is not valid. In addition, the proportional odds test rejects the null hypothesis of proportional odds ($\chi^2 = 901.654$, p-value<0.001).

4.4.1 Evaluation

This section evaluates the reduced model with dashboard usage. For this purpose, goodness-of-fit tests are considered. These tests compare observed with estimated probabilities. The deviance and pearson statistic are not valid since not all regressors are categorical. As a consequence, the asymptotics of these tests do not apply. A goodness of fit test for the multinomial model that allows continuous regressors is the Hosmer-Lemeshow statistic (Fagerland & Hosmer, 2012). More specifically, the outcome variable Y contains c unordered categories. Next, the data is divided into g groups by $\hat{\pi}_i$. The test statistic is calculated from the observed and estimated counts in the contingency table

$$C_g = \sum_{k=1}^{g} \sum_{j=0}^{c-1} (O_{kj} - E_{kj})^2 / E_{kj} \sim \chi_{df=(g-2)x(c-1)}.$$

The null hypothesis of this test is that the observed model is correct. Applied to this model, the null hypothesis is not rejected (χ^2 = 29.185, df = 24, p-value = 0.213). This result demonstrates that there is no evidence against a good fit of the model.

4.4.2 Assumptions

According to McFadden et al. (1976), an assumption is the independence of irrelevant alternatives assumption (IIA). As the distinct categories do not represent decisions and each student can only fall into a single category, the independence of irrelevant alternatives assumption holds. Statistical tests for this assumption exist, but these test are unsatisfactory for real-life data (Cheng & Long, 2007).
Garson (2014) states multiple assumptions of multinomial logistic regression. One assumption, the absence of multicollinearity is already checked in Section 3.2.2. A second assumption is the absence of perfect separation. When perfect separation occurs, logit coefficients tend go to infinity with large standard errors. As this is not the case in the estimates of the model, the assumption of no separation is valid. The last two assumptions are absence of outliers and over- or underdispersion. The nonsignificant goodness-of-fit statistic suggests no evidence against the latter two assumptions.

4.4.3 Sensitivity analysis

In this subsection a sensitivity analysis is performed. Just as in the linear regression model, an 'other' category is created for the missing values of *schooltype*. In addition, the missing values of *wavg* are set to 0. The reason is that both variables are not self-reported, but stem from the database of KU Leuven. This indicates that reason of misingness is the result of respectively a foreign high school system or dropout. Still, it is checked whether the conclusions depend on the decisions about the missing data that are made.

First, the same analysis as above is conducted on the dataset without altering the missing values. The likelihood ratio test of the reduced model with versus the reduced model without dashboard usage is still significant (χ^2=17.956, p<0.001). This is also the case for the data where only the 'other' category for missing values of *schooltype* was created (χ^2=26.123, p<0.001). Lastly, the dataset where only the missing values of *wavg* are set to 0 is analysed. In this data the likelihood ratio test also demonstrates that dashboard usage significantly increases the likelihood (χ^2=18.596, p<0.001). It follows that the conclusion does not depend on the decisions that have been made about the missing data.

4.5 Prediction students at risk

A last outcome variable is whether the student is at risk. A student at risk is defined as a student that has a weighted average grade lower than 8.5 (Mothilal, 2018). Hence, the *atrisk* variable is created by binning *wavg*. As a consequence, missing values of at risk are set to 1. A logistic regression model is fitted to predict whether the student is at risk. Next, the incremental predictive validity of dashboard usage is determined. The logistic regression model is fitted with a logit link and the same predictors as in the previous models:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1\text{dbuser} + \beta_2\text{schooltype} + \beta_3\text{math.hrs} + \beta_4\text{math.score} + \beta_5\text{physics} + \beta_6\text{chemistry} + \beta_7\text{pioneer} + \beta_8\text{mot} + \beta_9\text{tmt} + \beta_{10}\text{anx} + \beta_{11}\text{tst} + \beta_{12}\text{con} + \beta_{13}\text{advice}.$$

π_i equals the expected probability of being at risk. In order to have a realistic view on the predictive capacities, a test set is created. More specifically, 40% of the total data is at random assigned to a test set and not used for fitting the logistic regression model. The training data contains 47.69% students at risk, thus there is no need for adjustment of misclassification costs or over- or undersampling approaches.

First, backward variable selection is applied with the χ^2 statistic. The α is set to 0.10. The reduced model is the following:

$$logit(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1\text{dbuser} + \beta_2\text{schooltype} + \beta_3\text{math.hrs} + \beta_4\text{math.score} + \beta_5\text{chemistry} + \beta_6\text{pioneer} + \beta_7\text{tmt} + \beta_8\text{advice}.$$

The Nagelkerke pseudo R squared equals 0.245. Table 4.7 lists the coefficients, standard errors and p-values of the model.
The exponent of the coefficient of dashboard usage equals 0.537. This means that using the dashboard results in an increase of the odds of being at risk with a factor of 0.537, keeping all other independent variables fixed. To assess the impact of dashboard usage, a likelihood ratio test is applied to the reduced model with and without dashboard usage. The test demonstrates that dashboard usage results in a significant increase in likelihood ($\chi^2 = 15.984$, p<0.001).

Table 4.7: *Estimates, standard errors, t-values and p-values of the logistic regression model.*

	Estimate	SE	t-value	p-value
dbuser	−0.621	0.155	t = −4.013	p < 0.001
schooltype BSO	12.304	535.411	t = 0.023	p = 0.982
schooltype KSO	15.323	349.190	t = 0.044	p = 0.965
schooltype Other	0.722	0.280	t = 2.575	p = 0.011
schooltypeTSO	0.445	0.152	t = 2.926	p = 0.004
math.hrs6-7u	−0.918	0.174	t = −5.265	p < 0.001
math.hrs8u	−1.440	0.194	t = −7.425	p < 0.001
math.score>90%	−1.114	0.408	t = −2.729	p = 0.007
math.score60-70%	−0.559	0.177	t = −3.160	p = 0.002
math.score70-80%	−0.956	0.188	t = −5.095	p < 0.001
math.score80-90%	−1.402	0.238	t = −5.886	p < 0.001
chem>90%	−1.004	0.410	t = −2.448	p = 0.015
chem60-70%	0.115	0.180	t = 0.641	p = 0.522
chem70-80%	−0.010	0.182	t = −0.056	p = 0.956
chem80-90%	−0.457	0.226	t = −2.027	p = 0.043
tmt	−0.062	0.013	t = −4.913	p < 0.001
advice negative	0.682	0.177	t = 3.849	p < 0.001
advice positive	−0.444	0.144	t = −3.091	p = 0.002
advice unknown	0.447	0.169	t = 2.644	p = 0.009
pioneer	0.429	0.182	t = 2.359	p = 0.019
pioneer unknown	0.647	0.333	t = 1.943	p = 0.053
Constant	2.897	0.408	t = 7.102	p < 0.001
Observations	1,953			
Log Likelihood	−991.519			
Akaike Inf. Crit.	2,027.039			

Next, the predictive accuracy is assessed on the hold-out test set. A confusion matrix is created for both the reduced model with and without dashboard usage. The confusion matrix of the model with dashboard usage is presented in Table 4.8, the confusion matrix of the model without dashboard usage in Table 4.9.

Table 4.8: *Confusion matrix of the model with dashboard usage.*

	Predicted 0	Predicted 1	Total
Actual 0	800	111	911
Actual 1	236	159	395
Total	1,036	270	1,306

Table 4.9: *Confusion matrix of the model without dashboard usage.*

	Predicted 0	Predicted 1	Total
Actual 0	784	127	911
Actual 1	236	159	395
Total	1,020	286	1,306

The model without dashboard usage contains 16 more false negatives. Whether the difference is also statistically significant is assessed with the McNemar's chi-squared test. The null hypothesis of marginal homogeneity means that both models have the same accuracy. The test statistic is the following:

$$\chi^2 = \frac{(b-c)^2}{b+c} \sim \chi^2_{df=1}.$$

where b equals the number of times where model 1 was right and model 2 wrong and c the number of observations where model 2 is wrong and model 1 right. Let model 1 denote the model with dashboard usage and model 2 the model without dashboard usage. In this case b is equal to 39 and c is equal to 23. The difference in accuracy is statistically significant (χ^2 = 4.129, p=0.042). This implies that the null hypothesis of equal predictive accuracy of both models is rejected.

4.5.1 Evaluation

The goodness-of-fit is assessed in the same way as in the multinomial logistic model, with the Hosmer-Lemeshow statistic. The Hosmer-Lemeshow statistic with 50 groups equals 53.045, which does not provide evidence against the null hypothesis of a good fit (p=0.286). When the number of groups is varied (g= 5, 10, 25, 75, 100, 150), the result stays nonsignificant.

4.5.2 Assumptions

Since the coefficients do not approach infinity, there is no evidence of separation. The absence of multicollinearity is checked in Section 3.2.2 and is valid. The goodness-of-fit statistic indicates that there is no over or under dispersion present. The plot of the Cooks distances is shown in the Appendix F. The plot expresses that there are no highly influential observations.

4.5.3 Senstivity analysis

When the analysis is conducted on the dataset with no alterations of the missing data, dashboard usage results in a significant increase of likelihood in the reduced model (χ^2=15.14 , p< 0.001). A contingency table provides the means to evaluate the predictive accuracy. Model 1 was right in 23 cases where model 2 was wrong and model 2 was correct in 16 cases where model 1 was wrong. But the result is not significant; the McNemar's test points out that there is no evidence against equal predictive accuracy of the models with and without dashboard usage (χ^2=2.522, p=0.2623).

The same result applies for the model where only the missing values weighted average is set to 0. The likelihood ratio test is significant (χ^2=16.933, p< 0.001), while the McNemar's test is not (χ^2=2.522, p=0.112). Lastly, in the scenario where only the missing values of school type are handled, the likelihood ratio test is significant (χ^2=14.128 ,p< 0.001) and the McNemar's test not (χ^2=0.490, p=0.484).

4.6 Discussion

The dashboard that was sent out early in the academic year has predictive value for the weighted average grade later in the academic year on top on other known predictors. If a student uses the dashboard, his expected weighted average grade increases with 6.645 percentage points, keeping other covariates fixed. Furthermore, it also increases the odds of being at risk by a factor of 0.537. Hence the covariates demonstrate that the impact of using the dashboard is non-negligible. In addition, using the dashboard reduces the odds of being in the '< 30 CSE or dropout' group versus the '> 80 CSE' group by half, given that the other covariates stay the same. This finding is supported by the literature. Broos, Verbert, Van Soom et al. (2018) found that the most successful students in terms of CSE had a higher click-through rate compared to students with a medium CSE. The latter had in turn higher click through rates than students with a low CSE.

The analysis also confirms the findings of the literature study that demographic variables are important predictive factors. The hours of mathematics in high school, school type, score of mathematics and chemistry in high school and first-generation student status are present in each reduced model. Past research (Pinxten et al., 2017; Pinxten & Hockicko, 2016) has proved that time management, test strategy and motivation correlate moderatly with GPA. Still, only motivation and time management have incremental predictive validity on top of prior achievement (Pinxten et al., 2017). The present study only confirms the results of time management; time management is the only LASSI skill that is significant in all the analyses. Motivation is only present in the reduced linear regression model and not in the other reduced models. This is possibly related to the lack of portability of learning analytics models and differences in the available variables.

This finding has practical implications. The dashboard was sent out at the start of the academic year and thus long before the students are evaluated. As a consequence, dashboard usage potentially provides a tool to flag students at risk in the early academic year. Because collecting this variable requires limited resources and other indicators are scarce, it is particularly interesting. Hence, this study also addresses a gap in the learning analytics research addressed by Conijn et al. (2017). The authors state that there is lack of early predictors in learning analytics.

A limitation is the high amount of missing values. The effect of the choices made on the missing values are scrutinised in a sensitivity analysis. Still, better techniques exist to handle those missing values. A possible alternative is multiple imputation. Missing values are in this method replaced multiple times with draws out of an appropriate Bayesian predictive distribution. In this way, imputation of data that is missing at random does not artificially increase the precision. This technique also allows for a sensitivity analysis to check the robustness of the conclusions to missing not at random (Molenbergs & Kenward, 2007).

It would be imprudent to conclude that the increased achievement is the sole result of using the dashboard. Instead, it likely signals a certain disposition of a student that is linked to other factors. A possible underlying construct is student engagement. More specifically, dashboard usage possibly relates to cognitive and behavioural engagement. In the theory of Fredricks et al. (2004), cognitive engagement is defined as psychological investment in learning. Cognitive engagement also entails self-regulation and being strategic. A possible interpretation of visiting the dashboard is a manifestation of investment to enhance learning skills. In addition, it signals that the student keeps abreast of the information the university provides. This is related to behavioural engagement, which encompasses participation in academic activities. These constructs correlate moderately positive to academic success (Lei et al., 2018). Further research is required to investigate the relation between student engagement and dashboard usage.

In addition, further research must be done to investigate whether the findings of this study are portable to other programmes than science and engineering and other universities. As Conijn et al. (2017) noted in their study, lack of portability is an issue in learning analytics.

Chapter 5

Dashboard activity

A second objective of this thesis is investigating the activities of the students on the dashboard. An important part of this assignment is feature engineering to define fine-grained measures of dashboard activity. Next, these measures are related to the weighted percentage in September.

The dataset contains raw system output of actions that the students make. More specifically, the data contains information about the actions on the LASSI dashboard of 2155 science and engineering students. The actions are situated in the academic year 2017-2018. Since only 6.635% of the observations have missing values, missing data does not impose an issue in this dataset. Only 1.439% of the values of weighted percentage in September are missing. According to Bennett (2001), less than 10% missing data does not bias the results.

5.1 Event logs of the dashboard

Table 5.1 gives an overview of the types of event logs and logdata. When students click on the link, they first encounter a page with a general introduction. This introduction explains the intend and structure. Next they can click on five tabs, with each tab representing a LASSI scale (**tab**). The order of the tabs is the following: concentration, motivation, anxiety, test strategy and time management. Each tab follows the same structure. Students can access cards (**active card**) by clicking on the pane. There is one exception, when students have not clicked on a pane yet, they do not have to click on first pane. This pane contains a definition of the selected skill and the membership of the student in one of the five norm groups. In the plane below, students of certain programs can compare their score. More specifically, the CSE of the previous first-year students of their program is displayed in relation to the score on the LASSI scale. The third pane contains a collection of tips, recommended actions and references to remediate or improve their learning skill. Appendix G contains screenshots of the dashboard.

After a certain time a window pops up where the student can indicate how useful and clear the dashboard is (**feedback**). Alternatively, the student can skip the feedback. Students can also access other links via the dashboard (**link**). These links provide referrals to the academic counselor of KU Leuven, career and counseling services of TU Delft and a website that offers tips, training, exercises and videos for the specific learning skill. When students are inactive for 60 seconds, a ping is send to the system (**ping**). In addition, refreshes are logged (**refresh**). Appendix H contains the data of a random student. For privacy purposes, 'r1234567' replaces the identifier of the student.

Table 5.1: *Event logs of the dashboard.*

Logevent	#instances	#variants	variants
Active card	21073	15	card-intro, card-anx-now, card-anx-prev,card-anx-tips,...
Feedback	3452	12	clear-1, clear-2,clear-skip,useful-1,useful-2, useful-skip,..
Link	110	9	link academic counselor, link for extra information,...
Ping	58300	1	60-sec
Refresh	7227	1	x
Tab	37094	6	0,1,2,3,4,5

5.2 Time analysis

5.2.1 Analysis of case duration

A first analysis is the time the students spend on the dashboard. An issue in the data is the possibility that students leave the tab of the dashboard open but do other things. The system does not log these actions and hence, this is not visible in the data. This issue is solved by truncating the maximum time spend on a tab at 5 minutes. Of course this number is subjective. Students that visit the dashboard spend on average 2.874 minutes on the dashboard. The median is 1.800 minutes. In Figure 5.1 the distribution of the amount of seconds spend on the dashboard is displayed. It is clear that the data is highly skewed. The most extreme outliers are not shown in the figure; these are students that spend 44.933 and 42.233 minutes on the dashboard. Students that did not visit the dashboard receive a value of 0 minutes in the subsequent analysis.

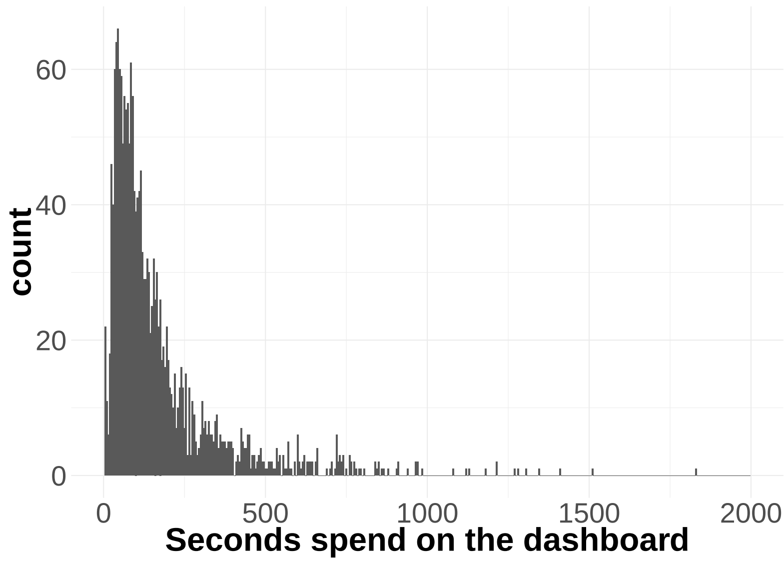


Figure 5.1: The amount of seconds that visitors spend on the dashboard.

The relation between the weighted percentage in September and the time spend on the dashboard is shown on the left of Figure 5.2. The blue line, the fit of the linear regression model, demonstrates that there is increasing trend present. Very few observations are present for high amounts of time spend on the dashboard. Nevertheless, they have a high influence on the slope. More specifically, they pull the slope down. Furthermore, the LOESS fit expresses that the increasing trend is only present for ordinary times spend on the dashboard. When the time spend takes extreme values, increases of time do not correspond with higher weighted percentages. As a consequence, the domain is restricted.

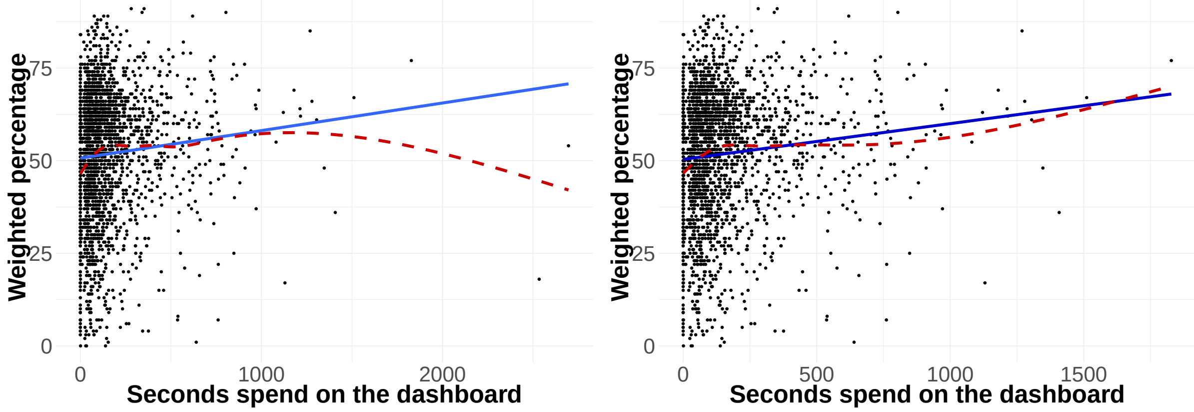


Figure 5.2: The relation between the amount of seconds that visitors spend on the dashboard and the weighted percentage in September. The left plot displays the whole domain, the right plot shows a restricted domain. The red line indicates the fit of locally estimated scatterplot smoothing, the blue line the fit of linear regression.

The right plot of Figure 5.2 presents a plot of the restricted domain. In this plot univariate outliers of the time spend on the dashboard are deleted. Outliers are in this case defined as observations that do not follow the general trend of the data. As a consequence, only the domain [0,2534] is considered. The sample size of this subset equals 2153. Since this domain entails 99.99% of the data, this is the trend of the majority of the cases.

A linear regression model is fitted to assess if the impact of time spend on the dashboard is significant in the restricted domain. The slope of the model equals 0.010. This means that for each minute spend, the expected weighted percentage increases with 0.600 percentage points. This result is statistically significant ($t=4.847$, $p<0.001$). The intercept equals 50.344, which models the expected weighted percentage for somebody that does not visit the dashboard. The fit is illustrated on the right of Figure 5.2 by the blue line. Appendix I displays the assumptions. The Cooks distance plot has one peak. This peak corresponds to a student that visited the dashboard for 1130 seconds, but his weighted percentage is 17. Since the Cooks distance is not alarming ($d=0.045$), alternation or deletion of the observation are not considered. The Shapiro-Wilk test suggests the residuals deviate from normality ($W= 0.969$, $p<0.001$), but no transformations are considered. The arguments are noted in Section 3.1.

It is important to note that predictions based on this simple linear regression model are not solid. The R squared equals 0.011. This means that the fit can only explain 1.10

% of the variance in the weighted percentage. The goal of this regression model is to illustrate the impact of time spend on the dashboard on weighted percentage in general, not to make predictions.

For exploratory purposes, a linear regression model is fitted on the whole dataset. The slope is a lot lower (β_1=0.007), but still significant (p<.001).

5.2.2 Analysis of time points

This section analyses the moment the student first visited the dashboard. The bar plot of the proportions is presented in Figure 5.3. There are two peaks in the bar plot. The most popular moments are 13 and 14 hour, closely followed by 21 and 22 hours. The reason for these peaks is probably because students of different programs received their dashboards at different time points. Engineers and engineers-architects received their dashboard at 13h, while several science programs received their link at 21h. Since the differences in expected weighted percentage of the program groups introduce noise to the data, separate analyses are performed.

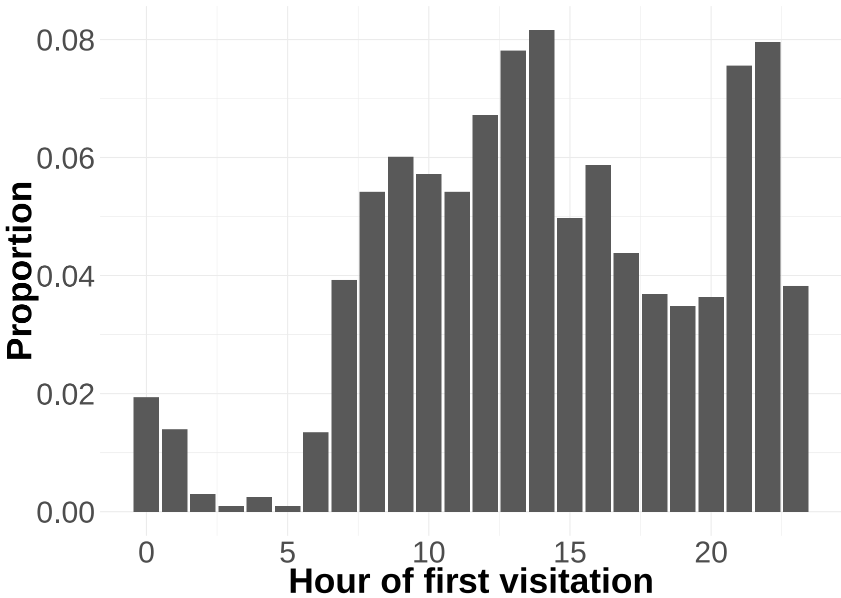


Figure 5.3: Bar plot of the proportions of first visitations that occurs at each hour.

Engineering and engineering-architecture

Engineers and engineers architects received their link to the dashboard at 13h. The sample size of this group equals 516. Since the time of the day is not on ratio scale, it is categorised into categories. The blocks and the frequencies are displayed in Table 5.2. The categorisation takes the size of the groups into account. In addition, the first group starts at 13h, when students receive the dashboard. Missing values of hours signal that the student did not visit the dashboard. Since this is informative, it is coded as 'None'. Now, the relation with the weighted percentage is explored. The boxplot is displayed in Figure 5.2.2. The differences in weighted percentage seem small.

Next, it is assessed whether the the weighted percentage differs significantly across the time categories. The Shapiro-Wilk test denotes that weighted percentage is not normally distributed (p<0.001). It follows that Anova is not appropriate. In addition, the Kruskal-Wallis test assumes that the shapes of the distributions of the ranks are similar. The plot of the distributions of the ranks is shown in the Appendix J. The plot suggests that the shapes of the distributions of the ranks are unequal. The median test provides a non-parametric alternative. The null hypothesis states that all medians are equal, the alternative hypothesis is that at least one of the medians is different from the others. The median test results in no rejection of the null hypothesis (χ^2=4.414, p=0.621). As a consequence, no contrasts are assessed.

Table 5.2: *Time categories and frequencies.*

Time	Hours	Frequency
None	No visitation	46
13h	13	67
14h	14	77
Afternoon	15,16,17	88
Evening	18, 19,20,21,22	86
Night	22,23,24,1,2,3,4,5,6	64
Morning	7,8,9,10,11,12	127

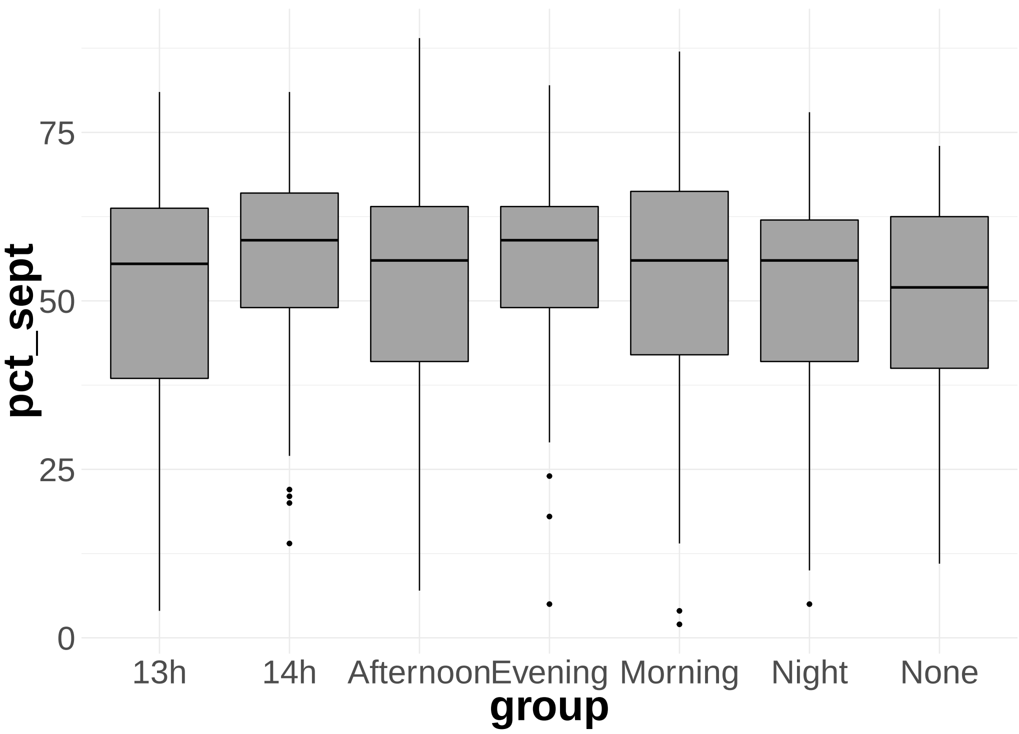


Figure 5.4: Boxplot of weighted percentage of September by time the engineering and engineering-architecture first visited the dashboard.

Science and engineering technology programs

Multiple programs received the link to the dashboard at 21h. These programs include bio-engineering, biochemistry, biology, physics, geography, engineering technology, informatics and mathematics. The sample size equals 1337. Again the students are divided into groups based on the moment they visited the dashboard. The frequencies are presented in Table 5.3.

Figure 5.5 illustrates the relationship between the moment of the first visitation and weighted percentage. The figure demonstrates that there are differences present. The assumption of normality is rejected by the Shapiro-Wilk test ($p<.001$), the assumption equal variances is rejected by the Brown-Forsythe test ($p=0.001$) and the assumption of equal variance of the ranks is rejected by the Brown-Forsythe test ($p<0.001$). A plot of the distribution of the ranks is presented in the Appendix J. The median test does not make any distributional assumptions and thus provides a valid alternative. The test rejects the null hypothesis of equal medians ($p< 0.001$).

Next, the use of pairwise contrasts enables investigating which differences are significant. The Holms correction is applied to correct for multiple comparisons. Appendix K displays the results. Students who visit the dashboard at 22h have a higher median weighted percentage than students that visit the dashboard at the morning, noon or do not visit the dashboard at all. Furthermore, students who do not visit the dashboard have a lower median weighted percentage compared to students that visit the dashboard in the early morning or evening. In addition, students who visit the dashboard in the early morning have a higher median than students that visited the dashboard in the morning. Lastly, the median weighted percentage is significantly higher of students who visit the dashboard in the evening compared to the morning and noon.

Table 5.3: *Time categories and frequencies.*

Time	Hours	Frequency
None	No visitation	46
21h	21	126
22h	22	131
Night	23,24,1,2,3,4,5,6,7	150
Early morning	8,9	168
Morning	10,11	139
Noon	12,13	146
Afternoon	14,15,16	197
Evening	17,18,19,20	177

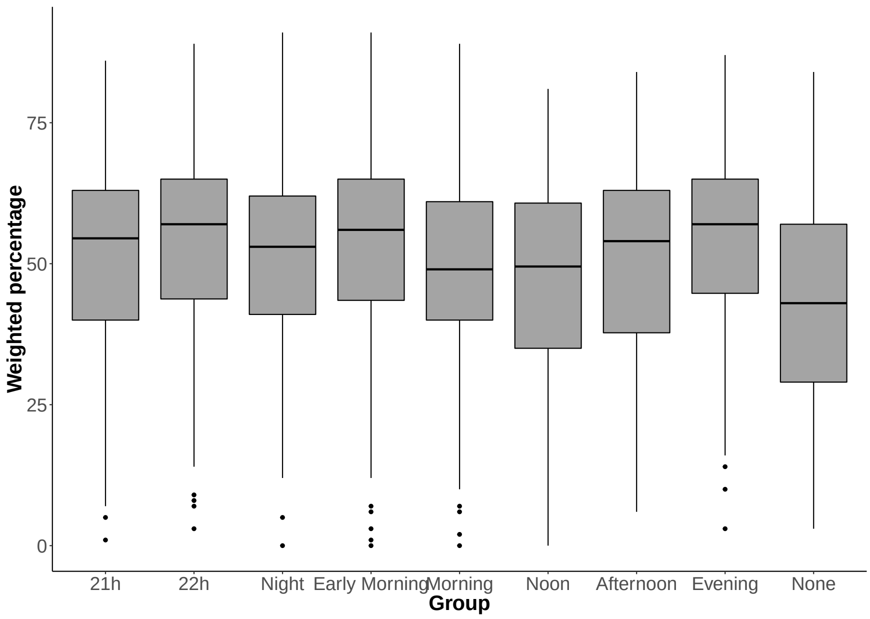


Figure 5.5: Boxplot of weighted percentage of September by time the science and engineering technology students first visited the dashboard.

5.2.3 Lag between sending and visiting the dashboard

Not every student immediately accesses the link to the dashboard. Even more, there is a large variance in the lag between sending out the link and the visitation of the dashboard in minutes (sd=443.9301). The median lag equals 16.62 hours. The distribution is displayed in Figure 5.6. The x-axis is limited to 120 hours, or 5 days. Still, 583 students waited more than 5 days to access the dashboard.

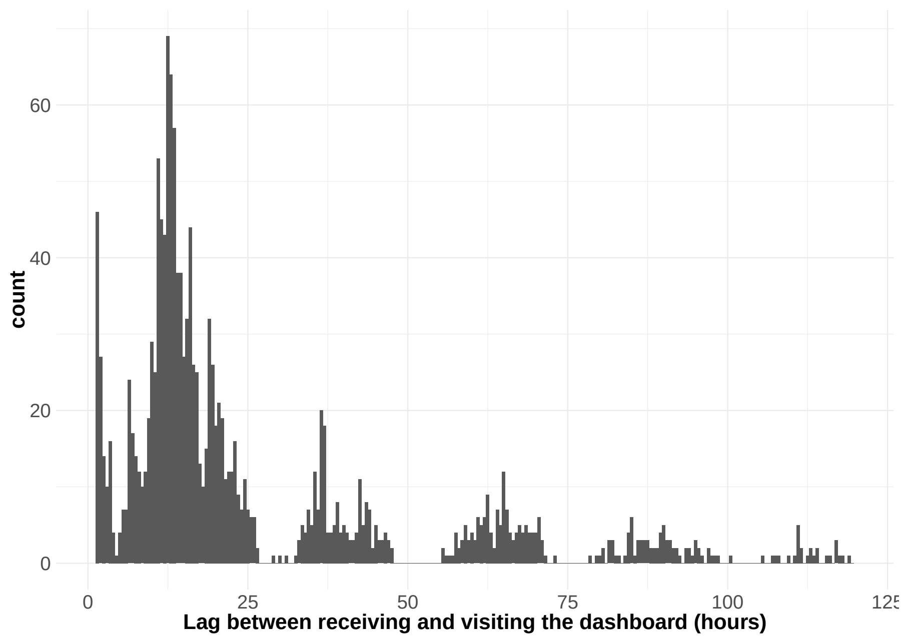


Figure 5.6: The lag between receiving and visiting the dashboard.

Next, the relation between the lag and weighted percentage is explored. The scatterplot of the full dataset is presented in the left of Figure 5.7. The LOESS fit expresses that the relationship is in negative at first. For very large values, the relationship becomes positive again. Still, the latter positive relationship is based on a small set of data. As a consequence, the domain is restricted. In addition, the dataset only consists of dashboard users.

In order to restrict the domain, the 3 IQR rule of Tukey (1977) is applied. Univariate outliers are defined as datapoints that are further away than 3 times the interquartile distance from the first or third quartile. One can discuss alternative methods, but this method has the advantage of robustness. The restricted domain entails [0,210.37], which contains 82.088 % of the datapoints (n=1769). The right plot of Figure 5.7 displays the relationship between the the lag and weighted percentage. The relationship is negative and far more linear. Still, the plot suggests that there is a lot of residual variance present.

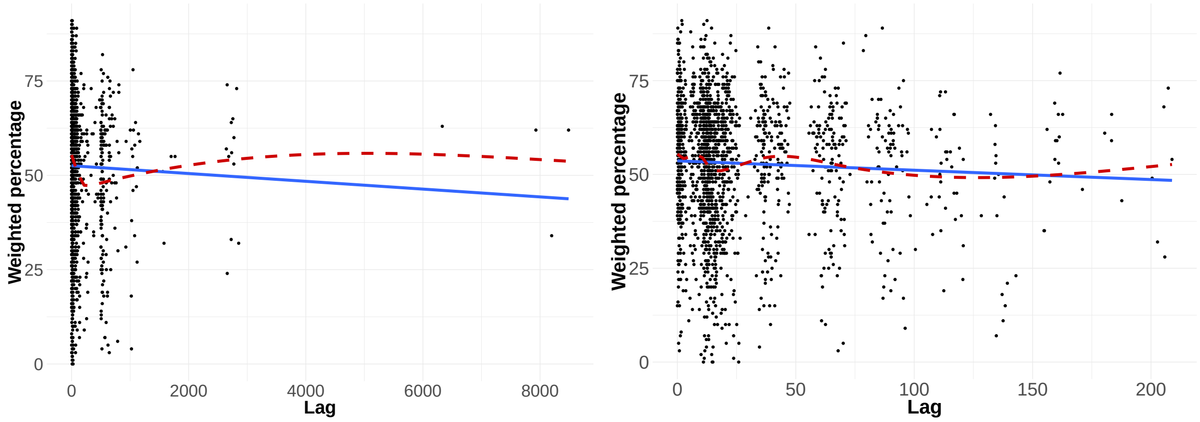


Figure 5.7: Scatterplot of the weighted percentage by the lag in hours. The right plot entails the whole sample, the right plot excluded the univariate outliers. Blue lines indicate the linear regression model, red lines weighted scatterplot smoothing

A linear regression model is fitted to predict the weighted percentage with the lag. The intercept equals 53.614 which models the weighted percentage of somebody that immediately visited the dashboard. The slope is -0.025. The slope is significantly different from 0 (t=-2.004, p= 0.045). This value means that for every extra hour between receiving and visiting the dashboard, the expected weighted percentage decreases with 0.025 percentage points. The R squared equals 0.002, which denotes that the model can only explain 0.2% of the variation in weighted percentage. The model is fitted to discover broad general trends instead of making predictions. The assumptions, except for normality, are valid and presented in the Appendix L. The Shapiro-Wilk test rejects the null hypothesis of normality (W=0.969, p< .001), but no transformations are considered. Section 3.1 provides the arguments of this decision.

5.3 Analysis of actions

5.3.1 Clicking on the tips

This subsection investigates the impact of clicking on the tips on weighted percentage. Only 6.821% of the students clicked on all the tips. Nevertheless, 43.758% of the students clicked on at least one tip. This is 46.915% of all dashboard users. The proportion of students that looked at the different tips is illustrated in Figure 5.8. A plausible reason why the concentration tip is much more popular than the other tips is because concentration is the first tab. Possibly an order effect is present. In the academic year 2018-2019, the ordering of the tips over the tabs is at random. In this case, the proportions of students that visited the different learning skill tips is similar. The bar plot is presented in the Appendix M.

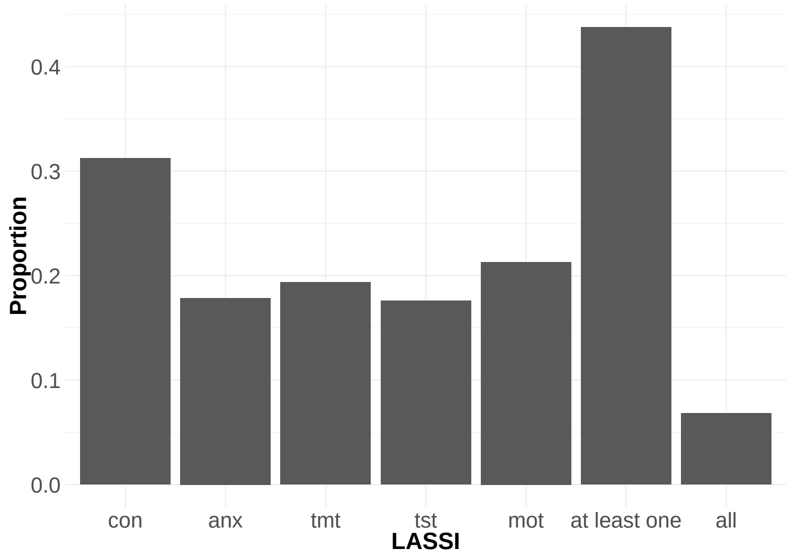


Figure 5.8: Proportions of students that clicked on the tips.

First, the effects of looking at the tips separately, looking at all tips or looking at at least one tip are explored in a boxplot (Figure 5.9). The persons that did not look at the tip or tips are subdivided into dashboard users and non dashboard users. The graph indicates that there are small differences in weighted average percentage between the groups.

Next, statistical analyses are conducted to investigate if the effects are significant. As noted in the literature study (Broos et al., 2017), persons with a lower score on a certain learning skill are more likely to access the tips of the corresponding skill. This effect is also present in this data. In addition, the authors found that students with better learning skills are more likely to visit the dashboard. This effect is significant in this data for motivation and test strategy. Furthermore, Pinxten et al. (2017) found that four of the learning skills of the present study correlated significantly with weighted GPA. To control for the effect of the learning skill, the learning skill is included as a covariate. In addition, interaction effects with the skill are included in the model. The interactions with dashboard usage and visitation of the tips are explored with heat maps in the Appendix N. The heat maps suggest that the effects are additive instead of interacting.

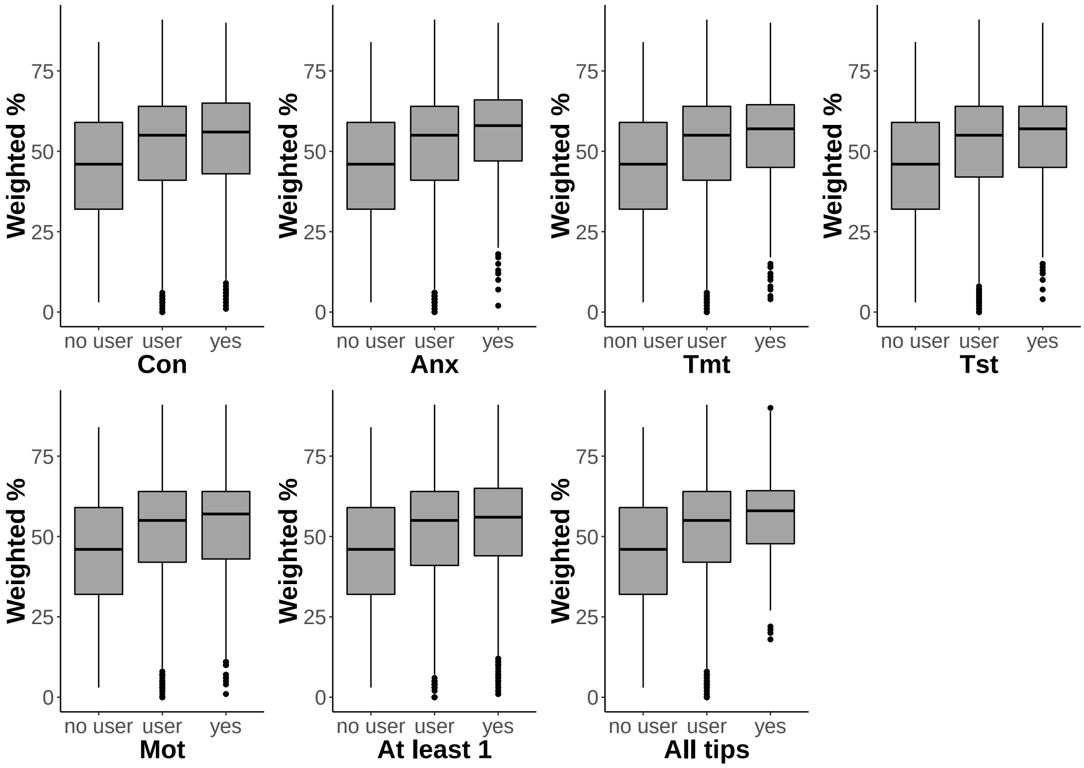


Figure 5.9: Boxplot of the effect of looking at the tips separately, looking at all tips or looking at at least one tip on weighted percentage.

The following regression models are fitted:

$$\text{Pct\_sept} = \beta_0 + \beta_1\text{skill} + \beta_2\text{dbuser} + \beta_3\text{tips} + \beta_{12}\text{skill*dbuser} + \beta_{13}\text{skill*tips}.$$

Hence, the model for students that did not visit the dashboard is the following

$$\text{Pct\_sept} = \beta_0 + \beta_1\text{skill}.$$

Students that visit the dashboard, but not the learning skill tips have the following model

$$\text{Pct\_sept} = (\beta_0 + \beta_2) + (\beta_1 + \beta_{12})\text{skill}.$$

Students that visit the tips of the skill are modelled in the following way

$$\text{Pct\_sept} = (\beta_0 + \beta_2 + \beta_3) + (\beta_1 + \beta_{12} + \beta_{13})\text{skill}.$$

After fitting the models, stepwise variable selection is conducted based on the F-test with as threshold α=0.10. Each of the reduced models only contains the main effects. It follows that visiting the dashboard or the tips does not influence the effect of the learning skill on the weighted percentage in September.

Table 5.4 displays the model of concentration. The F-statistic reveals that all predictors are jointly significant (F=30.267, p<.001). Looking at concentration tips has a significant effect on weighted percentage on top of the learning skill and dashboard usage. Looking at the tip increases the expected weighted percentage with 2.217 percentage points, given that the other covariates stay fixed.

Table 5.4: *Linear regression model of concentration.*

Covariate	coefficient	SE	t-value	p-value
concentration	0.609	0.078	7.786	<.001
dbuser	6.797	1.550	4.386	<.001
con_tips	2.178	0.818	2.664	0.008
Constant	28.108	2.570	10.938	<.001

Next, the effect of looking at the anxiety tips is investigated. Table 5.5 shows the model. The F-test expresses that the variables are jointly significant (F=19.940, p<.001). There is an additional predictive value of looking at the anxiety tips on top of the effect of anxiety and dashboard usage. When a student looks at the anxiety tips with all other predictors fixed, the expected weighted percentage increases with 4.616 units. Note that a high anxiety score corresponds with low anxiety.

Table 5.5: *Linear regression model of anxiety.*

Covariate	coefficient	SE	t-value	p-value
anxiety	0.276	0.072	3.851	<.001
dbuser	7.089	1.543	4.595	<.001
anx_tips	4.616	0.997	4.628	<.001
Constant	37.034	2.384	15.532	<.001

Table 5.6 displays the linear regression model of time management. The covariates are jointly significant (F=46.23, p<.001). Looking at time management tips has additional predictive validity. Given that the other regressors stay fixed, it corresponds with an increase of 3.367 percentage points in expected weighted percentage.

Table 5.6: *Linear regression model of time management.*

Covariate	coefficient	SE	t-value	p-value
time management	0.820	0.081	10.166	<.001
dbuser	6.823	1.524	4.478	<.001
tmt_tips	3.367	0.934	3.603	<.001
Constant	24.749	2.425	10.206	<.001

The covariates in the model of test strategy are jointly significant (F=35.830, p<.001). Table 5.7 indicates that looking at the test strategy tip has incremental predictive value on top of the other regressors. When a student looks at the tip, with the other covariates fixed, the expected weighted percentage increases with 3.364 units.

Table 5.7: *Linear regression model of test strategy.*

Covariate	coefficient	SE	t-value	p-value
test strategy	0.851	0.099	8.563	<.001
dbuser	5.990	1.541	3.888	<.001
tst_tips	3.364	0.976	3.448	0.001
Constant	20.437	3.160	6.467	<.001

Table 5.8 presents the model of motivation. The F-test demonstrates that the predictors are jointly significant (F=51.770, p< .001). In contrast to the other models, looking at the motivation tips does not have a significant additional predictive value. Since the coefficient is not significantly different from 0, the interpretation is not meaningful.

Table 5.8: *Linear regression model of motivation.*

Covariate	coefficient	SE	t-value	p-value
motivation	0.967	0.087	11.170	<.001
dbuser	7.038	1.517	4.640	<.001
mot_tips	1.656	0.899	1.843	0.066
Constant	17.328	2.812	6.162	<.001

Appendix O provides the residual plots to check the assumptions. The plots suggest that all assumptions are satisfied for all the models, except for the normality assumption. As noted in Section 3.1, no transformations are considered. In addition, the predictive value of the models is limited; the R squared ranges from 0.028 for the anxiety model to 0.069 for the motivation model. The goal of the regression model is to discover general trends in the data instead of making predictions.

5.3.2 Extra visitations

Many students visit the dashboard again after a while. More specifically, 26.91% of the students visits the dashboard more than once. Here an extra visit is operationalized as the case when the student revisits the homepage (and thus clicks again on the link), more than 60 minutes after the last action on the dashboard. Table 5.9 displays the frequency table of the visitations. Figure 5.10 plots the mean weighted average by the amount of visitations. The plot denotes an increasing trend when the number of visits increases. Starting from 4 visits the trend drops and becomes unstable. The low sample sizes starting from 4 visits provide an explanation (Table 5.9).

Table 5.9: *Amount of times students visited the dashboard.*

# Visits	0	1	2	3	4	5	6	7
Frequency	145	1430	469	88	14	4	4	1

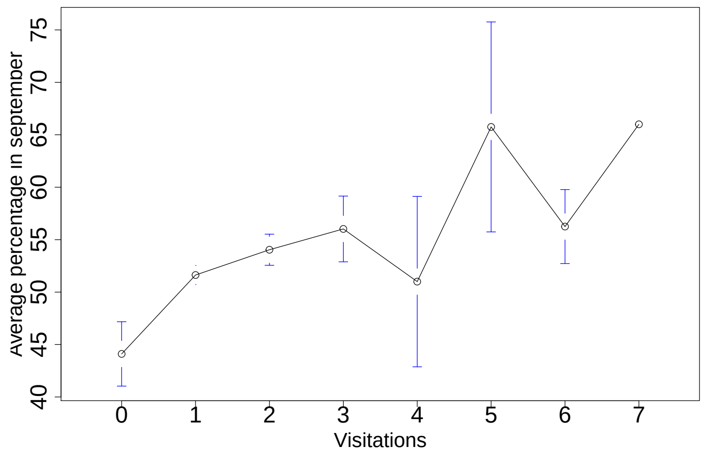


Figure 5.10: Average weighted percentage by amount of visitations. The blue lines illustrate the 95% confidence interval.

A linear regression model is fitted to assess whether the increasing trend is significant. The slope denotes that the expected weighted percentage increases with 2.847 percentage points for every extra visit. The effect of visits on weighted percentage is significantly different from 0 ($t=5.567$, $p< 0.001$). The intercept equals 48.242. Appendix P provides the residual plots and indicates the absence of violations of the assumptions. One exception is the normality assumption. The Shapiro-Wilk test additionally rejects normality ($W=0.970$, $p< 0.001$). With the same arguments as Section 3.1, a transformation is not considered.

5.3.3 Visitations in second semester

16.71% students visited the dashboard after the first semester. This entails 17.91% of all dashboard users. A plausible reason why so many students revisit the dashboard in February is because they received a new link together with their exam results. The relation with the weighted percentage in September is explored in Figure 5.11. The figure reveals there is a clear difference in weighted percentage in both groups. The Brown-Forsythe test suggests the variances are heterogeneous ($F=23.607$, $p< 0.001$). In addition, normality follows from the central limit theorem since the sample size is large (n_1=360, n_2=1795). It follows that the Welsch test is appropriate. Not surprisingly, the test expresses that the group means are significantly different ($t=6.349$, $p<0.001$).

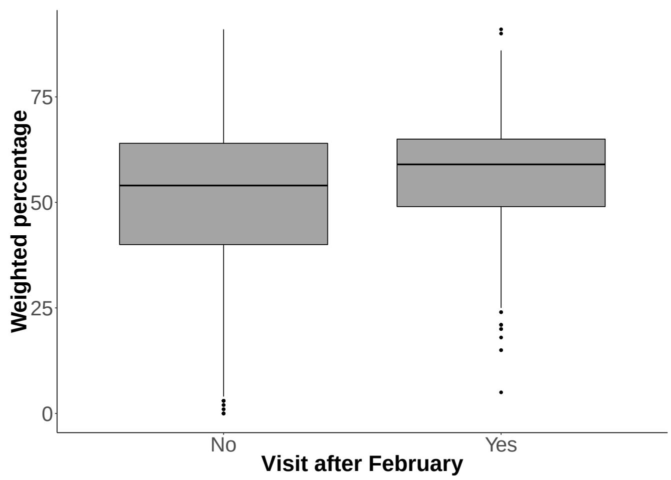


Figure 5.11: Weighted percentage grouped by whether the student visited after 1 February.

5.4 Discussion

In this chapter, multiple fine-grained measures of dashboard activity are defined that are linked to the weighted percentage in September. A first measure is the duration students spend on the dashboard. For every extra minute spend on the dashboard, the expected weighted percentage increases with 0.600 percentage points. This relation is present up to 42.23 minutes. This finding supports the results of Yu and Jo (2014), where total time on a learning management system is a significant predictor of student success. Since time online has contradicting results in the literature, it is not supported by other studies (Macfayden & Dawson, 2010). Next, the lag between receiving the link and visiting the dashboard has a significant negative effect on the expected weighted percentage in September. For lags up to 210.37 hours, the expected weighted percentage decreases with 0.025 for every extra hour between receiving and visiting the dashboard. This result supports the finding of Conijn et al. (2017) that less time until the first online session in a LMS is associated with higher grades.

Furthermore, the total number of visits has a significant positive effect. Each extra visit results in an increase of 2.847 percentage points of the expected weighted percentage. This finding supports the findings of MacFayden and Dawson (2010) and Conijn et al., (2017) where the number of visits has a positive association with the grade. In contradicts the result of Yu and Jo (2014) were the login frequency was not a significant predictor in the multiple regression model of academic achievement. In addition, students received the link again when they got their exam results. Students that revisit the dashboard in February have a significant higher mean percentage.

Next, there is a significant positive effect of visiting the concentration, test strategy, anxiety and time management tips on weighted percentage, while controlling for the effect of the corresponding learning skill and dashboard usage. Looking at the motivation tip has no incremental predictive validity on top of motivation and dashboard usage. Broos et al. (2017) found that motivation has a significant impact on the prediction whether students viewed any of the tips. A possible explanation is that looking at the motivation tips signals partly motivation, while this is already captured by incorporating motivation in the model. Motivation has a significant impact on weighted GPA (Pinxten et al., 2017).

Science and engineering technology students that check the dashboard at different times of the day have different median weighted percentages. Not unexpectedly, students who do not visit the dashboard have a lower median result than students who visit the dashboard in the evening, early morning and at 22h. Furthermore, the remainder of the results is possibly linked to the fact that students mostly have class in the morning, noon and afternoon. Students who visit the dashboard at 22h have a higher median result than students who visit the dashboard in the morning and at noon while they are possibly in class. In addition, students who visit the dashboard in the early morning have a higher median weighted percentage than students that visit in the morning. Lastly, the median of students who visit the dashboard in the evening is lower compared to students who visit the results at noon or in the morning. Further research is required to investigate this hypothesis.

Some of these findings can have practical implications since there is a lack of early predictors in learning analytics (Conijn et al., 2017). In addition, the deployment of a dashboard is low cost and the data is easy to collect. Hence, dashboard usage potentially offers an early warning signal for student counselors. Based on this information they can reach out to students. When a student has for example not visited the dashboard after a number of days, a student counselor can send him an invite for an appointment. Although the models based on the fine-grained measures do not have a strong predictive power, a possible practical implementation is combining this data with the early warning systems that are already in use.

A limitation in the analysis of case duration is that it is not logged when a student leaves a tab open, while he does other actions outside the dashboard. This introduces noise to the data. To handle this, a cutoff of maximum time on the dashboard without actions is set on 5 minutes. Still, it is possible that students leave before 5 minutes. In contrast, it is possible that students spend more than 5 minutes on a tab. The same analysis was performed with 15 minutes as a cut-off. Because this analysis results in the same conclusions, there exists some robustness.
The analysis of time points contains noise introduced by the fact that programs received their dashboard at separate time points. Separate analyses for programs that received their dashboards at different time points solves this issue. A drawback of this solution is the reduction in sample size. This is problematic since for both analyses the median test is used. This test is often criticised for his low power in small samples (Freidlin & Gastwirth, 2000). It follows that a possible alternative explanation for the nonsignificant effect in the engineers is the lack of power of the median test. In addition, since time is continuous, binning the time points in several groups is subjective. One can discuss to set breaks at different time points.
It is important to note that the R squared of each of the models did not surpass 0.06. This signals that the predictive value of the models is low. As noted in the text, the models have the goal to discover broad general trends instead of making accurate predictions. Further research is required in order to see if these small effects have incremental predictive validity on top of known factors. In addition, the portability of the conclusions to other programs and universities needs further research.

It would be imprudent to suggest that the positive effects of these measures are the sole result of the action performed on the dashboard. It is more likely that the way of using the dashboard signals a certain underlying disposition of the student. Further research is required to investigate which disposition this is. A possible hypothesis is that the dashboard activity is linked to student engagement. The following paragraphs link these measures to the student engagement theory of Fredricks et al. (2004). It is important to note that these links are not established yet and still require further research. One can for example examine if there exists a relationship with questionnaires of student engagement.

Time spend on the dashboard is possibly linked to cognitive engagement. Cognitive engagement is defined as investment; the thoughtfulness and willingness to make effort to master difficult skills and comprehend complex ideas. It ranges from simple memorisation to the use of self-regulated learning strategies in order to reach deep understanding and expertise (Fredricks et al., 2004). When a student spends considerable time on the dashboard, the students manifests investment to improve their learning skills. Furthermore looking at the tips can signal cognitive engagement, since the student performs a specific action in order to improve his learning skills. Cognitive engagement is in turn related to academic success (Lei et al., 2018).
The amount of visitations and whether the students revisits the dashboard after receiving their first exam results possibly relates to cognitive engagement. It is possible that these students encountered difficulties regarding to their learning skills and wanted to improve these skills. In order to do so, they visited the dashboard again to find information and concrete tips. Thus if this hypothesis proves to be true, these activities relate to cognitive engagement.
The lag between receiving the link and visiting the dashboard possibly linked to behavioural engagement. It signals that the student stays up to date on the information that is provided by the university. Behavioural engagement entails the behaviour of the student in the classroom and adherence to the school norms, rules and expectations (Nguyen, Cannata & Miller, 2016). In addition, the moment of the first visitation is possibly linked to behavioural engagement. The significant contrasts correspond to visiting the dashboard while they are possibly in a class versus visiting the dashboard early in the morning or in the evening after class. Behavioural engagement is moderately positive related to academic success (Lei at al., 2018). Still, further research is required to investigate the links between measures of dashboard activity and student engagement.

Chapter 6

Conclusion

This thesis provides the answer to two research questions. A first question is if dashboard usage has incremental predictive validity on top of other available data. A second research question is to define fine-grained measures that are linked to academic success.
In order to answer the first question, a regression model is fitted to predict the weighted average grade. In addition, a multinomial logistic regression model is implemented in order to predict the membership in CSE categories. Next, a logistic regression model is fitted to predict whether students are at risk. In all analyses variable selection is conducted. Afterwards a likelihood ratio test assesses the difference in likelihood of the reduced model with and without dashboard usage.

In all analyses the likelihood ratio tests with dashboard usage yields a significant result. It follows that dashboard usage significantly increases the likelihood of the models. Furthermore, the effect of dashboard usage is in each model significant. The linear regression model demonstrates that if the other variables stay fixed, dashboard usage results in an increase of 6.645 percentage points in expected weighted average grade. In addition, when a student uses the dashboard, the odds of being in the ">30 CSE or dropout" versus the "> 80 CSE" category is about two times smaller, given that the other variables stay fixed. Lastly, using dashboard usage results in an increase in the odds of being at risk with a factor 0.537.
It is also assessed whether the incorporation of dashboard usage results in more accurate model. This is indeed the case. Hence, dashboard usage has a non-negligible incremental predictive validity on top of known predictors. Still, caution is needed since the sensitivity analysis denotes that the significant increase in accuracy depends on decisions with respect to the missing data.

Several fine-grained measures are defined and related to the weighted percentage in September to answer the second research question. A first measure is the time spend on the dashboard. Students that spend more time on the dashboard have higher expected percentage. For every extra minute spend on the dashboard up to 42.23 minutes, the expected percentage increases with 0.6 percentage points. This indicates there is a clear trend present.
For an analysis of the effect of the moment of first visitation, the sample size is divided into two groups based on the moment of receiving the link. The effect of the moment of visitation is not present in the analysis of engineers. A second analysis contains a

larger sample size with students that received their dashboard at 21h. The the median test expresses that the medians of the students that visited their dashboard at different time points are unequal. Students that visit the dashboard at 22h have a higher median weighted percentage compared to students that visit the dashboard at the morning, noon or do not visit the dashboard at all. In addition, students that do not visit the dashboard have a lower median than students that visit the dashboard in the early morning or evening. Furthermore, students that visit the dashboard in the early morning have a significantly higher median than student that visited the dashboard in the morning. Lastly, students that visit the dashboard at the evening have a higher median weighted percentage in September compared to students that visit the dashboard at noon or in the morning.
Another fine-grained measure is the lag between receiving the dashboard and clicking on the link. For lags until 8.766 days, there exists a negative effect on expected weighted percentage. The expected weighted percentage decreases with 0.025 percentage points for each extra hour between receiving the link and visiting the dashboard.

Next, clicking on the tips is analysed. Dashboard usage and learning skills are included in the model to control for these possible confounding factors. Looking at the tips has significant incremental predictive validity, except for motivation. The following interpretations are conditional on the fact that the other covariates stay fixed. Looking at the concentration tip increases the expected weighted percentage with 2.217 percentage points. Furthermore, checking the anxiety tips results in an increase of 4.616 of the expected weighted percentage. If a student looks at the concentration tips, the expected value increases with 3.367 percentage points. Lastly, the expected weighted percentage increases with 3.364 percentage points if the student looks at test strategy tips.
A last analysis entails the timing and the number of visitations. The number of visitations has a significant effect on the expected weighted percentage; for each extra visit the expected weighted percentage increases with 2.847 percentage points. In addition, students that visit the dashboard after February have a significant higher mean compared to students that did not visit the dashboard after February.

Chapter 7

References

Agudo-Peregrina, A. F., Iglesias-Pradas, S., Conde-Gonzlez, M. A., & Hernndez-Garca., A. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior, 31*(1), 542-550.

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(3), 370-374.

Belsley, D. (1982). Assessing the presence of harmful collinearity and other forms of weak data through a test for signal-to-noise. *Journal of Econometrics, 20*(2), 211-253.

Bennett, D. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health, 25*(5), 464-469.

Boulton, C. A., Kent, C., & Williams H. T. P. (2018). Virtual learning environment engagement and learning outcomes at a bricks-and-mortar university. *Computers & Education, 126*, 129-142.

Broos, T., Peeters, L., Verbert, K., Van Soom, C., Langie, G., & De Laet, T. (2017). Dashboard for actionable feedback on learning skills: How learner profile affects use. *Proceedings of the 7th Workshop on Awareness and Reflection in Technology Enhanced Learning Co-located with the 12th European Conference on Technology Enhanced Learning (EC-TEL 2017), 1997*, 1-15.

Broos, T, Verbert, K, Langie, G, Van Soom, C,& De Laet, T. (2018). Multi-Institutional positioning test feedback dashboard for aspiring students lessons learnt from a case study in flanders. *ACM International Conference Proceeding Series*, 51-55.

Broos, T., Verbert, K., Van Soom, C., Langie, G., & De Laet, T. (2018). Small data as a conversation starter for learning analytics: Exam results dashboard for first-year students in higher education. *Journal of Research in Innovative Teaching & Learning*, 1-14.

Cheng, S.,& Long, J. (2007). Testing for IIA in the Multinomial Logit Model. *Sociological Methods and Research, 35*(4), 583-600.

Choy, S. (2001). *Students whose parents did not go to college postsecondary access, persistence, and attainment*. Washington, DC : National Center for Education Center, U.S. Dept. of Education, Office of Educational Research and Improvement : Educational Resources Information Center

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS.*IEEE Transactions on Learning Technologies, 10*(1), 17-29.

Cook, R. (2000). Detection of influential observation in linear regression. *Technometrics*, 42(1), 65-68.

Cytel inc. (2007). *Statistical software for exact nonparametric inference*. Camebridge, MA: Cytel inc.

Fagerland, M., & Hosmer, D. (2012). A generalized Hosmer-Lemeshow goodness-of-fit test for multinomial logistic regression models. *Stata Journal, 12*(3), 447-453.

Ferreira, S. A., & Andrade, A. (2016). Academic Analytics: Anatomy of an Exploratory Essay. *Education and Information Technologies, 21*(1), 229-243.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research, 74*(1), 59-109.

Freidlin, B., & Gastwirth, J. (2000). Should the Median Test be Retired from General Use? *The American Statistician, 54*(3), 161-164.

Fonteyne, Duyck, & De Fruyt. (2017). Program-specific prediction of academic achievement on the basis of cognitive and non-cognitive factors. *Learning and Individual Differences, 56*(C), 34-48.

Garson, D. G. (2014). *Logistic Regression: Binary and Multinomial*. Asheboro: Statistical Associates Publishing.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika, 75*(4), 649674, 2010.

Jung, Y., & Lee, J. (2018). Learning Engagement and Persistence in Massive Open Online Courses (MOOCS). *Computers & Education, 122*, 9-22.

Kutner, M. H., Li, W., Nachtsheim, C. J, & Neter, J. (2005). *Applied linear statistical models* (5th ed.). Boston:McGraw-Hill/Irwin.

Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: a meta-analysis. *Social Behavior and Personality, 46*(3), 517-528.

Little, R. (1988). *A Test of Missing Completely at Random for Multivariate Data with Missing Values.* Journal of the American Statistical Association, 83(404), 1198-1202.

Little, R. J. A, & Rubin, D. B. (2002). *Statistical analysis with missing data (2nd ed., Wiley series in probability and statistics).* London: Wiley.

Long, P. D., & Siemens, G. (2014). Penetrating the fog: Analytics in learning and education. *TD Tecnologie Didattiche, 22*(3), 132-137.

Macfadyen, & Dawson. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588-599.

Malikowski, S. R., Thompson, M. E., & Theis, J. G. (2007). A model for research into course management systems: Bridging technology and learning theory. *Journal of Educational Computing Research, 36*(2), 149173.

McFadden, D., Tye, W. and Train, K. (1976). An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model. *Transportation Research Record, 637*, 39-45.

Mohamad, S. K., & Tasir, Z. (2013). Educational Data Mining: A Review. *Procedia - Social and Behavioral Sciences, 97*, 320-324.

Molenberghs, G., & Kenward, M.G. (2007). *Missing Data in Clinical Studies.* John Wiley; New York.

Moore, M. G. (1989). Three types of interaction. *American Journal of Distance Education, 3*(2), 1-6.

Mothilal, R. K., De Laet, T., Broos, T., & Pinxten, M. (). Predicting first year engineering student success; from traditional statistics to machine learning.

Mothilal, K. R. (2018). *Statistical modeling of students' performance in an open-admission bachelor program in Flanders.* Leuven: KU Leuven. Faculteit Ingenieurswetenschappen.

National Math and Science Initiative. (2015). *The STEM crisis.* (Dallas, TX, National Math and Science Initiative

Nguyen, Tuan Dinh, Cannata, Marisa, & Miller, Jason. (2016). Understanding Student Behavioral Engagement: Importance of Student Interaction with Peers and Teachers. *Grantee Submission, 111*(2), 13.

O'Connell, A. (2006). *Logistic regression models for ordinal response variables (Quantitative applications in the social sciences).* Thousand Oaks (Calif.): Sage.

Pinxten, M., & Hockicko, P. (2016). Predicting study success of first-year Science and Engineering students at the University of Zilina An exploration of the added value of measuring incoming students' learning and study strategies. *2016 elektro 11th international conference, 18-23.*

Pinxten, M., Langie, G., Van Soom, C., Peeters, C., & De Laet, T. (2017). At-risk at the gate: Prediction of study success of first-year science and engineering students in an open-admission university in Flandersany incremental validity of study strategies? *European Journal of Psychology of Education*, 1-22.

Pursel, B. K., Zhang, L., Jablokow, K. W., Choi, G. W., & Velegol, D. (2016). Understanding MOOC Students: Motivations and Behaviours Indicative of MOOC Completion. *Journal of Computer Assisted Learning, 32*(3), 202-217.

Rafaeli, S. & Ravid, G. (1997). Online, web-based learning environment for an information systems course: Access logs, linearity and performance. *Proc. Inf. Syst. Educ. Conf, 97*, 9299.

Ruxton, Wilkinson,& Neuhuser. (2015). Advice on testing the null hypothesis that a sample is drawn from a normal distribution. *Animal Behaviour*, 107, 249-252.

Salmerón, R., García, C. B. & García, J. (2018). Variance Inflation Factor and Condition Number in multiple linear regression. *Journal of Statistical Computation and Simulation, 88*(12), 2365-2384.

Schwendimann, B.A., Rodrguez-Triana, M., Vozniuk, A., Prieto L. P., Boroujeni, M. S., Holzer, A., Gillet, D. & Dillenburg, P. (2016). Understanding learning at a glance: An overview of learning dashboard studies. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*,532-533.

Schmidt & Finan. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, 98, 146-151.

Spruyt, B., Kavadias, D. & Roggemans, L. (2014). Het selecteren van de uitverkorenen. Een onderzoek naar en discussie over het toelatingsexamen (tand)arts in Vlaanderen. *Tijdschrift Voor Onderwijsrecht En Onderwijsbeleid, No:2014*(5)

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*, 157167.

Trussel, J. & Burke-Smalley, L. (2018). Demography and student success: Early warning tools to drive intervention. *Journal of Education for Business, 93*(8), 363-372.

Tukey, J. (1977). *Exploratory data analysis* (Addison-Wesley series in behavioral science : quantitative methods). Reading: Addison-Wesley.

Tutz, G., Pnecker, W. & Uhlmann.L.(2015). Variable selection in general multinomial logit models. Computational Statistics and Data Analysis, 82(C), 207-222.

Van den Broeck, L., Pinxten, M., Langie, G., De Laet, T., Lacante, M. & Van Soom, C. (2018). Predicting the academic achievement of students bridging to engineering: The role of academic background variables and diagnostic testing. Journal of Further and Higher Education, 1-19.

Vargha, A. & Delaney, H. (1998). The Kruskal-Wallis Test and Stochastic Homogeneity. *Journal of Educational Statistics*, 23(2), 170-192.

Weinstein, C. E. & Palmer D. R. (2002). *LASSI users manual: For those administering the learning and study strategies inventory.* H & H Pub.

Wolff, A., Zdrahal, Z., Nikolov, A. & Pantucek, M. (2013). Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 145-149.

Yu, T. & Jo, I. (2014). Educational technology approach toward learning analytics: Relationship between student online behavior and learning performance in higher education. *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, 269-270.

Appendix A

Tests for normality and homogeneity of variance

A.1 Shapiro Wilk test

The Shapiro-Wilk test is a popular test to assess normality. The null hypothesis states that the sample follows a normal distribution. The alternative hypothesis is that this is not the case. When the sample of size N is ordered from small to largest, x_1 to x_N, the test statistic equals

$$W = \frac{(\sum_{i=1}^{nN} a_i x_i)^2}{\sum_{i=1}^{nN}(x_i - \bar{x})^2}.$$

where a_1 to a_N are weights depending on the expected values and the covariance matrix of the order statistics (Ruxton, Wilkinson & Neuhuser, 2015). The denominator can be interpreted as a variance measure of the sample. The numerator can be seen as a similar measure of variance in the case when the sample is drawn from an underlying normal distribution. The null hypothesis is rejected when W is below a critical threshold.

A.2 Brown-Forsythe test

The Brown-Forsythe test tests the null hypothesis of equal variances of groups. In contrast to the Hartley test, the test does not assume normality. The test statistic is constructed from the absolute deviations of the medians of the groups $d_{ij} = |Y_{ij} - \widetilde{Y}_i|$. The test statistic is the following

$$F_{BF}^{obs} = \frac{MSTR_{BF}}{MSE_{BF}} = \frac{\frac{\sum_{i=1}^{r} n_i(\bar{d}_{i.} - \bar{d}_{..})^2}{r-1}}{\frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i}(d_{ij} - \bar{d}_{i.})^2}{n_t - r}} \sim F_{df=(r-1)(n_t-r)}.$$

Appendix B

Median test

The median test is a non-parametric test that tests the equality of the medians of K distributions. The alternative hypothesis is that at least one of the medians is different. According to Cytel inc. (2007) the test is conducted in the following way. Let u_{ij} be the score of the ith individual of the jth group. First the pooled median of the whole sample δ is computed. Next, scores w_{ij} are assigned

$$w_{ij} = \begin{cases} 1 \text{ if } u_{ij} \leq \delta, \\ 0 \text{ if } u_{ij} > \delta \end{cases}$$

.

The total number of observations in the jth group that are at or below the median is computed as

$$w_j = \sum_{i=1}^{n_j} w_{ij}.$$

m is defined as the total number of observation in the pooled sample that are at or below the median

$$m = \sum_{j=1}^{K} w_j$$

The test is based on the following 2 x K contingency table (Table B.1) The entries in the upper row are the counts of the number of subjects in each group with responses at or below the median, while the counts in the seconds row are the number of subjects in each group with responses above the median.

Table B.1: *Data grouped in a 2 x K contingency table for the median test.*

Group	1	2	...	K	Row total
$\leq$ Median	w_1	w_2	...	w_K	m
$>$ Median	$n_1 - w_1$	$n_2 - w_2$	...	$n_K - w_k$	N-m
Column total	n_1	n_2	...	n_K	N

Conditional on fixing the margins, the probability of observing this contingency table under the null hypothesis is given by a hypergeometric function

$$h(\mathbf{w}) = \frac{\prod_{j=1}^{K} \binom{n_j}{w_j}}{\binom{N}{m}} \tag{B.1}$$

.

For any $\widetilde{\boldsymbol{w}} \in \mathbf{W}$, the test statistic for the median test is the Pearson Chi square statistic:

$$T = \sum_{j=1}^{K} \frac{(\widetilde{w}_j - n_j m/N)^2}{n_j m/N} + \sum_{j=1}^{K} \frac{(n_j - \widetilde{w}_j - n_j(N-m)/N)^2}{n_j(N-m)/N}.$$

.

The p-value is calculated as follows:

$p = \sum_{T \geq t} \mathrm{h}(\widetilde{\boldsymbol{w}})$.

Appendix C

Full linear regression model

	Estimate	SE	t-value	p-value
dbuser	1.292	0.195	$t = 6.635$	$p < 0.001$
schooltypeBSO	−4.889	3.426	$t = -1.427$	$p = 0.154$
schooltypeKSO	−4.493	1.959	$t = -2.293$	$p = 0.022$
schooltypeOther	−1.199	0.329	$t = -3.643$	$p < 0.001$
schooltypeTSO	−0.946	0.179	$t = -5.274$	$p < 0.001$
math.hrs6-7u	2.187	0.217	$t = 10.076$	$p < 0.001$
math.hrs8u	2.680	0.232	$t = 11.536$	$p < 0.001$
math.score>90%	1.996	0.392	$t = 5.096$	$p < 0.001$
math.score60-70%	0.677	0.215	$t = 3.144$	$p = 0.002$
math.score70-80%	1.538	0.226	$t = 6.801$	$p < 0.001$
math.score80-90%	2.079	0.263	$t = 7.916$	$p < 0.001$
fys>90%	1.447	0.372	$t = 3.892$	$p < 0.001$
fys60-70%	0.461	0.233	$t = 1.978$	$p = 0.049$
fys70-80%	0.925	0.242	$t = 3.819$	$p < 0.001$
fys80-90%	0.932	0.276	$t = 3.380$	$p = 00.001$
chem>90%	2.137	0.375	$t = 5.692$	$p < 0.001$
chem60-70%	0.638	0.220	$t = 2.897$	$p = 0.004$
chem70-80%	0.718	0.230	$t = 3.126$	$p = 0.002$
chem80-90%	1.264	0.266	$t = 4.746$	$p < 0.001$
mot	0.041	0.020	$t = 2.060$	$p = 0.040$
tmt	0.092	0.018	$t = 5.078$	$p < 0.001$
anx	0.026	0.014	$t = 1.950$	$p = 0.052$
tst	−0.031	0.021	$t = -1.506$	$p = 0.133$
con	−0.015	0.017	$t = -0.909$	$p = 0.364$
advice: Negative	−0.946	0.215	$t = -4.408$	$p < 0.001$
advice: Positive	0.498	0.163	$t = 3.060$	$p = 0.003$
advice: Unknown	−0.555	0.205	$t = -2.704$	$p = 0.007$
pioneer: Pionieer	−1.322	0.204	$t = -6.492$	$p < 0.001$
pioneer: Unknown	−1.176	0.409	$t = -2.878$	$p = 0.005$
Constant	1.300 (0.688	$t = 1.889$	$p = 0.059$	
Observations	3,119			
R^2	0.296			
Adjusted R^2	0.289			
Residual Std. Error	3.379 (df = 3089)			
F Statistic	44.713***	df = 29; 3089)		

Appendix D

Diagnostic plot of the full linear regression model

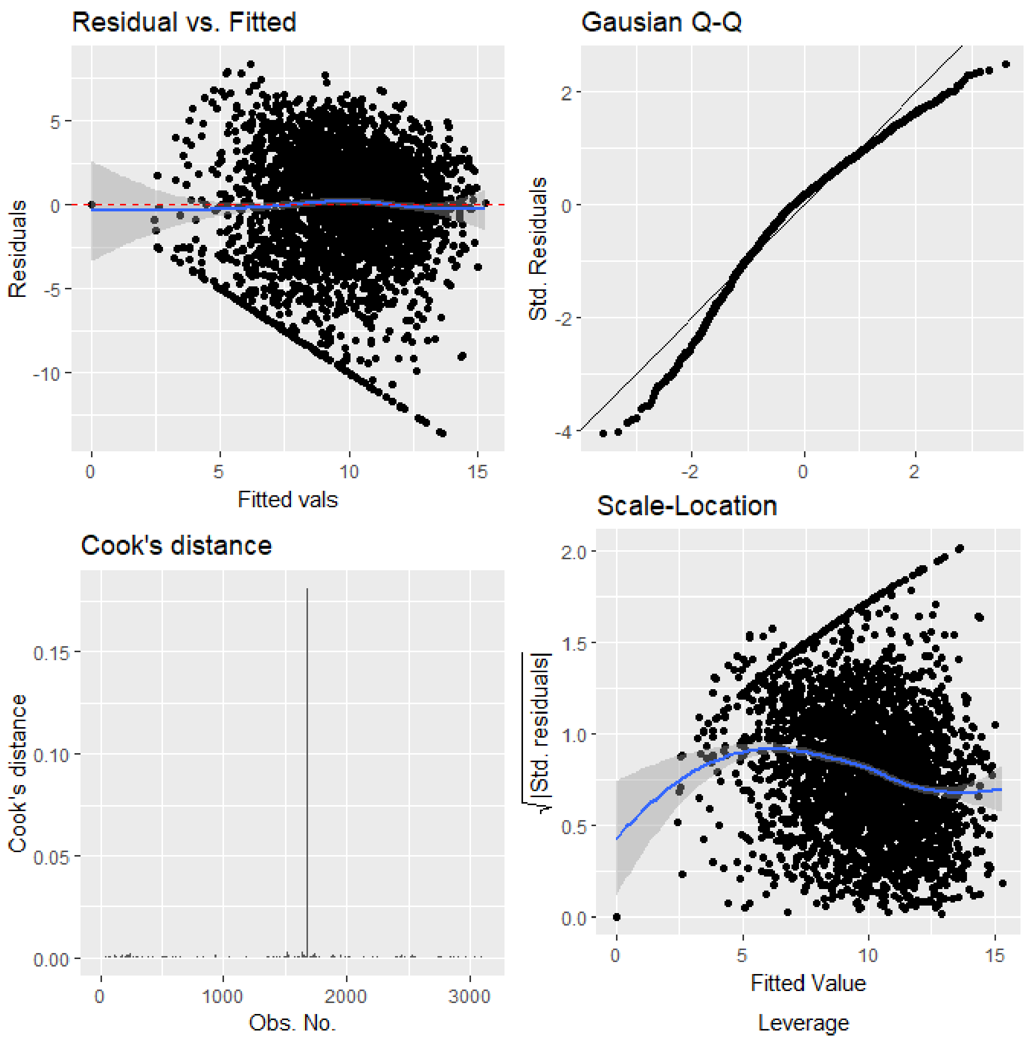


Figure D.1: Diagnostic plot of the full linear regression model with dashboard usage.

Appendix E

Reduced multinomial model

Table E.1: *Estimates, standard error and p-values of the full model with dashboardusage*

Analysis of Maximum Likelihood Estimates						
Parameter		cse_final	Estimate	SE	Chi-Square	P-val
Intercept		30-50% CSE	2.2996	0.5448	17.8140	<.0001
Intercept		50-80% CSE	2.0789	0.4725	19.3586	<.0001
Intercept		<30% CSE of dropout	3.9553	0.4467	78.4077	<.0001
dbuser		30-50% CSE	−0.2340	0.1920	1.4853	0.2230
dbuser		50-80% CSE	0.1123	0.1768	0.4037	0.5252
dbuser		<30% CSE of dropout	−0.6201	0.1506	16.9410	<.0001
schooltype	BSO	30-50% CSE	−2.2751	1251.5	0.0000	0.9985
schooltype	BSO	50-80% CSE	−1.7221	1014.5	0.0000	0.9986
schooltype	BSO	<30% CSE of dropout	10.0379	544.9	0.0003	0.9853
schooltype	KSO	30-50% CSE	−8.9896	346.0	0.0007	0.9793
schooltype	KSO	50-80% CSE	−9.5724	283.5	0.0011	0.9731
schooltype	KSO	<30% CSE of dropout	2.9709	1.2712	5.4622	0.0194
schooltype	Other	30-50% CSE	0.3963	0.3573	1.2298	0.2674
schooltype	Other	50-80% CSE	0.2409	0.3083	0.6105	0.4346
schooltype	Other	<30% CSE of dropout	0.9115	0.2671	11.6488	0.0006
schooltype	TSO	30-50% CSE	0.3346	0.1767	3.5853	0.0583
schooltype	TSO	50-80% CSE	−0.0951	0.1609	0.3494	0.5544
schooltype	TSO	<30% CSE of dropout	0.6973	0.1411	24.4097	<.0001
math_score	60-70%	30-50% CSE	−0.3447	0.2197	2.4620	0.1166
math_score	60-70%	50-80% CSE	−0.4018	0.1929	4.3397	0.0372
math_score	60-70%	<30% CSE of dropout	−0.6887	0.1798	14.6783	0.0001
math_score	70-80%	30-50% CSE	−0.9386	0.2316	16.4302	<.0001
math_score	70-80%	50-80% CSE	−0.9058	0.1983	20.8688	<.0001
math_score	70-80%	<30% CSE of dropout	−1.3532	0.1887	51.4197	<.0001
math_score	80-90%	30-50% CSE	−1.1609	0.2798	17.2157	<.0001
math_score	80-90%	50-80% CSE	−1.1037	0.2277	23.4938	<.0001
math_score	80-90%	<30% CSE of dropout	−1.8744	0.2369	62.6236	<.0001
math_score	>90%	30-50% CSE	−1.8326	0.6550	7.8286	0.0051
math_score	>90%	50-80% CSE	−1.0526	0.3619	8.4592	0.0036

Table E.1: *Estimates, standard error and p-values of the full model with dashboardusage*

Analysis of Maximum Likelihood Estimates						
Parameter		cse_final	Estimate	SE	Chi-Square	P-val
math_score	>90%	<30% CSE of dropout	−1.3423	0.3557	14.2408	0.0002
math_hrs	6-7u	30-50% CSE	−0.8644	0.2333	13.7230	0.0002
math_hrs	6-7u	50-80% CSE	−0.7610	0.2061	13.6343	0.0002
math_hrs	6-7u	<30% CSE of dropout	−1.4287	0.1817	61.8470	<.0001
math_hrs	8u	30-50% CSE	−0.9272	0.2489	13.8765	0.0002
math_hrs	8u	50-80% CSE	−0.6518	0.2147	9.2174	0.0024
math_hrs	8u	<30% CSE of dropout	−1.7970	0.1975	82.7862	<.0001
fys	60-70%	30-50% CSE	−0.1541	0.2299	0.4490	0.5028
fys	60-70%	50-80% CSE	0.3168	0.2122	2.2298	0.1354
fys	60-70%	<30% CSE of dropout	−0.1319	0.1891	0.4868	0.4854
fys	70-80%	30-50% CSE	−0.4063	0.2336	3.0245	0.0820
fys	70-80%	50-80% CSE	−0.1230	0.2154	0.3263	0.5678
fys	70-80%	<30% CSE of dropout	−0.6467	0.1947	11.0339	0.0009
fys	80-90%	30-50% CSE	−0.7757	0.2831	7.5059	0.0061
fys	80-90%	50-80% CSE	−0.1991	0.2402	0.6869	0.4072
fys	80-90%	<30% CSE of dropout	−0.4895	0.2260	4.6918	0.0303
fys	>90%	30-50% CSE	−2.0728	0.6494	10.1874	0.0014
fys	>90%	50-80% CSE	−0.8745	0.3576	5.9817	0.0145
fys	>90%	<30% CSE of dropout	−0.7290	0.3475	4.4011	0.0359
chem	60-70%	30-50% CSE	0.0114	0.2229	0.0026	0.9593
chem	60-70%	50-80% CSE	−0.0764	0.1906	0.1608	0.6884
chem	60-70%	<30% CSE of dropout	−0.2654	0.1764	2.2624	0.1325
chem	70-80%	30-50% CSE	0.0765	0.2290	0.1116	0.7383
chem	70-80%	50-80% CSE	−0.1660	0.1951	0.7237	0.3949
chem	70-80%	<30% CSE of dropout	−0.3923	0.1829	4.5996	0.0320
chem	80-90%	30-50% CSE	−0.4874	0.2820	2.9866	0.0840
chem	80-90%	50-80% CSE	−0.5449	0.2246	5.8871	0.0153
chem	80-90%	<30% CSE of dropout	−0.8821	0.2216	15.8436	<.0001
chem	>90%	30-50% CSE	−0.2928	0.4740	0.3817	0.5367
chem	>90%	50-80% CSE	−0.5918	0.3320	3.1775	0.0747
chem	>90%	<30% CSE of dropout	−1.4210	0.4032	12.4191	0.0004
advice	Negative	30-50% CSE	0.2758	0.2103	1.7194	0.1898
advice	Negative	50-80% CSE	0.2215	0.1889	1.3741	0.2411
advice	Negative	<30% CSE of dropout	0.8687	0.1732	25.1451	<.0001
advice	Positive	30-50% CSE	−0.6460	0.1596	16.3800	<.0001
advice	Positive	50-80% CSE	−0.2167	0.1292	2.8131	0.0935
advice	Positive	<30% CSE of dropout	−0.2637	0.1348	3.8251	0.0505
advice	Unknown	30-50% CSE	0.2466	0.1951	1.5985	0.2061
advice	Unknown	50-80% CSE	0.1335	0.1754	0.5799	0.4464
advice	Unknown	<30% CSE of dropout	0.6048	0.1665	13.1893	0.0003
pioneer	Pioneer	30-50% CSE	0.6137	0.2043	9.0273	0.0027
pioneer	Pioneer	50-80% CSE	0.1236	0.1898	0.4242	0.5149

Table E.1: *Estimates, standard error and p-values of the full model with dashboardusage*

Analysis of Maximum Likelihood Estimates						
Parameter		cse_final	Estimate	SE	Chi-Square	P-val
pioneer	Pioneer	<30% CSE of dropout	0.8709	0.1642	28.1200	<.0001
pioneer	Unknown	30-50% CSE	0.4583	0.4018	1.3010	0.2540
pioneer	Unknown	50-80% CSE	−0.0744	0.3785	0.0386	0.8442
pioneer	Unknown	<30% CSE of dropout	0.5672	0.3269	3.0101	0.0827
tmt		30-50% CSE	−0.0841	0.0163	26.7567	<.0001
tmt		50-80% CSE	−0.0315	0.0131	5.7350	0.0166
tmt		<30% CSE of dropout	−0.0577	0.0134	18.6460	<.0001
con		30-50% CSE	0.0247	0.0157	2.4826	0.1151
con		50-80% CSE	−0.0195	0.0126	2.3825	0.1227
con		<30% CSE of dropout	0.0095	0.0129	0.5454	0.4602

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	P-val
dbuser	3	24.6965	<.0001
schooltype	12	47.5576	<.0001
math_score	12	100.1728	<.0001
math_hrs	6	89.5402	<.0001
fys	12	46.5926	<.0001
chem	12	32.6780	0.0011
advice	9	86.1896	<.0001
pioneer	6	35.7316	<.0001
tmt	3	33.8339	<.0001
con	3	7.8359	0.0495

Appendix F

Cooks distances of the logistic regression model to predict atrisk

The plot indicates that there are no highly influential observations. The threshold to deem an observation influential according to Cook (2000) is in this case 1.548. Hence, no observation exceeds the threshold.

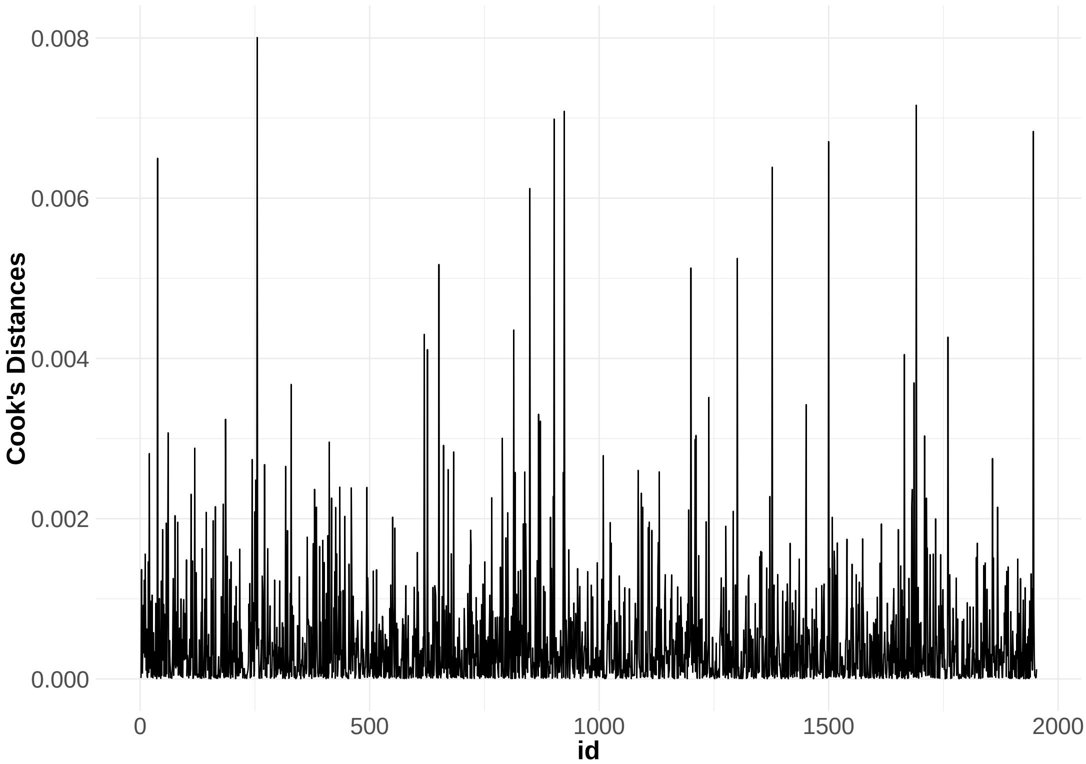
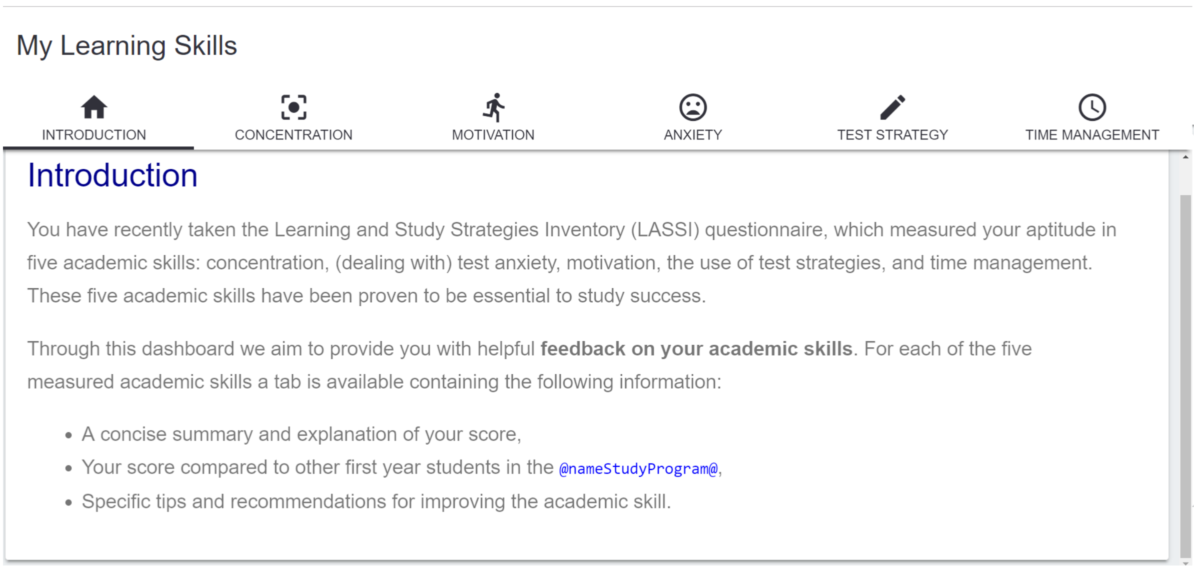
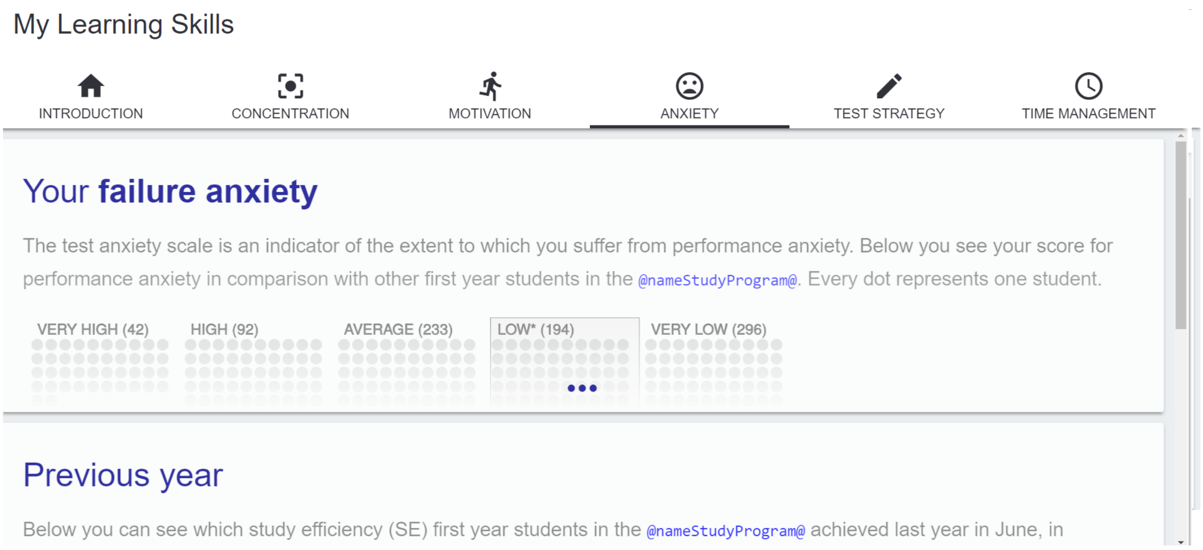


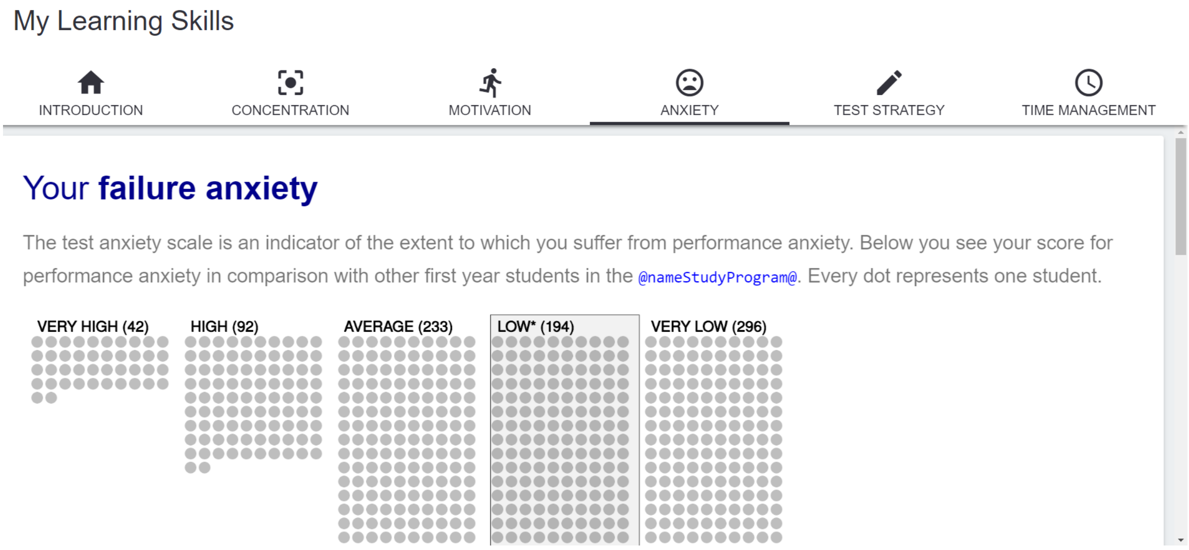
Figure F.1: Cook's distance plot of the reduced logistic regression model with dashboard use.

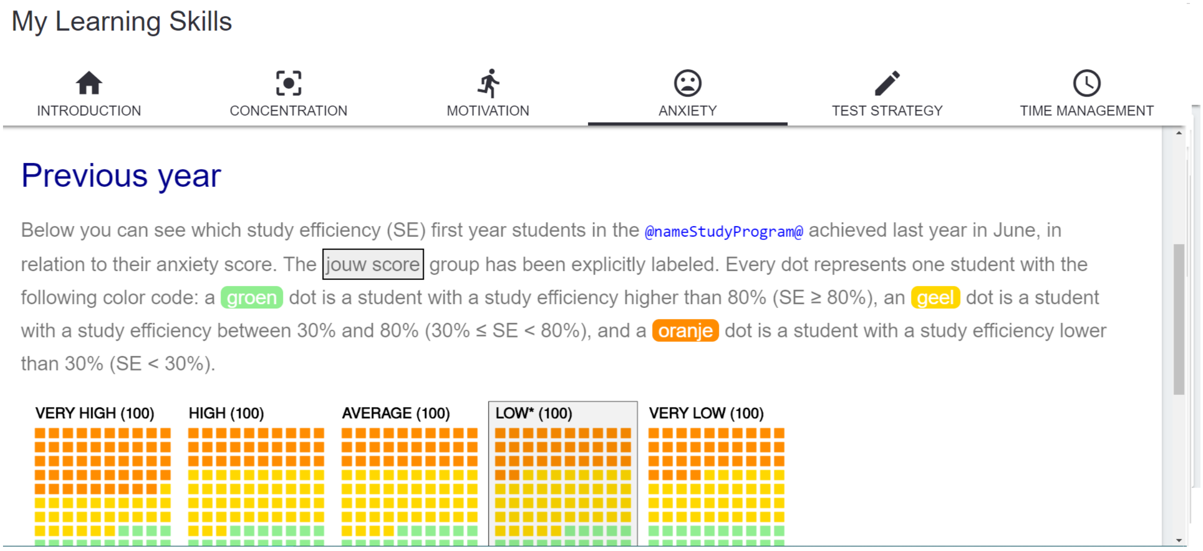
Appendix G

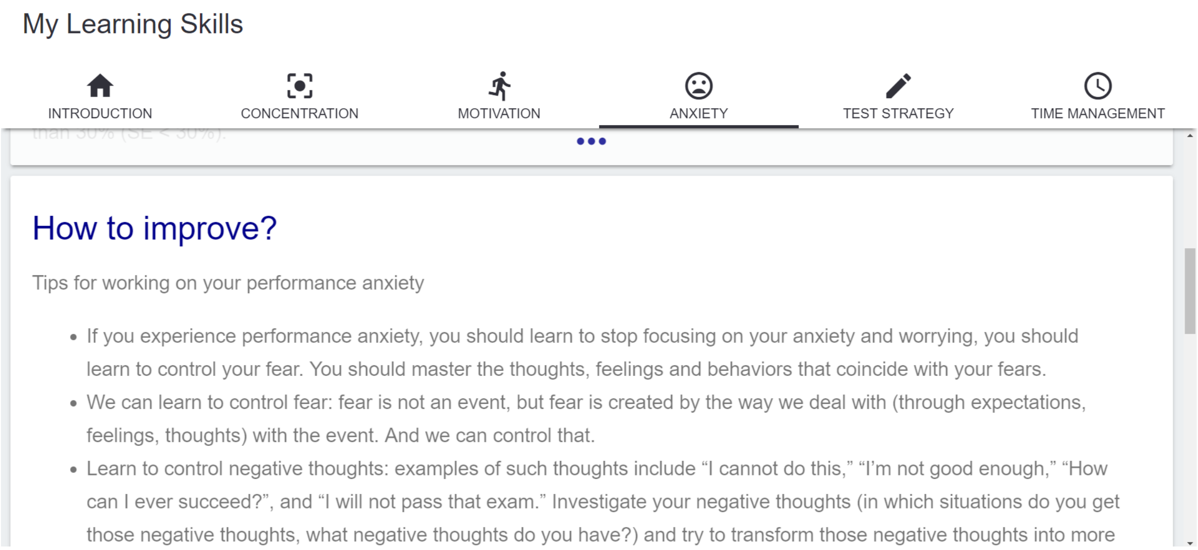
Screenshots of the LASSI dashboard






My Learning Skills
INTRODUCTION
CONCENTRATION
MOTIVATION
ANXIETY
TEST STRATEGY
TIME MANAGEMENT
Your failure anxiety
The test anxiety scale is an indicator of the extent to which you suffer from performance anxiety. Below you see your score for performance anxiety in comparison with other first year students in the @nameStudyProgram@. Every dot represents one student.
VERY HIGH (42)
HIGH (92)
AVERAGE (233)
LOW* (194)
VERY LOW (296)


My Learning Skills
INTRODUCTION
CONCENTRATION
MOTIVATION
ANXIETY
TEST STRATEGY
TIME MANAGEMENT
Previous year
Below you can see which study efficiency (SE) first year students in the @nameStudyProgram@ achieved last year in June, in relation to their anxiety score. The jouw score group has been explicitly labeled. Every dot represents one student with the following color code: a groen dot is a student with a study efficiency higher than 80% (SE ≥ 80%), an geel dot is a student with a study efficiency between 30% and 80% (30% ≤ SE < 80%), and a oranje dot is a student with a study efficiency lower than 30% (SE < 30%).
VERY HIGH (100)
HIGH (100)
AVERAGE (100)
LOW* (100)
VERY LOW (100)


My Learning Skills
INTRODUCTION
CONCENTRATION
MOTIVATION
ANXIETY
TEST STRATEGY
TIME MANAGEMENT
How to improve?
Tips for working on your performance anxiety
• If you experience performance anxiety, you should learn to stop focusing on your anxiety and worrying, you should learn to control your fear. You should master the thoughts, feelings and behaviors that coincide with your fears.
• We can learn to control fear: fear is not an event, but fear is created by the way we deal with (through expectations, feelings, thoughts) with the event. And we can control that.
• Learn to control negative thoughts: examples of such thoughts include “I cannot do this,” “I’m not good enough,” “How can I ever succeed?”, and “I will not pass that exam.” Investigate your negative thoughts (in which situations do you get those negative thoughts, what negative thoughts do you have?) and try to transform those negative thoughts into more

Appendix H

Actions on the dashboard of a random student

Obs	user	timestamp	logevent	logdata
1	r1234567	1509003753		
2	r1234567	1509003753		
3	r1234567	1509003753	refresh	x
4	r1234567	1509003753		
5	r1234567	1509003753		
6	r1234567	1509003753		
7	r1234567	1509003753		
8	r1234567	1509003753		
9	r1234567	1509003753		
10	r1234567	1509003753		
11	r1234567	1509003753		
12	r1234567	1509003753		
13	r1234567	1509003781	tab	1
14	r1234567	1509003813	ping	60−sec
15	r1234567	1509003816	feedback	clear−skip
16	r1234567	1509003818	feedback	useful−skip
17	r1234567	1509003822	active−card	card−con−prev
18	r1234567	1509003848	tab	2
19	r1234567	1509003849	active−card	card−mot−now
20	r1234567	1509003862	tab	3
21	r1234567	1509003863	active−card	card−anx−now
22	r1234567	1509003873	ping	60−sec
23	r1234567	1509003893	tab	2
24	r1234567	1509003894	active−card	card−mot−now
25	r1234567	1509003899	tab	4
26	r1234567	1509003900	active−card	card−tst−now
27	r1234567	1509003923	active−card	card−tst−prev
28	r1234567	1509003931	tab	5
29	r1234567	1509003933	active−card	card−tmt−now
30	r1234567	1509003933	ping	60−sec
31	r1234567	1509003959	tab	3

Obs	user	timestamp	logevent	logdata
32	r1234567	1509003961	active−card	card−anx−now
33	r1234567	1509003967	tab	1
34	r1234567	1509003968	active−card	card−con−now
35	r1234567	1509003973	tab	2
36	r1234567	1509003973	active−card	card−mot−now
37	r1234567	1509003978	tab	3
38	r1234567	1509003979	tab	4
39	r1234567	1509003980	active−card	card−tst−now
40	r1234567	1509003982	tab	5
41	r1234567	1509003983	active−card	card−tmt−now
42	r1234567	1509003993	ping	60−sec
43	r1234567	1509003998	tab	0

Appendix I

Assumptions of the model of total time spend on the dashboard

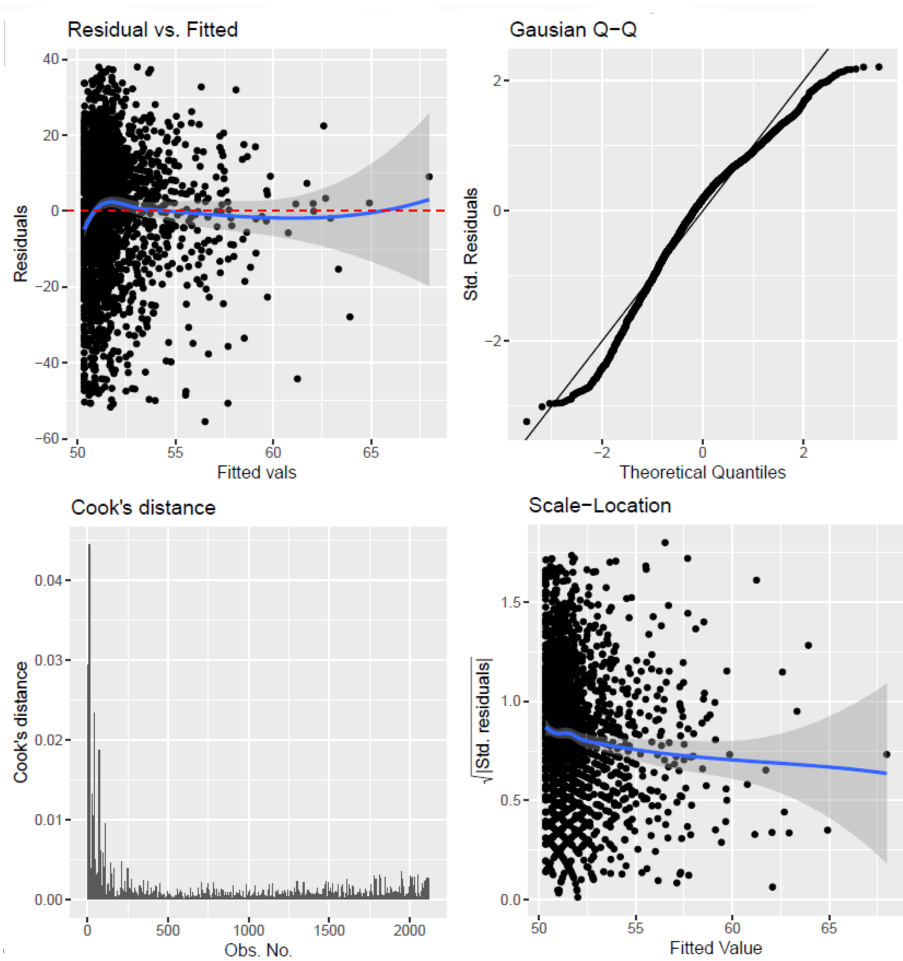


Figure I.1: Diagnostic plot of the linear regression model of the total time spend on the dashboard.

Appendix J

Distribution of the ranks

J.1 Distribution of the ranks of the engineers and engineers-architects

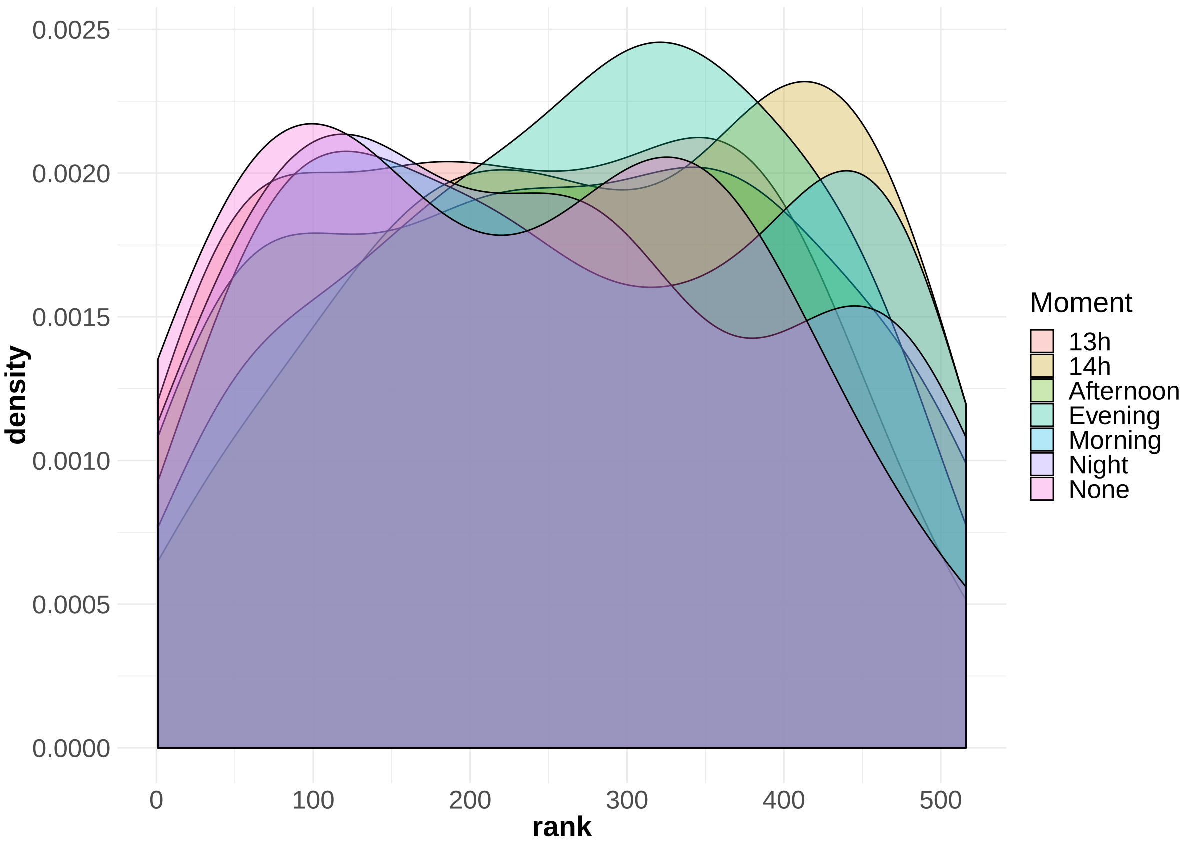


Figure J.1: Distribution of the ranks of the engineers and engineers-architects.

J.2 Distribution of the ranks of science and engineering technology students.

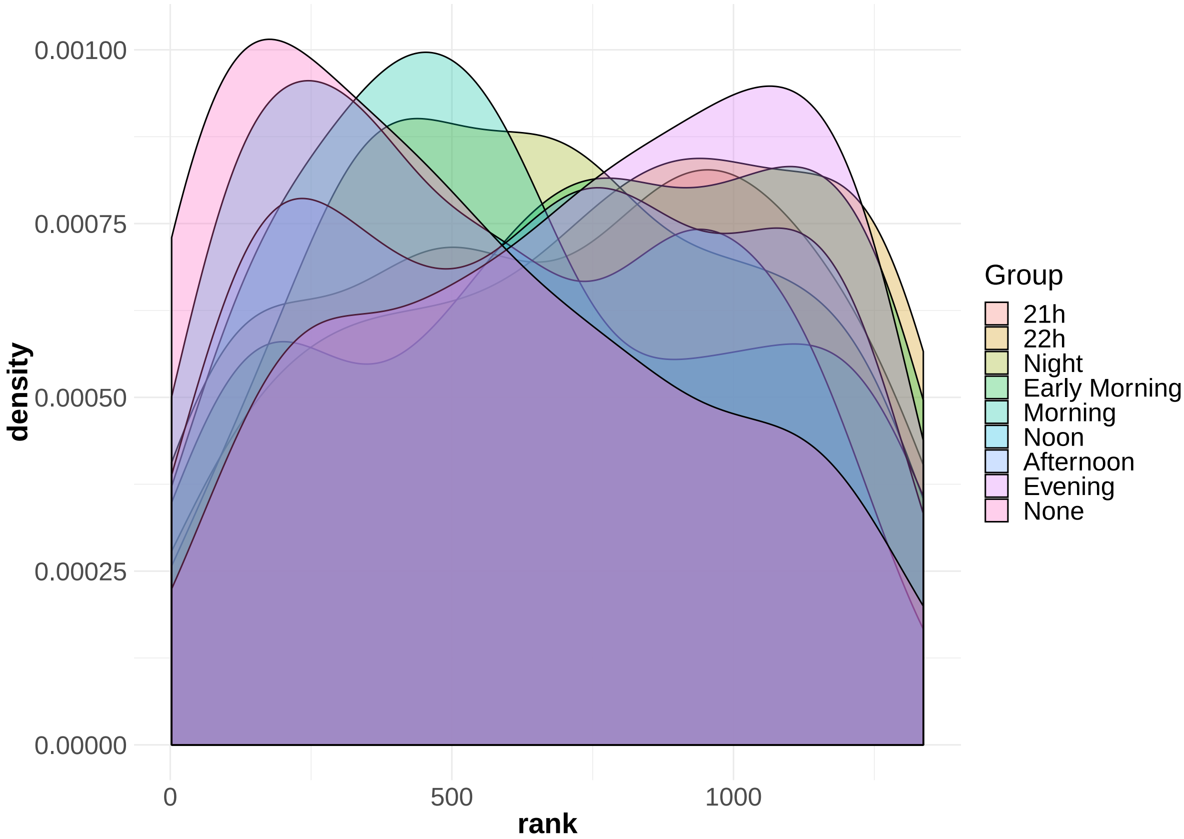


Figure J.2: Distribution of the ranks of science and engineering technology students.

Appendix K

Pairwise contrasts of timepoints

Table K.1: *Holms correction applied on the pairwise contrasts of time points.*

Contrast	p-value	P_k	Reject?
None- Evening	0.0000019	0.00139	yes
None-Early Morning	0.00001	0.00142	yes
Evening-Morning	0.0000720	0.00147	yes
None- 22h	0.0001	0.00152	yes
Morning-22h	0.00028	0.00156	yes
Noon-Evening	0.00044	0.00161	yes
Early Morning-Morning	0.00057	0.00167	yes
Noon- 22h	0.00076	0.00172	yes
None-Afternoon	0.00192	0.00179	no
Early Morning- Noon	0.00238		no
None-Night	0.00314		no
None- 21h	0.00428		no
Evening-Night	0.04070		no
22h- Night	0.05758		no
Noon- Afternoon	0.05868		no
Afternoon- Evening	0.06392		no
Noon-Night	0.07904		no
Morning-Afternoon	0.08117		no
Afternoon-22h	0.08916		no
Morning-21h	0.09565		no
Morning-Night	0.10441		no
Early Morning- Night	0.11219		no
Noon- 21h	0.11238		no
None- Morning	0.11842		no
None- Noon	0.13835		no
Early Morning- Afternoon	0.17391		no
Evening-21h	0.41181		no
21h-22h	0.44842		no
Morning- Noon	0.75279		no
Early Morning- 21h	0.78381		no
Afternoon- Night	0.81449		no
21h- Night	0.81475		no
Early Morning- 22h	0.83490		no
Early Morning- Evening	0.85669		no
Afternoon-21h	0.87914		no
Evening-22h	1.000		no

Appendix L

Assumptions of the model of the lag before visiting the dashboard

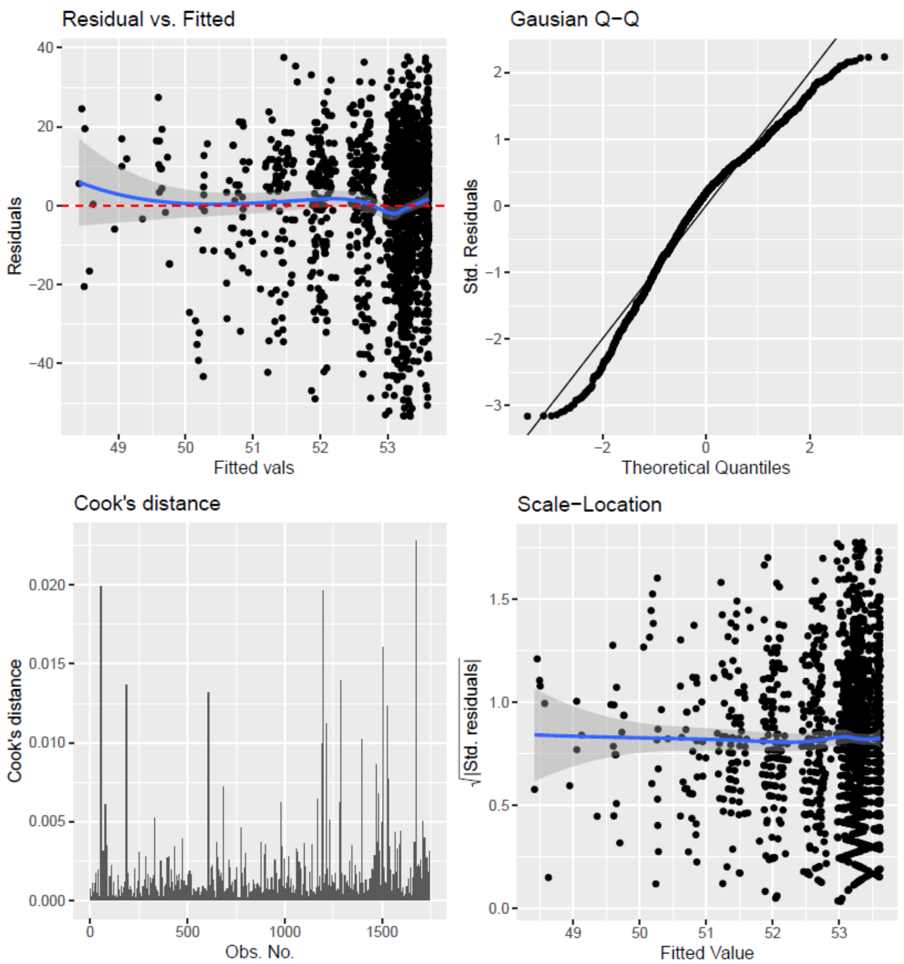


Figure L.1: Diagnostic plot of the model with the lag before visiting the dashboard.

Appendix M

Proportions of students that visited the tips in 2018-2019

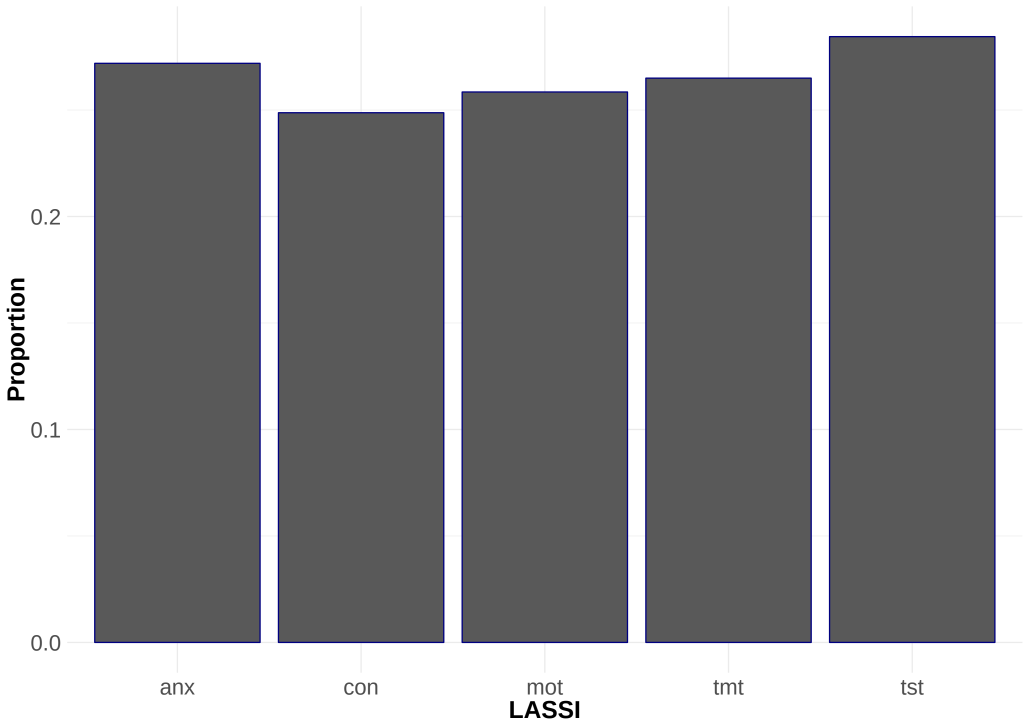


Figure M.1: Proportions of students that clicked on the tips in the academic year 2018-2019.

Appendix N

Heat maps of the interaction between dashboard usage and visiting the tips with Lassi

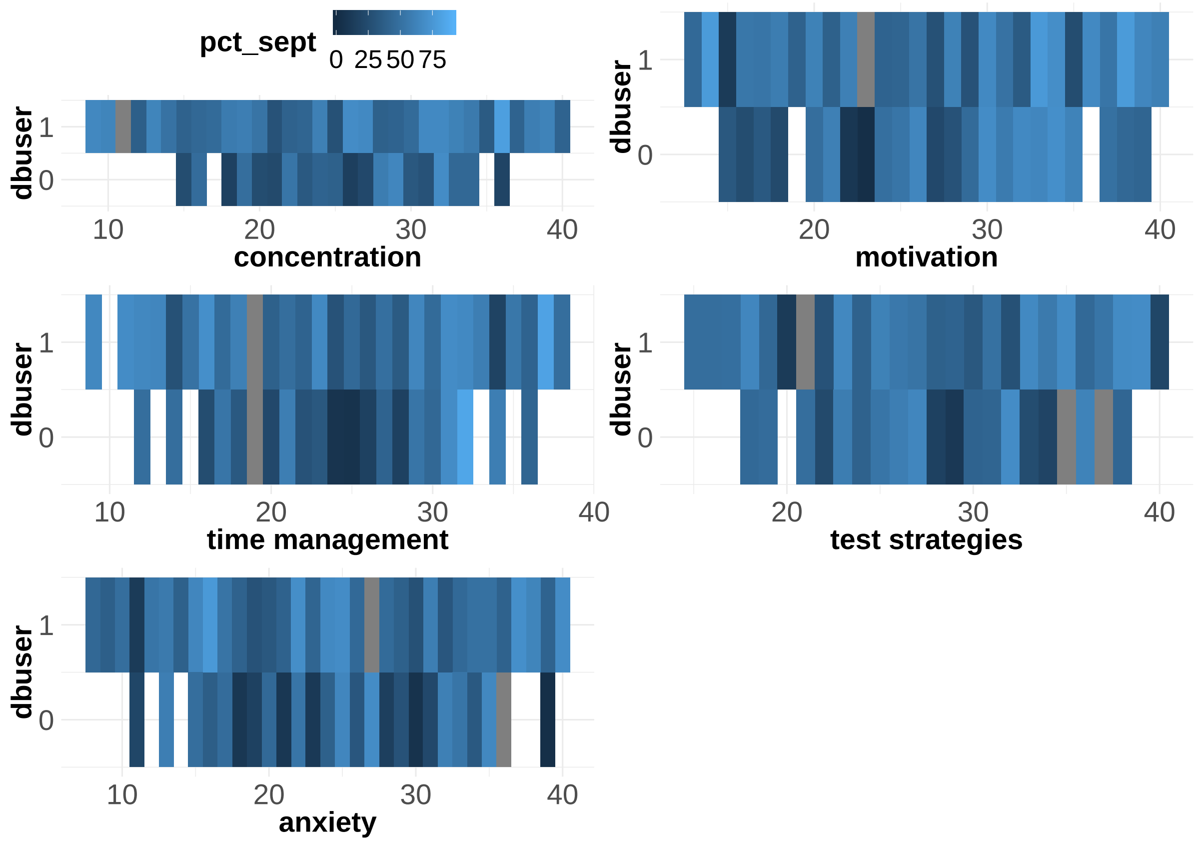


Figure N.1: Heat map of the interaction between learning skills and dashboard usage.

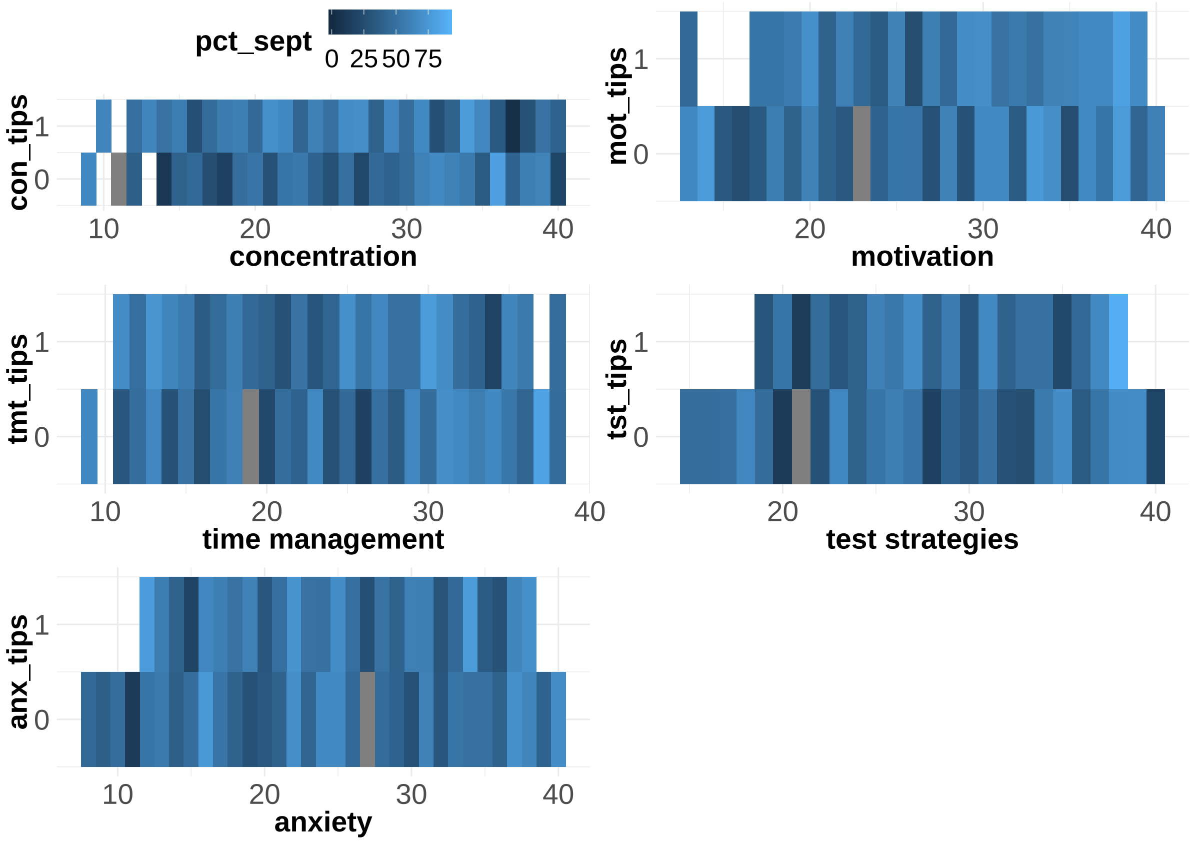


Figure N.2: Heat map of the interaction between learning skills and clicking on the corresponding tip.

Appendix O

Assumptions of the regression model of the tips

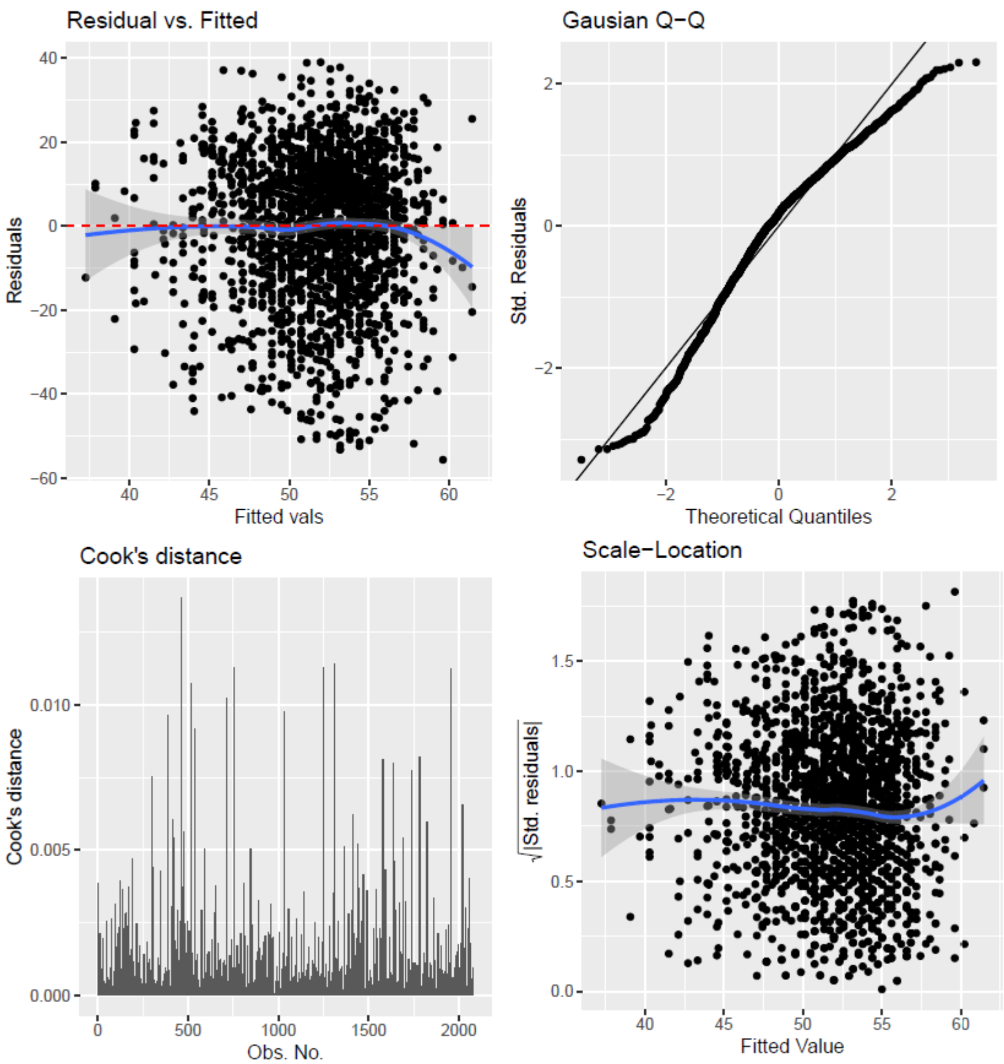


Figure O.1: Diagnostic plot of the linear regression model of concentration.

The Cooks distances plot of anxiety shows one spike. This corresponds to the observation of a student with very low anxiety (39), that did not visit the dashboard or the tips. The student has a weighted percentage in September of 3. Since the Cooks distance is not alarming (d=0.018), the observation is not altered or deleted.

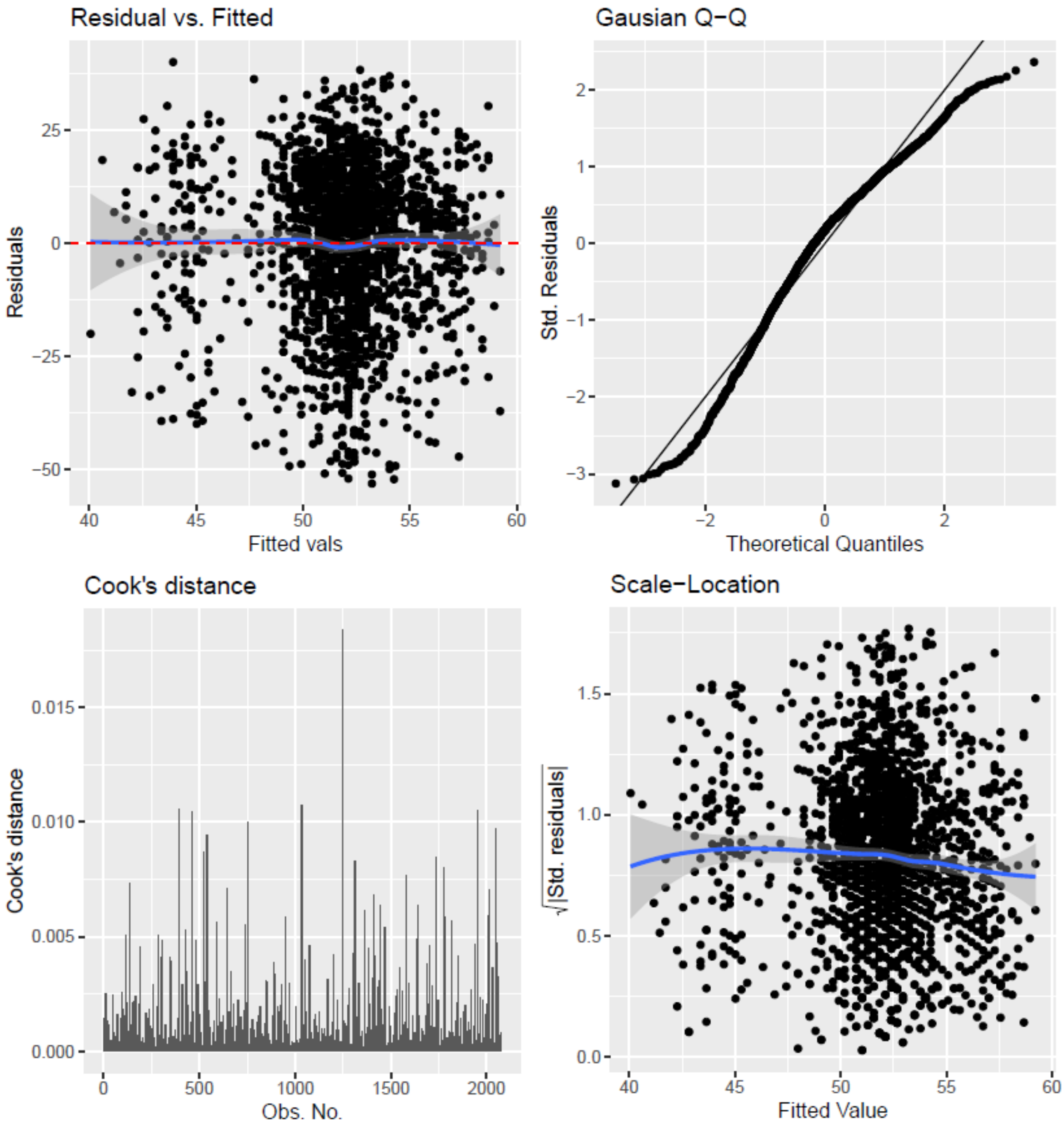


Figure O.2: Diagnostic plot of the linear regression model of anxiety.

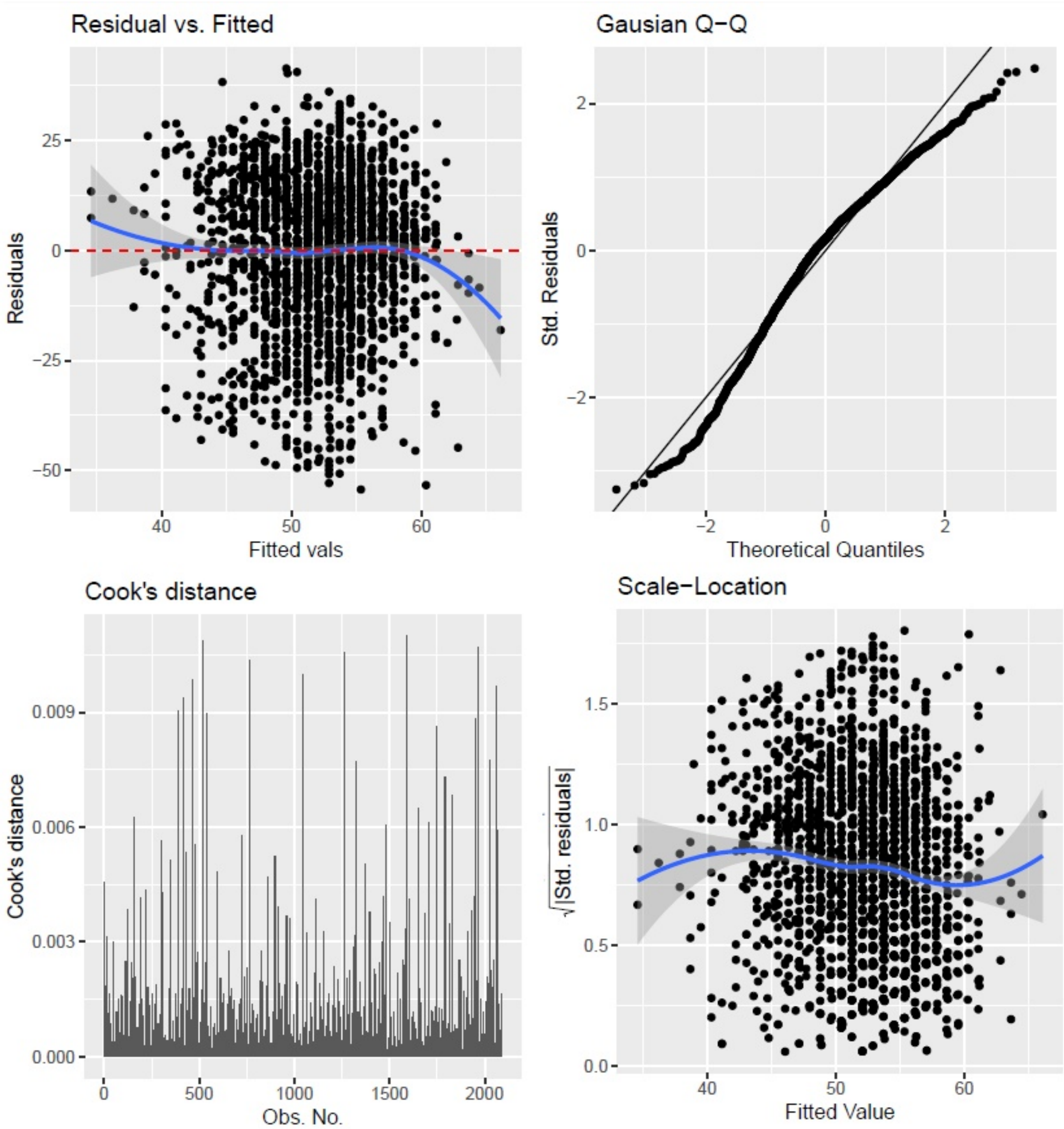


Figure O.3: Diagnostic plot of the linear regression model of time management.

The Cooks distances plot of test strategy shows one spike. This spike corresponds the same observation that caused a spike in the Cooks distances of anxiety. The student has very good test strategies (35), but did not visit the dashboard nor the tips. Still, the student has a weighted percentage in September of 3. The observation is not altered or deleted since the Cooks distance is not alarming (d=0.019).

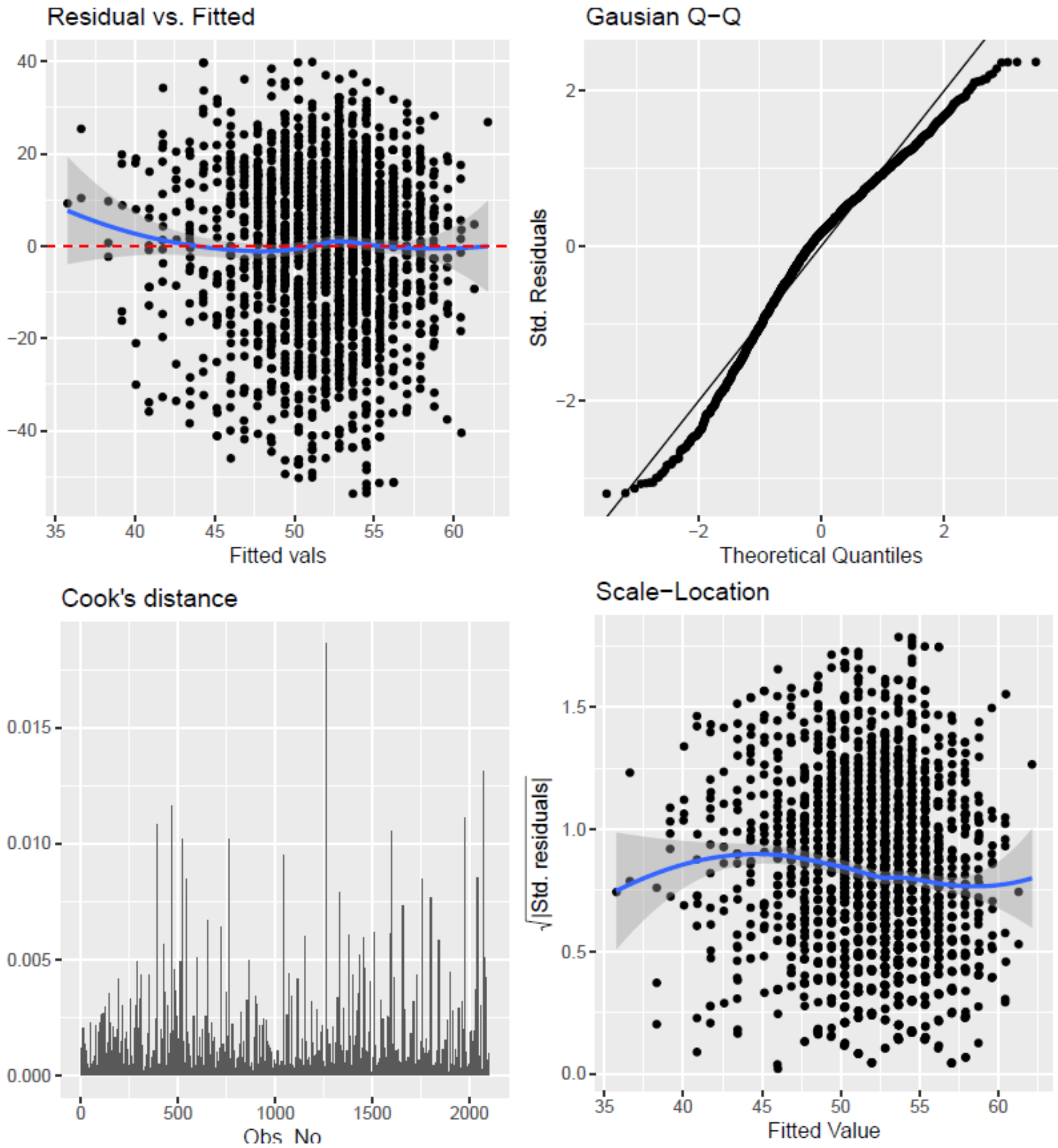


Figure O.4: Diagnostic plot of the linear regression model of test strategy.

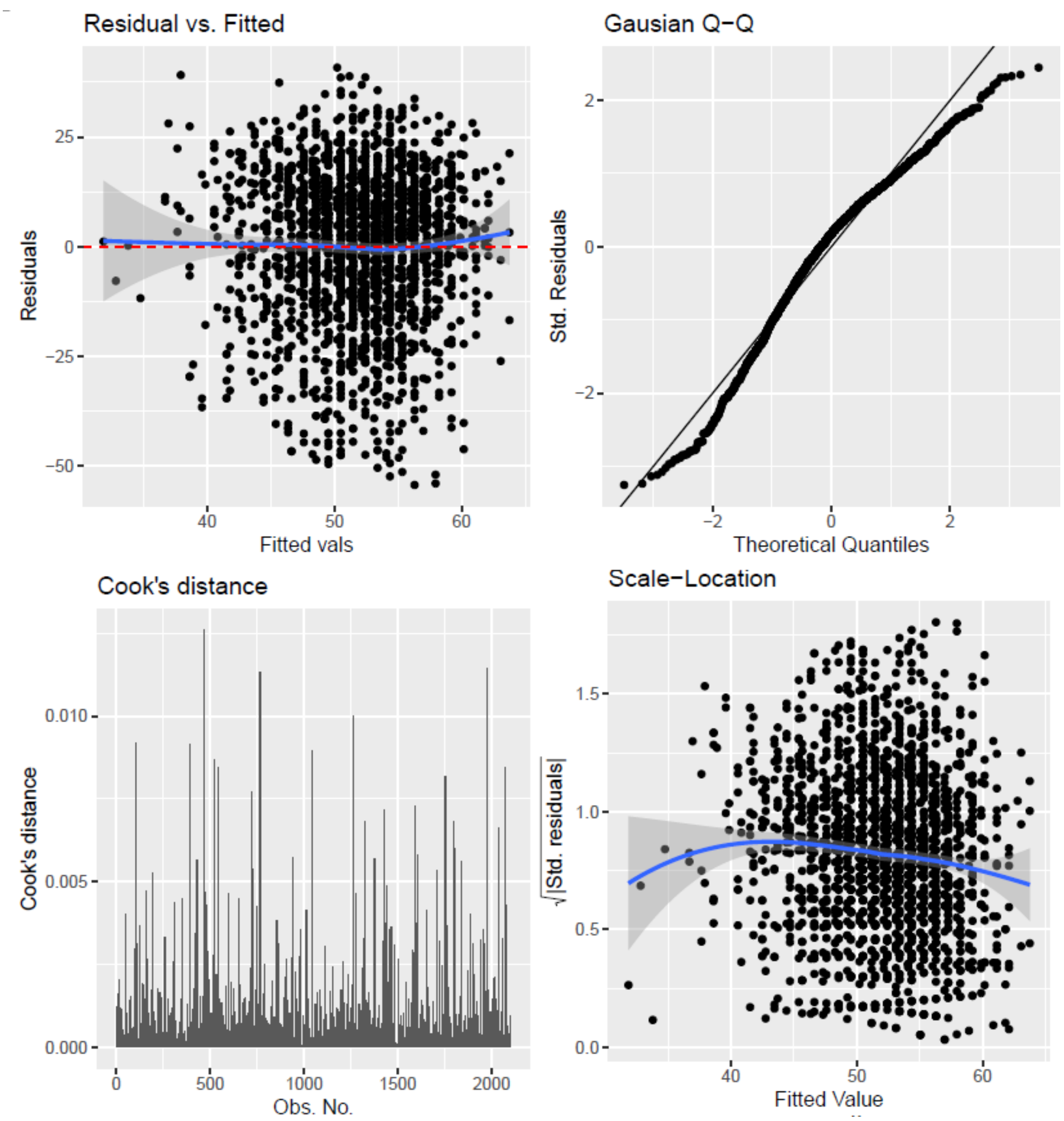


Figure O.5: Diagnostic plot of the linear regression model of motivation.

Appendix P

Assumptions of the regression model of the number of visitations

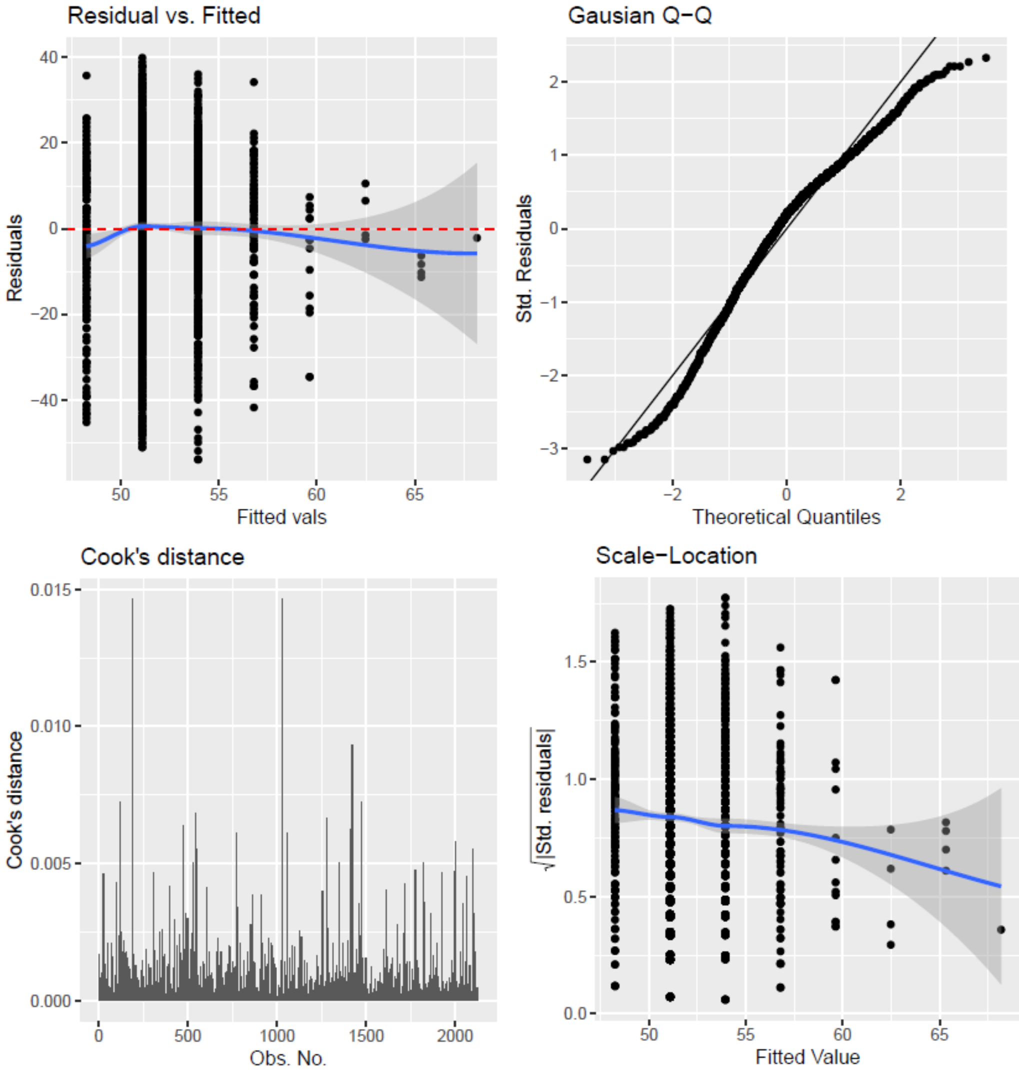


Figure P.1: Diagnostic plot of the linear regression model of the number of visitations.