Student differences in regulation strategies and their use of learning resources: implications for educational design

Bos, N.R.; Brand-Gruwel, S.

Note: To cite this publication please use the final published version (if applicable).

Student differences in regulation strategies and their use of learning resources: implications for educational design

Nynke Bos
University of Amsterdam
Faculty of Social and Behavioural Sciences
Amsterdam, The Netherlands
N.R.Bos@uva.nl

Saskia Brand-Gruwel
Open University of the Netherlands
Faculty Psychology and Educational Sciences
Heerlen, The Netherlands
Saskia.Brand-Gruwel@ou.nl

ABSTRACT

The majority of the learning analytics research focuses on the prediction of course performance and modeling student behaviors with a focus on identifying students who are at risk of failing the course. Learning analytics should have a stronger focus on improving the quality of learning for all students, not only identifying at risk students. In order to do so, we need to understand what successful patterns look like when reflected in data and subsequently adjust the course design to avoid unsuccessful patterns and facilitate successful patterns.

However, when establishing these successful patterns, it is important to account for individual differences among students since previous research has shown that not all students engage with learning resources to the same extent. Regulation strategies seem to play an important role in explaining the different usage patterns students' display when using digital learning recourses. When learning analytics research incorporates contextualized data about student regulation strategies we are able to differentiate between students at a more granular level.

The current study examined if regulation strategies could account for differences in the use of various learning resources. It examines how students regulated their learning process and subsequently used the different learning resources throughout the course and established how this use contributes to course performance.

The results show that students with different regulation strategies use the learning resources to the same extent. However, the use of learning resources influences course performance differently for different groups of students. This paper recognizes the importance of contextualization of learning data resources with a broader set of indicators to understand the learning process. With our focus on differences between students, we strive for a shift within learning analytics from identifying at risk students towards a contribution of learning analytics in the educational design process and enhance the quality of learning; for all students.

Categories and Subject Descriptors
• **Applied computing~Computer-assisted instruction**
• *Mathematics of computing~Exploratory data analysis*
• Theory of computation~Unsupervised learning and clustering

Keywords
Individual differences, regulation strategies, blended learning, cluster analysis, learning dispositions.

1. INTRODUCTION
All animals are equal, but some animals are more equal than others [32]. Although it initially seems that Orwell's criticism on the totalitarian political system of the Soviet Union has nothing to do with learning analytics, a closer look at this commandment suggests otherwise. Learning analytics, in general, treats all students as equal, while in fact some students are more equal than others.

The objectives relating to the use of learning analytics can globally be described to serve six goals: predicting course performance and discovering learner models; suggesting relevant learning resources to student; increasing reflection and awareness about the learning process; enhancing social learning environments by visualization of social interactions; detecting undesirable learning behaviors; and detecting affects of learning like boredom or confusion [37]. Although these issues are highly interrelated, the majority of the learning analytics research focuses on the prediction of course performance and modeling student behaviors [12] targeted on identifying students who are at risk of failing the course. This focus has a longer tradition within the educational data mining community, which could account for this overrepresentation. However, when modeling student behavior or predicting course performance to identify at risk students, learning analytics research focuses often on trace data from just one data source, for example the use of formative assessments or the number of comments in a forum or the hits in the Learning Management System (LMS). In doing so, learning analytics research ignores other course elements or other available trace data and draws conclusions based on just a fraction of the course. The risk of those isolated predictions is that they are detached from pedagogical experiences, practices [16, 26] and interventions [45], which reduce learning analytics to a series of clicks, and page visits. In order to avoid these isolated predictions all trace data of all available course elements and learning resources should be taken into account.

Second, learning analytics should have a stronger focus on improving the quality of learning for all students, not only identifying at risk students [26]. The predictors for failure or poor course performance, which are currently found, use predictive modeling techniques for at risk students and cannot be reasonably translated into recommendations to improve quality of learning

for all students [16]. If learning analytics wants to enhance the quality of teaching and learning, a shift from predictive modeling to identify students at risk towards pedagogical learning analytics interventions is needed [45].

With a shift in a focus from identifying at risk students or predicting failure towards improving quality of learning, learning analytics needs to become an element of the learning design process [45]. However, current learning analytics research provides us with a limited amount of information on how to improve the quality of education and have an impact on the design process. Learning data analysis could help us to identify successful students and their use of the specific learning resources. If we can subsequently identify ahead of time what successful and unsuccessful patterns look like [25] and adjust the course design according to those patterns we can redirect, and maybe even avoid, unsuccessful use of learning resources and facilitate successful patterns.

However, when establishing these successful patterns, it is important to account for individual differences among students. Current learning analytics research often takes course averages, for example the average amount of clicks within the LMS or the average time spent on online learning activities, as a target for predictive measures. However, this could lead to a false reference point for some groups of students since other groups of students can be overly active or inactive and hence influence the average activity [45]. When examining successful and or unsuccessful patterns it is important to provide aggregate measures for similar kind of students. However, it is not clear on which criteria we should aggregate these students. Learning analytics could determine how students interact with digital learning resources and establish successful or unsuccessful patterns of this interaction. Research shows that students do not interact with digital learning resources in the same way [14, 26, 27, 28, 29] and use different learning approaches when using digital learning recourses [16, 22]. Current learning data analysis uses measures, as hits in the LMS, which do not reflect individual user differences and several researchers propose that learning analytics data should be contextualized with a broader set of indicators [9, 16, 28, 44]. In doing so, trace data will reflect more that solely hits and clicks, but it opens up the opportunity to differentiate between differences in students their learning approaches and learning strategies [16]. So, adding more contextualized data to trace data from the different learning resources could mean a shift in the direction towards pedagogical learning analytics.

In summary, within learning analytics research a greater focus is needed on understanding student behavior so learning analytics can truly improve the quality of learning which goes further than targeting at risk students. To understand student behavior, current data sources need to be supplemented with contextualized data approaches to learning so individual differences can be accounted for. Or, as Orwell would put it, to get insight into why some students are more equal than others.

The current research aims to provide insight into student behavior by focusing on individual differences in approaches to learning and subsequently how these differences affect the use of (digital) learning resources. First we will explore individual differences in the use of learning resources in a blended learning setting and its connection with regulation of the learning process. The next section describes the methodology of the current research and subsequently the results are presentenced. The paper concludes with a discussion and lines for future research.

2. BACKGROUND
2.1 Individual differences
One common source of trace data, which is often used to model student behavior or predict course performance, is data from the LMS. Results on the strengths of these predictions are inconclusive although the majority of the results indicate that duration of use has no direct impact on course performance [30, 35, 47]. Although these studies do differentiate between different LMS variables, for example messages read, quizzes taken or time spent online, they do not account for individual differences in the use of these available digital learning resources. However, research suggests that students show some distinct usage patterns when offered different digital learning resources. For example, within a blended learning course students either rely heavily on one of the digital resources while ignoring other digital resources [24], do not use the resources at all [27, 31], or use it as a substitute for the face-to-face activities [8, 41].

Several studies conducted a cluster analysis based on trace data to identify these different usage patterns. For example [28] found four different clusters that reflect differences in the use of the digital learning resources: the no-users, the intensive-active users, selective users and intensive superficial users. Similarly [22] found, also based on cluster analysis, several different user profiles based on the use of digital learning resources and suggest that these differences might be related to differences in students' metacognition and motivation. More specific research shows [14] by performing a cluster analysis based on students' conceptions and approaches to learning, two different profiles: a cluster focused on understanding and a cluster focused on reproduction with subsequently differences in course performance.

Although the aforementioned studies show that students do differ in the use of the learning resources within a blended course. However there is little insight into why students do or do not use certain digital learning recourses and what the consequences of these (un)conscious choices are in relation to course performance, although research suggests that goal- orientation [28], approaches to learning [14] and the differences in instructional models [17] may be an important predictor of frequency and engagement of use.

2.2 Importance of self-regulation
Agency refers to the capacity to coordinate learning skills, motivation and emotions to reach the goals. Self-regulated learners exercise agency as they engage in a cycle of four main stages: *analyzing* the task; setting goals and *designing* plans; *engaging* in learning; and *adjusting* their approach to learning [44, 46]. One important aspect of student self-regulation of learning is the decision on if and if so, how to use the learning resources offered during a course: the learner agency [2]. The ability to self regulate the learning process is reflected by effective approaches and choices towards learning which are reflected in the capability a student has to handle a difficult task, practice and evaluate their learning and subsequently develop a deep understanding of subject matter [33]. The ability to self-regulate the learning process in an effective way is linked to academic success [34, 42]. Students' personal approaches to learning are intertwined with various other aspects of learning such as motivational aspects and regulation of the learning process [39] and goal orientation [28].

2.2.1 Regulation of the learning process
Not all students are able to regulate their learning in an effective way and some students rely on an external source to regulate their learning process. This concept is known as external regulation.

This external source could be, for example, the instructor who guides students through course material or the external source could be the learning objectives of the course. There are also students who suffer from a lack of regulation. These students have difficulties in regulating the learning process as a whole and do not find any support from internal or external sources. Research shows that a student their goal orientation plays an important role with regard to if and how a specific learning resource is being used [26]. Students with a performance goal orientation show a selective use of the learning resources, while students with a mastery goal orientation show an active choice in their learning resources. However, these differences in goal orientation and consequences for the use of learning resources are not confirmed by a similar study [4]. Moreover, studies that report these differences among students in terms of their use of learning resources acknowledge, all in retrospect, the importance of students' goal orientation, self-regulation and approaches to learning when shaping these profiles [22]. The question remains however if differences in regulation strategies actually *causes* differences in the use of learning resources.

The majority of the research about regulation strategies takes place in traditional settings of education. Fewer studies have been conducted on the role of regulation strategies and their implications for online or blended learning, although regulation strategies do have an impact on the use of digital learning resources. Students who are able to self-regulate their own learning are likely to use digital learning recourses differently than students who use an external regulation strategy. For example, in their research [11], show that students with a tendency to external regulate their learning have a higher amount of logons to the LMS compared to the group of students who were better at self regulating their learning. Similar results were found in [36] wherein students, within a blended learning course on statistics, show distinct differences in their use of the digital learning resources based on their regulation strategy. Also [29] investigated how students regulate the use of different learning resources throughout the course by temporal analysis. A cluster analysis showed that only a minority of the students (3%) regulated the use of the learning resources in line with the course phases and hence with the changing requirements of the course.

To sum up, self-regulation seems to play an important role and seem to have an impact on the use of learning resources. However, it remains unclear if these differences are actually caused by differences in regulation of learning. Further research into cause and effect of regulation strategies and the use of learning resources must determine if differences in regulation strategies causes these differences so successful patterns can be identified and adjust the course design accordingly.

2.3 Course design

The design of a course determines to a large extent if predictive modeling techniques will find significant predictors on the use of digital learning resources [17]. If a course is designed which requires a fair amount of LMS usage, a greater predictive value of LMS components will be found compared to a different designed course in which the LMS usage has a less prominent role [1, 16]. Within blended learning the dominant role still lays within face-to-face educational activities and the digital learning resources, among the LMS, has a less prominent, and often, supporting role.

Blended learning is often associated with student-oriented learning, in which students have varying degrees of control over their own learning process [24]. The current notion of blended learning is often an instructor-oriented approach in which the instructor determines the digital learning recourses that will be used during the course [18]. When blended learning design focuses on students and their choices to use the digital learning recourses, there is a large variety in the use of these recourses by students. Students use digital recourses in different ways, which were often not intended for in the educational design. For example, [20] find that when offering student optional learning resources in a blended course, students rely heavily on one supporting medium. They conclude that students do not create blended learning—a mix of different digital learning resources—and they suggest that students need explicit guidance in how to effectively combine learning resources. Also [27] finds three distinct usage patterns in their research on the usage of digital learning resources in a blended learning course: no-users, intensive users and incoherent users. They find a significant lower course performance of the no-users group. The authors provide no explanation for the causes of these differences in the use of the learning resource but suggest these differences might reflect students' dispositions as motivation, regulation strategies or metacognitive ability.

In their research on the use of lecture recordings [8], in which face-to-face lectures are recorded and made available afterwards, they find similar usage patterns: no-users, supplemental users and substitute users. What this study illustrates, is that within a blended learning setting the offline educational activities, for example face-to-face lectures, have a direct impact on the use of digital learning resources. When analyzing learning data within a blended learning setting one should take into account the course design; in this case the direct relation between the use of digital learning recourses and attendance to face-to-face activities, since the use of digital learning resources is directly influenced by supplemental or substitutional use. So, besides contextualizing learning analytics data with data about regulation strategies, learning analytics data should also consider attendance data for face-to-face activities within a blended learning setting to account for influences of the course design.

In summation, current use of learning analytics often uses trace data from one learning resource for predictive modeling to identify students who are at risk of failing a specific course. Learning data analysis should, ultimately, contribute to the quality of teaching and learning and should be an integrated part of the educational design process. With this direction towards pedagogical learning analytics we need to define what success looks like and how individual differences in the use of digital learning resources influences these pathways and subsequently influence the learning design process.

In line with recommendations made by [16, 45] to move beyond predictive analytics, we focus in this study on the differences in regulation strategies, and analyze how differences in regulation strategies reflect in the use of different learning recourses. When examining the use of different learning resources, we combine the use of offline learning resources (face-to-face activities) with the use of online, digital learning resources, since these two nodes of delivery are inextricably linked together in a blended learning setting.

This research aims to answer the following questions:

1. Can we identify different clusters of students based on differences in their regulation strategies?
2. Do these differences in regulation strategies reflect in differences in the use of (digital) learning resources?
3. What combinations of (digital) learning resources contribute most to course performance for each cluster?

4. Do differences in regulation strategies reflect in differences in course performance?

3. METHODS

3.1 Participants

The participants were 333 first year university Psychology students (243 female, 90 male, M_{age} = 20.17, SD_{age} = 1.66) attending an obligatory course on Biological Psychology. Students who took the course as an elective or had taken the course before were removed from the dataset.

3.2 The Blended Learning Course

The course consisted of 17 face to face lectures, with a 120-minute duration and a 15-minute break in half time over a period of 8 weeks. These lectures were university style lectures, with the instructor lecturing in front of the class. The face-to-face lectures were recorded and made available directly after the lecture had taken place and were accessible until the exam had finished. In the course design the recorded lectures were offered to students with the aim of supplementing the face-to-face lecture. If parts of the lectures were unclear, students could use the recorded lectures to revise these parts or revise the entire lecture if needed.

During the week several small workgroups were organized with mandatory attendance. Before these workgroups, students had to complete several assignments in the digital exercise book, which contains additional study materials, supplemented with formative assessments. Completing the formative assessments was mandatory, passing or failing these formative assessments was not. In total there were nine formative assessments available for students.

Within the LMS, students had access to extra study materials, like short introduction videos about certain concepts or additional reading materials available for download.

During the eight-week course there were two separate summative assessments. The first assessment covered the first four weeks of the course and had a focus on assessing the knowledge domain. The second assessment covered the last four weeks of the course and had a focus on assessing higher order thinking skills. The final grade for the course was calculated by taking the mean of both assessments.

Upon completion students received 6 European Credit Transfer and Accumulation Systems (ECTS).

3.3 Measurement instruments

Before the start of the course students were informed about the research and were asked to consent. Three students did not consent and were removed from the results.

In line with [28] we used multiple log indicators to capture ways students used the learning resources. In most cases the frequency of use and the duration of use were logged.

3.3.1 Attendance to face-to-face lectures

During the entire time frame of the lectures, student attendance was registered on an individual level by scanning student cards upon entry of the lecture hall. The scanning continued until 15 minutes after the lecture had started. The presence of the students was registered for all 17 lectures of the course.

3.3.2 Viewing of the recorded lectures

The viewing of the lecture recordings was monitored on an individual level and could be traced back to date, time, amount and part of the lecture viewed. For each lecture a separate recording was made, which made it possible to track the amount of minutes a student watched a specific lecture. Following the recommendations made by [23] the time on task measure was calculated based on data cleaning methods used by [19] wherein sessions shorter than two minutes were not considered to reflect actual use. Moreover, besides removing the outliers, the time-out chosen was four hours.

3.3.3 Formative assessments

For each formative assessment a log file within the LMS was created to determine if a student completed the formative assessment. Although passing or failing the formative assessments was not part of the design of the course, these grades were stored in the LMS. During the course students were obligated to complete 7 of the 9 formative assessments in order to pass the course. So besides the number of formative assessments completed, also the average score of the completed assessments was calculated.

3.3.4 LMS data

Two different types of LMS data were gathered. Except the previously mentioned digital resources, the recorded lectures and the formative assessments, the LMS also offered Powerpoint slides and additional reading materials (PDF) for download as well as some illustrative videos about certain topics. First the total amount of hits within the LMS was registered. These are hits as clicking on links to the recordings or formative assessments, clicking on announcements, checking grades, clicking on links to PDF files or links to certain video files. Second the total time spent in the LMS during the course was registered. This is the total time in minutes a student was logged on to the course in the LMS during the entire timeframe of the course. This measure was calculated by accumulating the time differences between logging on the course and subsequently logging of or logging on to another course.

3.3.5 Summative assessments

During the eight-week course, there were two separate summative assessments, which were scored on a scale from 1 to 10, with 10 the highest, and 5.5 as a pass mark. The first assessment covered the first four weeks of the course and the second assessment covered the last four weeks of the course. Both assessments contained 20 multiple-choice questions and 2 short essays questions. The final score for this course was calculated by taking the mean of these two assessments.

3.3.6 Inventory Learning Style (ILS)

The Inventory Learning Style (ILS) [39] is a self-report diagnostic instrument intended to measure aspects of study method, study motives and mental models about studying in higher education. The ILS consists of 120 items and contains four domains: processing strategies, regulation strategies, learning orientation and mental models of learning. For the purpose of the current study only the sub-scales of the domain regulation strategies were scored. These sub-scales are: self-regulation (11 items), external regulation (11 items) and lack of regulation (6 items). For a complete description of the ILS and each of its subscales we refer to [39].

The ILS was offered to students during the first week of the course. Completing the ILS was mandatory.

3.4 Data analysis

To establish differences in regulation strategies of students at the beginning of the course, we performed a two-step cluster analysis on ILS regulation strategy data. A two-step cluster analysis determines the natural and meaningful differences, formed in

clusters, which appear within the current population. The two-step method is preferred over other forms of cluster analysis when both continuous and categorical variables are used and when the amount of clusters is not pre-determined [10]. Cluster analysis was chosen over scoring the subscale regulation as one factor model since students tend to show variations in the way they regulate learning throughout the course, depending for example on the task at hand [43].

Next a MANOVA between the different clusters was conducted to determine significant differences in the use of (digital) learning recourses. The MANOVA was used to determine if certain clusters, based on regulation of the learning process, made a significant amount more use of certain learning resources than others.

Third a stepwise multivariate analysis was conducted for each cluster to determine the relative contribution of each of the different learning resources on course performance. This was done to determine which (combination of) learning resources contribute to the final grade for each separate cluster and differences between clusters.

Finally an ANOVA determined if there were any significant differences between the different clusters and course performance.

4. RESULTS

First, before determining how students differ in their regulation strategies, the reliability of the subscales of the ILS domain regulation strategies were calculated. The results can be found in table 1.

Table 1: Reliability of the ILS subscale Regulation Strategies

Subscale	Reliability (Cronbach's Alpha)
Self Regulation	.76
External Regulation	.71
Lack of Regulation	.73

4.1.1 Cluster analysis

Since the subscales show sufficient reliability the next step was to cluster students based on their reported regulation strategies. Using the two-step auto-clustering algorithm, 333 students were assigned to different clusters. The auto-clustering algorithm indicated that three clusters was the best model, because it minimized the Schwarz's Bayesian Criterion (BIC) value and the change in them between adjacent numbers of clusters (Table 2). The clustering criterion (in this case the BIC) is computed for each potential number of clusters. Smaller values of the BIC indicate better models. The improvement in the cluster solution, as measured by the BIC Change, is not worth the increased complexity of the cluster model, as measured by the number of clusters. The ratios of BIC change for the four cluster model is small, while a three cluster model shows clear distinct patterns.

Table 3 provides insight into the distribution of the three cluster solution based on the regulation strategies of the students. For each cluster the means are reflected as well as the means for the entire population.

Table 3 shows some distinct patterns in the ways students regulate their learning. Students in cluster 1 show no dominant regulation pattern, indicating that these students have no clear pattern to regulate their learning.

Table 2: BIC changes in de auto-clustering procedure.

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change[a]	Ratio of BIC Changes[b]
1	725.80		
2	643.95	-81.85	1.00
3	576.39	-76.57	.825
4	565.05	-11.35	.139
5	555.89	-9.46	.116
6	554.90	-.69	.008
7	557.10	2.21	-.03

a. The changes are from the previous number of clusters in the table.
b. The ratios of changes are relative to the change for the two-cluster solution.

Table 3: Distribution of regulation strategies for three clusters

Cluster number	1	2	3	All
N	128	95	110	333
Self-regulation	21.72	25.36	33.13	26.53
External Regulation	30.99	37.48	35.86	34.45
Lack of Regulation	12.96	19.09	12.71	14.63

Students in cluster 2 use a combination of two regulation strategies: lack of regulation and external regulation. They seek guidance in the learning process from external sources but when this external regulation fails, for example by absence of the instructor or unclear learning objects, they tend to show a lack of regulation.

Also cluster 3 shows a combination of two regulation strategies. They try to self-regulate their learning but when they fail they use an external source to in order to compensate for this.

Cluster 3 students are mainly able to self regulate the learning process, but when they fail to do, they use an external regulation strategy to compensate for this deficiency.

Cluster analysis indeed revealed some distinct patterns in the ways student regulate their learning with a group showing a tendency to use an external source to regulate their learning, a group who is mainly able to self regulate the learning process and a group showing no distinct preference in how they regulate their learning.

Different usage patterns

Next a MANOVA determined if there were any significant differences between the three different clusters and the use of different (digital) learning resources: lecture attendance, recorded lectures, hits in Blackboard, Blackboard duration and average score on formative assessments.

Table 4 shows the means of the use of learning recourses for each cluster. Cluster 2 students, mainly characterized by external regulation, show a greater use of the different learning resources. However, the differences in the use of the learning resources between the three clusters are not significant with $F(2,330) = 1.971$, $p = .141$ for lecture attendance, $F(2,330) = .046$, $p = .995$ for recorded lectures, $F(2,330) = 1.247$, $p = .289$ for hits in Blackboard, $F(2,330) = 3.206$, $p = .042$ for Blackboard duration and $F(2,330) = .279$, $p = .757$ for the average score on the formative assessments.

Table 4: Means of the use of learning recourses for the three clusters

		Lecture attendance		Recorded Lectures		Hits in Blackboard		Blackboard use (Minutes)		Formative Assessments (average score)	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Cluster 1	128	4.75	4.02	451.21	468.03	433.16	116.90	638.92	408.17	4.79	1.62
Cluster 2	95	5.54	3.97	467.12	416.71	452.54	142.89	642.14	426.46	5.30	1.27
Cluster 3	110	4.45	4.00	449.93	448.99	422.78	149.48	517.55	416.30	3.42	1.10

Table 5: Model summary for stepwise regression

	R			R²			Std. Error of the Estimate		
Cluster	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Model 1[a.]	.157	.283	.365	.025	.080	.133	1.5135	1.5147	1.7196
Model 2[b.]	.426	.395	.497	.181	.156	.247	1.3920	1.4588	1.6107
Model 3[c.]	.672	.482	.710	.452	.232	.504	1.1438	1.3991	1.3139
Model 4[d.]	.677	.490	.710	.458	.240	.504	1.1417	1.3996	1.3201
Model 5[e.]	.680	.495	.721	.463	.245	.520	1.1412	1.4026	1.3039

a. Predictors: (Constant), total face-to-face lectures
b. Predictors: (Constant), total face-to-face lectures, minutes recorded lectures
c. Predictors: (Constant), total face-to-face lectures, minutes recorded lectures, average score formative assessment
d. Predictors: (Constant), total face-to-face lectures, minutes recorded lectures, average score formative assessment, activity BB in minutes
e. Predictors: (Constant), total face-to-face lectures, minutes recorded lectures, average score formative assessment, activity BB in minutes, hits in BB

Although the cluster analysis revealed that there are distinct differences in the students their regulation strategies, these differences in regulation have no significant impact on the use of the different learning resources throughout the course.

4.1.2 Stepwise multi regression analysis

The next step in the analysis was to determine the relative contribution of each of the different learning resources on course performance. Since there were no clear indicators that one learning resource was likely to be of more value than another learning resource, a stepwise regression was used to find the set of different learning resources for each cluster. The summary of the stepwise regression with all the different learning resources for each cluster are shown in Table 5. SPSS output in Table 5 shows that lecture attendance was entered first in the regression analysis, explaining only 2,5% of the variance for cluster 1 students, 8% for cluster 2 students and up to 13,3% for cluster 3 students. For all clusters the activity in Blackboard in minutes and hits in Blackboard were not significant. All other variables were significant at the 0.05 level. Hence we use model 3 in our discussion of the results.

The overall models differ in their explained variance: 45,2% for cluster 1, 23,2% for cluster 2 and 50,4% for students in cluster 3 students. Remember that cluster 2 students mainly used an external regulation strategy and shows the lowest amount of variance explained caused by the use of the different learning resources. Cluster 3 students are students who try to self-regulate their learning process and this model explains the most of the variance in the use of the different learning resources. The difference in explained variance between the two groups of students is around 27%.

Students in cluster 1 and cluster 3 benefit the most from formative assessments. Students in cluster 2 benefits mostly from face-to-face lectures. Cluster 1 students also benefit from recorded lectures, while cluster 3 students benefit more from attending lectures. So besides differences in the explained variance for each subgroup, there are also differences in the types of learning recourses that has an added value for each cluster.

4.1.3 Course performance

The last step in the data analysis was to perform an ANOVA with cluster membership as a factor and with the final assessment as the dependent variable, to determine if differences in regulation strategies reflect in significant differences in course performance. A GT2 Hochberg performed the post-hoc analysis since the clusters differ in size.

There were no significant differences between groups for course performance as determined by the ANOVA ($F(2,330) = 1.018$, $p = .363$).

5. DISCUSSION

The current study aims to provide insight into differences in how students regulate their learning process and how differences in regulation of the learning process have an impact on the use of (digital) learning resources and subsequently contribute to course performance.

A cluster analysis showed three distinct patterns in the way students regulate their learning. One third of the students are mainly able to self-regulate their own learning process; one third of the students use an external source to regulate learning; one third of the students have no clear pattern when regulating their learning process; they switch between self-regulation, external regulation and lack of regulation during the learning process.

However, these differences in regulation strategies are not reflected in differences in the use of (digital) learning resources. Cluster 2 has a greater use of the different learning resources, but these differences are not significant. That students with different regulation strategies do not use the learning resources differently is confirmed by [36] where they found that both groups of students used the online learning resources to the same extent. Nonetheless, they found that differences in regulation strategies are reflected in course performance, when a high score on self-regulation correlates negatively with course

performance. This finding indicates that the structure of the course is beneficiary for students who report low self-regulated learning, but is a disadvantage for students who report high self-regulated learning. Although expected that students with better self-regulation strategies, would perform better in the current course, literature shows that student often ineffective self regulation to do so [3]. Moreover, students believe that an ineffective strategy is a good strategy, which itself may lead to poor self-regulation although reported otherwise [7].

Current results are confirmed by [27] who found two usage patterns in the use of learning resources: incoherent and intensive users. These two groups of users, however, did not show significant differences in their use of the learning resources. The current research finds the same pattern; although frequency and duration of use does not differ between self-regulated and externally regulated students, there are differences in how this use impacts course performance. For students with an external regulation strategy, 23% of the variability in course performance is due to the use of the different learning resources, while for self-regulated students this variability is 50%. These differences in explained variance could be caused by the *expertise reversal effect* [21]. The expertise reversal effect is a cognitive load framework that states that instructional techniques that are effective with inexperienced learners can lose their effectiveness when used by more experienced learners. A similar effect also will be obtained if novices must attempt to process very complex material, which will benefit the experienced learners. The students who report high self-regulated learning benefit more from the offered learning resources. This finding implies that not only duration of use has an impact on course performance but also the reported regulation strategy has an impact on the effectiveness of the learning resources.

This finding has two implications for learning analytics. First the contextualization of learning data with a broader set of indicators [9, 16, 26, 44] is crucial in establishing the impact of the learning data analysis since these conditions affect the learning process. The effect of internal conditions have also been stressed by [16] and current research shows that the use of the same learning resources to the same extent have different impacts on different groups of students. Second, although all clicks are equal, some clicks are more equal than others. Current learning analytics visualization trends use dashboards to mirror a student their activity with the class average. However, this class average is not as straightforward as previously assumed.

Besides differences in variability in course performance between clusters, we also established differences in the use of learning resources within each cluster. Cluster 1 and 3 show the most explained variance of the use of the different learning resources, however their composition of the variance is different. First, the similarity lies in the explained variance of the formative assessments, which is the highest for both clusters: 27% for cluster 1 students and 25,7% for cluster 3 students. These results are in line with [30] who found three predictive variables for course performance: number of forum postings, mail messages sent and assessments completed. This relative high variance is in line with the constructive alignment [6] of the formative assessments since they directly address the learning outcomes of the course and thereby reflect the level of the summative assessment. Next, the difference in the composition of the explained variance becomes clear when cluster 3 students benefit more from attending face-to-face lectures, while cluster 1 students benefit more from watching recordings of these lectures

online (15,5%) although students from both clusters use the learning resources to the same extent.

Regulation strategies thereby do not account for the previously reported differences in the use of digital learning resources by students [8, 20, 27, 28, 29] but does account for differences in effect of that use.

This research confirms, once again, the low predictive value LMS use has on course performance [17, 30, 35, 47]. For the three clusters, frequency and duration of LMS use were not significant in contributing to course performance. Since most mirroring techniques often use duration of LMS or frequency of logons to mirror student behavior, this is one more argument that this choice of mirroring should be examined critically. We found no significant relation between the amount of logons to the LMS and external regulated students. This finding is in contrast with previous research [11] wherein higher logons to the LMS were associated with external regulation of the learning process. The type of content the LMS offers during a course could explain this contradictory finding. In the current research, LMS content mainly consists of learning resources associated with learning activities and did not contain any resource that could be beneficiary for externally regulated learners, such as teacher-student interactions or a course catalog. When the LMS provides more information that supports external regulated students, like course content, it would elicit students to log to the LMS more, which could account for these differences. Once again, this shows that course design has an impact on predictive modeling techniques [17].

Surprisingly no significant differences were found between the three different clusters and course performance. This result is similar to [36], who found that self-regulation score correlates negatively with course performance, indicating that students who are able to self regulate their learning in general perform less well than students with less ability to self regulate their learning.

5.1 Implications for educational design

With modeling student behavior or predicting course performance to identify students who are at risk of failing the course, the focus is often on the choices of the instructor for course design and the choice for certain learning resources [16]. Current research shows that these choices of the instructor have a different impact on course performance for different groups of students. If instructors become more aware of these differences in learning approaches, they can effectuate a shift from blended teaching towards blended learning that is student orientated [24]. The current notion for blended learning is mostly aimed at putting technology into the learning environment without taking into account how that technology contributes to the learning outcomes [36] and supports individual differences [21].

The current research shows that not all students are able to self regulate their learning process. This lack of ability reflects in their ineffective use of the different learning resources, which eventually does not lead to a high quality of learning. As this and previous research shows [7, 8, 20, 27] students are not capable of making suitable choices with regard to their learning process and understand the value that certain learning resources have with respect to certain learning outcomes of the course. The use of pedagogical learning analytics interventions for students [45] in which learning analytics elicits students to become self-regulated learners and are made aware of the pedagogical intentions of the learning resources would benefit the quality of the learning process. Accurate monitoring of

learning is a crucial component of effective self-regulation of learning. In this regard teachers need to pay particular attention to the moments when students are analyzing the task and designing plans before engaging in learning. Learning analytics could foster these moments and enable self-regulated learning.

5.2 Limitations of the current research

The current course has a straightforward structure, in which students are offered guidance by means of formative assessments, the digital exercise book and online recordings of face-to-face lectures. The current course design is supporting students with an external regulation strategy. Previous research reports that students with a self-regulation strategy will not benefit from such a course design. Moreover the current course has a short duration, the time pressure is significant and students often find the subject matter hard which seldom gives students the opportunity to read additional literature or to do more that they are expected to do in a course. These attributes are characteristic for students who are able to self regulate their learning. The current course design forces students to use an external regulation strategy although some students are able to self regulate their learning. This causes a relative homogeneous set of learning strategies. Nevertheless, even with this caveat, the impact of the use of various learning resources clearly depends on how students regulate their learning.

The current research uses clicks on links and duration of use as a reflection of student effort, student engagement and participation [47]. With research into online learning it is always debatable whether these clicks and hits actually reflect use of digital resources or that a students clicks on a link and walks away and hence influencing the time on task measure. However, the time on task measure is debatable for all study activities, not only for the use of digital learning resources, since study sessions, and even attending class, can include e-mail, online shopping and so on [7].

Another limitation of the current research is the use of the ILS, since better instruments are available to measure regulation strategies. For example, in a follow up study the authors used the MSLQ, which is more in line with current approaches of assessing self-regulated learning [4, 5, 7, 40]. The reason for the current use of the ILS was rather straightforward; there is a long tradition within the faculty of psychology to administer the ILS to all freshmen (see for example (Busato, Prins, Elshout, & Hamaker, 1998; 2000) and is even embedded within the curriculum resulting in an 100% response rate.

The last limitation of the current study is the known calibration and inaccuracy problems with self-reports about study tactics [43]. As previously mentioned, students often consider themselves as self-regulated learners while the tactics they use to regulate their learning are ineffective. Moreover, even within a single course these self-reports about regulation of learning differ as a function of the task before them (multiple choice exam versus writing a paper) [44].

5.3 Recommendations for future research

This research showed that regulation strategies do not a have a direct impact on the use of (digital) learning resources. Nevertheless, it showed that differences in regulation strategies do have an effect on the explained variance for the learning resources in relation to course performance. The differences in explained variance could be caused by the expertise reversal effect. However, differences in explained variance could also be caused by the sequence in which students use the different learning resources. Current educational research, and especially learning analytics research, hardly ever considers sequences of the used learning resources as an important factor for course performance. Temporal analysis, in which methods as sequential pattern mining are being used, could establish if sequence of the used learning resources does account for differences between the actual use of the learning resources and differences in explained variance currently reported.

6. CONCLUSION

In this research we examined if regulation strategies could account for previously reported differences in the use of learning resources. We examined how 333 psychology students regulated their learning process and subsequently used the different learning resources throughout the course and established how this use of the learning resources contributed to course performance.

The results indicate that differences in regulation strategies do not account for differences in the use of (digital) learning resources. However, different regulation strategies do have an impact on the explained variance the different learning resources have on course performance meaning that some learning resources are more effective for some groups of students than others. Students with an external regulation strategy have the lowest explained variance on the use of learning resources in relation to course performance.

This study has several consequences for future practices of learning analytics and especially mirroring techniques. First, it gives recognition to the importance of contextualization of the learning data resources with a broader set of indicators to understand the learning process. Moreover, this research shows that mirroring students learning progress based on class average does not account for differences in impact that use has on course performance. Lastly, with our focus on differences between students, we strive for a shift of identifying at risk students towards a contribution of learning analytics in the educational design process and enhance the quality of learning; for all students.

7. ACKNOWLEDGMENTS

REFERENCES

[1] Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (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*, 542-550. doi:10.1016/j.chb.2013.05.031

[2] Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist, 40*(4), 199-209. doi:10.1207/s15326985ep4004_2

[3] Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology*

Research and Development, 56(1), 45-72. doi:10.1007/s11423-007-9067-0

[4] Beheshitha, S. S., Gašević, D., & Hatala, M. (2015). A process mining approach to linking the study of aptitude and event facets of self-regulated learning. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 265-269). ACM. doi:10.1145/2723576.2723628

[5] Biggs, J., Kember, D., & Leung, D. Y. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133-149. doi:10.1348/000709901158433

[6] Biggs, J. & Tang, C. (2011). *Teaching for quality learning at university.* Maidenhead, UK: McGraw-Hill Education.

[7] Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444. doi:10.1146/annurev-psych-113011-143823

[8] Bos, N., Groeneveld, C., Van Bruggen, J., & Brand-Gruwel, S. (2015). The use of recorded lectures in education and the impact on lecture attendance and exam performance. *British Journal of Educational Technology*. doi:10.1111/bjet.12300

[9] Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 92-101). ACM. doi:10.1145/2330601.2330629

[10] Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 263-268). ACM. doi:10.1145/502512.502549

[11] Cho, M. H., & Shen, D. (2013). Self-regulation in online learning. *Distance Education*, *34*(3), 290-301. doi:10.1080/01587919.2013.835770

[12] Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current state and future trends: A citation network analysis of the learning analytics field. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 231-240). ACM. doi:10.1145/2567574.2567585

[13] Drachsler, H., Stoyanov, S., & Specht, M. (2014). The impact of learning analytics on the dutch education system. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 158-162). ACM. doi:10.1145/2567574.2567617

[14] Ellis, R., Goodyear, P., Calvo, R., Prosser, M. (2008). Engineering students' conceptions of and approaches to learning through discussions in face-to-face and online contexts. *Learning and Instruction, 18*(3), 267-282. doi:10.1016/j.learninstruc.2007.06.001

[15] Entwistle, N., & McCune, V. (2013). The disposition to understand for oneself at university: Integrating learning processes with motivation and metacognition. *British Journal of Educational Psychology, 83*(2), 267-279. doi:10.1111/bjep.12010

[16] Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64-71. doi:10.1007/s11528-014-0822-x

[17] Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2015). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicating academic success. *The Internet and Higher Education*. doi:10.1016/j.iheduc.2015.10.002

[18] George-Walker, L.D., & Keeffe, M. (2010). Self-determined blended learning: a case study of blended learning design. *Higher Education Research & Development, 29*(1), 1-13. doi:10.1080/07294360903277380

[19] Gorissen, P., Van Bruggen, J., & Jochems, W. (2012). Usage reporting on recorded lectures using educational data mining. *International Journal of Learning Technology*, *7*(1), 23-40. doi:10.1504/ijlt.2012.046864

[20] Inglis, M., Palipana, A., Trenholm, S., & Ward, J. (2011). Individual differences in students' use of optional learning resources. *Journal of Computer Assisted Learning*, *27*(6), 490-502. doi: 10.1111/j.1365-2729.2011.00417.x

[21] Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational psychologist, 38*(1), 23-31. doi:10.1207/s15326985ep3801_4

[22] Kovanović, V., Gašević, D., Joksimović, S., Hatala, M., & Adesope, O. (2015). Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education, 27*, 74-89. doi:10.1016/j.iheduc.2015.06.002

[23] Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Does Time-on-task Estimation Matter? Implications on Validity of Learning Analytics Findings. *Journal of Learning Analytics, 2*(3)(in press), http://vitomir.kovanovic.info/?download=474

[24] Kupetz, R., & Ziegenmeyer, B. (2006). Flexible learning activities fostering autonomy in teaching training. *ReCALL*, *18*(01), 63-82. doi:10.1017/s0958344006000516

[25] Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, *57*(10), 1439-1459. doi:10.1177/0002764213479367

[26] Lodge, J., & Lewis, M. (2012). Pigeon pecks and mouse clicks: Putting the learning back into learning analytics. *Future challenges, sustainable futures. Proceedings ascilite Wellington*, 560-564.

[27] Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In Search of user profiles. *Computers & Education*, *57*(3), 2135-2144. doi:10.1016/j.compedu.2011.05.010

[28] Lust, G., Elen, J., & Clarebout, G. (2013). Students' tool-use within a web enhanced course: Explanatory mechanisms of students' tool-use pattern. *Computers in Human Behavior, 29*(5). doi:10.1016/j.chb.2013.03.014

[29] Lust, G., Elen, J., & Clarebout, G. (2013). Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education, 60*(1), 385-395. doi:10.1016/j.compedu.2012.09.001

[30] Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an 'early warning system' for educators: A proof of concept. *Computers & Education*, *54*(2), 588-599. doi:10.1016/j.compedu.2009.09.008

[31] Orton-Johnson, K (2009). I've stuck to the path I'm afraid: exploring student non-use of blended Learning. *British Journal of Educational Technology, 40*(5), 837-847. doi:10.1111/j.1467-8535.2008.00860.x

[32] Orwell, G. (1954). *Animal Farm: A Fairy Story.* London: Secker and Warburg.

[33] Perry, N. E., Phillips, L., & Hutchinson, L. (2006). Mentoring Student Teachers to Support Self-Regulated Learning. *The elementary school journal*, *106*(3), 237-254. doi:10.1086/501485

[34] Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of educational psychology*,*92*(3), 544. doi:10.1037/0022-0663.92.3.544

[35] Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. *Computers in Human Behavior*, *47*, 157-167. doi:10.1016/j.chb.2014.05.038

[36] Tempelaar, D., & Rienties, B. (2008). Explaining student learning preferences in a blended learning environment for learning statistics on the basis of student characteristics. *Student Mobility and ICT: Can ICT overcome the barriers to Life-long learning*, 51-60.

[37] Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011). Dataset-driven research for improving recommender systems for learning. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 44-53). ACM. doi:10.1145/2090116.2090122

[38] Verkroost, M. J., Meijerink, L., Lintsen, H., & Veen, W. (2008). Finding a balance in dimensions of blended learning. *International Journal on E-learning*,*7*(3), 499-522.

[39] Vermunt, J. D. H. M. (1992) *Leerstijlen en sturen van leerprocessen in het hoger onderwijs: naar procesgerichte instructie in zelfstanding denken [Learning styles and regulation of learning in higher education: towards process-oriented instruction in autonomous thinking]* (Amsterdam, Lisse: Swets & Zeitlinger).

[40] Watkins, D. (2001). Correlates of approaches to learning: a cross-cultural meta-analysis. In: R. J. Sternberg & L. F. Zhang (Eds.), *Perspective on thinking, learning, and cognitive styles* (p. 165-195). Mahwah, NJ, Lawrence Erlbaum Associates.

[41] Wiese, C., & Newton, G. (2013). Use of Lecture Capture in Undergraduate Biological Science Education. *Canadian Journal for the Scholarship of Teaching and Learning*, *4*(2), 4. doi:10.5206/cjsotl-rcacea.2013.2.4

[42] Winne, P.H. & Perry, N.E. (2000). Measuring self-regulated learning. In P. Pintrich, M. Boekaerts, & M. Seidner (Eds.), *Handbook of self-regulation* (p. 531-566). Orlando, FL: Academic Press.

[43] Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551-572. doi:10.1016/s0361-476x(02)00006-1

[44] Winne, P. H. (2006). How software technologies can improve research on learning and bolster school reform. *Educational Psychologist, 41*(1), 5-17. doi:10.1207/s15326985ep4101_3

[45] Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 203-211). ACM. doi:10.1145/2567574.2567588

[46] Woolfolk, A., Walkup, V., & Hughes, M. (2008). *Psychology in education.* Harlow, UK: Pearson-Longman.

[47] Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education,27,* 44-53. doi:10.1016/j.iheduc.2015.05.002