JOURNAL OF LEARNING ANALYTICS

Volume 1 - Issue 1

SOLAR
SOCIETY for LEARNING ANALYTICS RESEARCH

(2014). Editorial – Inaugural Issue of the Journal of Learning Analytics. *Journal of Learning Analytics 1*(1), 1–2.

Editorial: Inaugural Issue of the *Journal of Learning Analytics*

Dragan Gasevic, Negin Mirriahi,
Phil Long, and Shane Dawson
Editors, *Journal of Learning Analytics*

ABSTRACT: This article introduces the inaugural issue for the *Journal of Learning Analytics*. The article outlines the journal's aims and scope and summarizes the research and Hot Spot papers for the issue.

KEYWORDS: Inaugural issue, learning analytics, research, practice, Society for Learning Analytics Research, SoLAR

Editorial

Since the establishment of the Society for Learning Analytics Research (SoLAR) in 2011, the rapidly emerging learning analytics field warranted greater focus and opportunity to showcase the quality research and practice underway. As an initial step, SoLAR established the international conference on Learning Analytics and Knowledge. The *Journal of Learning Analytics* represents another important step supporting learning analytics researchers and practitioners.

On behalf of the Society for Learning Analytics Research, welcome to Volume 1, Issue 1 of the *Journal of Learning Analytics* (JLA). The JLA is a peer-reviewed, open-access journal published by SoLAR. This field-defining publication is the first journal dedicated to research investigating the challenges of collecting, analyzing, and reporting data with the specific intent to understand and improve learning. A core goal for the journal's development is to provide a space for promoting both research and practice. The first issue presents research papers and practitioner "hot spots." The publication represents and reflects the diversity of learning analytics work in formal and informal education contexts: from K–12, vocational and higher education, and workplace settings. The journal aims to connect researchers and developers with practitioners, to inspire, motivate, and disseminate the publication of new tools and techniques, to study learning transformations, and to provide evaluation and critiques of the conceptual, technical, and practice-based outcomes. The interdisciplinary focus of the journal recognizes that computational, pedagogical, institutional, policy, and social domains must be brought into dialogue with each other to ensure that interventions and organizational systems serve the needs of all stakeholders.

This first issue illustrates the diversity of learning analytics research and practitioner outcomes. The research papers address important challenges such as scaling-up learning analytics initiatives: the relationship between LMS/VLE usage and learning performance, the role of psychometric data to predict academic achievement, and the capacity to detect boredom through user log-data.

(2014). Editorial – Inaugural Issue of the Journal of Learning Analytics, *1*(1), 1–2.

In arguing for open learning analytics Jayaprakash et al., outline how models to predict students at-risk of academic failure can be applied across alternate educational contexts. This is an important point for the transferability of such models into the myriad of teaching and education settings that could benefit from such assessments. Interestingly, the authors found no significant difference in improved academic performance for students supported through a learning intervention versus students simply made aware of their potential risk.

The following research paper by Andergassen et al., describes a process for dealing with large LMS data and provides a novel insight into the analysis of such log-data. The authors examined some 250 million log-file entries to investigate patterns related to academic performance, finding that quantity and frequency of access along with completion of assessment exercises were predictive of overall academic performance.

In taking an alternate approach to the previous papers, Gray et al. examine the role of psychometric data for predicting academic performance. Psychometric factors of ability, personality, motivation, and learning strategies predicted academic performance. While the authors note that prior academic ability is a sound predictor of future academic performance, this is not the case for mature learners or groups with ethnic diversity.

In their investigation of the ASSISTments math tutoring system, Pardos et al. analyse the longitudinal and fine-grain data related to student affect and behavioural engagement and the relationship with exam performance. Through these analyses the authors were able to calculate the probability a learner is in a state of boredom, concentration, confusion, or frustration, or that the student is exhibiting off-task or gaming behaviours.

The practitioner *Hot Spots* section features two distinct yet complimentary reports of learning analytics in higher education institutions. Buerck and Mudigonda discuss the challenges they faced with implementing a top-down approach to academic analytics at their institutions to shifting to a more successful bottom-up learning analytics strategy. The second paper by Heath advocates for the consideration of contemporary privacy theories and student feedback to help inform institutional governance policies to address the privacy aspects and concerns of learning analytics initiatives.

In addition to the research and practitioner Hot Spots papers, this issue features an article by George Siemens, President of SoLAR. In his article, Siemens gives a brief overview of the main activities of SoLAR and describes the importance of the journal for the development of the field of learning analytics. The insight and creative analysis of learning data represented in this first issue of the *Journal of Learning Analytics* sets a tone and direction of critical insight coupled with practical advice. We hope you will share your work through JLA as we build the field together.

(2014). The Journal of Learning Analytics: Supporting and Promoting Learning Analytics Research. *Journal of Learning Analytics*, *1* (1), 3–4.

The *Journal of Learning Analytics*: Supporting and Promoting Learning Analytics Research

George Siemens
President, Society of Learning Analytics Research (SoLAR)
Learning Innovation and Networked Knowledge (LINK) Research Lab
University of Texas at Arlington
gsiemens@uta.edu

ABSTRACT: The paper gives a brief overview of the main activities for the development of the emerging field of learning analytics led by the Society for Learning Analytics Research (SoLAR). The place of the *Journal of Learning Analytics* is identified. Analytics is the most significant new initiative of SoLAR.

KEYWORDS: learning analytics, learning, education, field development

Welcome to the inaugural issue of the *Journal of Learning Analytics*. Articles that address big data and analytics have an obligatory introduction: data is everywhere, the amount of accessible data is growing rapidly, new sources of data are accelerating the already overwhelming quantity, and so on. The message is clear: we live in a world of data and our future promises even greater emphasis on analytics to understand data.

Analytics have arrived later to education than to government, healthcare, and business. While the education field has deep roots in data and analysis (research, after all, is primarily an exercise in making sense of data), the systemic use of analytics for improving teaching and learning is still emerging.

In 2010, a small group of us (Dragan Gasevic, Shane Dawson, Simon Buckingham Shum, Caroline Haythornthwaite, and I) initiated a conversation around the need for a conference on learning analytics. We approached other colleagues and eventually formed the steering committee[1] for the 1st International Conference in Learning Analytics and Knowledge (LAK). We were motivated by the growing influence of data in decision-making processes in teaching and learning settings. While the data focus was welcomed, it raised questions about the transparency of analytics methods, data access and ownership, as well as how analytics approaches themselves were being researched and validated.

A second challenge was on the nature of analytics practices. Often, computer scientists, machine learning experts, statisticians, and mathematicians had the technical capacity to make sense of large data sets, but lacked grounding in education and learning theory and literature. In contrast, learning scientists, psychologists, and sociologists had the theoretical lens to evaluate the social power structures

[1] https://tekri.athabascau.ca/analytics/node/5

(2014). The Journal of Learning Analytics: Supporting and Promoting Learning Analytics Research. *Journal of Learning Analytics*, *1* (1), 3–4.

and "soft domains" of learning, but lacked grounding in emerging data and analytics methods. We lamented this gap between groups:

> Advances in knowledge modeling and representation, the semantic web, data mining, analytics, and open data form a foundation for new models of knowledge development and analysis. The technical complexity of this nascent field is paralleled by a transition within the full spectrum of learning (education, work place learning, and informal learning) to social, networked learning. These technical, pedagogical, and social domains must be brought into dialogue with each other to ensure that interventions and organizational systems serve the needs of all stakeholders.[2]

From this perspective, the first LAK conference steering committee began to shape a research space that included an eclectic, at times challenging, mix of researchers. The committee emphasized the need to play at the margins of knowledge domains. Learning analytics is a bricolage field, incorporating methods and techniques from a broad range of feeder fields: social network analysis, machine learning, statistics, intelligent tutors, learning sciences, and others.

Since the first LAK conference, it has become increasingly clear that learning analytics is a research and practitioner domain. Across the spectrum of learning — from primary school through to corporate learning — data is playing a growing role in how learning occurs and how educators and administrators make decisions. At state, provincial, and national levels, interest in data and analytics in education has resulted in numerous government-sponsored reports in order to make sense of what analytics contributes to the education process.

The growing focus on data and analytics in education, the involvement of funding agencies (both research boards and private foundations), and the growing interest from researchers, have confirmed the importance of big data and analytics in education. The last four years have been significant for the field of learning analytics. In 2011, the Society for Learning Analytics Research (SoLAR)[3] was formed and took ownership of the annual LAK conference. We engaged in a series of projects to advance the field:

- A distributed doctoral lab to allow students to connect with other students and receive feedback and guidance from researchers.
- Learning Analytics Summer Institute (LASI) to serve as a forum for developing the field and to introduce doctoral students and academics to learning analytics. The first event was held at Stanford University in July 2013.
- LASI-Locals, a series of global analytics workshops connected to the Stanford LASI event. Nearly 1,000 participants attended LASI-Local or online events.

[2] https://tekri.athabascau.ca/analytics/about
[3] http://www.solaresearch.org/

(2014). The Journal of Learning Analytics: Supporting and Promoting Learning Analytics Research. *Journal of Learning Analytics*, *1* (1), 3–4.

- SoLAR founding universities — critical partners that provided support to develop SoLAR and the LAK conference. Founding members include: Stanford University, University of Michigan, Athabasca University, Open University UK, University of British Columbia, University of Hawaii, University of Texas at Arlington, University of South Australia, Marist College, University of Wisconsin-Madison, University of New England, University of Queensland, American Institutes for Research, and the University of Saskatchewan. These universities provided financial support to organize conferences, doctoral seminars, summer institutes, and numerous additional SoLAR initiatives.
- Open Learning Analytics (OLA). This project is based on a whitepaper[4] written by SoLAR members calling for an open architecture for learning analytics (an "Apache of analytics"). The project is ongoing and recent collaborations with the Apereo Foundation will move the concept to product in the near future.
- Learning Analytics Masters Program (LAMP). Many universities already offer a program in "big data and analytics." Few currently offer a learning analytics masters program or certificate. To facilitate the development of masters programs in learning analytics, SoLAR has initiated LAMP. LAMP will result in the creation of an open masters program that will be licensed for use and reuse by academic institutions.
- Affiliation and collaboration with our sister organization — International Educational Data Mining Society (IEDMS).[5] To improve research quality and collaboration with other organizations focused on data and analytics in education, SoLAR has engaged in several strategic partnerships with IEDMS, including LASI, OLA, and LAMP.

Finally, SoLAR's launch of the *Journal of Learning Analytics* is its most significant new initiative. Our vision for this journal is that it will serve as a critical node in the discourse around data and analytics in the learning process. As a scientific journal, we would like to reflect the messiness of science — a space where ideas and evidence are presented, challenged, verified, and refuted — a space where concepts of significance can grow and be extended by new researchers and researchers in related fields. Most importantly, the *Journal of Learning Analytics* is a space where the field can grow, where doctoral students can find inspiration, and where researchers can connect with peripheral domains.

The drivers of success of any academic journal lie behind the scenes: the editors, reviewers, and copy editors. Thank you for your stellar efforts.

[4] http://solaresearch.org/OpenLearningAnalytics.pdf
[5] http://www.educationaldatamining.org/

JOURNAL OF LEARNING ANALYTICS

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative

Sandeep M. Jayaprakash, Erik W. Moody, Eitel J.M. Lauría, James R. Regan, and Joshua D. Baron

Marist College, USA

Eitel.Lauria@marist.edu

ABSTRACT: The Open Academic Analytics Initiative (OAAI) is a collaborative, multi-year grant program aimed at researching issues related to the scaling up of learning analytics technologies and solutions across all of higher education. The paper describes the goals and objectives of the OAAI, depicts the process and challenges of collecting, organizing and mining student data to predict academic risk, and report results on the predictive performance of those models, their portability across pilot programs at partner institutions, and the results of interventions on at-risk students.

KEYWORDS: Learning analytics, open source, data mining, learning management systems, portability, retention, course completion

1 INTRODUCTION

Higher education, particularly in the United States, is facing major strategic challenges regarding course and degree completion rates as well as overall college retention. Across all types of four-year institutions, of those students starting bachelor degree programs in 2001, only 36% completed them within four years (U.S. Dept. of Education, 2009). As a result, the United States now ranks 12th in the world in the percentage of 25- to 34-year-olds with an associate's degree or higher (College Board Advocacy & Policy Center, 2010). Although not a panacea solution, the emergence of "big data" and analytics technologies within higher education has begun to provide new tools for addressing this developing national challenge (Long & Siemens, 2011).

At the forefront of these big data and analytics solutions is learning analytics, which has recently emerged in the education domain in the aftermath of the successful application of data mining techniques in business organizations. The goal of learning analytics[1] is to uncover hidden patterns in

[1] A distinction should be made between the terms *academic analytics* and *learning analytics*. There is more on this topic in Section 2. The name given to this research initiative when it was launched two years ago (Open Academic Analytics Initiative) is tied to an early definition of the term *academic analytics.* As the use of analytics in education is relatively new, there has been a natural evolution in the terminology used to describe it. The authors posit that the current definition of *learning analytics* better describes the kind of work carried out by this initiative. The main goal of this project is to improve the chances of student success in a specific course, which is also a central concern of *learning analytics*. We believe that this

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

educational data and use those patterns to attain a better understanding of the educational process, assess student learning, and make predictions on performance. The widespread deployment of learning management system platforms that log student interactions with educational software has made large data sets available, which enrich traditional student academic records and demographic data, facilitating new research in this domain.

The Open Academic Analytics Initiative (OAAI), supported by a grant from EDUCAUSE's Next Generation Learning Challenges program, which was funded in part by the Bill and Melinda Gates Foundation, has aimed to advance the field of Learning Analytics by exploring issues related to scaling this technology across all of higher education. In particular, the project has worked to address three core research questions:

1. What are the potential challenges, solutions, and benefits associated with developing a completely open-source early alert solution for higher education?
2. To what degree can predictive models be imported from the academic context (e.g., a four-year private liberal arts college) in which they were developed to new and potentially very different academic contexts (e.g., two-year community colleges)?
3. What intervention strategies are most effective in helping academically at-risk students succeed?

To examine these questions, the OAAI has been engaged in the development of a prototype open-source academic early alert system that feeds from the *Sakai CLE* (Collaboration and Learning Environment),[2] and includes open predictive models based on the *Pentaho Business Analytics* suite,[3] and intervention strategies that leverage Open Educational Resources (OER). In addition, the OAAI has conducted research on the portability of predictive models between institutions as well as the effectiveness of engaging students in online academic support communities as means to improve academic success. This paper will focus primarily on our research into predictive analysis, portability of the models across institutions and intervention effectiveness, but details associated with the technical aspects of OAAI's open learning analytics ecosystem are available on the Sakai Project Wiki.[4]

To investigate the scaling issues related to portability and intervention effectiveness, the OAAI began by creating a framework for the development of predictive models based on student data from Marist College, a mid-size comprehensive liberal arts institution located in New York State. These predictive models were created using student demographic data (e.g., gender, age), aptitude data (e.g., standardized high school test cores), and learning management system data (e.g., site visits, assignment submissions, partial contributions to the final course grade collected in the gradebook tool). This follows

project is only indirectly related to institutional goals such as college retention and cost savings, which are among the objects of study of *academic analytics*, in the current definition of this term.

[2] http://www.sakaiproject.org

[3] http://www.pentaho.com

[4] https://confluence.sakaiproject.org/x/8aWCB

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

the same initial approach that Campbell (2007) did at Purdue in his dissertation work, in which he developed models to detect students at risk of underperforming in a course. Comparing Marist's data to Purdue's has provided insights into the degree to which the predictive models can be imported from one type of institution to another.

The predictive models trained and tested with Marist College data were subsequently deployed at four partner institutions (two community colleges[5] and two HBCUs[6]), to further research issues of portability and intervention effectiveness. With the conclusion of the project, these predictive models have been released[7] under an open-source license in the standards-based Predictive Model Markup Language (PMML) as a means to facilitate use of and further enhancement of the models by others.

Predictive models do not influence course completion and retention rates without being combined with effective intervention strategies aimed at helping at-risk students succeed. To address this, the OAAI developed a concept called an Online Academic Support Environment (OASE) that leverages Sakai Project Sites to provide students with an online support community and resources aimed at aiding academic success. Resources include OER (open educational resources) content for remediation and study skill development, facilitation by a professional academic support specialist, and a student mentor who acts as a peer coach.

In the next section, the paper reviews related research showing how others have applied data mining techniques to assess student performance. Next, the paper describes the predictive modelling framework developed by the OAAI, and the results of model development and testing on Marist College data. Then the paper describes the deployment of an experimental academic early alert system, testing data from partner institutions, and the deployment and effectiveness of different intervention strategies at the aforementioned pilots. Finally, the paper provides a summary and conclusions, which include future research opportunities.

2 RELATED WORK

In the last decade, the discipline of analytics has permeated most layers of society and organizations. Defined succinctly as the discovery and communication of meaningful patterns of data, analytics has provided a data-driven approach to the way in which individuals and organizations conduct business and make decisions. Education, and in particular higher education, has not remained impervious to the lure of analytics. Many colleges and universities around the world have started to apply analytics to gain new insights on a variety of business and educational issues. The spectrum of possibilities is ample: enhancing decision making in the admissions process, increasing financial and operational efficiency,

[5] Community colleges in the USA are two-year public institutions providing higher education and lower-level tertiary education.
[6] Historical Black Colleges and Universities (HBCUs) are institutions of higher education in the USA established before 1964 with the intention of serving the black community.
[7] https://confluence.sakaiproject.org/pages/viewpage.action?pageId=75671025

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

improving fundraising programs, helping educators achieve a better understanding of their students learning process and abilities, and enhancing student performance across courses and disciplines.

As in the case of any new discipline, the use of analytics in education has adopted a variety of terms to describe aspects of research and practice. The extant literature on the use of analytics in education includes references to *academic analytics*, *learning analytics*, *predictive analytics*, *social learning analytics*, and *educational data mining*, to mention some of the most prominent terms used by researchers and practitioners. Two distinct research communities, Educational Data Mining and Learning Analytics and Knowledge, have developed because of the increasing interest in the application of analytics in education, and the booming amount of available data and software platforms and tools. Siemens and Baker (2012) chronicle the evolution of these research communities, their similarities and differences, and call for more collaboration and integration, given the overlaps in research interests, methodological approaches, and technologies between both communities.

The title given to the research initiative described in this paper (Open Academic Analytics Initiative) when it was proposed in early 2011 is a good example of the way in which the terms adopted to describe concepts and processes tied to analytics in education have evolved over time. If this project were to be launched today, it would almost certainly be called Open Learning Analytics Initiative. The distinction is subtle but substantive. In its inception, academic analytics was an overarching term, focusing both on institutional issues (e.g., enrollment management) and instructional issues. Over time, academic analytics has shifted its focus: most authors place its emphasis at an institutional level. Learning analytics has spun off as a more specific term, focused on instructional issues. A recent EDUCAUSE research report (van Barneveld, Arnold, & Campbell, 2012) tackles the issue of variability in terminology, providing a definition of learning analytics as "the use of analytic techniques to help target instructional, curricular, and support resources to support the achievement of specific learning goals" (p. 8), and uses the Course Signals project developed at Purdue (Arnold, 2010) as a typical example of a learning analytics project. We believe that this definition of learning analytics, which also reflects its commonly accepted use in the USA, encompasses the work of the OAAI, which focuses on early detection of at-risk students and subsequent intervention. In that spirit, the following paragraphs in section 2.1 provide a condensed review of the literature in learning analytics, in particular the work related to the use of data mining to predict academic performance. Section 2.2 follows with a short literature review on intervention theory.

2.1 Early Alert and Prediction of Academic Performance

Initial attempts to use machine learning and data mining techniques to predict academic performance and act upon it can probably be traced back to the early 2000s. Ma, Liu, Wong, Yu, and Lee (2000) used a scoring function based on association rules to identify potential weak students, and subsequently select the courses that each weak student is recommended to take. Chen, Liu, Ou, and Liu (2000) applied decision trees on web log data to profile student groups that exhibit similar behaviour to a

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

particular pedagogical strategy. A conceptual paper by Zaïane and Luo (2001) pointed to the enabling capabilities of these techniques on data generated by web-based course management platforms to help understand learners' behaviour through usage pattern recognition, supporting educators to evaluate the learning process better. Yu, Own, and Lin (2001) applied fuzzy association rules to capture relationships between web usage patterns of a learner, including the time spent online, amount of read and posted material, etc. Minaei-Bidgoli and Punch (2003) classified students using features extracted from logged data in a web-based system in order to predict their final grades. They combined multiple classifiers and weighted feature vectors using a genetic algorithm to optimize the prediction accuracy of the classifier. Laurie and Timothy (2005) used data mining as a strategy for assessing discussion forums in online courses, with the objective of enhancing the instructor's ability to evaluate the progress of a threaded discussion. Morris, Wu, and Finnegan (2005) used discriminant analysis on high school data (GPA and standardized test scores) to predict the successful completion of online courses. Romero, Ventura, & Garcia (2008) published a case study tutorial with the Moodle[8] learning management system to exemplify the application of data mining in learning management systems. The tutorial described how different data mining techniques and packages could be used in order to improve the course and the students' learning. Bravo, Sosnovsky, and Ortigosa (2009) profiled low performance in e-learning systems using a decision trees classifier trained with log data consisting of records of student actions within the system. The 2010 edition of the KDD Cup[9] challenged competitors to predict student performance on mathematical problems from logs of student interaction with Intelligent Tutoring Systems, another piece of evidence pointing at the growing interest in learning analytics as a rich applied research field.

There have been a number of academic initiatives at various colleges and universities preceding our work that sought to detect students in academic difficulty by using analytics. Campbell, deBlois, & Oblinger (2007) report on an experiment at the University of Alabama (UA) in 2002, where graduate students in a data-mining course who were given access to anonymized data of enrolled freshmen from 1999, 2000, and 2001 were able to develop predictive models of at-risk students using a variety of data mining techniques, including logistic regression, decision trees, and neural networks. The input data used to train the models included demographic and aptitude data (e.g., standardized high school scores and grades along with cumulative GPA in the freshman year). These models allowed the UA to identify 150–200 freshmen each year who were not likely to return for their sophomore year. In 2004, Northern Arizona University (NAU) launched an initiative that used multiple data sources to identify at-risk first-year students and to assess which proactive interventions have the best influence on their academic success and retention. The model measured utilization of services and resources (e.g., academic services recreational resources, social resources such as student organization membership, academic referrals, and advising sessions), levels of risk (e.g., standardized high school test scores, high school GPA), and

[8] http://www.moodle.org
[9] KDD Cup 2010. See http://www.sigkdd.org/kddcup

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

outcomes (measured by first-year student GPAs and enrollment retention status). The initiative has evolved into an early alert system called GPS.[10]

In his doctoral dissertation at Purdue University, Campbell (2007) used factor analysis and logistic regression on a set of student features derived from data extracted from Blackboard,[11] which included learning management system usage and student demographics. This research gave way to the development of Course Signals (Arnold, 2010) a prominent early intervention system originally developed at Purdue and currently owned and marketed by Ellucian.[12] Course Signals builds models from student data that predict which students may be struggling academically and subsequently provides proactive intervention. Reports on pilots between fall 2007 and fall 2009 showed significant improvement on course completion, and mastery of content learning outcomes, making Course Signals one of the most successful proofs of concept of the use of data mining and statistical techniques to develop early alert systems.

Barber and Sharkey (2012) report on the creation of a predictive model for the University of Phoenix to identify academically at-risk students. The model combines data from the learning management system, financial aid system, and student system to calculate a likelihood of any given student failing the current course. Other learning analytics projects include University of Maryland–Baltimore County's "Check My Activities" project (Fritz, 2011) and Grand Rapids Community College's Project ASTRO,[13] which altogether are indicative of the growing interest in the application of these technologies.

Nevertheless, the number of initiatives that have been able to transition from concept to implementation is still scarce, and the project described in this paper is one of the few to attain this status. In addition, only a small number of implementations have scaled up to more than a few institutions with most being implemented at just one (the one where it was developed).

This explains in part the amount of attention and recognition that the Open Academic Analytics Initiative (OAAI) has received, including two prestigious international awards.[14] As of its ending date of January 2013, the OAAI has successfully achieved all of its major project outcomes, including a) the development and deployment of an open-source academic early alert prototype system; b) the release of predictive models under an open-license; c) study of portability of predictive models from one academic context to another; d) research on the impact of different intervention strategies on student performance.

[10] Grade Performance Status, http://www4.nau.edu/ua/GPS/Faculty
[11] http://www.blackboard.com
[12] http://www.ellucian.com/signals
[13] http://projects.oscelot.org/gf/project/astro
[14] In March 2013, the OAAI was recognized by *Computerworld* as a 2013 Honors Laureate and Finalist in the Emerging Technology category, and in June 2013 by *Campus Technology Magazine* as one of only nine recipients for the Campus Technology Innovator Award (over 230 applied).

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

2.2 Intervention Theory

Our approach has been to build upon the success of the Course Signals system developed at Purdue University. We have used predictive analytical techniques to identify students at risk of course failure and subsequently researched the effectiveness of two different interventions designed to improve student outcomes**.** Course Signals addresses an issue that many instructors are aware of, but one that has largely gone unaddressed in the literature. Often students do not understand how well they are performing in a class until it is too late for those performing poorly to change their trajectory (Pistilli & Arnold, 2010). Our design was in no small part influenced by the EDUCAUSE/Gates Foundation grant that funded this project. The grant stipulated that effective techniques to improve student retention be investigated and demonstrated in socio-economically disadvantaged populations. We relied on Campbell's work at Purdue with Course Signals, which heavily referenced Tinto's and Astin's work, suggesting that positive interactions, good grades, and increased faculty–student interaction (email warnings) are among the most effective means of achieving better retention rates.

The work done at Purdue has shown that the use of relatively simple notification interventions can have a significant effect on student behaviour. In one course with 220 students, 55% of those who were initially identified as being at *high risk* for not completing the course moved into the *moderate risk* category because of receiving an intervention. More impressively, almost 25% moved from *high risk* to *no* or *low risk*, and of those who began at the "moderate risk" level, almost 70% rose to the no/low risk category. This seems to indicate that simply making students aware that they are at risk of not completing a course motivates them to seek help and change their academic behaviour. Interestingly, once the interventions stopped, they found that the students who had received the "notification interventions" continued to seek help and at a frequency of "30% more often than students in the control group" (Arnold, 2010).

As the field is so new, there is very little data available on the measure of retention — defined as continued enrollment or graduation — at a given institution. More recent data collected from the Signals program indicates that over three years, students who have taken between one and five Signals courses have significantly higher retention rates than students in a non-Signals control group. Furthermore, it was found that students who had chosen to participate in the Signals program had lower standardized testing scores at entrance than control subjects (Arnold & Pistilli, 2012). These findings represent the best data available on the longitudinal impact of an early alert intervention targeted toward at-risk students.

In a recent publication, Tinto discusses the importance of interventions that reach into the classroom. Campus-wide efforts to increase student engagement, such as clubs, social events, job fairs, etc., are more likely to reach traditional on-campus students than non-residential students. Often interventions must be employed through a course in order to reach minority populations (Tinto, 2012). The most effective attempts to affect retention positively occur through interactions with faculty (Tinto, 1982;

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Tinto, 1987). This point is supported by the Seven Principles of Good Practices, a series of approaches designed to improve student outcomes. All seven of the principles in one way or another will have a positive impact on student engagement (Chickering & Ehrmann, 1996). Principle one, encouraging contact between students and faculty, is perhaps the principle most directly achieved through the interventions deployed in this study. Our intervention achieves this through use of emails sent by the instructor to the student. These emails could be customized to address specific issues with which the student was struggling.

Tinto points out that early efforts to improve retention rates focused on selection and admission efforts: "Stop talking to faculty about student retention and focus instead on the ways their actions can enhance students' education" (Tinto, 2007, p. 9). Tinto correctly suggests that many institutions chose to address the issue of retention via the admission process. This approach ignores the fact that many students enrolled in higher education institutions are unprepared for the challenges that face them. Additionally some institutions specifically serve populations that are underprepared for the challenges of higher education. We must instead focus on what we can do within our institutions to improve retention rates of our admitted students. Tinto suggests that effective intervention will be specific to the setting in which it is attempted. Differences in the institution size, focus, and target population require unique approaches. Often institutions already have support services tailored to their student population needs. Many institutions offer student services, such as access to a writing centre, proofreading services, or math tutoring sessions. Unfortunately, these services often go underused by students who could benefit from them the most (Tinto, 2012). Effective interventions will connect existing services to students who may not even know they need these services. With this in mind, we encouraged students to take advantage of the resources offered at their institution and designed specifically for that institution's student body. The OASE intervention was developed to address this issue specifically by connecting the numerous services currently available at higher education institutions and the students who can most benefit from these services.

Campbell also references Astin's theory of student involvement, which suggests that most activities requiring a student to interact with his or her instructor improve student retention and academic performance (Astin, 1993; Astin, 1999). The receipt of an email from an instructor indicating that a student's performance is problematic creates a situation in which the student is more likely to address the issue directly with the instructor. These interactions develop the student's academic engagement, potentially resulting in better retention rates. Singell and Waddell (2010) provide a nice review of Tinto and Astin's theories regarding the complex issue of student retention.

The OAAI has adhered to the notification system concept used at Purdue by leveraging Sakai Project Sites to create an Online Academic Support Environment that provides a unique opportunity to engage students identified as "needing help" in an online community designed to help them succeed academically (more on this in section 4) . There is significant evidence from the past 20 years of research that high levels of student engagement or involvement with their institution is empirically linked to

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

higher rates of student retention (Cuseo, n.d.). Although the correlation between support groups and academic success has not been as widely researched, recent studies have also shown compelling evidence that students who participate in support groups had significantly higher first- and second-semester and cumulative GPAs than their peers who did not participate. In addition, students who participated in such groups were much more likely to persist into their sophomore year (79% vs. 39%) (Folger, Carter, & Chase, 2004).

There has been relatively little attention directed at the importance of timing in determining the effectiveness of any intervention. The logic is simple, the sooner an intervention can be deployed, the more time a student has to address the problem. A series of studies using absenteeism as an indicator of performance in the classroom have consistently pointed to the importance of providing feedback to the student early in the semester (Bevitt, Baldwin, & Calvert, 2010). Absenteeism is a reliable and easily collected indicator of performance, providing an opportunity for intervention as early as two weeks into the semester (Smith & Beggs, 2003; Colby, 2004). Two separate studies have confirmed, the earlier the intervention, the better the opportunity for the student to change his or her grade in a positive direction (Colby, 2004; Newman-Ford, Fitzgibbon, Lloyd & Thomas, 2008).

3 PREDICTIVE MODELLING FOR ACADEMIC RISK DETECTION

The predictive analysis goal of the OAAI was to detect, relatively early in the semester, those undergraduate students who were in academic difficulty in the course by using student data. This task was re-expressed as a binary classification process with the purpose of discriminating between students a) in good standing or b) academically at-risk.[15] Classification is a supervised learning task, where a set of input data samples, each of them labelled with a target class value, is used to train the classification model.

Figure 1 depicts the OAAI architecture. Four sources of data were considered: a) student demographic and aptitude data; b) course grades and course related data; c) Sakai-generated data on student interaction with the learning management system; d) partial contributions to the student's final grade collected by Sakai's gradebook tool (i.e., student grades on specific grading events, such as assignments and exams). The OAAI used Pentaho Business Intelligence, an open source analytics suite with data mining and data integration capabilities. Predictive models were developed using both Weka (Pentaho's data mining module) and IBM's SPSS Modeler, to preserve compatibility between data mining tools. Pentaho's data integration tool automated the sourcing of input data feeding the predictive modelling stage.

During the extraction process, data was anonymized to remove identifying student information. Data was subsequently rescaled, transformed, processed to handle missing values and outliers, and finally

[15] "At-risk" in this paper means academically at-risk.

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

consolidated into a single data set organized by institution, semester, course, and student. The unit of analysis was each course taken by a given student in a given semester, enriched with student demographic, aptitude and course data, Sakai-generated data, and students' grade contributions (grades on assignments, exams, etc). The target feature used to classify students in *good standing* and *at-risk* was derived from the course grade, using a C grade as a threshold of acceptable academic performance (students with less than a C are considered at risk).[16]

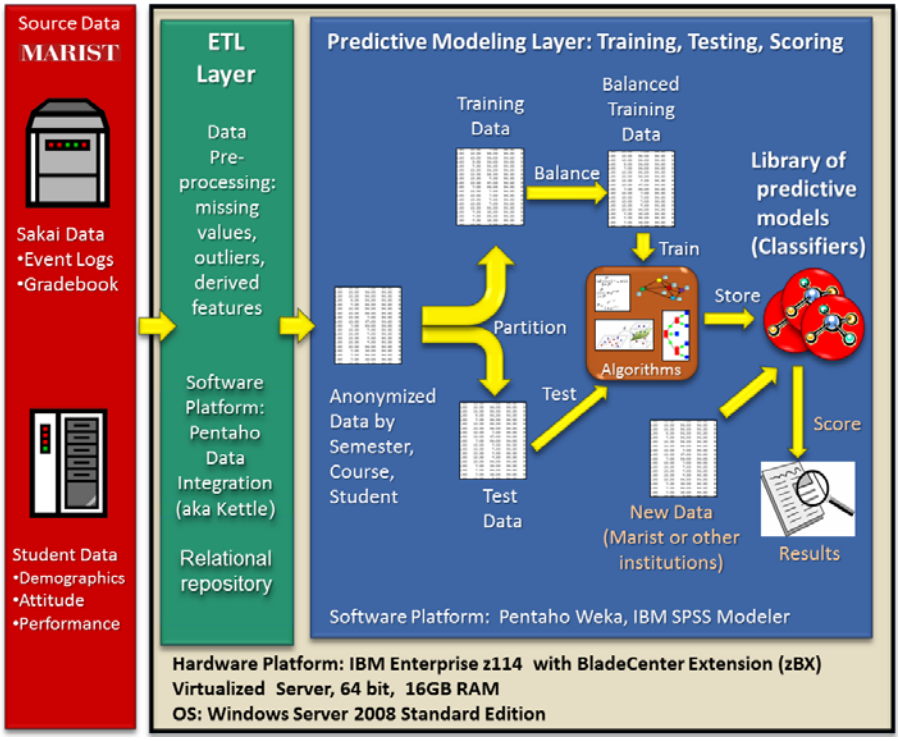

Figure 1: OAAI predictive modelling architecture

Sakai records student interactions with the learning management system on each of the tools that the instructor chooses to include as part of the course site. The ETL (extraction, transformation, and loading) process computed statistics on these Sakai-generated events, including the following: the number of Sakai course sessions opened by the student; the number of discussion forum threads read by the student; the number of discussion forum threads contributed by the student; the number of assignments submitted by the student; the number of assessment tests submitted by the student. An aggregated score was derived from Sakai's gradebook scores entered by the instructor when submitting to students their results on assignments, exams, projects, and other gradable events.

[16] Academic grading in the USA has traditionally used five letter grades (A, B, C, D, and F). A denotes the highest grade, and F denotes failure. Numerical values are applied to grades as follows: A=4, B=3, C=2, D=1, F=0. C is considered a passing grade in a course, and it is the minimum threshold for the average of undergraduate students' grades during their time at an institution (a requisite for graduation).

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Table 1: Input Data Set Used to Train and Test Predictive Models

Attritute Type	Attribute Name	Description	% Missing Values in Training Data
Predictors	ONLINE	online flag	0.00
	AGE	student's age	0.10
	GENDER	student's gender (self-reported)	0.00
	SAT_VERBAL	standardized verbal test	17.33
	SAT_MATH	standardized math test	17.36
	APTITUDE_SCORE	standardized (SAT) composite score or the converted ACT to SAT score (ACT is an alternative standardized test)	11.71
	FTPT	full-time or part-time student	0.00
	CLASS	freshman, sophomore, junior, senior	0.00
	CUM_GPA	cumulative GPA	0.00
	ENROLLMENT	course size	0.00
	ACADEMIC_STANDING	Deans's list or semester honors, regular, probation	0.00
	RMN_SCORE_PARTIAL (*)	score computed from partial contributions to the final grade submitted by instructor	0.00
	R_SESSIONS (*)	number of Sakai course sessions opened by the student	0.00
	R_CONTENT_READ (*)	number of times a section in the Lessons tool is accessed by a student	10.03
Target	ACADEMIC_RISK	at-risk, good standing	0.00
Discarded	R_FORUM_READ (*)	number of discussion forum threads read by the student	68.39
	R_FORUM_POST (*)	number of discussion forum thread posted by the student	68.83
	R_ASN_SUB (*)	number of assignments submitted by the student	64.16
	R_ASSMT_SUB (*)	number of exams submitted by the student	38.27

(*) calculated as a ratio by dividing by the average course value

Following standard supervised learning practice, the consolidated data set (see Table 1) was partitioned into training and test data subsets: a) two semesters of undergraduate data (fall 2010 and spring 2011, 9,938 samples) for the training data set; and b) one semester of undergraduate data (fall 2011, 5,212 samples) for the test data set. All records in both the training data set and the test data set were labelled with the target class value (ACADEMIC_RISK = at-risk, good standing). The label is used by the classification algorithm to supervise the learning of the model at training time. At test time, the label is used to compute the predictive performance of the classifier. Validated predictive models were stored for future scoring of incoming student data.

3.1 Machine Learning Algorithms for Predictive Modelling

Several machine-learning algorithms were considered to train predictive models. After evaluating a number of them, we settled for four well-known classifiers for comparison purposes: logistic regression, support vector machines using sequential minimal optimization (SVM/SMO), J48 decision trees, and

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Naïve Bayes, all of them robust classification methods that can handle categorical and continuous predictors. Weka provides implementations of all of them.[17]

Logistic regression is probably the most popular parametric method used in situations when the target variable is categorical (i.e., classification). Logistic regression models are generalized linear models that predict the outcome of a categorical dependent variable based on one or more predictor variables. In its simplest form, it can solve binary classification problems; in its multinomial form, it can be used to solve multivalued (2+ classes) classification problems. As the goal in this project was to detect at-risk students, we focused on binary logistic regression. The term *logistic regression* is usually applied in the literature (and in this article hereafter) to refer to those cases where the dependent variable is binary. Logistic regression models the probability of occurrence of a certain value of the target (class) variable as a logistic (or *logit*) function of the linear combination of the set of continuous or discrete predictor variables. In this sense, logistic regression is often referred to as a discriminative classifier because the probability of the class given the data can be viewed as directly discriminating the value of the class for any given configuration of predictor values.

Logistic regression makes no assumption about the distribution of the predictors, but the user must decide on the inclusion of predictors in the model, as well as any interaction and regularization terms. As in the case of linear regression, multicollinearity can have a negative effect on the parameter estimates, inflating their variance, and therefore affecting the model fit. For further information regarding logistic regression, see for example Neter, Kutner, Nachtsheim, and Wasserman (1996) and Larose (2006).

J48 is Weka's open source implementation of Quinlan's C4.5 decision tree, a non-parametric algorithm that learns rules from data. A decision tree is a graphical representation of rules inferred from input (i.e., training) data that constitute the basis for prediction. The set of rules learnt by the algorithm describe a class to which an object or event belongs (two class values in the case of this project: at-risk and good standing). Decision tree models use a recursive procedure to partition the training data progressively into groups according to a partition rule that maximizes the homogeneity of the dependent variable in each of the obtained groups. At each step of the procedure, the partition rule selects a predictor variable, to split the data file into the groups, stopping when pre-specified conditions are satisfied. The outcome of the learning process is a set of rules (or its associated tree-like representation) that describes the predictor features and their value ranges that specify a given class value. This makes decision trees highly expressive: good at both predicting and describing the nature of the prediction (i.e., the prediction is not the result of a black box). Quinlan's C4.5 algorithm uses an information theoretical metric (entropy reduction, also known as information gain) to determine the split criterion. For a detailed description of C4.5, see Quinlan (1993).

Support vector machines (or SVMs) are a state of the art family of supervised learning models proposed

[17] http://wiki.pentaho.com/display/datamining/classifiers

by Vladimir Vapnik (1995) that have become increasingly popular for classification, regression, and novelty detection tasks. SVMs are particularly well suited to analyze data with a large number of predictor attributes, and have therefore had considerable impact in text categorization and bioinformatics. The basic SVM is a discriminative, maximum-margin model based on the idea of classifying data into two categories by finding an optimal decision boundary (an *N*-dimensional hyperplane) that is as far away from the data in each of the classes as possible. The vectors near the hyperplane are the *support vectors*. Therefore, the basic SVM is a non-probabilistic, binary linear classifier. To deal with non-linear boundaries, an SVM maps data into a dimensional feature space where the data points can be categorized or predicted accurately, even if there is no easy way to separate the points in the original dimensional space. This involves using a kernel function to map the data from the original space into the new feature space. An SVM, like its close cousin the multilayer perceptron neural network model, does not provide output in the form of a function of its predictors. Thus, like neural networks, they are less expressive than other machine learning algorithms (more of a black box approach to prediction). For a detailed tutorial on SVMs, we recommend Burges (1998).

In this study, we used Platt's (1999) sequential minimal optimization (SMO) algorithm, which tackles the maximum-margin hyperplane optimization by decomposing the problem into 2-dimensional sub-problems that may be solved analytically, eliminating the need for a numerical optimization algorithm. In addition, in order to obtain posterior probability estimates for the classes, we set the parameter in Weka's implementation of (SVM/SMO) that fits logistic regression models to the outputs of the support vector machine.

Naïve Bayes classifiers (Friedman, Geiger, & Goldszmidt, 1997) are simplified Bayesian networks, graphical models based on the notion of conditional independence that encode the joint probability distribution of a set of variables in a compact manner using a directed graph to describe the probabilistic dependencies among variables. A Naïve Bayes classifier assumes that all predictor variables are conditionally independent given the class variable. This very strong independence assumption simplifies the computation of the likelihood of the data, reducing it to a product of the likelihood of each attribute given the class, and therefore significantly decreasing the amount of training data required to estimate the model parameters. The classifier learns the estimates of the class priors and the likelihood of the data (conditional probability of the data given the class). New examples can be assigned to the class value that yields the highest posterior probability, which is proportional to (likelihood $\times$ prior). This type of classifier is described as *generative* since the posterior probability distribution of the data given the class can be viewed as a random generator of data samples for a given class value. When the predictor attributes are discrete, or Gaussian with variance independent of the class, Naïve Bayes learners can be viewed as linear classifiers; that is, every such Naïve Bayes corresponds to a hyperplane decision boundary in predictor attribute space (Mitchell, 2005). Despite the oversimplified assumptions that give its name to the algorithm, Naïve Bayes classifiers exhibit excellent performance in many complex real-world situations. For further reading on the performance of Naïve Bayes, including a theoretical explanation on the optimality of the algorithm, see Rish (2001) and Zhang (2004).

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

3.2 Input Data Considerations and Data Quality Challenges

Data mining algorithms are affected by the quality and characteristics of the input data. Generally, a predictive model is usually as good as its training data. Although a careful choice of data mining algorithms can sometimes mitigate the poor quality of the data to be mined (Fisher, Lauría, Chengalur-Smith, & Wang, 2006), in general, no matter how robust the data mining algorithm is, it will fail to produce accurate models if faced with low-quality training data (Freitas, 2002). Therefore, for any data mining effort to be successful, it should be preceded by a data quality enhancement activity (Lauría & Tayi, 2003). The data collected at Marist College for training and testing purposes, particularly the log data from Sakai, was reviewed by staff at the IT Department prior to being submitted for data integration (ETL) and analysis, to verify its integrity and to ensure that technical problems did not result in erroneous data being collected. Marist College data presented a number of issues that had to be addressed before it could be effectively used in the predictive modelling stage.

Missing data: In the initial consideration of the input data set, missing data was present in a number of attributes (see Table 1 to check the percentage of missing values per attribute in the training data set). This was especially significant in those attributes related to Sakai usage (see the paragraph on variability in Sakai tools usage below). *Corrective action*: We used a cut-off of 20% missing data to discard those attributes in the input data set with a percentage of missing values above this threshold. For the rest of the attributes containing missing data, no pre-processing of missing data was made at ETL time as Weka's implementation of the machine learning algorithms used in this study provides built-in mechanisms to deal with missing data: a) Logistic regression and SVM/SMO use a filter (named ReplaceMissingValues in Weka's library) that is "trained" on the training data (i.e., it records the means on numeric attributes and modes on categorical attributes computed on the training data). These values are used to replace missing values in test instances; b) J48 (which inherits its missing data mechanism from C4.5) splits training instances with missing values into pieces. A piece going down a branch receives a weight proportional to the popularity of the branch, with weights adding up to 1. J48 uses a similar mechanism at prediction time for handling missing values as it does at training time. If the decision tree splits on an attribute that has a missing value in a given test instance, then predictions (probability distributions) from all sub-trees rooted at that point are combined with weights proportional to the number of training instances that supported each sub-tree; c) Naïve Bayes in turn ignores missing values altogether (i.e., it does not update statistics for an attribute when its value is missing in a training instance; at testing time, an attribute is omitted from the Bayes formula if its value is missing in the test instance). Some special considerations were made on those attributes related to Sakai usage (see paragraph below).

Variability in Sakai tools usage: Not all instructors use the same set of Sakai tools (e.g., traditional, on-the-ground courses do not usually include discussion forums). This produces null values in those records corresponding to courses where Sakai tools were not used and therefore data was never meant to be generated. *Corrective action:* We used the same guidelines that Campbell (2007) used in his research:

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

for a Sakai course tool to be counted, at least 50% of the students in the course should use the tool at least once. The missing data rule explained in the paragraph above discarded those attributes with 20% or more of missing values. A quick inspection of Table 1 shows that the attributes discarded are those corresponding to Sakai tools not typically used in undergraduate courses (discussion forums and submissions of assignments and exams are common in online courses, which represents a small percentage of the undergraduate course offering at Marist College). This is a design consideration: eliminating attributes from the analysis with a high percentage of missing values, or imputing missing data (using Weka's built-in mechanism to impute data on those algorithms that require complete data). The decision made was to use a low cut-off of missing data, and with this eliminate a number of attributes, particularly those corresponding to Sakai tool usage. We do acknowledge that this introduces a bias in the analysis, as we indirectly impose a selection of predictor attributes in the input data. We consider though that the alternative of imputing attributes with high percentages of missing data would introduce even more bias and, furthermore, would be conceptually incorrect: it is not the same to impute missing data omitted at random due to issues in the data collection (e.g., SAT_MATH, SAT_VERBAL) than to impute large amounts of missing data on attributes where it was never meant to be generated (R_FORUM_POST, R_ASN_SUB).[18]

Variability in assessment and student activity: Workload varies across courses and instructors (e.g., an instructor may be more demanding, or post materials more frequently, depending on the characteristics of the course, or its weekly schedule). This may have a confounding impact on the ability of the predictive models to capture behavioural patterns of similar students across different courses. These patterns may inaccurately portray differences in behaviour among students of similar characteristics in different courses that should otherwise reflect similar behaviour (e.g., a frequency of access to content material by a student in a course where the instructor posts course materials once every two weeks may be seen as an indication of less than satisfactory effort when compared to a course where the instructor posts materials twice a week). *Corrective action:* Frequencies of Sakai-generated events are replaced by ratios and proportions, normalizing those frequencies by dividing them by the average course frequency. For example the CONTENT_READ variable, measuring the number of times a section in the Lessons tool is accessed by a student becomes R_CONTENT_READ, measuring the number of times a section in the Lessons tool is accessed by a student divided by the average number of times a section in the Lessons tool is accessed by a student in that course.

Unbalanced classes: The proportion of academically at-risk students may vary across institutions. At Marist College, for example, the average percentage of students with poor grades is quite low (around 7% of the consolidated data set that lists courses taken by students exhibit grades below C). This poses an additional challenge as it yields input data that is unbalanced at the class value (*good standing*, *at-risk*). In situations where the distribution of classes is highly unbalanced, the number of samples of the

[18] Other research projects and initiatives had struggled with the same issues regarding course management tool usage and missing data. We had several fruitful discussions in this regard with John Campbell (Purdue) and Steve Lonn (U. Michigan). We gratefully acknowledge their input.

class of interest (at-risk students in this case) may be too small to provide useful information about what distinguishes students in good standing (the dominant class value). *Corrective action:* A stratified sampling approach was applied on the training data set to balance the proportion of classes and, therefore, improve the performance of the predictive model at detecting at-risk cases. This is accomplished either by oversampling the at-risk cases in the training data set, or sub-sampling the good-standing cases. Weka includes a re-sampling function that combines these two approaches by both oversampling the minority class and sub-sampling the dominant class. The overall sampling size can be controlled by setting a function parameter. The test data set is not oversampled; it keeps the original distribution of class values, as it represents the class distribution with which the trained classifier will be confronted when making predictions on new data.

During the data integration (ETL) process, extraneous data records (e.g., records without a corresponding final grade) were removed from the input data. Once the input data was integrated, including transformation of variables representing frequencies into ratios, the continuous attributes in both training and test data sets were analyzed to check for outliers (an outlier was defined as an observation distant 3+ standard deviations from the mean). Records containing outliers were eliminated. We did not consider this a relevant issue that could bias the analysis given the rather large size of the training and test data, and the fact that the number of records with outliers represented less than 2% of the data in both the training and test data sets.

3.3 Predictive Performance Assessment

Trained models in a binary classification setting are typically evaluated on test data using measures of predictive performance derived from the confusion matrix that yields counts of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). If the input data has an unbalanced distribution of class values, using predictive accuracy or error rate to measure predictive performance may be misleading, as those metrics are driven by the dominant class value (good students, in the case of Marist College). We therefore report predictive accuracy, but we focus on other metrics, namely recall, false positive (FP) rate, and precision, as reported by Weka, to measure the predictive performance of the classifiers. Recall (TP/(TP+FN)) measures the ability of the classifier to detect the class of interest (*at-risk*); FP rate (1-TN/(TN+FP)) measures the number of false alarms raised by the classifier; precision (TP/(TP+FP)) measures the fraction of instances predicted as positives that are actually positives.[19] A perfect classifier would be described as one having 100% recall (i.e., predicting all at-risk students as being at-risk) and 0% FP rate (predicting no good standing students as being at-risk). However, there is usually a trade-off between performance metrics, as there is a lower bound on the

[19] Sensitivity and specificity are alternative terms typically used in clinical trials but also used in other fields, including data mining to refer to recall and (1-FP Rate). Recall and FP Rate are more commonly used in the fields of pattern recognition, machine learning, and information retrieval. Sensitivity is also called true positive rate, whereas (1-FP Rate) is sometimes called true negative rate. As Weka reports recall an FP rate rather than sensitivity and specificity, we decided to stick to its terminology.

error rate that can be achieved by any classifier acting on a given attribute space (Duda, Hart, & Stork, 2001).

A receiver operating characteristics (ROC) graph is an appealing technique for comparing classifiers based on their predictive performance, that visually depict the trade-off between recall (detection of the class of interest) and false alarm rates of classifiers. ROC graphs are 2-D graphs in which recall (TP rate) is plotted on the Y-axis and FP rate is plotted on the X-axis. A point in ROC space is better than another if it is above and to the left (TP rate is higher, FP rate is lower, or both) of the first (Fawcett, 2006). ROC graphs are also helpful to compare a learnt classifier's performance with a random guessing strategy, which acts as a baseline. The diagonal line (Y = X) in a ROC graph represents the strategy of randomly guessing the class of interest, where the points along the diagonal are given by the frequency with which the random-guessing classifier guesses the class of interest. Evidently, any classifier that holds some value should yield a point in ROC space located above the diagonal representing random guessing.

3.4 Experimental Setup on Marist Data and Analysis of Results

Data from fall 2010, spring 2011, and fall 2011 was collected at Marist College and cleaned, recoded, and aggregated into data sets corresponding to courses taken by a student in a given semester using the record format described in section 3 and Table 1. Fall 2010 and spring 2011 data were used for training the classifiers (9,938 training samples); fall 2011 was reserved for testing purposes (5,212 samples).

Experiments were conducted to test the predictive performance of the classifiers following these guidelines:

a) Four baseline predictive models were trained using four classification algorithms (Logistic Regression, J48, SVM/SMO, Naïve Bayes) and the full non-balanced training data (no re-sampling applied on the training data, 9,938 training samples) for comparison purposes. Each model was subsequently tested using the test subset and predictive performance measures were computed.
b) Multiple balanced training data sets of varying size were created by varying the overall sampling size (25%, 50%, 75%, and 100% of the training data re-sampled). Five different training data sets were created for each balanced training size by varying the sampling seed, a total of 5 x 4 = 20 balanced training data sets. Models were trained with each of the 20 training data sets and 4 algorithms, for a total of 4 x 5 x 4 = 80 models. Each model was subsequently tested using the test subset and predictive performance measures were computed.

Models were grouped according to sampling size of the training data, and classification algorithm. On all five models corresponding to the same sampling size (25%, 50%, 75%, and 100% of the training data re-sampled), the predictive performance measures were summarized, computing a mean value, and a standard error. Table 2 and Figures 2 and 3 report the outcomes of this evaluation.[20] Clearly, balancing

[20] Preliminary results of a less extensive and systematic experiment using fall 2010 data were reported in Lauría et al., 2012.

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

the training data has a positive effect on the ability of the classifiers to detect at-risk students, as measured by the Recall metric. Logistic regression, SVM/SMO and Naïve Bayes outperform J48 in terms of Recall; the three algorithms exhibit a very stable behaviour when varying the overall sampling size. To try to explain this behaviour we should refer to the bias–variance trade-off in supervised learning.[21] Logistic regression, support vector machines (linear ones at least) and Naïve Bayes are all high bias/low variance learners. Their representational power is fairly low (all being linear models) and this tends to make them very stable, which in turn leads to low variance. Decision trees, on the other hand, are low bias (are more expressive, have a stronger representational ability) but high variance learners.[22] This means that small changes in the training data set can lead to radically different trees being produced.

Table 2: Predictive Performance on Marist Data

				Logistic Reg.		SVM/SMO		Naïve Bayes		J48 Dec. Tree	
% Resampled	Resampled Size		Metric	Mean (%)	SE (%)	Mean (%)	SE (%)	Mean (%)	SE (%)	Mean (%)	SE (%)
25	At Risk	1,228	Accuracy	86.84	0.17	86.26	0.84	84.06	0.91	85.92	0.32
			FP Rate	12.96	0.21	13.68	0.89	15.82	1.02	13.32	0.27
	Good Standing	1,256	Precision	32.50	0.30	31.78	1.34	28.08	1.08	29.72	0.71
			Recall	84.12	0.39	85.02	0.30	82.52	0.59	75.94	1.38
50	At Risk	2,469	Accuracy	87.02	0.16	84.96	0.93	82.96	0.88	86.98	0.25
			FP Rate	12.76	0.17	14.98	0.97	16.98	0.98	11.80	0.34
	Good Standing	2,500	Precision	32.90	0.28	29.64	1.51	26.68	0.99	30.92	0.50
			Recall	84.34	0.34	83.88	0.50	82.66	0.54	71.22	1.78
75	At Risk	3,701	Accuracy	86.96	0.19	84.40	0.82	83.02	0.67	87.98	0.19
			FP Rate	12.86	0.21	15.54	0.88	17.00	0.75	10.32	0.25
	Good Standing	3,752	Precision	32.92	0.30	28.78	1.35	26.68	0.76	31.96	0.43
			Recall	84.94	0.26	83.84	0.21	82.88	0.40	65.16	1.37
100	At Risk	4,934	Accuracy	87.04	0.12	84.76	0.72	83.32	0.50	88.98	0.16
			FP Rate	12.80	0.15	15.16	0.78	16.62	0.54	9.04	0.20
	Good Standing	5,004	Precision	32.96	0.16	29.24	1.20	27.00	0.58	33.88	0.43
			Recall	84.82	0.32	83.94	0.26	82.54	0.27	62.50	1.01
No resampling (unbalanced)	At Risk	657	Accuracy	94.20		93.20		92.40		93.70	
			FP Rate	1.20		2.90		5.00		2.20	
	Good Standing	9,281	Precision	66.70		51.00		46.00		56.90	
			Recall	31.10		40.60		57.50		39.20	

Training Data Set: 9938 samples Class Probability: At Risk = 7.08% ; Good Standing = 92.92%

Test Data Set: 5212 samples Class Probability: At Risk = 6.91% ; Good Standing = 93.09%

[21] The learning error can be decomposed into bias and variance components. Generally, there is a trade-off between the bias and variance of a supervised learning algorithm (Geman, Bienenstock, & Doursat, 1992). The inductive bias of a learning algorithm is the set of assumptions that the learner uses to predict outputs given inputs that it has not encountered (Mitchell, 1980). A linear boundary (e.g., a straight line in 2-D attribute space) has high bias (the linear assumption is rather inflexible), but a fluctuation in the training data set has a small impact on its predictive power, which means that is has low variance. Instead, a learning algorithm with low bias must be "flexible" so that it can fit the training data well (e.g., a curve that passes though all data points in a 2-D attribute space), but in doing so, it learns the irregularities of the training data as well, which reduces its ability to make predictions on new data (i.e., generalize). This means that a learning algorithm with low bias will typically have high variance: it is affected by fluctuations in the training data set.

[22] For a detailed account of the effect of class imbalance on C4.5 classifiers, see Drummond and Holte (2003).

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Although we balanced the classes through re-sampling at roughly 50%, keeping the class distribution approximately equal in all the re-sampled training data sets (25%, 50%, 75%, 100%), the fact that the total number of instances changes (actually increases as we increase the percentage of re-sampled data) leads to different trees being produced. Bigger trees are learnt on the larger training sets created through re-sampling. Note that the varying re-sampling percentage not only increases the size of the training data set, it also changes the characteristics of the training data set by duplicating samples of the minority class through oversampling, and eliminating samples of the majority class though sub-sampling. These trees are increasingly more expressive but have decreasing predictive power as they lose their ability to generalize on new instances.

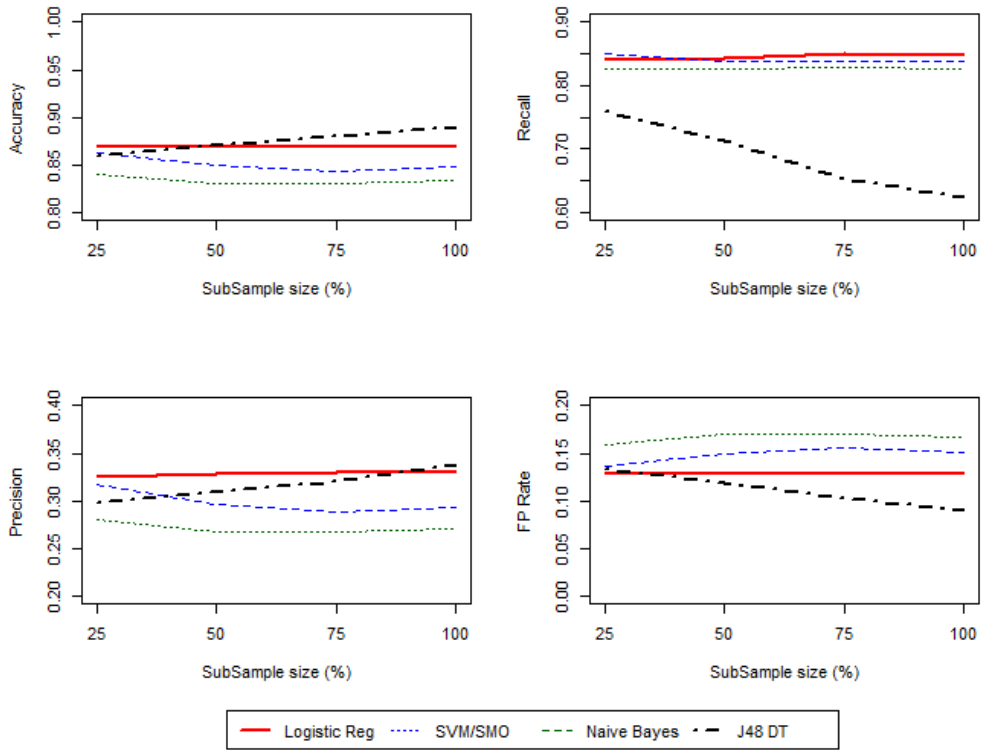

Figure 2: Predictive performance on Marist data

The different behaviour of the classifiers is reflected in Figure 2: for all three linear models (logistic regression, SVM/SMO, and Naïve Bayes), all metrics are fairly flat, as data set size increases. J48 exhibits a linear increase in accuracy and decrease in recall (which is tied to the decrease in FP rate and increase in precision).

Weka's handling of missing values may also be a source of variability in the behaviour of the classifier. Weka's re-sampling function oversamples the minority class (at-risk) and subsamples the dominant class. This change in the distribution of records belonging to each class also has an effect on the distribution of null values in the training data, which are then treated differently by the four learners

under consideration, as was described in section 3.2. This does not necessarily explain the direction of change (it would require a detailed analysis of the new proportions of missing data after each re-sampled training data set is produced), but it may point to some difference in behaviour among classifiers. In any case, the amount of missing data in the original (unbalanced) training data set is not very significant, as shown in Table 1. This leads us to believe that although Weka's handling of missing values may play a role, it is the inherent nature of the learners regarding the bias–variance trade-off that drives the behaviour of the classifiers when trained with different sized training data sets.[23]

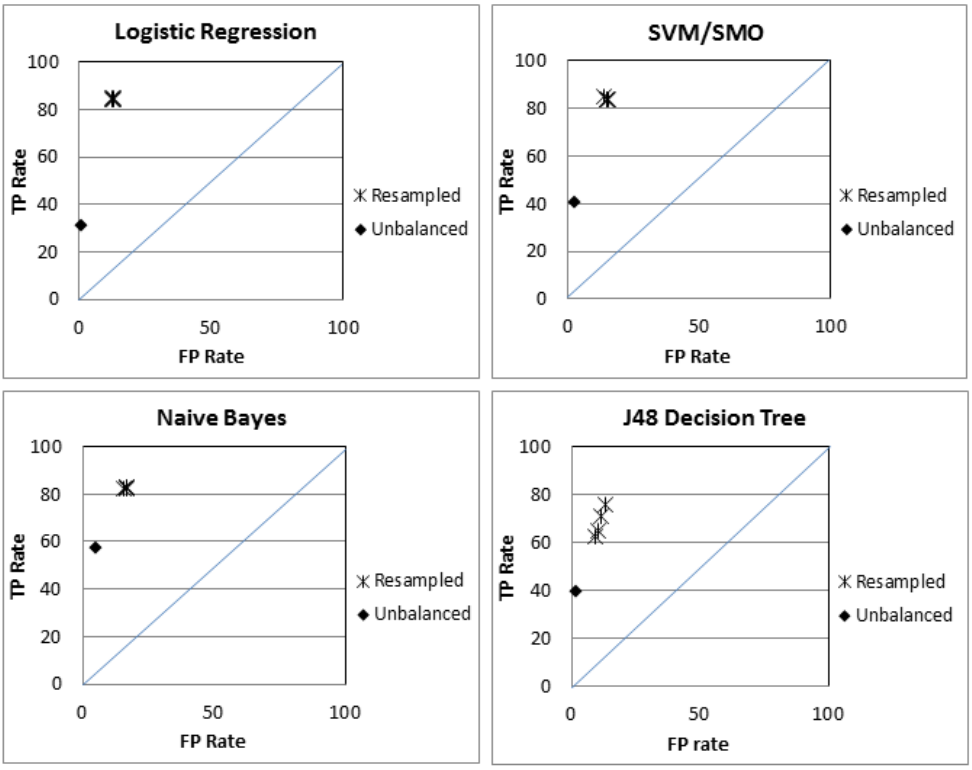

Figure 3: ROC graphs

The ROC graphs in Figure 3 show that the classifiers (in particular logistic regression, SVM/SMO and Naïve Bayes), when trained with balanced (re-sampled) data, exhibit more than acceptable performance, with value pairs (Recall, FP rate) located on the far northwest side of ROC space. Recall values in all three classifiers is high (more than 80%) while maintaining rather low FP values (less than 17%). These FP rates are relatively high when compared to the classifiers trained with unbalanced data (which exhibit single digit FP rates), which shows that balancing the training data through re-sampling improves detection of at-risk students but at the same time increases the number of false alarms. This issue requires further consideration as it has an impact on the number of students in good academic standing that are signalled by the system as being at-risk. Although this is not a matter of immediate

[23] This analysis came out of a discussion on this subject with Mark Hall, Weka's architect and one of its original core developers. The authors are grateful to Mark for his insightful comments and suggestions.

concern given that alerts are not automatically submitted to the students (they are submitted to the instructor instead), this does leave substantial room for improvement in terms of model development. The standard error in all four classifiers is small, which also means that the classifiers are not affected by random variations of the sampling seed. Overall, logistic regression seems to outperform the other classifiers, with a better combination of high Recall (above 84%), lower percentage of false alarms (FP Rate below 13%), and higher precision in predicting at-risk students (Precision close to 33%). SVM/SMO comes close, with similar performance metric values (Recall ≅ 84%, FP Rate ≅ 15%, Precision ≅ 29%).

To perform an assessment of the relevance of the predictors we picked one of the training data sets re-sampled at 50% and analyzed the learnt logistic regression model (we picked an arbitrary data set as we had checked before that the predictive performance of logistic regression is practically the same for all re-sampled data sets). We used SPSS Modeler to report the model outcomes as it provides a more detailed analysis along with a nice predictor importance chart.[24] According to this chart (see Figure 4), the score computed out of partial contributions to the final grade (RMN_SCORE_PARTIAL) appears to be the most relevant predictor, followed by cumulative GPA (CUM_GPA) and academic standing. The other predictors included in the chart are second tier. They include number of Sakai sessions logged in the semester (R_SESSIONS), online status (ONLINE_FLAG), full-time status (RC_FTPT), and student's class (RC_CLASS). The use of the RMN_SCORE_PARTIAL metric as a predictor seems promising if contributions to the final grade (such as assignments or tests) are available at prediction time.

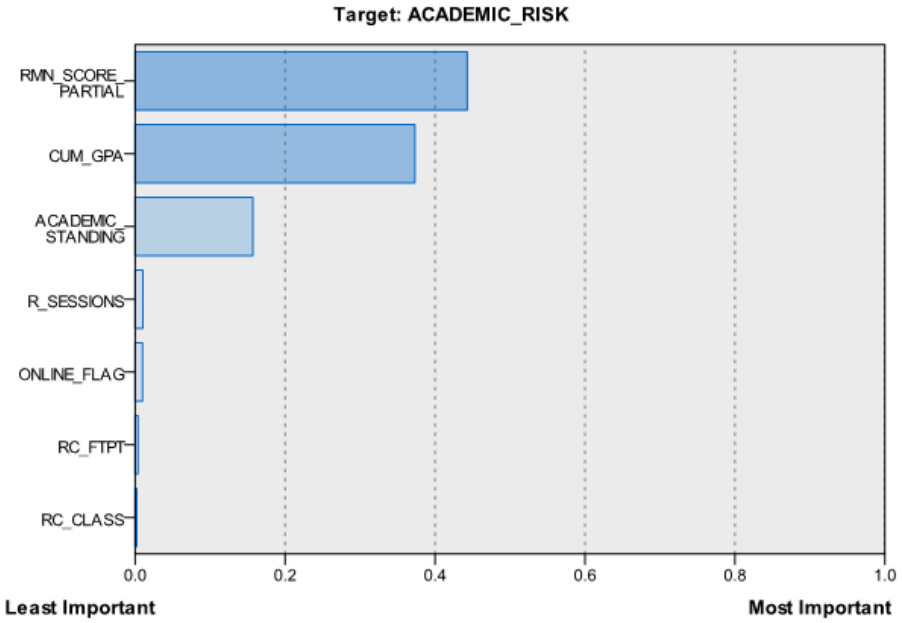

Figure 4: Predictor importance chart for logistic regression model

[24] The predictor importance chart in SPSS Modeler depicts the relative importance of each predictor in estimating the model. Since the values are relative, the sum of the values for all predictors on the display is 1.0. Predictor importance does not relate to model accuracy, it just relates to the importance of each predictor in making a prediction.

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Table 3 displays the outcome of the logistic regression model. Almost all regression coefficients are statistically significant (freshman and junior class indicators are the exception). As was expected, an increase of the partial grades score (RMN_SCORE_PARTIAL), cumulative GPA, and number of Sakai sessions (R_SESSIONS) decreased the expected probability of being at-risk relative to being in good standing, controlling for the other inputs. Regular students, compared to online students, have a large reduction (.319) in the expected ratio of the probability of being at-risk relative to being in good standing. Part-time students are expected to have a much higher (by a factor of 2.443) proportion of being at-risk relative to being in good standing than full time students. Sophomore students have a greater expected probability of being at-risk (1.34) relative to good standing. Finally, students in probation and regular students are much more likely to be at-risk (by a factor of 25.5 and 8.4 respectively) than honours students, controlling for the other predictors.

Table 3: Logistic Regression Model for At-Risk Students

Variable	Metric Slope (b)	Wald	df	p	Odds Ratio Exp(b)
Regular student (ONLINE_FLAG =0)	-1.143	24.80	1	<.001	.319
Part-time student (RC_FTPT=0)	.893	12.31	1	<.001	2.443
Cumulative GPA (CUM_GPA)	-2.354	297.68	1	<.001	.095
Partial grades score (RMN_SCORE_PARTIAL)	-.077	434.34		<.001	.926
Number of Sakai Sessions (R_SESSIONS)	-.146	5.06	1	.024	.864
Freshman (RC_CLASS=1)	-.134	.936	1	.333	.875
Sophomore (RC_CLASS=2)	.292	5.37	1	.020	1.340
Junior (RC_CLASS=3)	.023	.031	1	.861	1.023
Probation (ACADEMIC_STANDING=0)	3.243	54.84	1	<.001	25.598
Regular standing (ACADEMIC_STANDING=1)	2.132	28.49	1	<.001	8.428
Intercept	11.879	267.43	1	<.001	
Senior (RC_CLASS=4) and Honour/Dean's list (ACADEMIC_STANDING=2) are reference categories					
Chi-Square = 3859.12	df=10 p <0.001				

4 CONDUCTING PILOTS OF THE ACADEMIC ALERT SYSTEM AT PARTNER INSTITUTIONS

One of the factors associated with the scaling of learning analytics that the OAAI has researched is the portability of predictive models: how models developed for one academic context (e.g., a large research university), can be effectively deployed in another (e.g., community college). During the summer of 2011 we ran correlations on Marist data between students' grades and the same set of predictors used by Campbell (2007) in his original dissertation research, which included student demographic (e.g., age), aptitude (e.g., SAT scores), and learning management system usage (e.g., number of sessions initiated by the student). Although Marist College and Purdue University differ in a number of ways (e.g., different institutional type, size, and instructional approaches), and use different learning management

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

systems (Blackboard in the case of Purdue, Sakai at Marist College) they have similar key institutional characteristics that provided a good initial test of the model's portability. These include percentage of students receiving federal Pell[25] Grants (Marist 11%, Purdue 14%), percentage of Asian/Black/African American/Hispanic students (Marist 11%, Purdue 11%), and ACT[26] composite 25th/75th percentile (Marist 23/27, Purdue 23/29) (U.S. Dept. of Education, 2010). In addition, both Blackboard and Sakai can log similar types of events generated by student interaction with the learning management system (e.g., assignments posted, contributions to discussion forums). We compared the predictors that were correlated with student grades (as done by Campbell) as a means to understand the degree to which the models differed. In general, we found the same statistically significant elements as Purdue with similar correlation strengths. These initial findings on portability were included in a paper presented at the 2012 international Learning Analytics and Knowledge (LAK) conference (Lauría, Baron, Devireddy, Sundararaju, & Jayaprakash, 2012).

Building on these early results on portability, and using the predictive models trained with Marist data depicted in Section 3, we set out to pilot the OAAI prototype open-source academic early alert system at four partner institutions: two community colleges (Cerritos Community College and College of the Redwoods, both in California), as well as two HBCUs (Savannah State University and North Carolina Agricultural and Technical State University).

Table 4: Demographics and Retention Rates at Marist College and Pilots

Institution	Insitution Type	Undergraduate Enrollment	Male : female ratio	Student : Faculty ratio	Pell Grant awardee population	Retention rates within 150% of normal time of program
Marist	4-year, private	5,442	41:59	15:1	16%	80%
Savannah	4-year, public, HBCU	4,386	46:54	23:1	78%	30%
Cerritos	2-year, public , community college	21,335	45:55	30:1	45%	20%
Redwoods	2-year, public , community college	6,874	43:57	25:1	44%	4%
NCAT	4-year, public, HBCU	9,206	46:54	18:1	61%	41%

Table 4 depicts differences in demographics and retention rates between Marist College and the four partner institutions (U.S. Dept. of Education, 2010). At Marist College, 12% of the students are minorities and only 16% of the population receives Pell grants. Savannah State has a 94% Black/non-Hispanic student population with 78% of the students receiving Pell Grants; 22% of students at College

[25] Federal Pell Grants are limited to students in financial need.
[26] The ACT (American College Testing) college readiness assessment is a standardized test for high school achievement and college admission in the United States.

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

of the Redwoods are minorities and 44% receive Pell Grants; Cerritos College's student body is 41% Hispanic and 45% of the students receive Pell Grants; and North Carolina AT&T has an 89% Black/non-Hispanic student population and 61% of the students are awarded Pell grants. In addition, the retention rates at these institutions are considerably lower when compared to Marist College.

The purpose of the pilots was twofold: a) research the portability of predictive models developed at Marist (see section 3) as means to understand how their accuracy changes as they are deployed in different academic contexts (four-year vs. two-year institutions) and how to deal with such changes; b) explore the impact that different "intervention" strategies, such as participating in an Online Academic Support Environment (OASE), have on course completion and content mastery (i.e., final course grades) outcomes.

4.1 Description of the Academic Alert System

The pilots were conducted in spring 2012 and fall 2012. Three times during each of those semesters (at approximately four week intervals, 25%, 50% and 75% of the semester completed) anonymized data was collected from a pre-established set of courses at each of the partner institutions (note: NCAT did not participate in the spring 2012 pilot). Most courses were in the 15-18 week range while a few were shorter and in the 9-week range. We chose mostly freshman/first year introductory courses (often referred to as "gateway courses" as students who do not succeed in these tend not to continue in the course subject matter). The courses were face to face and spanned a wide range of subjects, including math, sciences, business, and art. Class sizes ranged from 20 to 60 students.

A logistic regression model trained with fall 2010/spring 2011 Marist College data was applied on the pilot data to identify students who were potentially at risk of not completing their course. Once the prediction process was completed an Academic Alert Report (AAR) corresponding to the collected data (25%, 50% and 75%) was produced for each partner institution, listing the students in the pilot data who had been identified as at-risk. The generated AARs were placed in a secure location (a Sakai Project Site). Instructors accessed their corresponding AARs from the specified site, and recovered the identity of each course/ student record using an encrypted Student Identification Key. Figure 5 depicts the workflow for generation and distribution of AARs.

Students identified as at-risk, were subjected to two different intervention strategies: "Awareness Messaging" and participating in an "Online Academic Support Environment (OASE)." To explore the impact of these different intervention strategies, different sections of the same course were designated as either control group (received no intervention) or treatment groups receiving either "Awareness Messaging" intervention or OASE intervention. In most cases, we had the same instructor teaching three sections of the same course and used one section as the control and the other two for the two treatment groups. Although not perfect, this approach helped to control for variations in teaching style.

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

- Awareness Messaging Intervention: Students identified in an AAR who were assigned to classes in the "Awareness Intervention" group received a message indicating that they were at risk of not completing the course successfully along with guidance on what they might do to improve their chances of success.

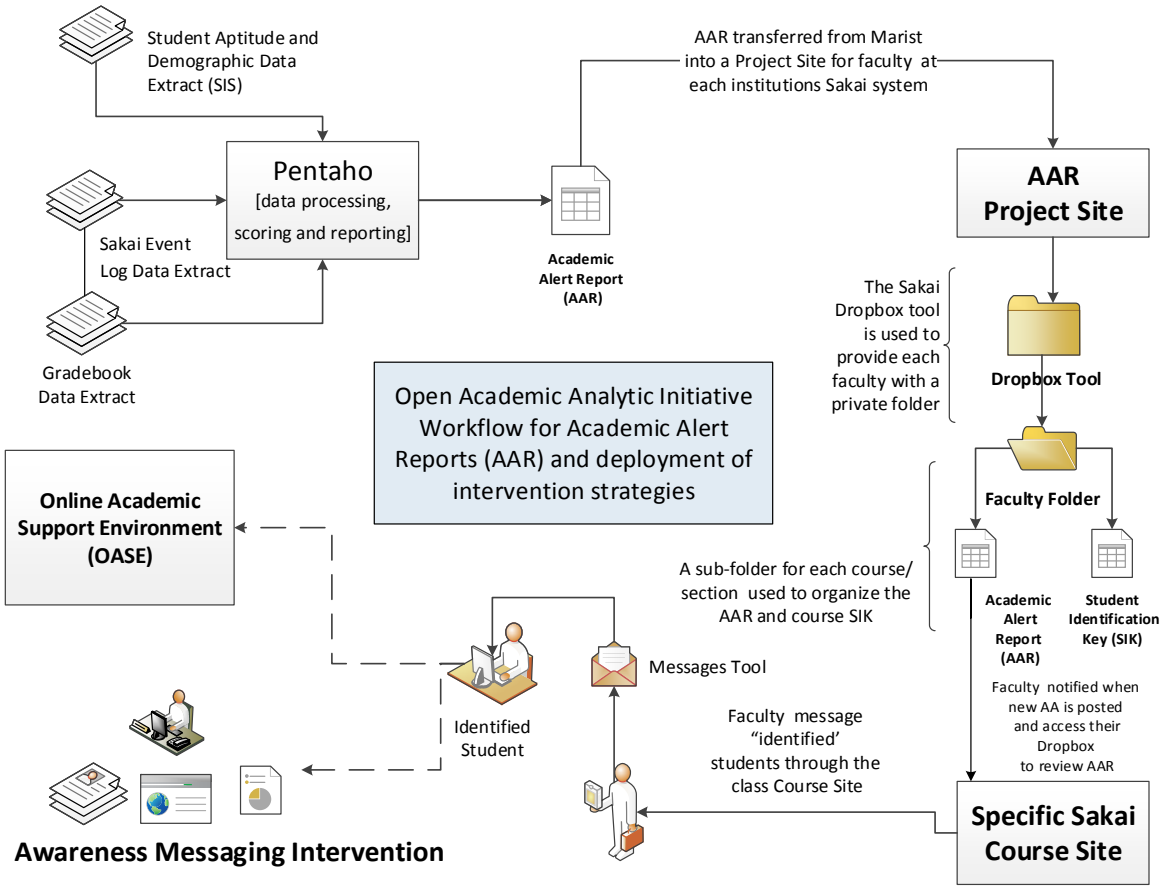

Figure 5: Workflow for AAR generation and distribution

- Online Academic Support Environment Intervention: Students identified in an AAR who were assigned to classes in the "Online Academic Support Environment Intervention" group received a very similar message to the other intervention group except that instead of specific recommendations, the students were encouraged to join the institutions Online Academic Support Environment, a Sakai-based online support site in which they are given access to Open Educational Resources (OER) instructional materials (e.g., Khan Academy videos, Flat World Knowledge textbooks, etc.). In addition to these materials, they are provided with a range of mentoring from peers and professional support staff.

In both types of interventions, the text of the messages is standardized across instructors and the text becomes increasingly serious in tone as students receive their second and third message.

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Ease of Use: We recognize that any effective intervention cannot be overly burdensome on an instructor. The willingness of an instructor to commit to the use of the intervention system is critical to the success of the intervention. Given the demands currently put upon adjunct instructors and full-time faculty it is critical that any effective intervention is efficient. If the intervention requires too much additional effort on behalf of the instructor, many will simply choose not to use the system. At three different intervals during the semester (25%, 50%, and 75% of the semester completed) students identified with low, medium, and high levels of probability of being at-risk were sent to the instructor, who ultimately determined to whom to forward an email message. This provided the instructor with an opportunity to consider special circumstances such as illness or the fact that they may have already spoken to the student. Ultimately, the instructor has the best opportunity to judge who could benefit from an early intervention. It is also especially critical in the developmental stages to provide instructors with an opportunity to preview and customize[27] any alerts sent to their students. The system is a tool that the instructor can use to help better stay in touch with their student's progress, and it offers an opportunity to interact with the student, improving academic engagement. These are both critical to establishing higher levels of student engagement and ultimately increasing student retention.

Early Alerts: The opportunity to provide early feedback to the student about their progress in the course gives the student a greater opportunity to change what may be ineffective strategies. In many universities and colleges, students do not receive a high level of feedback until midterm grades are returned. Unfortunately, in many courses, by the time the student has received their midterm grades, it is too late to improve their grade significantly, often resulting in a poor grade or even course failure. The use of numerous metrics employed by the predictive model allows for feedback to students much earlier in the semester. The importance of early feedback has not been well addressed in the literature perhaps because techniques to do so have been limited. Now with the advent of powerful learning analytics techniques and the use of automated alert systems, instructors have the opportunity to provide feedback to their students with enough time for the students to change their behaviour.

4.2 The Interventions Design Framework

The Awareness Intervention was modelled on the service provided in the Purdue Signals program. The work done at Purdue has shown that this use of relatively simple *notification interventions* can have a significant effect on student behaviour. The OAAI has added to the notification system concept, or what we refer to as *awareness intervention,* by leveraging Sakai Project Sites to create an "Online Academic Support Environment" (OASE) to provide a unique opportunity to engage students identified as at-risk and needing help in an online community designed to help them succeed academically.

[27] The first part of each message was standardized; the end of the message was then something the instructor could customize.

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

The past two decades of research into the design of effective online courses has resulted in compelling evidence that the type and frequency of interactions between the learner and instructor, the learner and content and the learner and his or her peers correlates closely with overall course satisfaction, engagement and obtainment of learning objectives (Cuseo, n.d.). The OASE Design Framework is organized around this simple but powerful concept of *learner interactions* with the goal of creating a compelling online environment in which learners will feel part of an academic support community.

These types of interactions allowed the students to develop strategies that have the potential of carrying over from course to course. They were able to access a variety of materials pertaining to study skills, time management, stress reduction tips, etc. They were also able to access general subject-specific materials to help with a variety of classes, i.e., algebra, statistics, writing, researching. These materials allowed students to get answers to general questions that can provide support for most course work.

A team consisting of an instructional designer, student support staff, and academic advising experts at Marist College designed a general framework for the Online Academic Support Environment (OASE), based on research and best practices on creating online support communities. Team members then worked with partner institutions to apply this framework locally to create a customized OASE site that leveraged local support resources and met the needs of their particular student population and academic context. Some of the core design elements for the OASE are to:

- *Promote Awareness of Academic Support Services* - The site was facilitated by an institutional representative, possibly an academic advisor or support staff, who answered questions and helped direct students to campus-based or online resources provided by the institution (e.g., tutoring services, writing labs etc.). In addition to making students aware of these resources, such interactions helped students feel engaged with their institution.
- *Promote Peer-to-Peer Engagement* - The site was co-facilitated by advanced students who acted as peer mentors and provided a more experienced student perspective on issues of campus and academic life. For example, they managed an online “Student Lounge” discussion forum in which students would engage in discussions that were most relevant to them (e.g., how to deal with test anxiety). In other cases, student-developed videos on academic success issues (e.g., Cerritos College’s iFALCON program2), were made available.
- *Provide Access to Self-Assessment Tools* - Students were given access to a range of self-assessment tools, such as the Learning and Study Strategies Inventory (LASSI), to help them become more aware of their strengths and weaknesses as a learner as well as their preferred learning style. Recommendations to either seek out in-person assistance or review educational materials were provided based on the results of these assessments.
- *Provide Access to Educational Scaffolding Content* - Students were provided with a range of open content as a means to improve study skills, refresh content knowledge, or engage in skill remediation. For example, students were given access to Flat World Knowledge’s textbook, available for free online under a Creative Commons license, titled College Success[3] to better understand how

to take notes or improve their time management skills. In other cases, students were directed to open-source tutoring software available through the Carnegie Mellon Open Learning Initiative for remediation purposes. The OAAI also collaborated with other Open Educational Resources and related Next Generation Learning Challenge-funded projects to work to incorporate their content into the OASE.

These interactions were focused on two primary supports: a) assisting learners in finding resources (e.g., remedial content, tutors, academic advising services, etc.) and b) facilitating online discussions on timely topics related to student academic success. It should be noted that the goal of these online discussions will not generally be to answer specific content-related questions but rather to direct students to where they can obtain such support.

Once at-risk students had been identified, this information was used to alert them that they were at risk of failing the course. At each university, a Site Coordinator from the participating institution played a critical role as liaison between investigators, university administration, regulators, and instructors. Instructors were provided with an orientation about the study purpose, their role in the study, and how to use the alert system. All study details were disclosed with respect to privacy, data storage, the volunteer nature of the study, and consent for instructors and students. A total of 3,176 students were assigned to one of three groups (control Awareness & OASE). Each course was assigned to one of the three groups. To the extent possible, instructors were assigned to sections representing each of the three groups. At three different points during the semester (25%, 50% and 75% of the semester completed), Academic Alerts were automatically sent to the instructors. After reviewing the AAR list, instructors sent the email messages to the students they felt were struggling in their course.

Students in the Awareness group were sent emails with messages like the following:

- *Based on your performance on recent graded assignments and exams, as well as other factors that tend to predict academic success, I am becoming worried about your ability to complete this class successfully.*
- *I am reaching out to offer some assistance and to encourage you to consider taking steps to improve your performance. Doing so early in the semester will increase the likelihood of you successfully completing the class and avoid negatively impacting your academic standing.*

Additionally, Instructors were encouraged to recommend the following:

- Ask the student to visit you during office hours.
- Set up an appointment with a tutor, academic support person, or consider participating in a study group.
- Access web-based resources such as online tutoring tools.
- Take practice exams, complete additional exercises and homework questions.

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

Students in the OASE group received the same messages plus access to the resources described in the OASE Design Framework, such as academic support services like The Kahn Academy, Flat World Knowledge textbooks, as well as access to mentoring from peers and professional support staff. At the end of the semester, we collected data on a number of measures, including course grade, content mastery, and course withdrawal.

4.3 Analysis of Predictions on Pilot Data

Table 5 reports the predictive performance of the logistic regression model (the model of choice for the AARs) at each partner institution in spring 2012 and fall 2012. This evaluation included assessing the model's performance at three points during the semester (25%, 50%, and 75% of the semester completed), which corresponds to when Academic Alert Reports were provided to instructors, to evaluate how the model's performance improved as more LMS event and gradebook data became available. It is important to highlight that the model was deployed for prediction at institutions representing vastly different educational contexts as compared to Marist, both in terms of demographics and retention rates (please refer to Table 4 for details).

Looking at the predictive performance of the model in each of the AARs, using Recall as an indicator (percentage of at-risk students that were identified), it is clear that the results are considerably higher than random chance. In three out of four partner institutions (Savannah, Cerritos, Redwoods) the average Recall across all AARs was approximately 74.5%, with highs of 84.5% (Redwoods, AAR2) and lows of 61% (Cerritos, AAR1). If we restrict the analysis to the AARs generated in fall 2012, the results are even better: an average of 75.5%. When comparing these values to the predictive performance of the model tested with Marist data (Table 2), we find only a 10% difference on the average. Given that we expected a much larger difference between how the model performed when tested with Marist data and when deployed at community colleges and HBCUs, this was a surprising and encouraging finding. It should be noted, however, that NCAT Recall values are way below the aforementioned scores (an average of 52%), a fact that deserves further consideration.

These findings seem to indicate that predictive models developed based on data from one institution may be scalable to other institutions, even those that are different regarding institutional type, student population, and instructional practices. We believe this very interesting finding may be the result of the specific elements of the models that have shown to be the most powerful predictors. The attributes that are most predictive of student outcomes are cumulative GPA and the aggregated score (RMN_SCORE_PARTIAL) summarizing partial contributions to the final grade, as reported by the LMS gradebook. Given that these two attributes are such fundamental aspects of academic success, it is not surprising that the predictive model has fared so well across these different institutions. If this explanation is correct, it does point to the importance of instructors using the gradebook within their LMS if they wish to take advantage of learning analytics. Although grades and cumulative GPA are well documented predictors in the extant literature (see Dziuban and Moskal (2011) for a recent reference),

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

and were also identified by Purdue's Signals project (Arnold, 2010), there is no documented precedent of the use of partial contributions to the student's final grade extracted from the LMS gradebook tool to determine academic risk in the early stages of the semester. Our research further corroborates earlier studies with respect to the relevance of good grades as predictors of future academic performance but, more importantly, it points to the use of this data-driven approach to allow the instructor to make predictions of student performance much earlier in the semester (2–3 weeks into the course), compared to what the instructor might be able to do through visual inspection (typically after the midterm). It also indicates that models may not import well into institutions where partial contributions to the final grade and/or cumulative GPA are not available (e.g., non-credit training programs).

Table 5: Predictive Performance on spring 2012 and fall 2012 Pilot Data

	College	AAR run	# Students	Accuracy	FP Rate	Precision	Recall
Spring 2012	Savannah	AAR1	504	67.26%	35.36%	61.48%	70.54%
		AAR2	504	74.40%	32.50%	67.15%	83.04%
		AAR3	504	79.37%	18.21%	77.03%	76.34%
	Cerritos	AAR1	502	61.95%	43.69%	47.41%	72.32%
		AAR2	601	71.88%	27.49%	59.62%	70.78%
		AAR3	649	75.19%	25.12%	62.50%	75.76%
	Redwoods	AAR1	195	67.69%	40.48%	52.78%	82.61%
		AAR2	195	78.97%	13.49%	72.58%	65.22%
		AAR3	195	77.95%	14.29%	70.97%	63.77%

	College	AAR run	# Students	Accuracy	FP Rate	Precision	Recall
Fall 2012	Savannah	AAR1	425	68.47%	38.34%	58.19%	78.49%
		AAR2	425	72.59%	30.04%	65.17%	76.16%
		AAR3	425	73.41%	26.88%	65.13%	73.84%
	Cerritos	AAR1	465	65.38%	32.35%	49.49%	61.01%
		AAR2	465	70.75%	27.78%	55.96%	67.92%
		AAR3	465	73.98%	24.51%	60.11%	71.07%
	Redwoods	AAR1	182	83.63%	16.52%	71.21%	83.93%
		AAR2	182	83.82%	16.52%	72.06%	84.48%
		AAR3	182	85.63%	13.04%	76.56%	83.05%
	NCAT	AAR1	719	64.12%	31.25%	26.53%	45.45%
		AAR2	719	71.07%	24.83%	35.29%	54.55%
		AAR3	719	75.10%	20.14%	40.82%	55.94%

A number of issues require further study. We found variability in predictive performance across institutions and in pilot runs in different semesters (spring 2012 vs. fall 2012). Fall 2012 outcomes at NCAT were rather poor, with Recall values in the 45–56% range. In addition, we noticed that the average false positive rate at partner institutions (percentage of false alarms) was larger than the average value obtained when testing the model with Marist College data (an average of 26%, with highs of 43% for the

pilots versus roughly 13% for Marist College). A possible explanation can be found by considering the difference in retention rates between institutions: the model trained with Marist College data was fine-tuned to detect at-risk students (only 7% of the student population at Marist College). It could be inferred that such model, applied on a student population where the proportion of at-risk students is much higher would raise a higher rate of false alarms. Thus, although our findings are encouraging regarding portability, important questions remain regarding scaling up models across academic settings that are more diverse. Portability values were higher than expected, but when data is available, it is reasonable to assume that models with better predictive power can be learnt using training data from the same institution.

4.4 Analysis of Intervention

4.4.1 Impact on At-Risk Students Academic Success

The study described above was conducted over two semesters in the spring and fall of 2012.[28] The assessment conducted in the spring included three institutions: Cerritos Community College, College of the Redwoods, and Savannah State University. The study conducted in the fall included the previously mentioned institutions as well as North Carolina Agricultural and Technical State University. The two treatment groups ("awareness messaging" and OASE) were comprised of students who had received at least one intervention based on any of the three Academic Alert Reports provided during the course of the semester. To identify control subjects, which by definition did not receive interventions, we selected those students who had been identified as having an average "risk level" of three or higher across all three Academic Alert Reports. Students were categorized into academic risk categories based on the predictive model's ability to generate failure probability scores (likelihood of not completing the course successfully) as a part of prediction output. These probability values were classified into four ranges: 1) no risk: probability range of 0%–50%; 2) low risk: probability range of 50%–75%; 3) medium risk: probability range of 75%–90%; and 4) high risk: probability of 90% and above.

In the spring of 2012, 1,739 students were enrolled in the OAAI study. Four-hundred and fifty-one of these students were identified as being at-risk. Participating students were then divided into one of three groups (Awareness: n = 193, *M* = 77.47, *SEM* = 0.97; OASE: n = 179, *M* = 77.5, *SEM* = 1.05; control group: n = 79, *M* = 75.17, *SEM* = 1.32). A one-way ANOVA was conducted revealing a significant difference between groups (*F* (2,448), 8.484, *p* = .000*, see Figure 6). Post-hoc analysis showed no differences between the two treatment groups; however, there were statistically significant differences between control and Awareness (*p* = .000*) and control and the OASE group (*p* = .000*).

A similar one-way ANOVA was conducted collapsing across 696 students identified as at-risk in the spring and fall semesters. Students were assigned to one of the three experimental groups (Awareness: n = 277, *M* = 70.81, *SEM* = 0.77; OASE: n = 254, *M* = 71.14, *SEM* = 0.83; control group: n = 165, *M* =

[28] Preliminary findings using spring 2012 data were reported in Lauría, Moody, Jayaprakash, Jonnalagadda, and Baron (2013).

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

66.33, *SEM* = 0.96). Again, a significant difference was found (*F* (2,693) = 8.025, *p* = .000*, see Figure 7). Post-hoc analysis once again showed no differences between the two treatment groups, but confirmed a statistically significant difference between control and Awareness (*p* = .002*) and control and the OASE group (*p* = .002*). In a course with a final grade range of 75, students in the two treatment groups have shown a 6% improvement over non-intervention controls.

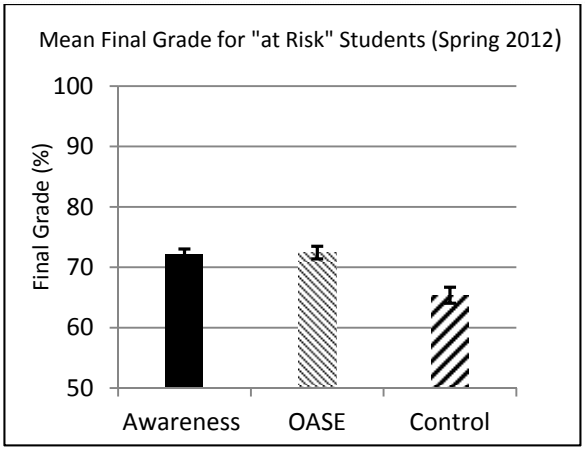

Figure 6: Impact on general student academic success, spring 2012 data (error bars represent SEM)

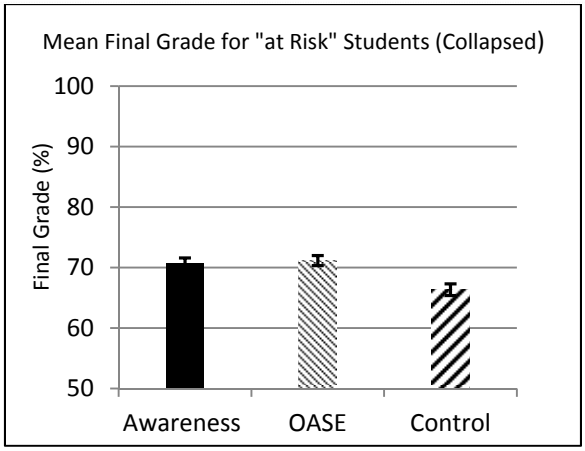

Figure 7: Impact on general student academic success, spring and fall 2012 data collapsed (error bars represent SEM)

4.4.2 Impact on the Academic Success of At-Risk Students Receiving Pell Grants

The Pell Grant is awarded to students who can demonstrate an "exceptional" financial need. Pell Grant status is considered a reliable predictor of a student's socio-economic status. There is considerable evidence showing that students with lower socio-economic status have lower GPAs and graduation rates (Stinebrickner & Stinebrickner, 2003; Griffith, 2008; Day, Dworsky, Fogarity, & Damashek, 2011). In an effort to isolate at-risk students identified as having a lower socio-economic status, we further refined the groups from the previous ANOVA analysis to include only students awarded Pell Grants.

In the spring of 2012, 326 students were identified as being "at-risk" and had also been awarded a Pell grant. These students were then divided into one of three groups (Awareness: n = 138, *M* = 71.52, *SEM* = 1.14; OASE: n = 132, *M* = 71.74, *SEM* = 1.22; control group: n = 57, *M* = 63.77, *SEM* = 1.35). A one-way ANOVA was conducted revealing a significant difference between groups (*F* (2,324), 8.35, *p* = .000*, see Figure 8). Post-hoc analysis showed no differences between the two treatment groups; however, there were statistically significant differences between control and Awareness (*p* = .001*) and control and the OASE group (*p* = .000*).

A similar one-way ANOVA was conducted collapsing across 499 students identified as at-risk and receiving Pell grants in the spring and fall semesters. These students were assigned to one of three groups (Awareness: n = 200, *M* = 70.10, *SEM* = 0.92; OASE: n = 187, *M* = 70.08, *SEM* = 0.98; control

group: n = 112, *M* = 65.45, *SEM* = 1.10). A significant difference was found (*F* (2,693) = 8.025, *p* = .000*), see Figure 9). Post-hoc analysis once again showed no difference between the two treatment groups, but confirmed a statistically significant difference between control and Awareness (*p* = .002*) and control and the OASE group (*p* = .002*).

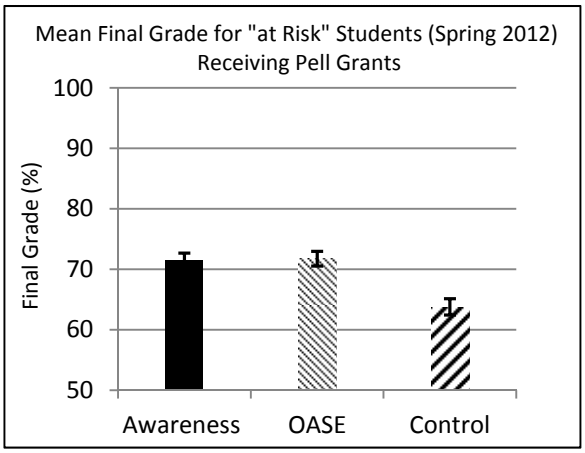

Figure 8: Impact on academic success of "at Risk" students receiving Pell grants, spring 2012 data (error bars represent SEM)

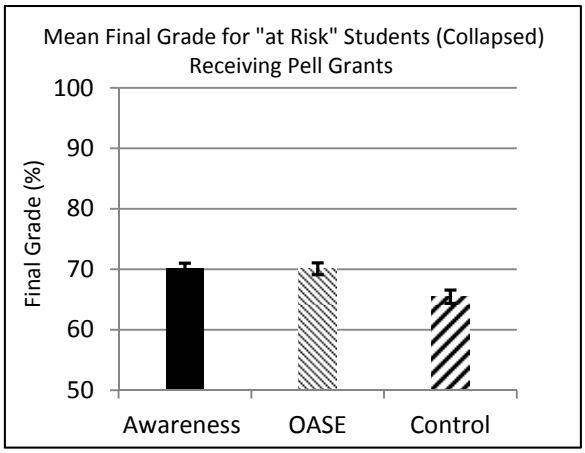

Figure 9: Impact on academic success of "at Risk" students receiving Pell grants, spring and fall 2012 data collapsed (error bars represent SEM)

4.4.3 Impact on Withdrawal Rates

Among other outcome measures, we examined withdrawal rates. As with all previous comparisons, no differences were seen between the two treatment groups (Awareness and OASE). These two groups were collapsed for additional analysis. Chi-square analysis was conducted on at-risk students from the spring and fall semesters independently and on the two semesters collapsed. The spring data showed that 20.9% of students withdrew in the intervention group, whereas only 13% of control subjects withdrew. When spring data was analyzed alone, a significant difference was not found; however, there was a trend indicating potential differences between the control and treatment groups (χ^2 (1) = 3.108, *p* = .079, see Figure 10).

The fall data showed that 25.6% of students withdrew in the intervention group, whereas only 14.1% of control subjects withdrew. This semester, a significant difference was found indicating higher rates of withdrawal among treatment subjects (χ^2 (1) = 11.044, *p* = .079, see Figure 11). When the data from both semesters was collapsed, 25.6% of intervention students withdrew while only 14.1% of control students withdrew. Chi-square analysis once again found that students in the treatment groups had proportionally higher rates of withdrawal than control subjects among those who had been identified as at-risk of failure (χ^2 (1) = 14.611, *p* = .000*, see Figure 12).

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

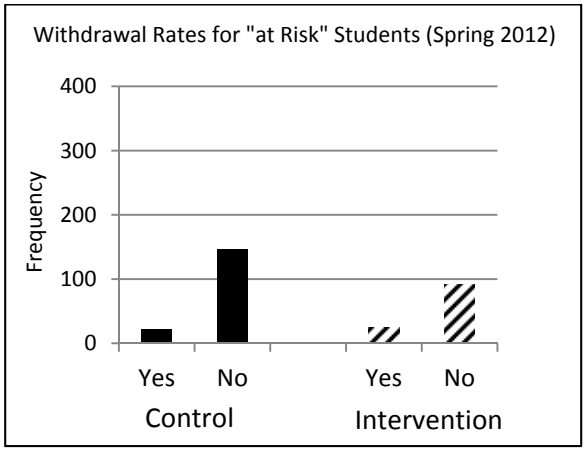

Figure 10. Impact on withdrawal rates in at-risk students, spring 2012 data

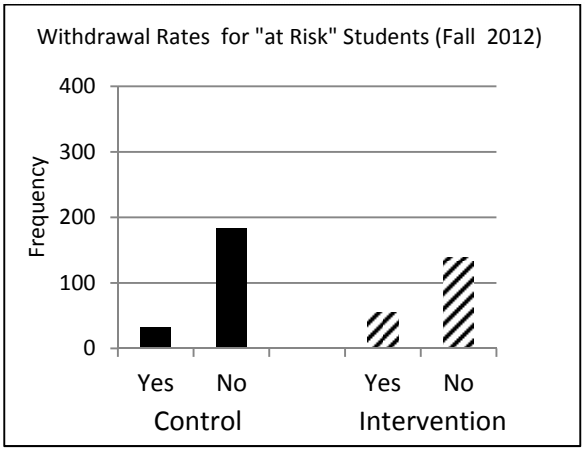

Figure 11. Impact on withdrawal rates in at-risk students, fall 2012 data

Early research on the Purdue Signals project found an increase in withdrawals early in the course, which stabilized, and was not found to be significant (Arnold, 2010). More recent research on Signals has found that withdrawal rates increased in sections of Agronomy and Psychology but decreased in sections of Statistics (Pistilli, Arnold, & Bethune, 2012). The difference in withdrawal rates might be explained by some students who have chosen to withdraw soon in the course, rather than attempting to complete and failing. The mixed results suggested we may see a change in withdrawal rates but the direction was unclear. The data collected in the spring indicated that withdrawal rates were higher in the treatment groups, but not significantly. Fall data, however, indicated that withdrawal rates in the treatment groups were significantly higher. The data collected in both semesters were higher in the treatment groups; however, these results were inconsistent, as seen in the Purdue research (Arnold, 2010; Pistilli, et al., 2012).

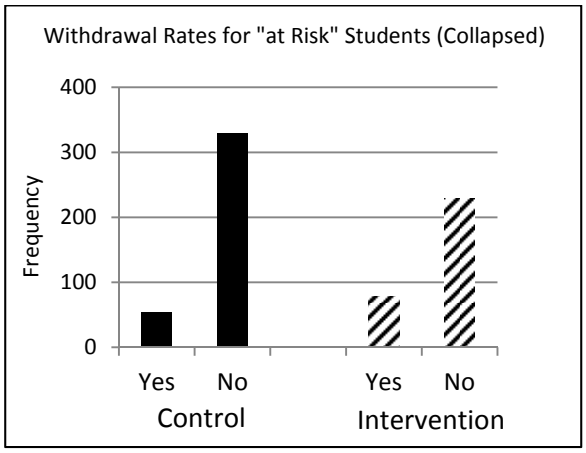

Figure 12. Impact on withdrawal rates for at-risk students, spring and fall 2012 data collapsed

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

4.4.4 A Discussion on the Ethics of the Study on Intervention

Many ethical issues we addressed when designing the study were typical of any design that involves the ongoing participation of human subjects. The primary ethical concerns at the outset of the study were of privacy and consent. Participation in the study, which was fully described to the students, was completely voluntary; students were informed they were free to discontinue participation at any point in time. Great efforts were taken to ensure that confidentiality could not be breached, as unique identifiers were generated for each subject by the participating institution. No study personnel had access to personal identifiers. Slade and Prinsloo (2013) have addressed the issue of ethical issues relating to learning analytics, specifically such complex issues as motivation and access.

One issue we encountered after the data had been collected arose when it was observed that the two intervention groups had higher rates of withdrawal than the control group. It must be recognized that any effort to identify an at-risk student will result in some amount of error. In some cases, at-risk students will not be identified by the model. These students will not be offered an intervention that may have been beneficial to them. In other cases, the model will identify students who, in fact, are not at risk of failure. Some of these students may choose to withdraw in fear that they may not pass the course. This is the inevitable type-one vs. type-two error quandary encountered by any attempt to provide an intervention to a segment of the population (Singell & Waddell, 2010). This issue forces us to develop the most accurate predictive models possible, as well as to take steps to reduce the likelihood that any intervention would result in the unnecessary withdrawal of a student. In an effort to avoid unnecessary withdrawal, we selected an intervention that was simple (email), could be easily customized, and most importantly required review by the instructor prior to sending. In recognition of the potential misidentification of an at-risk student, we left the decision to send a warning email with the instructor.

One instance in which a withdrawal might be considered a positive outcome is if a student is unable to improve his or her grade sufficiently for a myriad of reasons (over-scheduled, illness, overwhelmed). If students are made aware of their likely failure then they may be able to withdraw early enough to avoid a negative impact on their transcript. For example, at Marist College withdrawal in the first half of the semester results in a W (withdrawal) whereas students who withdraw in the second half of the semester receive an F. Due to inconsistencies in how withdrawal data was reported by the participating institutions, information about the timing of withdrawal is incomplete and therefore unreliable. A preliminary analysis of withdrawal timing, excluding questionable data, found no differences in both the control and intervention groups in the first half of the semester relative to withdrawal rates in the second half of the semester. Beyond this observation, it is difficult to conclude much about the effects of the interventions on the timing of withdrawal behaviour. What is clear, is that the issue of withdrawal timing is important, and needs to be investigated explicitly.

5 CONCLUSION AND FUTURE RESEARCH

This paper reports on the research findings of the Open Academic Analytics Initiative, which we believe

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

contributes to our collective understanding of the issues related to scaling of learning analytics across all of higher education. Specifically, our research shows:

a) The feasibility of implementing an open-source early-alert prototype for higher education, and provides a detailed account of the challenges and design criteria used in implementing such a system.
b) The strength of scores derived from partial contributions to the student's final grade as predictors of academic performance.
c) How these predictive models can help the instructor detect students at academic risk earlier in the semester.
d) Initial evidence that predictive models can be imported from the academic context in which they were developed to different academic contexts while retaining most of their predictive power.
e) That there may be benefits associated with customizing imported predictive models using local institutional data as a means to enhance their predictive power further.
f) That relatively simple intervention strategies designed to alert students early in a course that they may be at risk academically can positively impact student learning outcomes such as overall course grades.
g) That there are no apparent gains between providing students with an online academic support environment and simply making students aware of their potential academic risk.
h) That interventions can have unintended consequences, such as triggering students to withdraw from courses, often early in the semester, as means to avoid academic and financial penalties.

Predictive models were trained and tested using Marist College data; those models were then applied on pilot runs using data from several partner institutions. The research tested the portability of those models, and the success of intervention strategies in improving at-risk student outcomes. The results are promising, as they seem to point to a higher portability of learning analytics models than initially anticipated. These results had a subsequent positive impact on the effectiveness of interventions on students at academic risk. We hope that these results will encourage researchers from other institutions to develop similar strategies of early detection of and intervention in academic risk.

Based on our work to date, our research team has begun to discuss and identify areas of research that we believe will be important to the field of learning analytics as it begins to be deployed more widely. These research questions, outlined below, could form the basis for a national research agenda in this new and emerging field.

What is the importance of an early alert and how does learning analytics facilitate early alerts?
The importance of timeliness has been identified as critical to the success of any intervention intended to change the trajectory of an at-risk student (Kim, Newton, Downey, & Benton, 2010). If a student becomes aware of their risk of failure after too many grades have been recorded, the likelihood that any change in effort will lead to a grade change decreases. Many students only start to consider that they

(2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

might be in trouble after a poor mid-term grade, and for many this is simply too late for a measurable recovery. Learning analytics can provide an alert within the first five weeks of a course, giving both the instructor and student time to address the issue before too many grades have been recorded.

Does learning analytics allow us to identify students who might not complete a course that the typical instructor would miss?
Learning analytics solutions are often seen as ways to improve student success in large lecture-style courses (100+ students) in which it can be very difficult for an instructor to identify, early on in the course, which students may not succeed. As learning analytics is deployed at institutions with smaller class sizes, as was the case with many of the OAAI course pilots, it will be important to understand the "value added" of learning analytics over what an instructor is capable of doing on his or her own. Although many of the instructors in our pilots noted that they found the identification of at-risk students very helpful, it remains unclear if they would have identified the same students that our model identified had they attempted to do so on their own. Thus, we believe it will be important to conduct further studies in which instructor predictions are compared to model predictions.

What are the characteristics of students who seem to have "immunity" to the treatment (those who received interventions but never improved) versus those who were effectively treated after just one intervention?
From our initial research, we have found that students seem to fall into one of two broad categories: those who improve after receiving just one "treatment" or intervention and those who do not improve regardless of the number of "treatments" received. Very few students who did not improve after the first intervention went on to improve after the second or third. Our theory is that some students respond very well to the "treatment" and thus improve after just one intervention while other seem "immune" to the "treatment" and do not improve regardless of how many treatments they receive. Understanding why this is the case and what characteristics are associated with these two categories of students would help us to understand better how to deploy interventions most effectively.

How portable are predictive models designed for one type of course delivery (e.g., face-to-face) when they are deployed in another delivery format (e.g., fully online)?
We are particularly interested in exploring the issue of portability regarding face-to-face and fully online programs given how much more LMS usage takes place in the later mode of instruction. It may be that models developed based on face-to-face courses do not import well to fully online courses or at least that such models could be significantly improved if they were customized for fully online courses.

ACKNOWLEDGEMENTS

This research is supported by EDUCAUSE's Next Generation Learning Challenges, funded through the Bill & Melinda Gates Foundation and The William and Flora Hewlett Foundation. It is also partially supported by funding from the National Science Foundation, award numbers 1125520 and 0963365. The

authors would like to thank Nicole Maziarz for proofreading the document.

REFERENCES

Arnold, K. (2010). Signals: Applying academic analytics. *EDUCAUSE Review,* http://www.educause.edu/ero/article/signals-applying-academic-analytics

Arnold, K.E., & Pistilli, M.D. (2012). Course Signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK'12),* Vancouver, Canada, 29 April–2 May. New York, NY: Association of Computer Machinery, 267–270. doi:10.1145/2330601.2330666

Astin, A.W. (1993). *What matters in college? Four critical years revisited*. San Francisco: Jossey-Bass.

Astin, A.W. (1999). Student involvement: A developmental theory for higher education. *Journal of College Student Development, 40*(5), 418–429.

Barber, R., & Sharkey, M. (2012). Course correction: Using analytics to predict course success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK'12),* Vancouver, Canada, 29 April–2 May, New York, NY: Association of Computer Machinery, 259–262. doi:10.1145/2330601.2330664

Bevitt, D., Baldwin, C., & Calvert, J. (2010). Intervening early: Attendance and performance monitoring as a trigger for first year support in the biosciences. *Bioscience Education E-journal, 15*, doi:http://journals.heacademy.ac.uk/doi/abs/10.11120/beej.2012.20000053

Bravo, J., Sosnovsky, S., & Ortigosa, A. (2009). Detecting symptoms of low performance using prediction rules. *Proceedings of the 2nd Educational Data Mining Conference (EDM'09),* Universidad de Cordoba, Cordoba, Spain, 1–3 July, 31-40.

Burges, C.J.C. (1998). A tutorial on support vector machines for pattern recognition: Data mining and knowledge discovery 2, 121–167. http://research.microsoft.com/en-us/um/people/cburges/papers/svmtutorial.pdf

Campbell, J.P. (2007*).* Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study. (Doctoral dissertation, Purdue University). (UMI No. 3287222).

Campbell, J., deBlois, P., & Oblinger, D. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review* (July/August), 41–57. http://net.educause.edu/ir/library/pdf/erm0742.pdf

Chen, G., Liu, C., Ou, K., & Liu, B. (2000). Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology. *Journal of Educational Computing Research, 23*(3), 305–332. doi:10.2190/5JNM-B6HP-YC58-PM5Y

Chickering, A.W., & Ehrmann, S.C. (1996). Implementing the seven principles: Technology as lever. *American Association for Higher Education and Accreditation Bulletin, 49*, 3–6.

Colby, J. (2004). Attendance and attainment. *Fifth Annual Conference of the Information and Computer Sciences: Learning and Teaching Support Network (ICS-LTSN)*, 31 August–2 September, University of Ulster. doi:www.ics.heacademy.ac.uk/italics/Vol4-2/ITALIX.pdf

College Board Advocacy & Policy Center (2010). The college completion agenda 2010 progress report,

Executive summary. Retrieved February 17, 2011 from http://completionagenda.collegeboard.org/sites/default/files/reports_pdf/progress_executive_summary.pdf

Cuseo, J. (n.d.) Academic advisement and student retention: Empirical connections & systemic interventions (Marymount College). Retrieved February 13, 2011 from https://apps.uwc.edu/administration/academicaffairs/esfy/CuseoCollection/Academic%20Advisement%20and%20Student%20Retention.doc

Day, A., Dworsky, A., Fogarity, K., & Damashek, A. (2011). An examination of post-secondary retention and graduation among foster care youth enrolled in a four-year university. *Children and Youth Services Review, 33*(11), 2335–2341. http://EconPapers.repec.org/RePEc:eee:cysrev:v:33:y:2011:i:11:p:2335-2341ral

Drummond, C., & Holte, R. (2003). "C4.5, class imbalance and cost sensitivity: Why under-sampling beats over-sampling," *Proceedings of Workshop on Learning Imbalanced Datasets II, ICML*, Washington DC, 21 August, 1–8.

Duda, R.O., Hart, P.E., Stork, D.G. (2001). *Pattern classification*, 2nd ed. New York, NY: John Wiley & Sons.

Dziuban, C., & Moskal, P. (2011). "A course is a course is a course: Factor invariance in student evaluation of online, blended and face-to-face learning environments." *The Internet and Higher Education*, doi:10.1016/j.iheduc.2011.05.003

Fawcett, T. (2006). "An introduction to ROC analysis," *Pattern Recognition Letters, 27*, 861–874.

Fisher, C., Lauría, E., Chengalur-Smith, S., & Wang, R. (2006). *Introduction to information quality.* Bloomington, IN: AuthorHouse.

Folger, W., Carter, J.A., & Chase, P.B. (2004). "Supporting first generation college freshmen with small group intervention." *College Student Journal, 38*(3), 472–476.

Freitas, A.A. (2002). *Data mining and knowledge discovery with evolutionary algorithms*. New York, NY: Springer.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning, 29,* 131–163.

Fritz, J. (2011). Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers. *Internet and Higher Education, 14*(2), 89–97.

Geman, E., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation, 4*, 1–58. http://stuff.mit.edu/afs/athena.mit.edu/course/6/6.435/OldFiles/www/Geman92.pdf

Griffith, A.L. (2008). Determination of grades, persistence and major choice for low-income and minority students. Working Paper #110. Cornell Higher Education Research Institute. http://www.ilr.cornell.edu/cheri/workingPapers/2008.html

Kim, E., Newton, F.B., Downey, R.G., & Benton, S.L. (2010). Personal factors impacting college student success: Constructing College Learning Effectiveness Inventory (CLEI). *College Student Journal, 44*(1), 112–125. http://www.sunyrockland.edu/Members/xshi/assigned-reading-materials-1/files/personal-factors-impacting-college-student-success.pdfEducationAL

Larose, D.T. (2006). *Data mining methods and models.* Hoboken, NJ: Wiley. doi:10.1002/0471756482

Lauría, E., Baron, J., Devireddy, M., Sundararaju V., & Jayaprakash, S. (2012*)*. Mining academic data to improve college retention: An open source perspective. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK'12),* Vancouver, Canada, 29 April–2 May. New York, NY: Association of Computer Machinery, 139–142. doi:10.1145/2330601.2330637

Lauría, E., Moody, E., Jayaprakash, S., Jonnalagadda, N., & Baron, J. (2013). Open Academic Analytics Initiative: Initial research findings. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (LAK'13),* Leuven, Belgium, 8–12 April. New York, NY: Association of Computer Machinery, 150–154. doi:10.1145/2460296.2460325

Lauría, E., Tayi, G.K. (2003). A comparative study of data mining algorithms for network intrusion detection in the presence of poor quality data. *Proceedings of the 8th International Conference on Information Quality*, Cambridge, Massachusetts, USA, 7–9 November, 190–201.

Laurie, P.D., & Timothy, E. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education, 45*(1), 141–160.

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5). http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education

Ma, Y., Liu, B., Wong, C., Yu, P., & Lee, S. (2000). Targeting the right students using data mining. *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining (KDD 2000)*, Boston, Massachusetts, USA, 20–23 August, 457–464. ftp://ftp.cse.buffalo.edu/users/azhang/disc/disc01/cd1/out/papers/kdd/p457-ma.pdf

Minaei-Bidgoli, B., & Punch, W. (2003). Using genetic algorithms for data mining optimization in an educational web-based system. *Proceedings of Genetic and Evolutionary Computational Conference,* Chicago, Illinois, USA, 12–16 July, 2252–2263.

Mitchell, T.M. (1980). The need for biases in learning generalizations. (Report CBM-TR 5-110). New Brunswick, NJ: Rutgers University Department of Computer Science.

Mitchell, T. (2005). Generative and discriminative classifiers: Naïve Bayes and logistic regression.. http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf

Morris, L.V., Wu, S., & Finnegan, C. (2005). Predicting retention in online general education courses. *The American Journal of Distance Education, 19*(1), 23–36. doi:10.1016/j.iheduc.2005.06.009

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied linear regression models,* 3rd ed. Chicago, IL: Irwin.

Newman-Ford, L.E., Fitzgibbon, K., Lloyd, S., & Thomas, S.L. (2008). A large-scale investigation into the relationship between attendance and attainment: A study using an innovative, electronic attendance monitoring system. *Studies in Higher Education, 33*(6), 699–717.

Pistilli, M.D., & Arnold, K.E. (2010). Purdue Signals: Mining real-time academic data to enhance student success. *About Campus, 15*, 22–24. doi:10.1002/abc.20025

Pistilli, M.D., Arnold, K.E., & Bethune, M. (2012). Signals: Using academic analytics to promote student success. *EDUCAUSE Review*, July/Aug 2012, online.

Platt, J.C. (1999). Using analytic QP and sparseness to speed training of support vector machines. In M.S.

Kearns, S.A. Solla, D.A. Cohn (eds.), Advances in neural information processing systems, vol. 11. Cambridge, MA: MIT Press.

Quinlan, J.R. (1993). C4.5: Programs for machine learning. San Mateo, CA: Kaufman.

Rish, I. (2001). An empirical study of the Naïve Bayes classifier. *Proceedings of Workshop on Empirical Methods in Artificial Intelligence*, Seattle, Washington, USA, 4–10 August, 41–46.

Romero, C., Ventura, S., & Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education, 51*(1), 368–384.

Siemens, G., & Baker, S.J. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK'12),* Vancouver, Canada, 29 April–2 May. New York, NY: Association of Computer Machinery, 252–254. doi:10.1145/2330601.2330661

Singell, D.L., & Waddell, G.R. (2010). Modeling retention at a large public university: Can at-risk students be identified early enough to treat? *Research in Higher Education, 51*(6), 546–572.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1509–1528.

Smith, E.M., & Beggs, B.J. (2003). A new paradigm for maximizing student retention in higher education. *IEE Engineering Education Conference,* Southampton, UK, 6–7 January. doi:www.ulster.ac.uk/star/resources/paradigm.pdf

Stinebrickner, R., & Stinebrickner, T.R. (2003). Understanding educational outcomes of students from low-income families: Evidence from a liberal arts college with a full tuition subsidy program. *The Journal of Human Resources, 38*(3), 591–617. doi:10.3368/jhr.XXXVIII.3.591

Tinto, V. (1982). Limits of theory and practice in student attrition. *The Journal of Higher Education*, *53*(6), 687–700.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition.* Chicago: University of Chicago Press.

Tinto, V. (2007). Research and practice of student retention: What next? *Journal of College Student Retention, 8*(1), 1–19.

Tinto, V. (2012). Enhancing student success: Taking the classroom success seriously. *The International Journal of the First Year in Higher Education, 3*(1), 1–8.

U.S. Department of Education, National Center for Educational Statistics, Postsecondary Education Data System. (2009). Graduation rates of first time postsecondary students...1996 through 2004. Retrieved February 15, 2011 from http://nces.ed.gov/programs/digest/d09/tables/dt09_331.asp

U.S. Department of Education, National Center for Education, Integrated Postsecondary Education Data System. (2010). Retrieved February 15, 2011 from http://nces.ed.gov/collegenavigator/

van Barneveld, A., Arnold, K.E., & Campbell, J.P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative*. http://educause.edu/ir/library/pdf/ELI3026.pdf

Vapnik, V. (1995). *The nature of statistical learning theory*. New York, NY: Springer-Verlag.

Yu, P., Own, C., & Lin, L. (2001). On learning behavior analysis of web based interactive environment.

Proceedings of ICCEE, Oslo/Bergen, Norway.

Zaïane, O., & Luo, J. (2001). Web usage mining for a better web-based learning environment. *Proceedings of Conference on Advanced Technology for Education* Banff, Alberta, Canada, 60–64.

Zhang, H. (2004). The optimality of Naïve Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS),* 562–567. http://www.aaai.org/Library/FLAIRS/2004/flairs04-097.php

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results

Monika Andergassen, Felix Mödritscher, and Gustaf Neumann
Information Systems and New Media
Vienna University of Economics and Business (WU)
Austria
monika.andergassen@wu.ac.at

ABSTRACT: Learner-centric research on factors influencing learning results has focused, among other things, on student characteristics, demographic data, and usage patterns in learning management systems (LMSs). This paper complements the existing research by investigating potential correlations between learning results and LMS usage during exam preparation, focusing on practice and repetition. Based on 250 million log-file entries used to analyze student interactions within specific courses and overall in the LMS, results show positive, albeit modest, correlations between usage variables and final exam grades. Regarding practice, the number of learning days and the number of days between the first and the last learning sessions correlate better than the coverage of different learning materials. The findings for repetition indicate that it is more beneficial to transfer learning to new tasks than to repeat the same items many times. The study not only looks at single usage variables but also examines the distribution of the descriptive and dependent variables and uses visualization techniques and quantiles to deal with outliers. This paper describes the largest empirical study of learner interactions in blended learning courses conducted so far (at least according to the authors' knowledge) and including techniques for processing and analyzing large datasets about LMS usage.

KEYWORDS: Learning analytics, web usage mining, learning management systems, blended learning courses, practice, repetition, correlation analysis, regression analysis

1 INTRODUCTION

In the context of higher education, learning management systems (LMSs) are particularly important for coping with the phenomenon of mass education (Johnson et al., 2012). Consequently, LMS platforms that provide online and blended learning courses collect large amounts of interaction data from students (cf. Ferguson, 2012). Learning Analytics (LA) attempts to exploit user-generated data through Business Intelligence techniques in order to support different stakeholders, from students and teachers to LMS developers and service providers, and to predict learning performance in educational settings (Siemens et al., 2011). Additionally, large sets of learning-related data can be of interest for research purposes, i.e., to understand how students learn and to advance an LMS.

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

Several universities in Austria teach massive cohorts of students in the first semesters of their study programs. At the Vienna University of Economics and Business (WU), for instance, some courses in the Bachelor study programs have 1,000 students or more per class. As a result, the LMS has become an essential instrument for managing large groups of students and for providing multiple-choice tests to assess student performance and learning outcomes. This paper examines correlations between learning results and LMS usage variables in relation to the practice and repetition of course material. It draws on previous research focusing on seasonal effects in LMS usage data (Andergassen, Neumann, & Mödritscher, 2013) and dependencies between LMS usage in terms of page hits, user sessions and learning results (Mödritscher, Neumann, & Andergassen, 2013). In order to do this, the log-files of the Learn@WU LMS are analyzed and compared with other data sources, namely, the final exam grades achieved by students of three selected courses. Thanks to the considerable number of users and the intense use of the LMS, large datasets were available for this study.

We proceed in the following way. First, we examine the theoretical foundations of LMS usage, with a particular focus on practice and repetition. We also indicate the limitations of the existing research. Next, we point out the aims and objectives of our research, and describe the research methodology for the empirical study, in which we apply Web Usage Mining techniques and statistics to show dependencies between LMS usage variables and learning results. The “Results and Discussion” section summarizes the most relevant findings gained from this study. The last section concludes the paper and gives an outlook for future research.

2 FOUNDATIONS OF RESEARCH ON LMS USAGE AND LIMITATIONS

2.1 LMS Usage in Learning Analytics Research

In recent years, research has first proposed and then pushed the development of Learning Analytics, which seems to reveal a promising insight into students’ technology usage behaviour within educational processes. Furthermore, it also provides the basis for improving the situation for various stakeholders, from learners and teachers to service providers, developers, and organizations (Siemens et al., 2011). The latter, in particular, have started to investigate analyzing LMS usage systematically because of data-driven research.

According to Chatti, Dyckhoff, Schroeder, and Thüs (2012), research in the field of Learning Analytics principally focuses on the adaptation of learning environments (40%), monitoring and analysis (33%), assessment and feedback (13%) and the prediction of student performance (12%). Other objectives, such as reflection or supporting competence development are (still) underrepresented. For LMS technology, LA seems to be useful for identifying didactic or technical flaws (e.g., inadequate or incorrect task assignments in courses or usability problems with the system); predicting trends (e.g., the learning outcomes in connection with different factors); detecting patterns in LMS usage data (e.g., at-

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

risk vs. highly gifted students); and supporting learning and competence development (e.g., by regularly providing performance indicators and visual elements).

Research so far has also included topics such as course persistence; for instance, correlations between demographic data and course persistence (Doherty, 2006). Furthermore, cumulations of activities throughout semester tertiles has served to identify different types of persistence in a course, such as "low-extent-users," "late users" and "online-quitters" (Hershkovitz & Nachmias, 2011). Similarly, changes in LMS activity have been used to identify students at risk of course attrition (Wolff, Zdrahal, Nikolov, & Pantucek, 2013). Whitmer's (2012) study of learning results investigates links between LMS usage variables and student characteristics, i.e., the demographic data of the highly diverse student population and their grades. Other approaches include the comparison of web users and mobile users according to their usage behaviour within the LMS, or the identification of seasonal effects by analyzing activity patterns in log-files (Andergassen et al., 2013). With regard to didactic design, research has considered topics such as the impact of online discussions on learning (Khan, Clear, & Sajadi, 2012; Wise, Zhao, & Hausknecht, 2013). Early research results by Mödritscher, Neumann, et al. (2013), whose study explored correlations between the number of different learning days and the overall learning time, indicate that practice and repetition play an important role in determining final exam performance.

2.2 Practice and Repetition

The analysis of LMS usage highly depends on the didactical model implemented in the online and blended learning courses (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Mödritscher, Andergassen, Law, & García-Barrios, 2013). Courses with LMS platforms are often used to distribute learning materials and to provide self-assessment tests. Consequently, it is possible to track students' repetition and practicing of the course contents within the LMS itself through LA approaches.

Once information has been received by the working memory, practice is important in order to establish this data in the long-term memory (Willingham, 2004). In his meta-study, Cotton (1989) examined practice in relation to time, comparing time factors with achievements. She investigated differences between the time for learning allocated by teachers and the real times of student engagement, including dead time as well as learning times above and below the relevant experience levels. Her findings were that the allocated time showed a slightly positive relationship with learning results (grades), while the time-on-task had a positive relationship and the times of learning a strong one.

Another large body of research reports on positive correlations between spacing effects and learning (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Dunlosky et al., 2013; Thalheimer, 2006; Vlach & Sandhofer, 2012; Wells & Hagman, 1989). Thalheimer (2006) reviews research on spacing effects and related learning factors and finds, among other things, that repetition is effective for learning, spaced repetition is generally more efficient than non-spaced repetition and spacing is beneficial for long-time retention. Similar findings are reported by Cepeda et al. (2008) and Vlach and Sandhofer (2012).

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

Regarding repetition, Rawson and Dunlosky (2011) differentiate between recalling and relearning. With a focus on the mnemonic strategy of retrieval practice, they examined the effects of initial learning criterion (i.e., a required number of correct recalls) and relearning (i.e., the repetition of test items after an extended time period) on the durability and efficiency of learning. Here, it was shown that learning efficiency is influenced by the number of trials with correct recalls, while durability depends on the rate of relearning. Increasing the initial learning criterion showed a strong effect on retention in the absence of repetition. This effect diminished markedly through relearning. Relearning was thus shown to be more costly in terms of time but also reduced the number of attempts to reach the initial learning criterion.

2.3 Discussion and Limitations of Former Findings on LMS Usage

Summing up this section, early research results from the field of LA have yielded promising insights into LMS usage. However, more research is needed about LMS usage in general and about practice and repetition in particular. Firstly, most existing research either deals with small datasets or small samples (Andergassen et al., 2013; Wise et al., 2013), investigates LMS activities without differentiating the activity types (Hershkovitz & Nachmias, 2011), or is restricted to one course only (Whitmer, 2012). Although Whitmer (2012) gives evidence that technology usage can be more useful than learner characteristics to predict student performance, it should be pointed out that these findings are based on one online course with 377 students and that the LMS usage of students beyond this course has not been considered at all. Moreover, the study only examines a limited number of demographic variables (e.g., being from a racial/ethnic under-represented minority); five LMS usage categories (i.e., administration, assessment, content activity, engagement activity, overall course activity); and only one usage indicator (i.e., the number of hits).

Secondly, although research has pointed to the importance of practice and repetition in learning, most existing empirical results for this topic go back to the time before widespread LMS usage (Wells & Hagman, 1989); are based on rather small setups (Mödritscher, Andergassen, et al., 2013; Vlach & Sandhofer, 2012); or incorporate only small sets of usage variables (Cepeda et al., 2008; Cotton, 1989; Mödritscher, Neumann, et al., 2013). In addition, an important aspect is the focus on a real-world setting given by the log-files of an LMS platform, whereas most of the research mentioned above uses experimental settings. While the outcome of practicing an exercise item — as measured by Rawson and Dunlosky (2011) — cannot be logged by a regular log-file entry, our log-file analysis gives insights into students' real-world learning behaviour. The research presented in this paper thus aims to complement existing research by tackling these issues.

2.4 Aims and Objectives

In this study, therefore, we address the following: a) a substantial set of LMS usage variables; b) all

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

activities of learners in a platform not restricted to one course; and c) several courses that have high numbers of participants and can be assigned to different knowledge domains. Drawing on the studies mentioned above, usage data makes it possible to measure activities both within and outside a specific course; interactions with specific LMS applications such as quizzes and exercises; and the repetition of actions, spacing of actions, and learning time within the LMS. This article reports on our research efforts on analyzing LMS usage and on an empirical study that we have conducted at our university. In particular, we investigate potential correlations between student e-learning usage patterns and final exam grades. The following research questions are addressed in this paper:

1. How are final grades related to practice and repetition in LMS usage while preparing for an exam?
2. How does preparing for a specific exam relate to the overall usage of an LMS in the exam preparation period?
3. Does exam preparation vary between different course domains, and which commonalities can be found?

The empirical study is based on the LMS of the Vienna University of Economics and Business (WU), called Learn@WU. The WU ranks among the largest business and economics universities worldwide and is one of the largest universities in Austria. The Learn@WU system covers all university courses (about 5,000 courses per year), which are predominantly held in a blended learning mode. One of the traditional problems of the (public) university is the heterogeneous knowledge of first-year students, which results in different learning paces among students with different backgrounds. This was one of the reasons for a strong emphasis on self-assessment facilities in the LMS in the first place. To improve the students' throughput and to decrease the period of study for the most gifted students, the university introduced half-semester terms.

Thus, in their first year of study, students can pass a semester-long course in half a semester with the help of e-learning. If a student fails the exam, he/she can enroll in the course again in the second half of the semester. Failure rates of 50% or more are not unusual in these exams. The university offers three exam weeks per semester for all beginner courses: one at the beginning, one in the middle and one at the end of the semester. The exam weeks are preceded by a so-called "exam preparation week." In these weeks, the LMS usage rates are the highest in the whole year, with up to 3.8 million page views per day. Students solve up to 600,000 self-assessment exercises per day (Mödritscher, Neumann, et al., 2013). With these figures, Learn@WU is one of the most intensely used e-learning systems worldwide.

Against this background, our engagement in LA aims to gain a better understanding of the student learning processes in this study phase. The overall goal from a mid- and long-term view is to improve the courses in terms of learning processes and didactic models based on the empirical data from the LMS.

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

3 RESEARCH METHODOLOGY

3.1 Context and Data Sources

The research was conducted based on data from the 14 days of the exam preparation week and the exam week in November 2012. The raw data comprised LMS server log-files containing 250 million entries, which means an average of about 17 million hits per day.

The research layout focuses on the following: 1) a comparison of LMS usage in three blended courses from different knowledge domains, comprising elementary business, law, and IT topics; and 2) a comparison of LMS usage within such a course with the usage in all other areas of the LMS. The selected courses are from the beginning phase of the Bachelor program. Usually, students attend these courses in parallel with a set of additional courses. From the three courses, we analyzed the overall LMS usage of all students who attended the final exam of these courses, using data retrieved from a period of the 14 days immediately before the exam. Due to the high number of self-test learning materials, the usage of the LMS during this time span is the highest in the semester. All three courses are half-semester courses concluding with a final exam, which consists of a paper and pencil multiple-choice test. The course C1 comprised n=883 participants, C2 comprised n=389 participants, and C3 comprised n=578 participants. Typically, the failure rate in these courses is relatively high (39.8% in C1, 45.8% in C2, and 46.0% in C3).

3.2 Web Usage Mining

In order to analyze student LMS usage, we followed the Web Usage Mining process suggested by Srivastava, Cooley, Deshpande, and Tan (2000), which consists of three major steps.

The first step, ***data preprocessing***, deals with the selection and transformation of data. As displayed in Figure 1, we selected all entries from the raw data of the students that participated in at least one of the three courses examined in our study (i.e., the log-files from the 14 days before the exam, which were available in a slightly extended Combined Log format). This led to three datasets containing 2.3 million entries (course C1), 1.2 million entries (course C2), and 1.6 million entries (course C3), respectively.

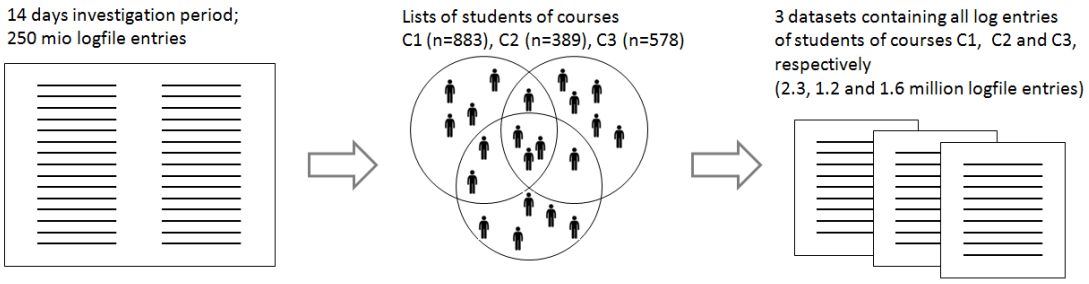

Figure 1: Generation of datasets for all log-file entries of students attending at least one of the three courses

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

We then sorted the log-file entries according to each student's actions within the specific course and outside this specific course (Figure 2). Overall, the distribution of student activity in the LMS can be summarized as follows: 788,517 activities (i.e., page views) in C1 vs. 386,913 activities in 166 other courses; 229,260 activities in C2 vs. 444,238 activities in 124 other courses; 731,694 activities in C3 vs. 456,849 activities in 139 other courses. Next, the log and exam data were connected, and all data was anonymized.

The second step, ***pattern discovery***, aims to extract patterns of LMS usage and to generate condensed data structures, such as usage variables and descriptive statistics. In our case, we a) extracted the user sessions from the log-files and b) defined and calculated usage variables for practice and repetition within the targeted course and within the LMS for each student. The third step, ***pattern analysis***, applies analysis techniques to the condensed data on usage patterns. For our study, we applied inferential statistical methods from correlation and regression analysis, combined with visual representations of the analysis results.

In all three phases of this Web Usage Mining process, we proceeded in an explorative way. To be precise, we followed an iterative process, whereby the definition of usage variables led to inferential statistics about their correlation with final grades, and the results of that led to the definition of new sets of variables or the need to apply other methods of analysis, and so on.

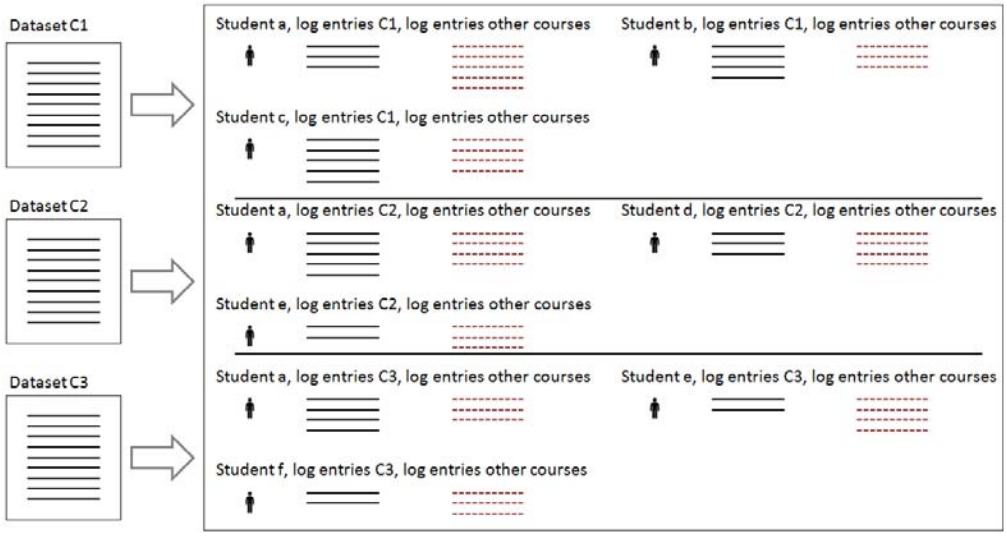

Figure 2: Sorting of log-file entries according to each student's actions within the specific course and outside the specific course

3.3 Iterative Selection and Refinement of Usage Variables and Analysis Methods

We started with a set of variables that operationalizes practice and repetition. The variable selection

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

was approached from a time-related and a learning-material-related perspective. Regarding ***practice***, the user sessions made it possible to calculate learning times by adding together the times between the time stamps of log-file entries in the sessions. Furthermore, the time stamps allowed us to detect the number of different learning days. Two kinds of learning materials for measuring student progress in practicing were investigated, namely, self-assessment exercises and sample exam questions. These materials, which are the most intensely used on Learn@WU, are particularly intended for self-assessment and preparation for the final exam.

The usage of each interactive exercise and exam question is captured in the log-files by means of a unique identifier. The amount of usage can therefore be calculated as a further indicator for practice. Through the HTTP methods GET and POST in the log-file entries, it is possible to distinguish between activities related to the instruction or solution pages of an exercise.

Regarding ***repetition***, the repetitive use of these learning materials was counted, again based on their unique identifiers. Time gaps between the repeated solving of self-assessment exercises and sample exam questions were also investigated. Since the log-files and user sessions contain information about courses through the URLs, it was possible to compare usage within and beyond the three individual courses. While the literature suggests having longer gaps between learning sessions (i.e., from one day to several weeks) to foster long-term retention (cf. Cepeda et al., 2008; Dunlosky et al., 2013), this study focuses on short-term retention in order to pass exams in a Bachelor program. Time gaps were calculated based on days with an expected range from 0 to 14 days. A more fine-grained analysis would have been possible but was not conducted in this paper.

As a next step, descriptive statistics such as frequencies, means, and medians were produced and correlation coefficients were calculated. Some variables were complemented with or evolved into more fine-granular variables. Over several iterations, we defined and refined a set of 83 LMS usage variables and calculated them for all students of the three courses. Due to the limited space, this paper concentrates on those variables that revealed the highest correlations with final grades. Table 1 gives an overview of these variables. The usage variables in lines 2 to 10 are defined as follows:

- **topic | other:** The students usually attend many parallel courses in each semester. They might not distribute their learning evenly among the courses, and learning intensely for one course might negatively affect the learning in other courses. To study such possible displacement competition, the identifier {topic|other} refers to the measurement of all usage variables regarding student activities within a course under investigation (*topic*) and overall LMS usage and thus student activities in other courses (*other*). For instance, the variable “duration” was calculated for each student’s visiting duration within the course C1 (or C2 and C3, respectively), and the visiting duration in other courses. Thus, two variables were calculated for the duration time: *topic.duration* and *other.duration*.

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

- **excs | exam:** The identifier {excs|exam} indicates the investigated learning materials, namely, self-assessment exercises and sample exam questions. For instance, the variable "different" was calculated for each student's number of unique solved self-assessment exercises (*excs*) and sample exam questions (*exam*).

Table 1: Overview of the LMS usage variables extracted from raw data

Dependent Variable	**Description**
Points, grade	Points and grade achieved in the final exam
Variables as indicators for practice	
{topic|other}.duration	Absolute duration (hours) of each student's activities. This variable is calculated by adding up the time intervals between each of a student's learning activities (page hit).
{topic|other}.days	Absolute number of different days each student used the LMS. The variable *topic.days* counts each day that contains at least one activity by the student within the course. The variable *other.days counts e*ach day that contains at least one activity by the student in other courses. Thus, the range varies from 0 (no activities) to 14 days (investigation time span).
{topic|other}.dayspan	Time span (days) between first and last activity. This variable complements the variable {topic|other}.days by indicating whether a student has distributed his learning time evenly throughout the days, or, for example, has only started to prepare shortly before the exam.
topic.{excs|exam}_different	Absolute number of unique interactive solved exercises and sample exam questions. The three investigated courses contain up to about 1,500 self-assessment exercises and sample exam questions. The variable measures how many different items a student has solved.
topic.{excs|exam}_coverage	Similar to **_different*, but as a percentage.
topic.{excs|exam}_avg_considertime	Average time used to consider exercises and exam questions (i.e., work towards the solution). The investigation distinguishes between page views of the exercise problem and page views of the exercise solution. The variable **_avg_considertime* is the ratio of the total time spent on viewing a problems page versus the sum of all the exercises/exam questions solved by the student.
Variables as indicators for repetition	
topic.{excs|exam}_repeat_coverage	Percentage of unique self-assessment exercises and sample exam questions that were solved more than once
topic.{excs|exam}_repeat_factor	Ratio of total solved self-assessment exercises and sample exam questions versus different solved items
topic.{excs|exam}_avg_repeat_gap	Average time differences between repetitions in solving self-assessment exercises and sample exam questions

The variables **.duration*, **.days* and **.dayspan* were calculated both for in-course (topic) and out-of-course (other) activities. The variables **_difference*, **_coverage* and **_avg_considertime* were only

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

calculated for in-course activities. The results of these show some differences between self-assessment exercises (*excs*) and sample exam questions (*exam*), since the variables relating to the entire LMS usage besides the specific course (*other.**) are close to null in these cases. As a result, we conducted the final pattern analysis with a set of 12 variables to investigate practice in LMS usage and a set of six variables to investigate repetition.

The correlation of the variables with the final exam grades was calculated using Pearson correlations. Moreover, the significance levels were determined as follows. Null-hypotheses about each independent variable's correlation with the final grades were formulated. The significance level of 5% was corrected using the conservative Bonferroni method, thus leading to a new significance level of 0.42% for the 12 variables representing practice and 0.83% for the 6 variables representing repetition. Since the correlation coefficients only give a limited impression of the nature of the dependencies, bag plots and scatter plots were created to get a more detailed picture. Bag plots are bivariate extensions of box plots, and make it possible to analyze the distribution of two variables in relation to each other. Scatter plots display the values of two variables on the axis of a Cartesian coordinate system and are useful in identifying dependencies between different variables in a visual manner.

Scatter plots can be prone to outliers, for instance because of the smoothing of the non-parametric regression curve. Therefore, quantile box plots, a visualization technique for graphically depicting groups of numerical data through five-number summaries (i.e., the median, the lower and upper quartiles, as well as the sample minimum and maximum) for different quantiles of a usage variable, were applied. Furthermore, quantile regression comprises a method for regression analysis that aims to estimate either the conditional median or other quantiles of the response variable, and is thus more robust against outliers in the response measurements (Koenker, 2008) that we were facing in our dataset. For this kind of regression analysis, we used the R package "quantreg" (Koenker, 2013). We created quantile regression plots, which consist of box plots for the dependent variable (points), and contrasted the quantile regression and ordinary least square (OLS) estimates of selected covariates.

Finally, we observed a particularity in the data of course C1. The exam for C1 consists of two partial exams. If a student does not reach a minimum of points in either partial exam, he is assigned 0 points for the entire exam. This caused a highly uneven distribution of final points in our plots, giving the illusion that many students were not able to gain any points at the exam. Therefore, we filtered out students with 0 points, which reduced the number of participants in C1 from n=883 to n=786.

3.4 Research Ethics

Research ethics are important within LA research, including issues such as learner rights and data ownership. However, still lacking is an ethical framework that defines, for instance, the rights of learners in relation to their data, including opting out of the analytics record and giving informed consent for data usage to researchers (Ferguson, 2012).

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

To counter this, Slade and Prinsloo (2013) suggest six principles towards an ethical framework. They include the following points: 1) LA should function as moral practice; 2) students should be seen as agents and thus as collaborators in developing their learning; 3) data collected about performance should be seen as temporal, dynamic constructs and thus should have an expiry date; 4) student success should be seen as a multidimensional phenomenon far beyond learning analytics metrics; 5) transparency about data usage should be offered by the university; and 6) higher education needs to use LA better to understand and develop outcomes for students.

In our study, the following measures were taken to ensure the rights of the students. Users of the Learn@WU platform were informed about data storage and data analysis by the university upon registration. Any analysis of student-related data occurs in an anonymized form to protect student privacy as far as possible. As in the current case, the unique identifiers assigned to the dataset through k-anonymization make a re-identification of the students theoretically possible. Data that allow the back-tracing of individual students without their informed consent is not published at all.

Regarding the sixth principle of Slade and Prinsloo, our aim in LA research is to advance the LMS in terms of quality and effectiveness of teaching and learning. In order to do this, student-related data may need to be collected, but only to the extent absolutely necessary.

4 RESULTS AND DISCUSSION

4.1 Practice

4.1.1 Learning time, learning days, and dayspan correlate positively, albeit modestly, with final exam points

There is a positive relationship between the total duration, the learning days, and the dayspan in the course on the LMS and final points. Table 2 summarizes these variables for all three courses, including the Pearson correlation coefficients. The table shows that all correlation coefficients are based on a good sample size (i.e., the number of all participants in the courses) and most of them also have good significance levels when considered individually, i.e., *p*-values smaller than 0.0042, indicating that the correlation is unlikely to result from random chance. This also holds when considering the 12 practice-related variables and null-hypotheses interdependently by applying the Bonferroni correction method. Tables 2 and 3 highlight the *p*-values above the corrected significance level of 0.42%.

The correlation is the highest in course C2, with Pearson correlation coefficients of 0.41 for *topic.days* and 0.40 for *topic.dayspan*. The numbers show indications, albeit on a modest level, that the more different days a student prepares for the exam on average, the better are his/her results. For C1 and C3, these correlations are lower (C1: 0.29 and 0.25; C3: 0.24 and 0.21).

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

It can also be observed that student behaviour regarding learning days and dayspan varies between the courses. While in C1 the students study on 11 days with a median dayspan of 12, and thus almost every day of the investigation period, the learning days and dayspan in C3 (6 and 10 days, respectively) and C2 (5 and 9 days, respectively) are lower. The weaker correlations in course C1 indicate that, for this course, other factors play a more important role for the final grades than regular online learning.

Table 2: Mean, Median, and Pearson correlation coefficients for the variables *.duration, *.days and *.dayspan; all three courses (Bonferroni-corrected significance level: 0.0042)

Duration (hours)	**C1** (n=786)		**C2** (n=389)		**C3** (n=576)	
	Topic	*Other*	*Topic*	*Other*	*Topic*	*Other*
Mean	25.05	5.11	6.00	18.39	12.62	13.57
Median	24.17	1.62	3.74	15.72	8.54	8.20
Pearson corr. coeff.	**0.3006** (p<2.2e-16)	0.0889 (**p=0.0127**)	0.2672 (p=8.7e-8)	0.0328 (**p=0.5195**)	0.2379 (p=7.0e-9)	0.0848 (**p=0.0417**)
Learning days	**C1** (n=786)		**C2** (n=389)		**C3** (n=576)	
	Topic	*Other*	*Topic*	*Other*	*Topic*	*Other*
Mean	10.30	8.88	5.21	9.47	6.57	9.05
Median	11.0	10.0	5.0	11.0	6.0	10.0
Pearson corr. coeff.	0.2861 (p=4.4e-16)	0.2063 (p=5.3e-9)	**0.4119** (p<2.2e-16)	0.1504 (p=0.0029)	0.2390 (p=5.9e-9)	0.2400 (p=5.1e-9)
Dayspan	**C1** (n=786)		**C2** (n=389)		**C3** (n=576)	
	Topic	*Other*	*Topic*	*Other*	*Topic*	*Other*
Mean	10.90	10.30	7.29	10.40	8.18	9.91
Median	12.0	11.0	9.0	12.0	10.0	11.0
Pearson corr. coeff.	0.2476 (p=1.9e-12)	0.2056 (p=6.0e-9)	**0.4037** (p<2.2e-16)	0.2038 (p=5.1e-5)	0.2105 (p=3.3e-7)	0.2221 (p=6.8e-8)

4.1.2 Displacement competition between courses is not observed

To study the displacement competition of online learning activities, we calculated the correlation of *other.duration* (online time spent by a student preparing for an exam outside the topic of the exam), *other.days* and *other.dayspan* with the points achieved in the exam. As shown in Table 2, the correlations with learning results are either very low (*other.duration*) or rather positive (*other.days*, *other.dayspan*). Moreover, the learning duration outside the examined courses shows little significance, as the *p*-values for *other.duration* are above the Bonferroni-corrected significance level of 0.0042 in all courses. The positive correlation coefficients indicate no displacement competition between online learning for one course and online activities in other courses.

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

4.1.3 Coverage of exercises and exam questions correlates positively, albeit modestly, with final exam points

Regarding learning materials, the correlation between the usage of self-assessment exercises and sample exam questions and the achieved results was investigated. In absolute numbers, course C1 contains 272 self-assessment exercises and 184 sample exam questions; course C2 contains 593 exercises and 105 sample exam questions; and course C3 contains 1,368 exercises and 120 sample exam questions. This is represented by the usage variable **_different* of Table 3.

Table 3: Mean, Median, and Pearson correlation coefficients for the variables topic.*_different, topic.*_coverage and topic.*_avg_considertime; all three courses (Bonferroni-corrected significance level: 0.0042)

Different	**C1** (n=786)		**C2** (n=389)		**C3** (n=576)	
	Excs	*Exam*	*Excs*	*Exam*	*Excs*	*Exam*
Mean	101.0/272	63.3/184	93.8/593	43.5/105	333/1368	41.9/120
Median	93/272	47/184	2/593	34/105	193/1368	10/120
Pearson corr. coeff.	0.2489 (p=1.5e-12)	0.3132 (p<2.2e-16)	0.2390 (p=1.9e-6)	0.2243 (p=7.9e-6)	0.3388 (p<2.2e-16)	0.2610 (p=1.9e-10)
Coverage	**C1** (n=786)		**C2** (n=389)		**C3** (n=576)	
	Excs	*Exam*	*Excs*	*Exam*	*Excs*	*Exam*
Mean (percent)	37.03	34.39	15.53	41.41	23.91	34.89
Median (percent)	34.19	25.54	0.33	32.38	13.85	8.33
Pearson corr. coeff.	0.2489 (p=1.5e-12)	0.3132 (p<2.2e-16)	0.2390 (p=1.9e-6)	0.2243 (p=7.9e-6)	**0.3388** (p<2.2e-16)	0.2610 (p=1.9e-10)
Avg. considertime	**C1** (n=786)		**C2** (n=389)		**C3** (n=576)	
	Excs	*Exam*	*Excs*	*Exam*	*Excs*	*Exam*
Mean (seconds)	173.0	166.0	35.9	38.7	37.8	31.4
Median (seconds)	171.0	178.0	15.0	37.9	36.3	31.1
Pearson corr. coeff.	0.0846 (**p=0.0177**)	**0.3742** (p<2.2e-16)	0.06013 (**p=0.2368**)	0.1480 (p=0.0034)	0.1007 (**p=0.0155**)	0.1667 (p=5.6e-5)

C3 has the largest set of self-assessment exercises, which explains the rather small median percentage. The usage variable **_coverage* expresses the percentage of learning materials (exercises and sample exam questions) solved by the students when preparing for the final exam within the two weeks of investigation. Table 3 shows positive correlations between *topic.excs_different* and *topic.exam_different* (and **_coverage*, respectively) and final exam points. The correlation is strongest between *topic.excs_coverage* and final points in the course C3. The variable *topic.excs_avg_considertime* is less

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

significant with *p*-values of up to 0.2368. The Bonferroni correction only has effects on the average considertime for exercises in all courses, whereby the correlation is very low in these cases.

To obtain a better understanding of the learning material coverage, we first provide an overview of the medians (*topic.excs_coverage*, *topic.exam_coverage*) and achieved points. Such dependencies can be plotted as bag plots. The bag plots in Figure 3 show the variable *topic.excs_coverage* on the *x*-axis, and number of points attained in the exam on the *y*-axis. In course C1, a student has to achieve 71 points to pass the test, in C2 30 points and in C3 24 points. The bag plots mark the medium 50% of values, which are included in the box plot between the first and third quartile. The outer (light blue) area corresponds to the length of the whiskers. The points located outside the blue area are outliers and plotted as individual points.

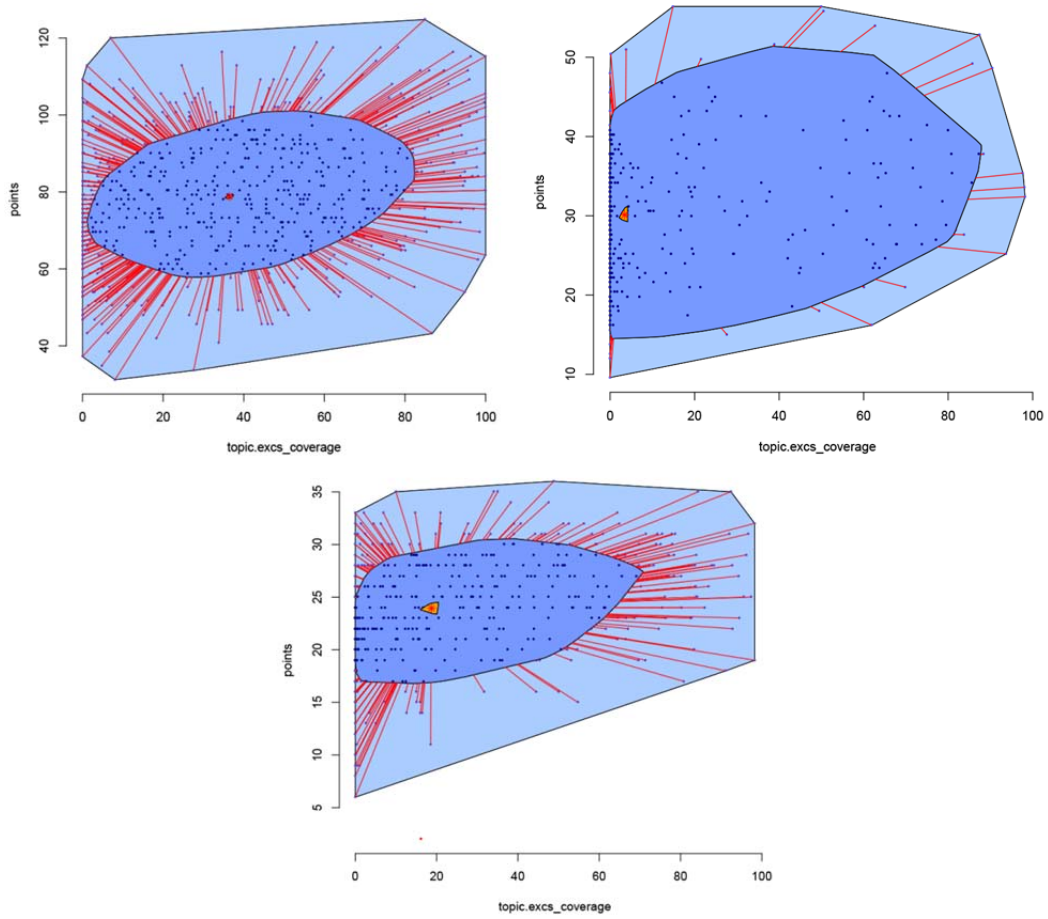

Figure 3: Bag plots with topic.excs_coverage and final points for C1 (top left), C2 (top right), and C3 (bottom)

The bag plots show that the median student solves about 40% of self-assessment exercises and gains about 80 points in course C1 (Figure 3, top left). The range between the first and third quartile of students in the inner (dark blue) area shows that 50% of students get between 60 and 100 points,

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

solving between about 3% and 80% of exercises on the topic in the LMS. In the light blue area at the bottom left, there are the students who have not solved many exercises but also have not gained many points. The top right of the plot depicts those students who have solved many exercises with good results. At the bottom right, the students who have solved high percentages of exercises for the topic but have not passed the exam are situated. Some students also gained good results without substantial use of the LMS (top left).

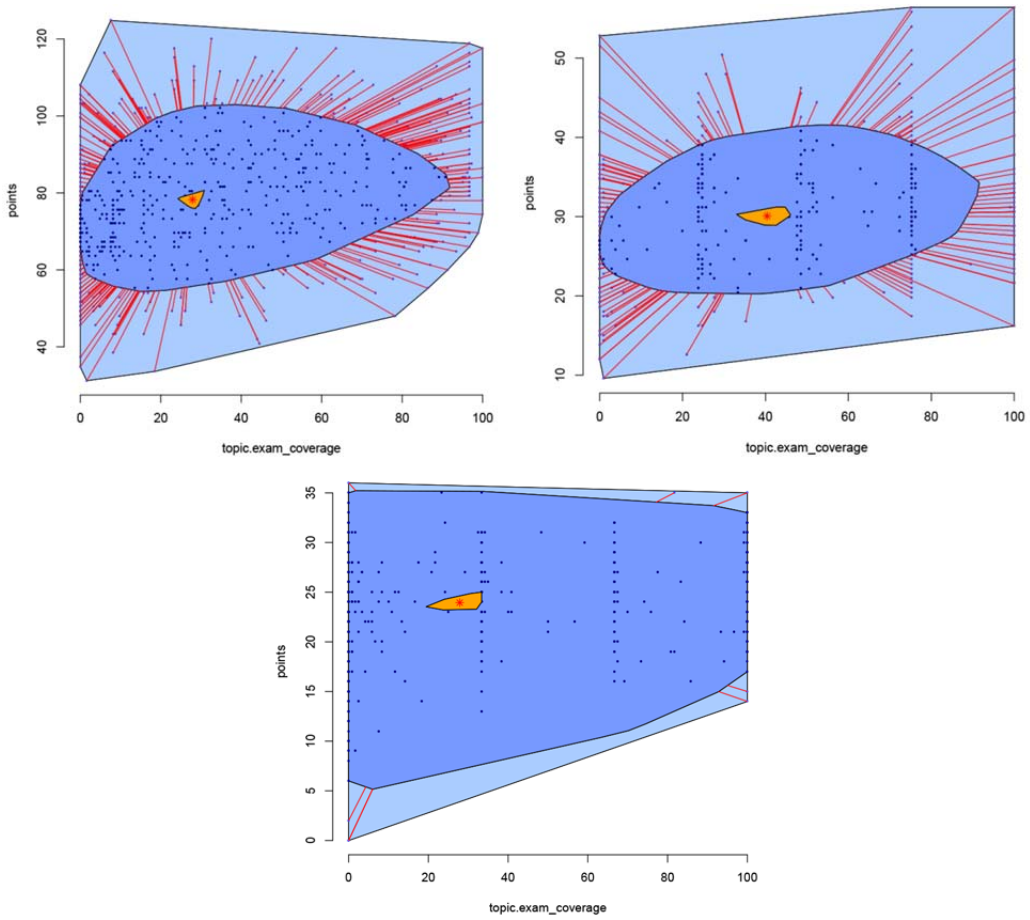

Figure 4: Bag plots with topic.exam_coverage and final points for C1 (top left), C2 (top right), and C3 (bottom)

The median of *topic.excs_coverage* is very low for C2 compared to C1 (0.33% vs. 34.19%), although the correlation coefficients are similar (0.24 vs. 0.25). For all three courses, the 50% population bubble reaches about 80% coverage of exercises, despite the fact that C3 offers about 5 times more exercises than C1 (Table 3). Thus, while the average coverage of solved exercises in course C2 is much smaller than in C1 and C3, the coverage of sample exam questions is much higher. Figure 4 shows this relationship. Furthermore, many students solve 100% of the sample exam questions in course C2, in contrast to C1 and C3. Solving sample exam questions also correlates higher with the final points in C2 than solving exercises.

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

4.1.4 Oscillation effect and regression of exercise coverage and final exam points

The scatter plots of Figure 5 visualize the relationship between the usage variable *topic.excs_coverage* and final points in more detail. The scatter plots can be read as follows. As above, the *x*-axis shows the values of the explanatory variable, which in our case is *topic.excs_coverage*. The *y*-axis shows the values of the outcome variable, which in our case are the final points. The green straight line is the linear regression line (OLS). In our study, this rises in each plot and thus marks positive relationships. The red solid line is the non-parametric regression line, and the red dotted lines are the nonparametric regression lines for the first and third quantiles. The scatter plots also contain box plots to show dependencies between the explanatory and outcome variables. For instance, the plot of C1 (top left) can be read as the average student solving about 38% of self-assessment exercises and gaining 78 points.

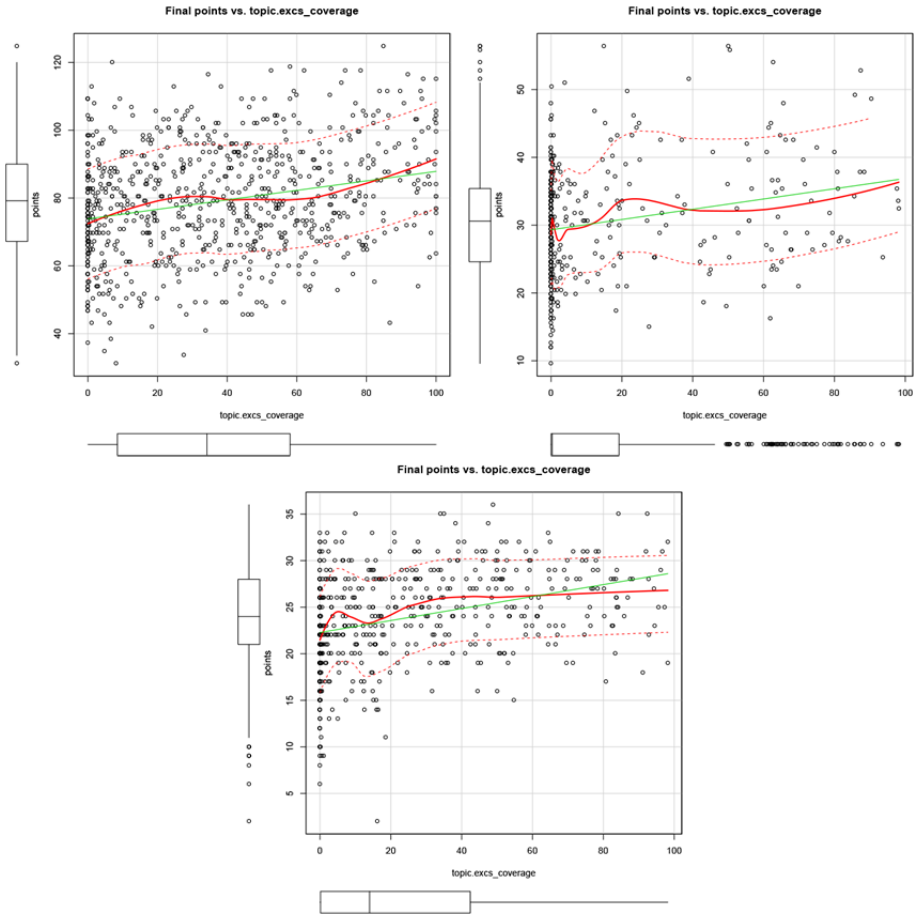

Figure 5: Scatter plots with topic.excs_coverage and final points for C1 (top left), C2 (top right), and C3 (bottom)

Course C1 shows a nearly linear distribution, while some oscillation effect is present in courses C2 and C3. Moreover, it becomes apparent that in courses C2 and C3, many students do not solve a significant fraction of the self-assessment exercises during exam preparation. Nevertheless, their final points cover

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

the entire range from very low to very high scores, and some candidates achieve good results without significant LMS usage. The non-parametric regression curve of C3 (bottom) becomes horizontal towards the end, indicating that after a certain number of exercises, saturation is reached where solving more exercises does not correlate with better results. The right hand side of the scatter plot is in a sparely populated area, so there is no strong significance. However, the diminishing gain from additional exercises is clearly visible. Furthermore, the regression curve is oscillating in courses C2 and C3.

These observations might be caused by either outlier or smoothing effects in the scatter plot splines. Therefore, we investigated the phenomena further by introducing quantile box plots and quantile regression plots. The quantile box plots and quantile regression were made for the explanatory variable *topic.excs_coverage* and final points in course C2. The quantile box plot (Figure 6, left) shows the quantiles of *topic.excs_coverage* in relation to the final points. It only contains the values of the students who solved at least one interactive exercise (n=222/389 students). In the quantile box plots, the usage variable *topic.excs_coverage* is divided into 10 quantiles. As before, we observe a significant oscillation in the box plot of C2, this time at the 0.3 quantile. The students in the high quantiles (e.g., at 0.9) of *topic.excs_coverage* do not score significantly higher than those of the low quantiles.

While the analysis with the multiple box plots used quantiles for more robust results on the explanatory variable *topic.excs_coverage*, quantile regression (Figure 6, right) uses the quantiles of the outcome variable to study the effects of solved exercises for successful and unsuccessful students. Table 4 lists the OLS regression coefficients, the quantile regression coefficient at the 0.1, 0.5, and 0.9 quantiles for the explanatory variable *topic.excs_coverage* and the outcome variable for all three courses. The values of the quantile regression coefficients differ significantly from the OLS regression coefficient in course C2. In the 0.9 quantile, each additional percentage of solved exercises would lead to 0.12 more points in a literal interpretation.

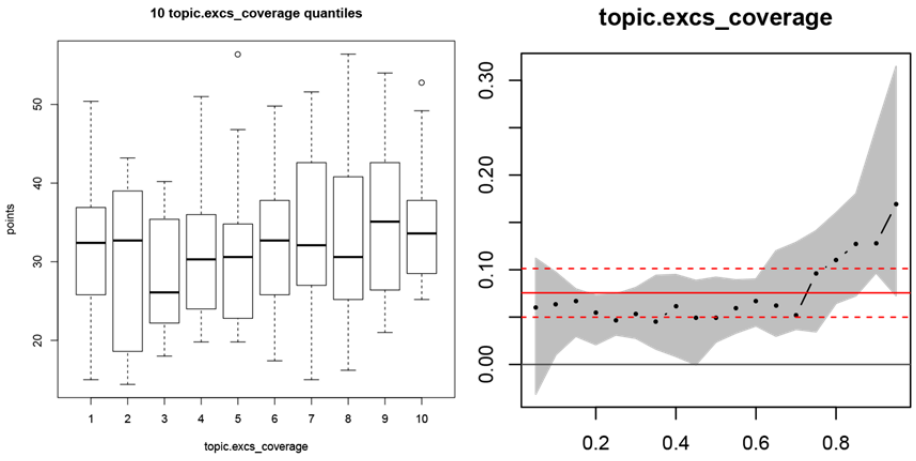

Figure 6: Quantile box plot (left; n=222) and quantile regression (right; n=389) for the explanatory variable topic.excs_coverage and the depending variable final points; course C2

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

Table 4: OLS regression and quantile regression coefficients for the usage variable topic.excs_coverage

Total points	*OLS regression (coefficient β, std. error σ)*	*Quant. regression at 0.1 quantile*	*Quant. regression at 0.5 quantile*	*Quant. regression at 0.9 quantile*
C1 topic.excs_coverage (percent)	β=0.14, σ=0.0195 (p=1.5e-12)	0.1682	0.1204	0.1321
C2 topic.excs_coverage (percent)	β=0.0756, σ=0.0156 (p=1.9e-6)	0.0636	**0.0491**	**0.1277**
C3 topic.excs_coverage (percent)	β=0.0668, σ=0.0077 (p<2.0e-16)	0.0914	0.0582	0.0438

The quantile regression plot of Figure 6 (right) can be read as follows: The quantiles of the outcome variable (points achieved) are listed on the x-axis and the coefficient magnitudes on the y-axis. The red solid and dotted lines mark the OLS coefficient and the OLS confidence interval. The OLS coefficient is equal throughout all the quantiles and shows the correlation between *topic.excs_coverage* and final points (one additional percent of solved exercises relates on average to 0.14 points for C2). The black dotted line marks the connection between quantile regression coefficients; the grey area is the confidence interval. If the black line is outside of the OLS confidence interval, then it can be concluded that there is a strong deviation between the OLS regression and the quantile regression.

It can be clearly observed in Figure 6 that the higher quantiles (students with higher grades) benefit strongly from solving self-assessment exercises. This deviation only starts at the 0.7 quantile. For the higher quantiles, the quantile regression line is steeper than the OLS regression line, indicating that for better performing students, the variable *topic.excs_coverage* had a more positive correlation. This positive deviation from the OLS regression is also reflected in the quantile regression coefficients of Table 4.

Summing up, learning through practicing self-assessment exercises does positively correlate to final points. However, there are differences in the three courses. In course C2, where the exercises demand abstract thinking, we can observe oscillation effects. Furthermore, many students do not practice through solving self-assessment exercises. Particularly in course C2, solving sample exam questions is a more widely used practice among students than solving self-assessment exercises. To gain the highest grades, however, it is beneficial to solve high percentages of self-assessment exercises.

The time taken to solve exercises differs depending on the course domain

Finally, we investigated how long students take to solve exercises and sample exam questions. As the numbers in Table 3 show, the usage variables *topic.*_avg_considertime* vary strongly between the courses. While the times are similar in C2 and C3 and range between 31 and 39 seconds on average, they are much higher in C1, at about 2.5 to 3 minutes. This can be explained by the nature of the

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

subjects. While materials in C1 include many items where calculations have to be made, the exercises and sample exam questions in C3 are mainly text-based.

4.2 Repetition

While the focus of the previous section was on results regarding practice, this section presents results regarding indicators for repetition in LMS usage.

4.2.1 Repeated solving of exercises and exam questions does not strongly correlate with final grades

With a focus on the learning materials, Table 5 summarizes the mean values and Pearson correlation coefficients of variables including the repeat coverage, the factor of the repetition, and the average repeat gap. Again, we have corrected the significance level for the six repetition-related variables according to the Bonferroni method, which has led to the insight that the correlation coefficients for the exam-related variables in C2 and for *exam_repeat_avggap* in C3 are less significant.

Table 5: Mean, Median, and Pearson correlation coefficients for the variables topic.*_repeat_coverage, topic.*_repeat_factor and topic.*_repeat_avggap; all three courses (Bonferroni-corrected significance level: 0.0083)

Repeat coverage	**C1** (n=786)		**C2** (n=389)		**C3** (n=576)	
	Excs	*Exam*	*Excs*	*Exam*	*Excs*	*Exam*
Mean	12.45	16.28	6.47	17.59	5.92	10.58
Median	5.88	8.15	0.00	6.67	1.33	0.83
Pearson corr. coeff.	0.1494 (p=2.6e-5)	0.1960 (p=3.0e-8)	0.1837 (p=2.7e-4)	0.0560 (**p=0.2704**)	0.2121 (p=2.7e-7)	0.1725 (p=3.1e-5)
Repeat factor	**C1** (n=786)		**C2** (n=389)		**C3** (n=576)	
	Excs	*Exam*	*Excs*	*Exam*	*Excs*	*Exam*
Mean	1.27	1.10	0.81	1.15	0.98	0.68
Median	1.22	1.07	1.00	1.00	1.12	1.00
Pearson corr. coeff.	0.1334 (p=1.8e-4)	0.1164 (p=0.0011)	0.1757 (p=5.0e-4)	0.0875 (**p=0.0849**)	0.2242 (p=5.1e-8)	0.2375 (p=7.5e-9)
Repeat avggap	**C1** (n=786)		**C2** (n=389)		**C3** (n=576)	
	Excs	*Exam*	*Excs*	*Exam*	*Excs*	*Exam*
Mean(hours)	28.45	25.76	7.28	21.60	18.70	16.80
Median(hours)	14.12	6.66	0.00	0.00	2.60	0.00
Pearson corr. coeff.	0.1926 (p=5.3e-8)	0.2075 (p=4.3e-9)	0.1410 (p=0.0053)	0.0812 (**p=0.1097**)	0.1955 (p=2.2e-6)	0.1049 (**p=0.0116**)

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

The numbers suggest that on average, not much repetition is done when preparing for the exams. This can also be noted in the bag plots in Figure 7 (*topic.excs_repeat_factor*). The repeat coverage shows that, in courses C2 and C3, repeating exercises correlates more strongly with grades than repeating sample exam questions. As these two courses have more exercises, the repeat coverage is lower than in C1. Furthermore, the average repeat gap for exercises in C2 is much shorter than for C1 or C3. It becomes apparent that "coverage"-variables (see the previous section) have higher correlations with the final points than the "repeat"-variables. This indicates that solving many different exercises and exam questions is more effective than repeating exercises and exam questions.

Therefore, although the variables reveal differences in the example usages between the courses, they do not serve as good predictors for results in an overall analysis. Below, parametric regression and quantile regression analysis is used once again for a more detailed view.

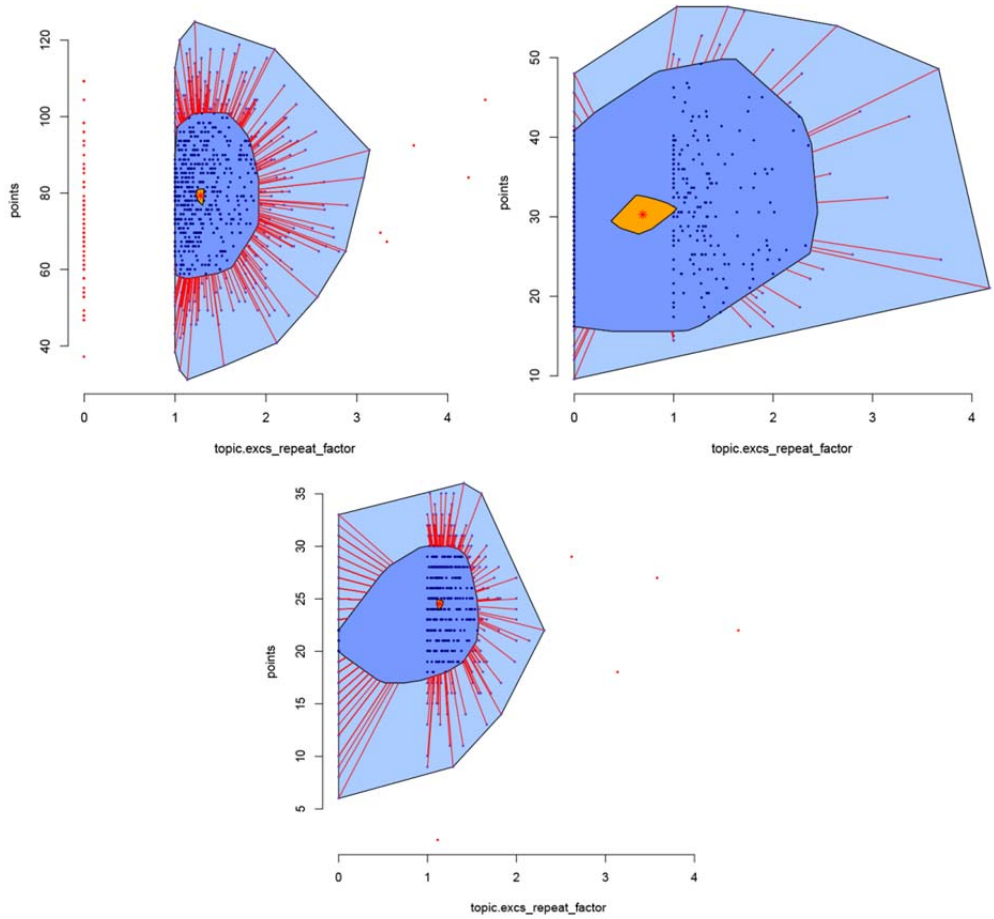

Figure 7: Bag plots with topic.excs_repeat_factor and final points for C1 (top left), C2 (top right) and C3 (bottom)

4.2.2 Repeat factor: A "bend" in repeated solving of exercises

The dependencies between the repeat *topic.excs_repeat_factor* and final points can be plotted as bag plots (Figure 7). In course C1, the inter-quartile 50% of students (dark blue area) has a repeat factor of

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

between 1 and 1.9 and final points ranging from 58 to 100 points. The inter-quartile bags of courses C2 and C3 also contain items with a repeat factor of 0, meaning that these students did not solve any exercises. This observation is similar to the one made in the previous section. Indeed, the variable *topic.excs_repeat_factor* builds upon the coverage of exercises (*topic.excs_repeat_coverage*) attempted by the students.

There are, however, some characteristics in the distributions. The scatter plots of Figure 8 show that repeat factors of 1 (C1) or slightly above (C2 and C3) relate well to high final points. A repeat factor of 1 in this context indicates that, on average, a student solved exercises once; a repeat factor of 1.3 indicates that, on average, a student solved exercises 1.3 times, etc. After these factors, the curves drop in all three courses.

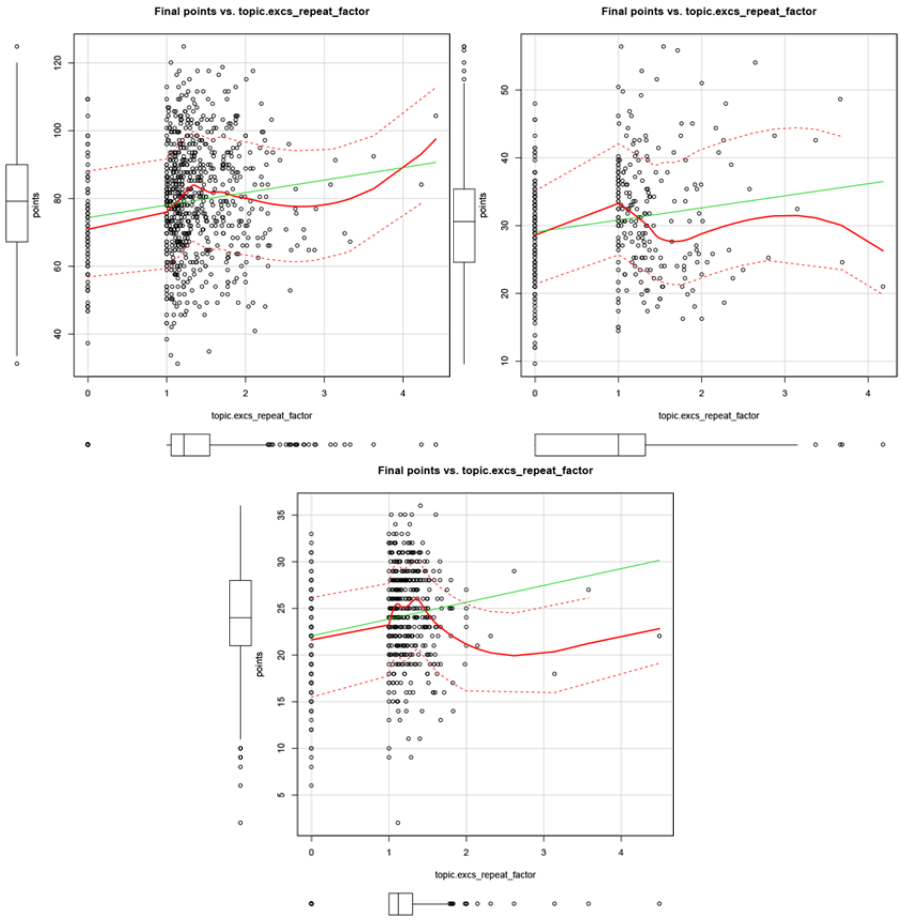

Figure 8: Scatter plots with topic.excs_repeat_factor and final points for C1 (top left), C2 (top right), and C3 (bottom)

The quantile box plot of Figure 9 (left) further analyzes the characteristics of the scatter plots and bag plots, using course C2 as an example. The quantile box plot only contains the values of those students whose *topic.excs_repeat_factor* is greater than 0, and thus represents 222 values. Some similar

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

characteristics to the scatter plot (Figure 8 top right) can be observed. There is a small rise from the first to the second quantile. Then the median drops, with a small exception in the sixth quantile. The best result is reached in the second quantile with a repeat factor of between 1.01 and 1.07, and in the tenth quantile with a repeat factor of between 2.07 and 4.18. It becomes apparent that the quantile box plot differs slightly from the scatter plot, which seems to be less robust against outliers (particularly on the far right).

Finally, Figure 9 (right) shows the quantile regression plot for the final point quantiles and the explanatory variable *topic.excs_repeat_factor*. As can be seen in Table 6 and Figure 9, the quantile regression deviates strongly from the OLS regression. In other words, students with low grades benefit much less from repetition than the students in the high quantiles.

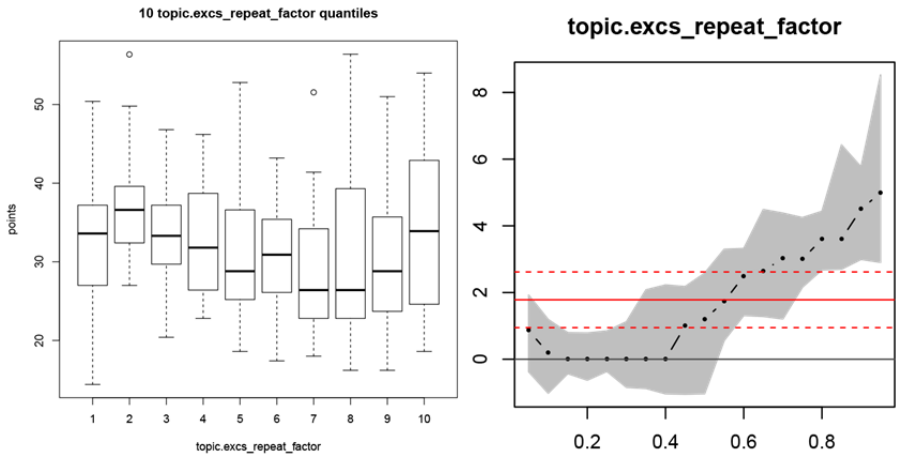

Figure 9: Quantile box plot (left n=222) and quantile regression (right; n=389) for the explanatory variable topic.excs_repeat_factor and the depending variable final points; course C2

Table 6: OLS regression and quantile regression coefficients for the usage variable topic.excs_repeat_factor

Total points	*OLS regression (coefficient β, std. error σ)*	*Quant. regression at 0.1 quantile*	*Quant. regression at 0.5 quantile*	*Quant. regression at 0.9 quantile*
C1 topic.excs_repeat_factor	β=3.687, σ=0.978 (p=1.8e-4)	4.218	3.600	2.057
C2 topic.excs_repeat_factor	β=1.783, σ=0.508 (p=5.0e-4)	**0.189**	**1.200**	**4.500**
C3 topic.excs_repeat_factor	β=2.009, σ=0.364 (p=5.1e-8)	2.000	1.122	1.956

4.2.3 Due to low levels of repeating activity, correlations regarding spacing between repetitions and final grades are of limited interest

The variables *topic.excs_repeat_avggap* and *topic.exam_repeat_avggap* as indicators of spacing between repeating an exercise or sample exam question have limited informative value regarding final

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

grades. As analyzed above, the coverage of repeated learning materials and the numbers of repetitions are quite low, and therefore any variables deduced from there cover small proportions of items too. Furthermore, Table 6 shows low Pearson correlation coefficients.

5 CONCLUSION AND FUTURE WORK

This paper analyzed dependencies between practice and repetition in an LMS and final exam results. The LMS usage was analyzed based on 250 million log-file entries collected from the Learn@WU platform over a time span of 14 days during the exam preparation phase. Three courses of different knowledge domains were investigated, and variables were calculated for usage both within and outside the selected courses. Regarding practice, we found positive, albeit modest, correlations between the amount of the learning time, learning days, dayspan of learning, and the coverage of self-assessment exercises and sample exam questions and the final exam points in each course. Repeated solving of exercises and exam questions does not correlate strongly with grades; the best results are achieved when the items are practiced once or twice only. The repeating activity in our test data is rather low. The correlations between final grades and repeat factor, as well as the spacing between repetitions of learning material items, should therefore be treated with caution.

Our analysis covers the exam preparation time of blended learning courses. We know from questionnaire data that students in the courses under investigation used the LMS for more than 60% of their learning time, but we have no detailed analysis of their offline activities. With regard to the literature, the findings are in line with Cotton (1989) in terms of learning time having a positive, albeit modest, relationship with student achievement. However, the correlations are below the effect of time-on-task activity (d=0.38) reported by Hattie (2009). Furthermore, while the positive relationship found between distributing learning over several days (learning days and dayspan) and final grades confirms earlier findings by Vlach and Sandhofer (2012), the relationship is again smaller than Hattie's (2009) spaced versus mass practice (d=0.71). The deviating numbers might be because our study only considers online activities and thus no in-class time and offline learning. As well, the study only considers the brief period of 14 days of exam preparation.

Regarding repetition, our findings only partially correlate with the literature. Where Wells and Hagman (1989) and Thalheimer (2006), for instance, find that, generally, repetition is necessary to achieve proficiency, our findings only show modest correlations between final grades and the repetition of exercises and sample exam questions.

On the other hand, Wells and Hagman also argue that, depending on the goal of learning, multiple repetitions are helpful for the learning of a specific task, while fewer specific task repetitions in combination with a broader task variety promote transfer to other tasks. Our findings indicate that it is more beneficial to transfer learning to new tasks than to repeat items often. Indeed, the exercises and sample exam questions in the investigated courses cover a broad variety of tasks and therefore promote

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

the transfer of learning to different tasks. The correlations are higher between *topic.different* (as a representation of task variety) and final grades than between *topic.{excs|exam}_repeat_factor* (as a representation of the repeated solving of exercises) and final grades in all three courses. However, when compared to Wells and Hagman's results, both the task variety and the repetition of items only show modest correlations with final exam grades. The low correlations could therefore be an indicator of the need to advance the didactic design of the LMS by providing more possibilities for teachers to design spaced repetition events. Examples include the writing of summaries and elaborative interrogation (cf. Dunlosky et al., 2013).

The overall learning material-usage data also did not correlate as strongly with the student's grades as we would have expected. Among others, one reason might be that students have different learning strategies and start with a heterogeneous array of knowledge. Consequently, in order to understand this better, it is not sufficient to look at a single variable. In contrast, it is necessary to examine the distribution of the descriptive and dependent variables and to use visualization techniques and quantiles to deal with outliers. Based on scatter plots, non-parametric regression, and quantile regression, we were able to identify saturation effects and even negative effects (oscillation effects) on higher numbers of online learning activities.

Regarding the relationship between preparing for a specific exam and the overall usage of an LMS, no displacement effects were found. Thus, a student who spends time in the course classes and in other courses is still likely to perform well in the final exam. Indeed, the correlations were small and rather positive between LMS usage within a specific course and final grades and between LMS usage in other courses and final grades. However, there are some differences between the courses. Oscillation effects in the scatter plots of the self-assessment exercise coverage in the three courses indicate that sometimes students achieve good exam grades with little LMS usage. The prevalent use of either self-assessment exercises or sample exam questions also differs among the courses. Moreover, the time spent on the individual exercises varies strongly between the courses, which might be due to the different course domains.

There are certain limits to the interpretation of the observed variables. The detailed explanation of these variables requires deeper knowledge about the didactic design of the courses. The data is most meaningful for the professors teaching these courses, because they can correlate observed behaviour in class with the planned online learning activities. In contrast, our data does not cover all the students' learning activities, but only their online ones. It is therefore impossible to draw conclusions about the total learning effort and strategies of the students, or about the quality or suitability of the learning design or materials. Another limitation is that in order to investigate dependencies between LMS usage and final exam grades, only students who received points in the final exam were considered, and thus the activities of other users (e.g., the lecturer or students without a final grade) were filtered out at this stage.

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

Future work will head into two directions. Although the three blended courses examined in this paper cover a variety of knowledge domains, we plan to extend the analysis over a full semester and across other courses. It would also be interesting to perform a longitudinal study to measure the stability of the results over multiple exams and semesters and to observe the impact of spacing over longer periods. Here, we would face the challenge of having to process significantly more data but we also see the potential for learning more about usage variables and pedagogical factors that correlate to learning in positive or negative ways.

ACKNOWLEDGEMENTS

We thank Dr Thomas Rusch for his valuable input on the application of statistical methods.

REFERENCES

Andergassen, M., Neumann, G., & Mödritscher, F. (2013). The four seasons: Identification of seasonal effects in LMS usage data. Presented at the *Alpine Rendez-Vous 2013: Workshop on Data Analysis and Interpretation for Learning Environments (DAILE '13)*, Villard-de-Lans, France.

Cepeda, N.J., Vul, E., Rohrer, D., Wixted, J.T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102. doi:10.1111/j.1467-9280.2008.02209.x

Chatti, M.A., Dyckhoff, A.L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5/6), 318–331. doi:10.1504/IJTEL.2012.051815

Cotton, K. (1989). *Educational time factors* (No. Close-Up #8). Northwest Regional Educational Laboratory. Retrieved from http://educationnorthwest.org/webfm_send/564

Doherty, W. (2006). An analysis of multiple factors affecting retention in web-based community college courses. *The Internet and Higher Education.* doi:10.1016/j.iheduc.2006.08.004

Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J., & Willingham, D.T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. doi:10.1177/1529100612453266

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304. doi:10.1504/IJTEL.2012.051816

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. In L.H. Meyer, S. Davidson, H. Anderson, P.M. Fletcher, P.M. Johnston, & M. Rees (Eds.), *Tertiary assessment & higher education student outcomes: Policy, practice & research* (pp. 259–275). Wellington, New Zealand: Ako Aotearoa.

Hershkovitz, A., & Nachmias, R. (2011). Online persistence in higher education web-supported courses. *The Internet and Higher Education*, *14*(2), 98–106. doi:10.1016/j.iheduc.2010.08.001

Johnson, J., Shum, S.B., Willis, A., Bishop, S., Zamenopoulos, T., Swithenby, S., … Helbing, D. (2012). The FuturICT education accelerator. *The European Physical Journal Special Topics*, *214*(1), 215–243. doi:10.1140/epjst/e2012-01693-0

Khan, T.M., Clear, F., & Sajadi, S.S. (2012). The relationship between educational performance and online access routines: Analysis of students' access to an online discussion forum. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 226–229). New York, NY: ACM. doi:10.1145/2330601.2330655

Koenker, R. (2008). Censored Quantile Regression Redux. *Journal of Statistical Software*, *27*(6), 1–25.

Koenker, R. (2013). *Quantile Regression in R: A Vignette*. Retrieved from http://www.econ.uiuc.edu/~roger/research/rq/vig.pdf

Mödritscher, F., Andergassen, M., Law, E.L.-C., & García-Barrios, V.M. (2013). Application of learning curves for didactic model evaluation: Case studies. *International Journal of Emerging Technologies in Learning (iJET)*, *8*(S1), 62–69. doi:10.3991/ijet.v8iS1.2357

Mödritscher, F., Neumann, G., & Andergassen, M. (2013). Dependencies between e-learning usage patterns and learning results. In *Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies*. Presented at the i-KNOW '13, Graz, Austria.

Rawson, K.A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology. General*, *140*(3), 283–302. doi:10.1037/a0023956

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., … Baker, R.S.J.d. (2011). *Open learning analytics: An integrated & modularized platform. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*. SOLAR Society for Learning Analytics Research. Retrieved from http://solaresearch.org/OpenLearningAnalytics.pdf

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*. doi:10.1177/0002764213479366

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P.-N. (2000). Web usage mining: Discovery and applications of usage patterns from Web data. *SIGKDD Exploration Newsletter*, *1*(2), 12–23. doi:10.1145/846183.846188

Thalheimer, W. (2006). *Spacing learning events over time: What the research says*. Retrieved from http://www.work-learning.com/catalog.html

Vlach, H.A., & Sandhofer, C.M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*(4), 1137–1144.

Wells, R., & Hagman, J.D. (1989). *Training procedures for enhancing reserve component learning, retention, and transfer*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a217450.pdf

Whitmer, J.C. (2012). *Logging on to improve achievement: Evaluating the relationship between use of the learning management system, student characteristics, and academic achievement in a hybrid large enrollment undergraduate course*. University of California.

Willingham, D.T. (2004). Practice makes perfect — but only if you practice beyond the point of perfection. *American Educator*, *28*(1), 31–33.

(2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1* (1), 48–74.

Wise, A.F., Zhao, Y., & Hausknecht, S.N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 48–56). New York, NY: ACM. doi:10.1145/2460296.2460308

Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013). Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 145–149). New York, NY: ACM. doi:10.1145/2460296.2460324

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

A Review of Psychometric Data Analysis and Applications in Modelling of Academic Achievement in Tertiary Education

Geraldine Gray, Colm McGuinness, Philip Owende, Aiden Carthy
Institute of Technology Blanchardstown, Ireland

geraldine.gray@itb.ie

ABSTRACT: Increasing college participation rates, and diversity in student population, is posing a challenge to colleges in their attempts to facilitate learners achieve their full academic potential. Learning analytics is an evolving discipline with capability for educational data analysis that could enable better understanding of learning process, and therefore mitigate these challenges. The outcome from such data analysis will be dependent on the range, type, and quality of available data and the type of analysis performed. This study reviewed factors that could be used to predict academic performance, but which are currently not systematically measured in tertiary education. It focused on psychometric factors of ability, personality, motivation, and learning strategies. Their respective relationships with academic performance are enumerated and discussed. A case is made for their increased use in learning analytics to enhance the performance of existing student models. It is noted that lack of independence, linear additivity, and constant variance in the relationships between psychometric factors and academic performance suggests increasing relevance of data mining techniques, which could be used to provide useful insights on the role of such factors in the modelling of learning process.

KEYWORDS: Learning analytics, educational data mining, psychometrics, classification, academic performance, ability, personality, Big-5, motivation, learning style, self-regulated learning, learning dispositions

1. INTRODUCTION

It is increasingly evident that significant numbers of college students do not complete the courses in which they enrol, particularly courses with lower entry requirements (ACT, 2012; Mooney et al., 2010). Enrolment numbers to tertiary education are increasing, as is the academic and social diversity in the student population (HEA, 2013; OECD, 2013). This adds to the challenge of both identifying students at risk of failing and provisioning the appropriate supports and learning environment to enable all students to perform optimally (Mooney et al., 2010). Tertiary education providers collect an ever-increasing volume of data on their students, particularly activity data from virtual learning environments and other online resources (Drachsler & Greller, 2012). As a result, the application of data analytics to educational settings is emerging as an evolving and growing research discipline (Sachin & Vijay, 2012; Siemens & Baker, 2012), with the primary aim of exploring the value of such data in providing learning

professionals, and students, with actionable information that could be used to enhance the learning environment (Siemens, 2012; Chatti et al., 2012). A key challenge for learning analytics is the need to develop capability to explore and identify data that will contribute to improving learning models, including data not currently gathered systematically by tertiary education providers (Buckingham Shum & Deakin Crick, 2012; Tempelaar et al., 2013).

Learning is a latent variable, typically measured as academic performance in assessment work and examinations (Mislevy, Behrens, & Dicerbo, 2012). Factors affecting academic performance have been the focus of research for many years (Farsides & Woodfield, 2003; Lent, Brown, & Hacket, 1994; Moran & Crowley, 1979). It remains an active research topic (Buckingham Shum & Deakin Crick, 2012; Cassidy, 2011; Komarraju, Ramsey & Rinella, 2013), indicating the inherent difficulty in both measurement of learning (Knight, Buckinham Shum, & Littleton, 2013; Tempelaar et al., 2013), and modelling the learning process, particularly in tertiary education (Pardos et al., 2011). Cognitive ability remains an important determinant of academic performance (Cassidy, 2011), often measured as prior academic ability. Demographic data, such as age and gender, have been cited as significant (Naderi et al., 2009), as are data gathered from learner activity on online learning systems (Bayer et al., 2012; López et al., 2012). In addition to the data systematically gathered by providers, other factors can be measured prior to commencing tertiary education, which could be useful in modelling learner academic performance. For example, models predicting academic performance that include factors of motivation (e.g., self-efficacy, goal setting) with cognitive ability yield a lower error variance than models of cognitive ability alone, particularly at tertiary level (reviewed in Boekaerts, 2001; Robbins et al., 2004). Research into personality traits, specifically the BIG 5 factors of openness, conscientiousness, extroversion, agreeableness, and neuroticism, and their impact on academic achievement in tertiary education, suggests some personality factors are indicative of potential academic achievement (Chamorro-Premuzic & Furnham, 2004, 2008; De Feyter et al., 2012). For example conscientiousness, which is associated with persistence and self-discipline (Chamorro-Premuzic & Furnham, 2004), is correlated with academic performance, but not with IQ, suggesting conscientiousness may compensate for lower ability (Chamorro-Premuzic & Furnham, 2008). Openness, which is associated with curiosity, can be indicative of a deep learning style (Swanberg & Martinsen, 2010). Learning style (deep or shallow) and self-regulated learning strategies are also relevant, and have been shown to mediate between other factors (such as factors of personality and factors of motivation) and academic performance (Biggs et al., 2001; Entwhistle, 2005; Swanberg & Martinsen, 2010).

This paper reviews a range of psychometric factors that could be used to predict academic performance in tertiary education (section 2). It lays emphasis on factors that can be measured prior to, or during learner enrolment in tertiary education programmes. The unique focus is to facilitate, and inform, early engagement with students potentially at risk of failing (e.g., Arnold & Pistilli, 2012; Lauría et al., 2013). Furthermore, results from learner profiling during student induction can provide useful feedback to the learner on preferred approaches to learning tasks, and development of a personalized learning environment. A review of pertinent data analysis techniques is presented in section 3, with an emphasis

on empirical modelling approaches prevalent in educational data mining. Section 4 outlines the benefits of greater collaboration between educational psychology and learning analytics.

2. PSYCHOMETRIC VARIABLES RELEVANT TO ACADEMIC PERFORMANCE

The following discussion of student-centred factors focuses on four key areas: aptitude, temperament, motivation, and learning strategies. These were chosen based on being directly or indirectly related to academic performance and measurable in the early stages after student enrolment. The following sections outline the available evidence on correlations between individual attributes and academic achievement. All studies cited were based on tertiary education.

2.1. Cognitive Ability: How It Is Measured and Its Correlation with Academic Performance

Cognitive ability tests were originally developed to identify low academic achievers (Jensen, 1981; Munzert, 1980). The first such test measured general cognitive intelligence, *g*, as identified by Spearman (1904, 1927). Test results for an individual across a range of cognitive measures tend to correlate providing good evidence for a single measure of intelligence (Jensen, 1981; Kuncel, Hezlett, & Ones, 2004). In addition to general cognitive intelligence, there is widespread evidence for a multi-dimensional construct of intelligence comprising of a range of sub-factors (Flanagan & McGrew, 1998). Abilities in such sub-factors vary from one individual to another, and vary within an individual across factors, in other words, an individual can have higher ability in one sub-factor than in another (Spearman, 1927, p. 75). Recently the Cattell-Horn-Carroll (CHC) theory of cognitive abilities has gained recognition as a taxonomy of cognitive intelligence (McGrew, 2009). The CHC is based on ten broad cognitive categories, summarized in Table 1.

Cognitive ability tests have been criticized based on what is being measured. Sternberg (1999) asserts that intelligence tests measure a developing expertise rather than a stable attribute, and the typically high correlation between intelligence scores and academic performance is because they measure the same skill set rather than developing a causal relationship. In an analysis of a range of IQ studies measuring trends across two generations, Flynn (1987) identified a significant rise in IQ from one generation to the next. Since the observation (Flynn effect) is unlikely to be due to genetic changes in such a short period, it would appear to be the result of acquired skills that improve performance in IQ tests by subjects with the same IQ as the parent generation. This view is supported by other studies that compare children in Western and non-Western standards of education. These have shown that children tended to score well on tests that measured skills valued by their parents (Sternberg (1999 p. 8). It is notable that correlations between general intelligence and academic performance are stronger at secondary school level than in tertiary level education (Bartels et al., 2002; Cassidy, 2011; Colom & Flores-Mendoza, 2007; Eysenck, 1994; Matarazzo & Goldstein, 1972). Therefore, prior academic

performance, such as High School Grade Point Average (HSGPA),[1] and/or standardized tests, such as American College Testing (ACT)[2] scores or Scholastic Aptitude Test (SAT)[3] scores, are frequently used as measures of ability when modelling academic performance in tertiary education.

Table 2 illustrates that correlations between ability and academic performance in tertiary education are consistent and relatively strong for studies of standard students. For example, a meta-analysis of 109 studies conducted by Robbins et al. (2004) found a moderate correlation between academic performance and SAT scores (r=0.388, 90% CI [0.353, 0.424]) and a marginally higher correlation between academic performance and HSGPA (r=0.448, 90% CI [0.409, 0.488]). Eppler and Harju (1997) found that correlations between academic performance and SAT scores were not as strong for mature students. Brady-Amoon and Fuertes (2011) attribute their insignificant correlation (r=0.16, n=271) to the fact that study participants included a more diverse group of students from a variety of ethnic backgrounds, thereby supporting the findings of Schmitt et al. (2009) that the interaction between prior academic ability and GPA differs for students from different ethnic groups. The lower correlations reported by Ning and Downing (2010) (r=0.1, $p<0.05$, n=581) could be attributed to their measure of prior academic performance, which was based on A level[4] scores in just two subjects. The relatively high level of correlation reported by Cassidy (2011) could also be attributed to a difference in how prior academic performance is measured. Cassidy used GPA accrued in the first year of study as a measure of prior academic performance in order to predict students' final GPA aggregate.

2.2. Temperament: Definition and Relevance to Academic Performance

Theories of temperament focus on aspects of personality discernible at birth (Boeree, 2006; John et al., 2008). Historically, research that links temperament with academic achievement has lacked a well-defined referential framework for the interactions between temperament and academic performance. Studies have varied in their perspective of personality, with diverse views on the relevant traits to be considered as measures of temperament, such as factors of persistence, factors relating to motivation and/or moral factors such as honesty (de Raad & Schouwenburg, 1996). While many factors are associated with temperament, factor analysis by a number of researchers, working independently and using different approaches, has resulted in broad agreement of five main personality dimensions (Ackerman & Heggestad, 1997; Boeree, 2006; John et al., 2008). These are commonly referred to as the Big Five (Cattell & Mead, 2008; Goldberg, 1992, 1993; Tupes & Cristal, 1961) or the related Five-Factor Model (Costa & McCrae, 1992). The five factors — openness, agreeableness, extroversion, conscientiousness, and neuroticism — are described in Table 3.

[1] High School Grade Point Average (HSGPA) is a secondary school, end-of-year, aggregate measure of academic performance, which can be a combination of continuous assessment results and end of term exams.
[2] ACT tests are based on high school curriculum in English, Mathematics, Reading, and Science (www.act.org).
[3] SAT measures general intelligence in addition to maths and verbal subscales (Frey & Detterman, 2003). Frey and Detterman (2003) found SAT scores to be highly correlated with IQ (r=0.82, $p<0.001$).
[4] Hong Kong's secondary school termination exam. Students can select from a range of subjects.

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

Table 1: CHC's ten broad factors of cognitive ability (McGrew, 2009)

Factor	Symbol	Description
Fluid Intelligence	Gf	Ability to solve problems independently of knowledge learned.
Crystallized Intelligence	Gc	Acquiring and organizing knowledge and skills, and ability to use such knowledge in solving problems.
Visual processing	Gv	Ability to process and analyze visual information.
Auditory Processing	Ga	Ability to process and analyze auditory information.
Processing Speed	Gs	Ability to perform automatic cognitive tasks quickly (measured in minutes).
Reaction Time/ Decision Speed	Gt	Speed at which an individual can react to a stimulus, or make decisions (measured in seconds).
Short-Term Memory	Gsm	Ability to hold information with immediate awareness and reuse within a few seconds.
Long-Term Retrieval	Glr	The ability to store and retrieve information over a longer period.
Quantitative Knowledge	Gq	The ability to understand quantitative concepts and relationships, and work with numeric symbols. This is a measure of mathematical knowledge acquired, as distinct from mathematical reasoning (Gf).
Reading-Writing	Grw	Basic reading and writing skills (considered by Cattell-Horn to be part of Gc).

Table 2: Correlations between cognitive ability and academic performance

Study	n	Age	Academic Performance	g	SAT/ACT	Prior ability
Brady-Amoon & Fuertes (2011)	271	m=21.26	GPA			0.16
Cassidy (2011)	97	m=23.5	GPA			0.519**
Chamorro-Premuzic & Furnham (2008)	158	m=19.2	GPA	0.24*		
Conrad (2006)	300	m=19.48	GPA-self-reported		0.28*	
Duff et al. (2004)	146	17-52	GPA			0.274*
Eppler & Harju (1997)	212	m=19.2	GPA		0.37***	
Eppler & Harju (1997)	25	m=29.8	GPA		0.09	
Furnham & Zhang (2006)	64	[20-55]	Mean exam results	0.22		
Kaufman et al. (2008)	315	m=23.5	GPA			0.28
Kobrin et al. (2008)	151,316	18+	GPA		0.35	0.36
Ning & Downing (2010)	581	m=20.24	GPA			0.1*
Robbins et al. (2004)	Meta-analysis		GPA		0.39	0.448

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

While the Big Five concept is empirical rather than a theory of personality (Srivastava, 2010), good reliability and consistency has been reported (de Raad & Schouwenburg, 1996; John et al., 2008).

Table 3: Big Five personality factors (McCrae & Costa, 1991; Goldberg, 1992)

Big Five Factor	Traits of high scorers	Traits of low scorers
Extroversion	Enjoys human interaction, cheerful, outgoing.	Cautious, likes to be alone, can lack enthusiasm.
Neuroticism	Temperamental, moody, nervous, finds stress difficult to cope with.	Calm, even-tempered, unafraid.
Openness	Openness to new ideas and imagination, broad minded, tolerant, intellectual curiosity.	Likes routine and familiarity, factually orientated, practical.
Agreeableness	Kind, trusting, warm, unselfish.	Stubborn, rude, uncooperative.
Conscientiousness	Organized, thorough, reliable.	Relaxed, lazy, careless.

2.2.1. Relating Personality to Academic Performance

Chamorro-Premuzic & Furnham (2004) found that personality attributes measured using the big five construct accounted for up to 30% of the variance in academic performance at tertiary level education. There is a consensus across studies that conscientiousness is the best personality-based predictor of academic performance (O'Connor & Paunonen, 2007; Swanberg & Martinsen, 2010; Trapmann et al., 2007). Many researchers have cited conscientiousness as compensating for lower cognitive intelligence (see Chamorro-Premuzic & Furnham, 2004, 2008), and it is a consistent predictor of academic performance across assessment type (Allick & Realo, 1997; Kappe & van der Flier, 2010; Shute & Ventura, 2013).

Some significant correlations between openness and academic performance have been reported, but correlations with academic performance are not as high (see Table 4). Openness is considered by Chamorro-Premuzic and Furnham (2008) to be a mediator between ability and academic performance. Openness in turn is mediated by learning approach, with open personalities being more likely to adopt a deep learning strategy, which in turn improves academic performance (Swanberg & Martinsen, 2010). Sub-factors of openness, namely intellectual curiosity, creativity, and open-mindedness, have been associated with effective thinking and learning dispositions (Buckingham Shum & Deakin Crick, 2012; Tishman, Jay & Perkins, 1993). Knight et al. (2013) argues that assessment design should nurture such dispositions. Kappe and van der Flier (2010) found that open personalities tend to do better when assessment methods are unconstrained by submission rules.

The relationship between neuroticism and academic performance is not as strong, and like openness, is influenced by assessment type. Neuroticism can have a negative impact on academic performance in stressful examination conditions such as end-of-year exams with time limitations (Hembree, 1988). Where academic performance is measured under less stressful conditions, such as continuous assessment work, the relationship between neuroticism and academic performance is less well-defined

(Chamorro-Premuzic & Furnham, 2009, p. 75). Kappe and van der Flier (2010) found neuroticism to be positively correlated with academic performance (r=0.18, p<0.05, n=133) when assessment is free from time constraints and supervision.

Table 4: Correlations between temperament and academic performance

Study	n	Age	Academic Performance	Conscien-tious	Open	Extrovert	Neurotic	Agreeable
Chamorro-Premuzic & Furnham (2008)	158	18-21	GPA	0.37**	0.21**	0.16	-0.05	0.02
Chamorro-Premuzic & Furnham (2003)[+]	70	17-21	grades	0.33**	-0.06	0.05	-0.28**	0.34**
Conrad (2006)	300	m=19.48	GPA self-reported	0.35*	-0.02	0	-0.6	0.11
Dollinger et al. (2008)	338	m=21.9	GPA	0.26*	0.03	0.02	0.05	0.16*
Duff et al. (2004)	146	17-52	GPA	0.21	0.06	0.06	-0.13	0.115
Gray & Watson (2002)	300	18-21	GPA	0.36*	0.18*	-0.09	0	0.15*
Kappe & van der Flier (2010)[++]	133	18-22	GPA	0.46**	-0.08	0.05	-0.06	0
Kaufman et al. (2008)[+++]	315	m=23.5	GPA	0.18	0.12	0.03	0.07	0.06
Komarraju et al. (2011)	308	18-24	GPA self-reported	0.29**	0.13*	0.07	0	0.22**
O'Connor & Paunonen (2007)	meta-analysis		various	0.24	0.05	-0.05	-0.03	0.06
Trapmann et al. (2007)	Meta-analysis		GPA	0.216	0.083	0.011	-0.044	0.041

*p<.05, **p<.01, m=mean, n=number of participants

\+ Figures based on 1st year exam results. Correlations for agreeableness were lower for 2nd and 3rd year results (0.06 and 0.03 respectively).

++ Matched exam technique to personality type.

+++ Measured Emotional Stability, the reverse of Neuroticism.

Research is inconsistent regarding the remaining two personality dimensions of extroversion and agreeableness and their relationship with academic performance. Introverts tend to have better study habits and are less easily distracted (Entwistle & Entwistle, 1970 as cited in Chamorro-Premuzic &

Furnham, 2009, p. 78), while extroverts tend to perform better in class participation, oral exams, seminar presentations, and multiple-choice style questions (Furnham & Medhurst, 1995; Kappe & van der Flier, 2010). In their meta-analysis of a number of studies investigating personality as a predictor of academic performance, O'Connor and Paunonen (2007) concluded agreeableness is not associated with academic performance. Farsides and Woodfield (2003) found that agreeableness, while not related to academic performance, was linked to other performance indicators such as attendance record. Chamorro-Premuzic and Furnham (2003) agreed, and found high correlations between academic performance and agreeableness were not replicated in later years of the study, but agreeableness was correlated with absenteeism in first year of study.

2.3. Motivation and Correlations with Academic Performance

Ryan and Deci (2000) define motivation simply as being "moved to do something." Defining how learners are motivated to behave in a certain way, and more specifically to learn, is more complex, and is characterized by a range of complementary theories that aim to explain both the level of individual motivation and the nature of the motivation (Steel & Konig, 2006). Current theories in turn encompass a number of factors, some of which are relevant, directly or indirectly, to academic performance (Robbins et al., 2004). Informed by the categorization of motivation theories relevant to academic achievement proposed by Robbins et al. (2004), the following sections discuss three such theories relating to expectancy, goals, and needs.

2.3.1. Expectancy Theory of Motivation

Expectancy models of motivation explore the extent to which a person regards outcome as being a consequence of behaviour. Levels of expectancy motivation are therefore influenced by the extent to which a person believes he or she is in control of the outcome (Cassidy, 2011). There are two strands of expectancy motivation (Eccles & Wigfield, 2002; Pintrich & DeGroot, 1990):

1. *Outcome Expectation* refers to a belief that a particular behaviour will lead to a particular outcome, e.g., active engagement in class work results in better grades;
2. *Self-Efficacy* refers to a person's belief that they can achieve that outcome (e.g., I can actively engage in class and so I can achieve better grades). High self-efficacy is associated with setting more challenging goals, a willingness to work hard, and persistence with a task.

Table 5 gives a summary of correlations found between expectancy motivation and academic performance. A meta-analysis of a range of studies recorded correlations varying between 0.38 and 0.5 (Brown et al. 2008). A number of studies identified self-efficacy as a useful predictor of academic performance (Brady-Amoon & Fuertes, 2011; Cassidy, 2011; Yusuf, 2011). Indirect relationships between self-efficacy and academic performance mediated either by other motivational factors or learning strategies are also cited (Breiman, 2001; Yusuf, 2011). On the other hand, Pintrich & DeGroot (1990) found that self-efficacy was not significantly related to performance when cognitive engagement variables such as engagement in the learning process, self-regulation, and learning strategies were also

considered, thereby concluding that self-efficacy facilitates cognitive engagement, but cognitive engagement itself is more directly linked to academic performance.

2.3.2. Goal Theory of Motivation

High self-efficacy is associated with a student setting challenging goals in terms of academic achievement. Such achievement goals fall into two categories: performance goals, where an individual is looking for favourable feedback, and learning goals, where an individual desires to increase competency (Covington, 2000; Dweck, 1986; Dweck & Leggett, 1988; Eccles & Wigfield, 2002; Eppler & Harju, 1997). Performance-oriented goals are associated with a tendency to engage in tasks in which a student is guaranteed to excel, and avoid tasks that may highlight incompetence (Dweck, 1986). This approach can inhibit a student from challenging and enhancing existing competencies. It is also associated with superficial cognitive processing and inefficient use of study time (Covington, 2000). Learning goals are motivated by the need or desire to increase existing competencies and master new skills and, therefore, tend to be more challenging in nature (Covington, 2000). Learning goals are associated with high self-efficacy, a belief that ability is dynamic, and a belief that increased effort will result in increased success (outcome expectancy). This is regarded as an important learning disposition (Buckingham Shum & Deakin Crick, 2012). Interestingly, Dweck (1986) found that there was no relationship between a child's academic ability (at age 14) and his or her goal orientation. Instead, goal orientation was influenced by the perception of ability as being fixed (resulting in a performance-goal orientation) or dynamic (resulting in a learning-goal orientation).

Table 5: Correlations between expectancy motivation and academic performance

Study	n	Age	Academic performance	Self-efficacy	Outcome Expectancy
Brady-Amoon & Fuertes (2011)	271	m=21.26	GPA	0.22*	
Bruinsma (2004)	117	18	Y1 credits	0.26**	
Cassidy (2011)	97	m=23.5	GPA	0.397***	0.195
DiBenedetto & Bembenutty (2013)	113	18+	Module grade	0.37**	0.08
Diseth (2011)	177	m=21.21	Specific exam	0.44**	
Klassen et al. (2008)	261	m=23.3	Self-reported GPA	0.36**	
Komarraju & Nadler (2013)	257	18+	GPA	0.3**	
Robbins et al. (2004)	Meta-analysis		GPA	0.496	

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

Studies have found learning goals to be more strongly correlated with academic performance than performance goals (see Table 6). A contributing factor to the exception in the study conducted by Diseth (2011) could be in how academic performance was measured. Unlike the other cited studies, Diseth

(2011) was based on an exam grade (A-F) from a single six-hour exam. Eppler and Harju (1997) found a statistically significant difference in the average GPA of students with high learning goals (some of whom also had high performance goals) and those with both low learning goals and low performance goals, with learning goals accounting for 9% of the variance in academic performance. They also found older students to be stronger in their endorsement of learning goals, while younger students tended towards performance-oriented goals.

2.3.3. Self-determination theory (needs-based motivation)

Self-Determination Theory (SDT) focuses on our innate psychological need for competency (Deci & Ryan, 2000) and aims to explore the difference in the types of goals learners adopted, and the justification. SDT distinguishes between intrinsic motivation, where motivation arises from enjoyment of activity, and extrinsic motivation, where the outcome is attractive (Deci & Ryan, 2000). It has been argued that this is one factor represented as a continuum from an intrinsic, behaviour-oriented state, to an extrinsic, goal-oriented state (Apter, 1989; Atherton, 2009; Entwhistle, 2005). Alternatively, SDT has been viewed as two separate factors that can both be present (Dweck & Leggett, 1988; Eppler & Harju, 1997). Individuals can alter between intrinsic or extrinsic motivation, depending on the time or situation, but will generally be predisposed to one or the other (Apter, 1989). Cury et al. (2002) found that both performance and learning goals are associated with improving a student's level of intrinsic motivation. For more detailed discussions, see Apter (1989), Entwhistle (2005), and Ryan and Deci (2000).

Table 6: Correlations between goal orientation and academic performance

Study	n	Age	Academic Performance	Learning goals	Performance goals
Diseth (2011)	177	m=21.2	Specific exam	0.21**	0.39**
Dollinger et al. (2008)+	338	m=21.9	Exam performance	0.21**	
Eppler & Harju (1997)	212	m=19.2	GPA	0.3***	0.13
Eppler & Harju (1997)	50	m=29.8	GPA	0.28*	0.08
Robbins et al. (2004)+	meta-analysis		GPA	0.179	
Wolters (1998)	115	m=19.1	Average grade	0.36***	-0.21*

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

+These studies cited correlations for achievement goals in general, rather than learning or performance goals specifically.

Correlations with academic performance tend to be higher for intrinsic motivation than extrinsic motivation, but self-determination is not as strong, or as consistent, a predictor of academic performance as either self-efficacy or learning goals (see Table 7). Goodman et al. (2011) found both intrinsic and extrinsic motivation to be significantly correlated with academic performance; however, the selection of participants in this study could have introduced bias. Students were invited to take part by email, with responders being entered into a prize draw. There was a 6.3% response rate. Komarraju, Karau, and Schmeck (2009) found significant correlation between intrinsic motivation and academic

performance in a study of participants from a variety of disciplines. The study included three sub-factors of intrinsic motivation from the Academic Motivations Scale (AMS): motivation to know (r=0.17, p<0.01), motivation to accomplish (r=0.22, p<0.01), and motivation to experience stimulation (r=0.13, p<0.05). In a later study, Komarraju and Nadler (2013) found the correlation between intrinsic motivation and GPA was not significant when using a shorter 4-item scale to measure intrinsic motivation, the Motivated Strategies for Learning Questionnaire (MSLQ, Pintrich et al., 1991). Kaufman et al. (2008), in a study of non-standard students from a diversity of ethnic backgrounds and using a 60-item motivation scale, did not find correlations to be significant, suggesting that factors impacting on academic performance can vary for different student groups.

Table 7: Correlations between self-determination and academic performance

Study	n	Age	Academic Performance	Intrinsic motivation	Extrinsic motivation
Bruinsma (2004)	117	m=18	Y1 credits	0.09	
Goodman et al. (2011)	254	[17-29]	GPA	0.281**	0.205**
Kaufman et al. (2008)	315	m=23.5	GPA	0.08	-0.05
Komarraju et al. (2009)	308	18-24	self-reported GPA	0.2**	0.11
Komarraju & Nadler (2013)	257	m=20.48	self-reported GPA	0.11	0.05
Wolters (1998)	115	m=19.1	Average grade	0.14	0.05

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

2.3.1. Impact of motivation on academic performance

While many studies cite correlations between academic performance and various measures of motivation, particularly self-efficacy, learning goals, and intrinsic motivation, evidence supporting causal relationships between motivation and academic performance are less consistent, and are influenced to some extent by the selection of factors included in any specific study. For example, Chamorro-Premuzic and Furnham (2003) and Breiman (2001) found motivation was a mediator between conscientiousness and performance, while Komarraju et al. (2009) found conscientiousness mediated between intrinsic motivation and performance. Komarraju et al. (2009) also noted that motivation did not account for any additional variance on academic performance beyond what was already explained by the Big Five. Brown et al. (2008) on the other hand, in a study not including personality factors, found that self-efficacy had a causal relationship with academic performance. In a meta-analysis covering a range of psychosocial and study skills impacting on academic performance at the tertiary level, excluding personality factors, Robbins et al. (2004) found self-efficacy and achievement motivation to be the best predictors of GPA attained by learners. A number of studies investigating both personality and motivation argue that personality-based factors are a better predictor of academic performance than motivation (De Feyter et al., 2012; Komarraju et al., 2009). However Zuffianó et al. (2013) found that self-efficacy significantly contributed to the explained variance in academic performance over and above ability and personality. It also has a more practical value in that self-efficacy beliefs are more easily

changed than ability or personality. This would suggest that while correlations exist between factors of personality and motivation, factors of personality, particularly conscientiousness, and factors of motivation, particularly self-efficacy and achievement goals, each have value, and are worth further consideration in models of student learning.

2.4. Defining learning strategies and their relationship with academic performance

A number of studies have found that the relationship between academic performance and temperament or motivation is mediated by a student's approach to the learning task itself. Important factors include learning style (e.g., Bruinsma, 2004; Chamorro-Premuzic & Furnham, 2008; Diseth, 2011; Sins et al., 2008) and self-regulation (e.g., Nasiriyan et al., 2011; Ning & Downing, 2010). The following sections discuss both learning styles and self-regulation.

2.4.1. Learning style constructs

Many constructs and frameworks exist for learning styles: instructional preference, information processing style, and cognitive personality style (see Coffield et al. (2004) for a detailed review). Approaches to learning have their foundation in the work of Marton and Säljö (2005) who classified learners as shallow or deep. Deep learners aim to understand content, while shallow learners aim to memorize content regardless of their level of understanding. Later studies added strategic learners as a third category (Entwhistle; 2005, p. 19) whose priority is to do well, and will adopt either a shallow or a deep learning approach, depending on the requisites for academic success. Both personality and self-determined motivation are indicative of personal approaches to learning. Openness, conscientiousness, and intrinsic motivation are correlated with a deep learning approach, while neuroticism, agreeableness, and extrinsic motivation are associated with a shallow learning approach (Busato et al., 1999; Duff et al., 2004; Marton & Säljö, 2005).

Table 8: Correlations between learning orientation and academic performance

Study	N	Age	Academic Performance	Deep	Shallow	Strategic
Cassidy (2011)	97	m=23.5	GPA	0.308**	-0.013	0.316**
Chamorro-Premuzic & Furnham (2008)	158	m=19.2	GPA	0.33**	-0.15	0.18*
Duff et al. (2004)	46	m=24.3	GPA	0.097	-0.054	0.153
Snelgrove (2004)	289	18+	GPA	0.20*	-0.13	0.17*
Swanberg & Martinsen (2010)	687	m=24.5	Single exam	0.16	-0.25	

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

Many studies concur with a negative correlation between a shallow learning approach and academic performance (see summary in Table 8). Some studies show higher correlations between academic performance and a deep learning approach (e.g., Chamorro-Premuzic & Furnham, 2008; Snelgrove,

2004), while others cite marginally higher correlations with a strategic learning approach (e.g., Cassidy, 2011; Duff et al., 2004). Volet (1996) found the importance of learning approach varied with assessment type. A lack of correlation between a deep learning approach and academic performance is in itself an insightful result, as it suggests an assessment design that fails to reward an important, malleable learning disposition (Buckingham Shum & Deakin Crick, 2012; Knight et al., 2013), and hence, may elicit secondary, follow-up actions.

2.4.2. Self-regulated learning

Self-regulated learning is recognized as a complex concept that overlaps with a number of other concepts including temperament, learning approach, and motivation, specifically self-efficacy and goal setting (Bidjerano & Dai, 2007; Boekaerts, 1996). While many students may set goals, the ability to self-regulate learning can be the difference between achieving, or not achieving, the goals set (Covington, 2000). Self-regulated learners take responsibility for setting and achieving their own learning goals by planning their learning, having effective time management, using appropriate learning strategies, continually monitoring and evaluating the quality of their own learning, and altering their learning strategies when required (Schunk, 2005; Zimmerman, 1990). Such learners regard learning as a process they can control, but their motivation factors can vary (Pintrich & DeGroot, 1990). To be motivated to self-regulate, a learner must be confident in setting goals and organizing study, and also be confident that study efforts will result in good marks (high self-efficacy). Such learners must also accept delayed gratification as self-regulation requires students to focus on long-term gains for their effort (Bembenutty, 2009; Komarraju & Nadler, 2013; Zimmerman, 1990; Zimmerman & Kitsantas, 2005). Volet (1996) argues that self-regulated learning is more significant in the tertiary level than earlier levels of education because of the shift from a teacher-controlled environment to one of self-regulated study.

A number of studies cite significant correlations between academic performance and factors of self-regulation (see Table 9 for a summary). For example, a longitudinal study of first year students (n=581) found academic performance to be more strongly correlated with self-testing strategies ($r=0.48$, $p<0.001$) and monitoring levels of understanding ($r=0.42$, $p<0.001$) than effort management ($r=0.24$, $p<0.01$) (Ning & Downing, 2010). Conversely, in a study of undergraduates across all years of study, Komarraju and Nadler (2013) found effort management to have a higher correlation with academic performance ($r=0.39$, $p<0.01$) than other factors of self-regulation. They also found that monitoring and evaluating learning aspects of self-regulation did not account for any additional variance in academic performance over and above self-efficacy, but study effort and time did account for additional variance. In a longitudinal study on the causal dilemma between motivation and self-regulation, De Clercq et al. (2013) concluded that a learning goal orientation resulted in a deep learning approach, which in turn resulted in better self-regulation. A study comparing the relative importance of both learning approach (deep or shallow) and learning effort found that learning effort had a higher impact on academic performance than learning approach (Volet, 1996).

2.5. Regression Models of Academic Performance Based on Psychometric Variables

Table 10 presents examples of hierarchical regression models that have attempted to explain variance in academic performance. Relatively high levels of model accuracy related to studies that include factors of cognitive ability combined with either factors of personality or motivation, along with some additional factors such as age and time spent studying. Cassidy (2011) accounted for 53% of the variance in a regression model including prior academic performance, self-efficacy, and age (n=97). However, the high model accuracy may be due to the measure of prior academic performance used (first year GPA). Chamorro-Premuzic and Furnham (2008) accounted for 40% of the variance in a regression model that included prior academic ability, personality factors, and a deep learning strategy. A similar proportion of variance (44%) was reported by Dollinger et al. (2008) in a regression model including prior academic ability, personality factors, academic goals, and study time. Not all studies concur with these results.

Table 9: Correlations between academic performance and self-regulation

Study	N	Age	Academic Performance	Effort regulation	Time management	Self-regulation
Bidjerano & Dai (2007)	217	m=22	GPA, self-reported	0.23**	0.33**	
Dollinger et al. (2008)	338	m=21.9	exam performance		0.21**	
Goodman et al. (2011)	254	[17-29]	GPA	0.276**		
Komarraju & Nadler (2013)	257	18+	GPA	0.39**	0.31**	0.14*
Ning & Downing (2010)	581	m=20.24	GPA	0.24**		0.42***
Snelgrove (2004)	289	18+	GPA			0.26*
Sundre & Kitsantas (2004)	62	18-24	Single MCQ			0.35**

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

Both Kaufman et al. (2008) and Swanberg and Martinsen (2010) accounted for lower levels of variance when modelling non-standard students. Kaufman et al. (2008) reported accounting for 14% of the variance in a model with prior academic performance, personality factors and self-determined motivation, when modelling students from a variety of ethnic backgrounds. Swanberg and Martinsen (2010) accounted for 21% of variance in a model with prior academic performance, personality, learning strategy, age, and gender when modelling students with an older average age (m=24.8). Lower variances were also reported in studies not including ability. Komarraju et al. (2011) accounted for 15% of the variance in a model including personality and learning approach. Eppler and Harju (1997) accounted for 11% of the variance in a model including factors of motivation and work commitments, while Bidjerano and Dai (2007) also accounted for an 11% variance in a model including factors of personality and self-regulation. These results suggest that ability is an important determinant of academic performance, particularly in models of standard students. Authors also found that psychometric variables accounted for additional variance beyond that accounted for by prior academic performance (Cassidy, 2011;

Chamorro-Premuzic & Furnham, 2008; Dollinger et al., 2008; Kaufman et al., 2008; Swanberg & Martinsen, 2010).

3. ANALYSIS TECHNIQUES USED ON EDUCATIONAL DATA

Statistical models have dominated data analysis in the social sciences, including educational psychology (Dekker et al., 2009; Freedman, 1987; Herzog, 2006). For example, the studies cited in section 2 primarily used correlation (78% of the studies) and regression (54% of the studies), with some papers citing path analysis results (14%) and structural equation models (11%). Statistical modelling has a sound theoretical basis, allowing verifiable conclusions to be drawn from model coefficients; therefore, statistical models have made, and will continue to make, a valuable contribution to the understanding of learners and the learning process. However, such models are based on assumptions, including assumptions of normality, independency, linear additively, and constant variance (Nisbet et al., 2009). It is evident from current knowledge of the factors influencing academic performance, that such factors are interdependent (Prinsloo et al., 2012). While each factor measures unique attributes, overlaps occur in the constructs being measured. In addition, there is evidence to suggest variance is not constant for all attributes. For example, De Feyter et al. (2012) found that low levels of self-efficacy had a positive, direct effect on academic performance for neurotic students only, and for stable students, average or higher levels of self-efficacy had a direct effect on academic performance. In addition, Vancouver and Kendall (2006) found evidence that high levels of self-efficacy can lead to overconfidence regarding exam preparedness, which in turn can have a negative impact on academic performance. Similarly, Poropat (2009) cites evidence of non-linear relationships between factors of personality and academic performance, including conscientiousness and openness. Duff et al. (2004) observed that because academic performance is itself a complex measure, calculated as an aggregate of a variety of assessment types, this weakens the result of correlation analysis with other learning dimensions. While recognizing the continuing importance of statistical models, Freedman (1987) and Breiman (2001) argued that alternative-modelling approaches should be considered when dimensionality is high, and relationships are complex such as in the social sciences. Cox, in a response to Breiman's paper, notes the importance of the probabilistic base of standard statistical modelling, but agrees with Breiman that in some circumstances, an empirical approach is better (Breiman, 2001, p. 18). It is therefore pertinent to ask if data mining's empirical modelling approach can add value to psychometric data analysis, in particular their relevance to models of academic achievement.

Data mining is a relatively young field that has evolved primarily to aid the extraction of information from the vast amounts of data accumulated in databases and data repositories in many domains (Larose, 2005). The wide range of analytical techniques used in data mining emanate from a variety of disciplines including database systems, statistics, machine learning, visualization, logic, spatial analysis, signal processing, image analysis, information retrieval, and natural language processing, thereby making data mining itself a diverse, interdisciplinary field of study (Han & Kamber, 2006). Data mining uses inductive reasoning to find strong evidence of a conclusion. While suited to big data analysis, it does not provide the statistical certainty offered by traditional statistical modelling (Nisbet et al., 2009).

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

Table 10: Regression models with beta values for significant attributes.

				Ability		Personality				Motivation				Learning Style			Self-regulation		Other factors		
Study	n	Age[+]	Vari-ance	g	prior	C	O	N	A	SE	AG	IM	AM	De	Sh	St	Effort	Time	Age	Gen-der	Job
Bidjerano & Dai (2007)	217	22	18%	0.28													0.27				
Bidjerano & Dai (2007)	217	22	11%			0.14											0.31				
Chamorro-Premuzic & Furnham (2008)	158	19.2	40%	0.29		0.49								0.21							
Cassidy (2011)	97	23.50	53%		0.54					0.26									0.36		
Dollinger et al. (2008)	338	21.9	43%	0.44	0.32													0.21			-0.1
Duff et al. (2004)	146	[17-52]	34%		0.39	0.37													0.3		
Eppler & Harju (1997)	262	21.2	21.5%	0.3								0.34									-0.14
Eppler & Harju (1997)	262	21.2	11%									0.32									-0.16
Kaufman et al. (2008)	315	25.9	14%		0.24	0.12						0.15	0.16								
Komarraju et al. (2011)	308	[18-24]	15%			0.33	0.14	0.19	0.15												
Swanberg & Martinsen (2010)	687	24.5	21%		0.3										0.15	-0.17				-0.14	

g=general cognitive intelligence; C=Conscientiousness; O=Openness; N=Neurotic; A=Agreeableness; SE=self-efficacy; AG=academic goals; IM=Intrinsic Motivation; AM=Achievement Motivation; De=Deep; Sh=Shallow; St=Strategic.

[+] Mean age, except where a range of ages is given.

Algorithms typically used on educational data include the following: a) clustering techniques to identify homogenous subgroups in a dataset; b) association analysis to identify values that frequently co-occur; c) classification techniques to build models that predict membership of predefined classes in a dataset; and d) visual analytics to facilitate human analysis via interactive visual representations of the data (Baelpler & Murdoch, 2010; Romero & Ventura, 2007). A review of mining approaches used in educational data mining by Baker and Yacef (2010) identified a recent predominance of classification techniques, which are reviewed in the following section.

3.1. Classification Algorithms Used on Educational Data

A Decision Tree (DT) algorithm identifies patterns in a dataset as conditions, represented visually as a decision tree (Quinlan, 1986). For example, the following two conditions depict a branch of depth two that capture characteristics of instances in a class "grade=good": "*if Conscientiousness > 5.6 and Self-Efficacy > 6.3 then Grade = Good*." The size of the tree (rule depth) is configurable, influencing the specificity of the resulting model (Quinlan, 1986). Simpler implementations (e.g., C5.0) limit each branch to value ranges from a single attribute, making this a linear classifier with a further restriction that each condition is an axis-parallel hyperplane (Tan et al., 2006). Less restrictive implementations can incorporate a greater range of patterns (e.g., CART, Breiman et al., 1984). Model interpretability makes decision trees a popular choice (Han & Kamber, 2006).

Rule-based classifiers define class membership based on a set of *if…then…* rules. Basic implementations generate models similar to a decision tree model (Tan et al., 2006) despite the difference in search strategies used. Rule-based classifiers implement a depth first search; decision trees implement a breath first search (Gupta & Toshniwal, 2011). However, rule-based classifiers can be extended to incorporate fuzzy rules with less precise conditions, allowing an instance to match more than one class. For example "*if Conscientiousness is 'very' good and Self-Efficacy is 'fairly' good then atRisk = False*" uses the fuzzy sets "very" and "fairly" instead of specific value ranges. This non-deterministic model of the data can represent more complex, non-linear class boundaries (Otero & Sánchez, 2005; Tang et al., 2012).

Models based on Bayes Theory include Naïve Bayes and Bayesian Networks. Naïve Bayes builds a model of probabilities based on both the distribution of classes in a dataset, and the distribution of attribute values present in each class. It then applies Bayes theorem to estimate the probability of class membership for any given combination of attribute values (Ng & Jordon, 2001). For example, a result could be "*P(atRisk=false | gender=female and self-efficacy=0.7) = 0.063*; *P(atRisk=true | gender=female and self-efficacy=0.7)=0.0001."* Naïve Bayes works well with a variety of data types (Tan et al., 2006), and can converge to its optimal accuracy quickly, making it suitable for relatively small datasets (Ng & Jordon, 2001). However, Naïve Bayes simplifies the learning task by assuming all attributes are independent. If this assumption is invalid, conditional probabilities between attributes can be modelled as a Bayesian Network (Bekele & Menzel, 2005). Bayesian Knowledge Tracing (BKT), based on a Bayesian Network, is a popular method for estimating student knowledge based on their behaviour on intelligent

tutoring systems. BKT models the probability that a student has learned a skill based on the estimated likelihood that a correct answer is either a guess or knowledge learned, and an incorrect answer is either a slip or lack of knowledge (Baker et al., 2011).

A Neural Network (NN) is an empirical classifier that can approximate any function mapping input values to an output value. Inspired by the biological neural system, a neural network is a network of nodes, connected by weights, which when multiplied by input values and summed, will approximate an output value (Han & Kamber, 2006). Each node can optionally apply an activation function to its output, such as a logistic function, to model a non-linear mapping from inputs to output. Training a network involves adjusting weights to bring the calculated output closer to the actual output. The resulting model may not be optimal, particularly when the solution is non-linear (Tan et al., 2006). Nonetheless, NNs performance has been found to be comparable with other statistical approaches, particularly when approximating complex patterns based on numeric input values (Sargent, 2001; Groth, 2000).

A Support Vector Machine (SVM) models class membership by approximating a hyperplane that defines a linear boundary between two classes (Cortes & Vapnik, 1995). In cases where the class boundary is non-linear, a kernel function can transpose the dataset to a higher number of dimensions, which may provide a linear class boundary (Nisbet et al., 2009, p. 13). Training an SVM is a convex optimization problem to which a globally optimal solution can be found (Tan et al., 2006). While SVMs are limited to numeric attributes and binary classification tasks, Dixon and Brereton (2009) found SVMs outperformed other learners when modelling datasets that are not normally distributed.

k-Nearest Neighbour (*k*-NN) uses instances from the original dataset to classify a new row of data, and so works with the full dataset rather than a generalized model (Tan et al., 2006; Cover & Hart, 1967). For example, a student would be classified according to the class membership of the *k* rows in a dataset most similar to the characteristics of that student, where *k* is a configurable parameter determining neighbourhood size. Decisions made are local, and decision boundaries can be irregular in shape, making k-NN suitable to datasets not easily generalizable because of pattern complexity (Tan et al., 2006).

Ensembles aggregate the predictions of a collection of classification models (Breiman, 1996; Banfield et al., 2004). Individual models within an ensemble can differ based on the subset of data used to train each model, and/or the algorithms used to build each model. There is also a variety of ways to aggregate predictions including averaging, using a voting strategy, or training a learner to identify which model to use for a given instance (Tan et al., 2006, p. 276). While resource intensive in terms of training time, ensembles tend to outperform individual classifiers, particularly when the accuracies of individual learners are relatively poor and their incorrect predictions are uncorrelated (Tan et al., 2006).

3.1.1. Review of Model Performance

Table 11 summarizes a selection of educational data mining studies, the algorithms used, and the

accuracies achieved. A distinction is made between models of log data capturing student actions over time and models of static data, such as prior academic performance, demographic data, and psychometric factors, measured at a point in time. Many publications on student modelling focus on log data gathered from Virtual Learning Environments (VLEs) hosting educational resources and student interaction, or Intelligent Tutoring Systems (ITS) aimed towards curriculum adaptation to each learner by monitoring progress and measuring skill levels (Baker & Yacef, 2010; Tempelaar et al., 2013). Less focus has been given to modelling non-temporal data from outside virtual or online learning environments.

Both Pardos et al. (2011) and Minaei-Bidgoli et al. (2003) recommended an ensemble to predict performance on an ITS, particularly for larger datasets. However, in a comparison of ensembles with individual classifiers to track student knowledge, Baker et al. (2011) concluded that an ensemble was not statistically significantly better than the best individual classifier, a BKT model. Bekele and Menzel (2005), Conati et al. (2002), Jonsson et al. (2005) and Mayo and Mitrovic (2001) argue that Bayesian networks are particularly suited to student models because of the inherent uncertainty in interpreting student behaviour, and the incompleteness of any dataset attempting to capture all factors relevant to classifying students. However Yu et al. (2010) found that while Bayesian networks were suitable for modelling the temporal nature of data from an online learning tool, when data was converted into a single vector per student, more traditional classification approaches gave more accurate results, such as a decision tree ensemble. Romero et al. (2008) achieved the best accuracy using fuzzy rule learning when modelling Moodle (VLE) usage data converted to a single vector per student. Similarly, Merceron and Yacef (2005) achieved high accuracy using a decision tree to predict exam performance based on a single student vector aggregated from their behaviour on an ITS.

In a comparison of models based on prior academic performance and demographic data, Herzog (2006) found decision trees and neural networks had similar performance to logistic regression when modelling datasets with little co-linearity between variables, but outperformed logistic regression when modelling datasets with greater dependencies between variables. Additionally, both decision tree and neural network models identified significant predictor variables that had shown little statistical significance in a regression model. In a comparison of DT, logistic regression and SVM, Lauría et al. (2013) reports comparable performance when modelling prior academic performance, demographic data, and ITS usage data. Gray et al. (2013) agreed that model performance was comparable when modelling students as a single group, but found models capable of representing complex patterns (SVM, NN and k-NN) outperformed other models (DT, logistic regression, Naïve Bayes) when modelling subgroups split by age. Bergin (2006) achieved good accuracy with Naïve Bayes when modelling a small dataset of prior academic performance and psychometric data, and observed that while an ensemble had marginally higher accuracy than Naïve Bayes, it did not justify the additional effort involved in compiling the ensemble.

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

Table 11: Data Mining models for predicting academic performance in tertiary education

Study	Algorithm	Accuracy	n	Class label	Demographic Data	Prior Education	Psychometric data	ITS
Bergin (2006)	Ensemble (stackingC)	82%	102	weak/strong		x	x	
Gray et al., (2013)	SVM	82%	636	weak/strong		x	x	
Herzog (2006)	Decision Tree (C5.0)	83%	4564	degree completion time	x	x		
Dekker et al. (2009)	Decision Tree (J48)	79%	1002	drop out		x		
Lauría et al. (2013)	Decision Tree	87%	6445	weak/strong	x	x		x

Study	Algorithm	Accuracy	n	Class label	VLE	ITS
Baker et al. (2011)	Bayesian Network (BKT)	AUC: 0.7029	76	next question correct		x
Merceron & Yacef (2005)	Decision Tree (C4.5)	87%	224	pass/fail		x
Minaei-Bidgoli et al. (2003)	Ensemble	94%	227	pass/fail		x
Pardos et al. (2011)	Ensemble (Neural Networks)	AUC:0.77	5,422	Performance on ITS		x
Romero et al. (2008)	Fuzzy Rule (MaxLogit-Boost)	62%	438	module performance 4 bins	x	

n=number of instances; AUC=Area under the Curve

4. BENEFITS OF GREATER COLLABORATION BETWEEN EDUCATIONAL PSYCHOLOGY AND LEARNING ANALYTICS

Notably for this literary review, a limited number of educational data mining studies have investigated the role of psychometric factors in models of learning (Buckingham Shum & Deakin Crick, 2012; Shute & Ventura, 2013). Bergin (2006) found that adding self-efficacy and study hours improved model accuracy, but due to the small sample size (n=58) could not draw reliable conclusions from the findings. Lauría et al. (2012) also achieved good model accuracy when modelling psychometric data with prior academic

performance and other demographic attributes. Gray et al. (2013) suggested that while good accuracies can be achieved without the addition of psychometric data, the inclusion of psychometric data could offer better insights into factors influencing academic performance. In addition, including psychometric data in models of learning can provide useful feedback on the learning dispositions that assessment design rewards (Nelson et al., 2012). Buckingham Shum and Deakin Crick (2012) argues for greater recognition of learning dispositions (e.g., persistence, curiosity, awareness of learning,) as important dimensions of learning that should be assessed in conjunction with discipline knowledge. Shute and Ventura (2013) concur, and observe that important competencies such as persistence, openness, and self-efficacy are not currently taught or assessed, despite evidence of their importance. Furthermore, Knight et al. (2013) argues that learning analytics should be more than just generating models, it should become part of the learning process itself, for example, supporting learners in self-regulating their learning through feedback on actions taken. Such developments necessitate that analytics tools acquire psychometric data to capture learner disposition and approaches to learning task. Interestingly, evidence from Shute and Ventura (2013) suggests some learner dispositions can be inferred from their online behaviour (e.g., persistence and creativity).

Learning analytics can offer benefits over and above traditional data analysis methods prevalent in the social sciences, including a greater range of modelling approaches, scalability, analysis of relevant trace data and a quick feedback cycle. Studies cited above suggest data mining algorithms can offer additional insights over and above standard statistical modelling (e.g., Herzog, 2006). In addition, increased use of technology has resulted in a wealth of digital trails generated by learners, providing large volumes of trace data collected during the learning process (Knight et al., 2013). Many data mining algorithms have implementations adapted for this big-data environment, for example, Decision Tree (Ben-Haim & Tom-Tov, 2010), k-NN (Liang et al., 2009), Neural Networks (Gu et al., 2013), SVM and regression (Luo et al., 2012), and supporting tools are available (Prekopcsák et al., 2011), facilitating quick analysis and feedback (Siemens & Long, 2011). Recent developments in learning analytics frameworks (e.g., the learning warehouse, Buckingham Shum & Deakin Crick, 2012) illustrate the potential for learning analytics to support automation of the full life cycle from data gathering through to deployment of recommendations and interventions based on analysis results.

5. CONCLUSION

This review has collated evidence on the importance of psychometric factors in the modelling of academic achievement in tertiary level education. While not accounting for all of the variance in the noted academic performance, learner ability, personality, motivation, and self-regulation have significant relationships with academic performance, and overlap with noteworthy learning dispositions. Since such attributes can be measured prior to student engagement in course work, they facilitate early recognition of learners at risk of failing, inform appropriate interventions, and provide early input to personalized learning environments. Prior academic performance is a good predictor of academic performance for standard students, but it does not perform as well for mature learners or learner groups with ethnic diversity. Conscientiousness is also a strong personality-based predictor of academic

performance, while self-efficacy is the best motivation-based predictor of academic performance. Self-regulation, particularly study time and study effort, are also significant. On these bases, there has been extensive work done by educational psychologists on the evaluation of psychometric predictors of academic performance using parametric models. However, there is evidence that datasets that include psychometric variables are complex in terms of redundancy and non-linearity of relationships, and therefore could be suited to the empirical modelling approaches used in data mining.

To date, the complementary disciplines of learning analytics and educational data mining have focused predominantly on analyzing data systematically gathered in educational settings, which at the tertiary level includes factors of prior academic performance, demographic data, such as age and gender, and data gathered by logs recording student behaviour in online learning environments. Though both are relatively new disciplines, initial results are encouraging across a variety of analysis techniques. However, there is scope for more research investigating the contribution of additional data that could be gathered by tertiary education providers, as well as how this data should be modelled to enhance current student models, and offer actionable feedback on the learning process. Further work is needed to determine if greater inclusion of psychometric data in algorithmic models of student learning can add value to the knowledge learned from these models.

REFERENCES

Ackerman, P.L., & Heggestad, E.D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*(2), 219–245.

ACT (2012). ACT 2012 retention completion summary tables, Technical report, www.act.org.

Allick, J., & Realo, A. (1997). Intelligence, academic abilities, and personality. *Personality and Individual Differences, 23*(5), 809–814.

Apter, M.J. (1989). *Reversal theory: Motivation, emotion and personality*. London: Routledge.

Arnold, K.E., & Pistilli, M.D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Second International Conference on Learning Analytics and Knowledge*, ACM, Vancouver, British Columbia, Canada.

Atherton, J.S. (2009). Learning and teaching: Motivation. Retrieved from http://www.learningandteaching.info/learning/motivation.htm#Levels%20of%20Motivation

Baelpler, P., & Murdoch, C.J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning, 4*(2), 1–9.

Baker, R.S.J.d., Pardos, Z.A., Gowda, S.M., Nooraei, B.B., & Heffernan, N.T. (2011). Ensembling predictions of student knowledge within intelligent tutoring systems. *Proceedings of Nineteenth International Conference on User Modeling, Adaptation, and Personalization,* Girona, Spain, 13–24.

Baker, R.S.J.d., & Yacef, K. (2010). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.

Banfield, R.E., Hall, L.O., Bowyer, K.W., & Bhadoria, D. (2004). A comparison of ensemble creation techniques. *Multiple Classifier Systems, Lecture Notes in Computer Science,* Springer, Berlin, Heidelberg, 223–232.

Bartels, M., Rietveld, M.J.H., Van Baal, G.C., & Boomsma, D.I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research, 5*, 544–553.

Bayer, J., Bydzovská, H., Géryk, J., Obsvac, T., & Popelnsky, L. (2012). Predicting drop-out from social behaviour of students. *Proceedings of the Fifth International Conference on Educational Data Mining,* Chania, Greece, 103–109.

Bekele, R., & Menzel, W. (2005). A Bayesian approach to predict performance of a student (BAPPS): A case with Ethiopian students. *Proceedings of the International Conference on Artificial Intelligence and Applications*, Vienna, Austria.

Bembenutty, H. (2009). Academic delay of gratification, self-regulation of learning, gender differences, and expectancy-value. *Personality and Individual Differences, 26*, 347–352.

Ben-Haim, Y., & Tom-Tov, E. (2010). A streaming parallel decision tree algorithm. *Journal of Machine Learning Research, 11*, 849–872.

Bergin, S. (2006). Statistical and machine learning models to predict programming performance, PhD thesis, Computer Science, National University of Ireland, Maynooth.

Bidjerano, T., & Dai, D.Y. (2007). The relationship between the big-five model of personality and self-regulated learning strategies. *Learning and Individual Differences, 17*, 69–81.

Biggs, J., Kember, D., & Leung, D. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Education Psychology, 71*, 133–149.

Boekaerts, M. (1996). Self-regulated learning at the junction of cognition and motivation. *European Psychologist, 1* (2), 100–112.

Boekaerts, M. (2001). Bringing about change in the classroom: Strengths and weaknesses of the self-regulated learning approach. *EARLI presidential address*, Centre for the Study of Education and Instruction, Leiden 2300 RB, The Netherlands.

Boeree, G. (2006). Personality theories. Retrieved from http://webspace.ship.edu/cgboer/perscontents.html

Brady-Amoon, P., & Fuertes, J.N. (2011). Self-efficacy, self-rated abilities, adjustment and academic performance. *Journal of Counseling and Development, 89*(4)*,* 431–438.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*, 123–140.

Breiman, L. (2001). Statistical modelling: The two cultures. *Statistical Science, 16*(3), 199–231.

Breiman, L., Friedman, J., Stone, C.J., & Olshen, R.A. (1984). *Classification and regression trees*. Boca Raton, FL: Chapman and Hall/CRC.

Brown, S.D., Tramayne, S., Hoxha, D., Telander, K., Fan, X., & Lent, R.W. (2008). Social cognitive predictors of college students' academic performance and persistence: A meta-analytic path analysis. *Journal of Vocational Behaviour, 72*, 298–308.

Bruinsma, M. (2004). Motivation, cognitive processing and achievement in higher education. *Learning and Instruction, 14*, 549–568.

Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. *Second International Conference on Learning Analytics and Knowledge*, Vancouver, British Columbia, Canada.

Busato, V.V., Prins, F.J., Elshout, J.J., & Hamaker, C. (1999). The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. *Personality and Individual Differences, 26,* 129–140.

Cassidy, S. (2011). Exploring individual differences as determining factors in student academic achievement in higher education. *Studies in Higher Education, 37*(7), 1–18.

Cattell, H.E.P., & Mead, A.D. (2008). The Sixteen Personality Factor Questionnaire (16PF). In *The SAGE handbook of personality theory and assessment*, Vol. 2, *Personality measurement and testing* (chapter 7). Thousand Oaks, CA: SAGE Publications Ltd.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality, 37*, 319–338.

Chamorro-Premuzic, T., & Furnham, A. (2004). A possible model for understanding the personality–intelligence interface. *British Journal of Psychology, 95*, 249–264.

Chamorro-Premuzic, T., & Furnham, A. (2008). Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and Individual Differences, 44*, 1596–1603.

Chamorro-Premuzic, T., & Furnham, A. (2006). *Personality and intellectual competence*, Sussex UK: Psychology Press, 1st ed., 68–92.

Chatti, M.A., Dychhoff, A.L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning. Special Issue on State of the Art in TEL,* 318–331.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). Should we be using learning styles? What research has to say to practice. Learning and Skills Research Centre, UK, www.LSRC.ac.uk

Colom, R., & Flores-Mendoza, C. (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence, 35*, 243–251.

Conati, C., Gertner, A., & Vanlehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *Use Modeling and User-Adapted Interaction, 12*, 371–417.

Conrad, M.A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality, 40*, 339–346.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297.

Costa, P.T.J., & McCrae, R.R. (1992). *Revised NEO personality inventory (NEO- PI- R) and NEO Five-Factor inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Cover, T.M., & Hart, P.E. (1967). Nearest neighbour pattern classification. *IEEE Transactions on Information Theory, 13*(1).

Covington, M.V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology, 51*, 171–200.

Cury, F., Elliot, A., Sarrazin, P., Da Finseca, D., & Rufo, M. (2002). The trichotomous achievement goal model and intrinsic motivation: A sequential mediational analysis. *Journal of Experimental Social Psychology, 38*, 473–481.

De Clercq, M., Galand, B., & Frenay, M. (2013). Chicken or the egg: Longitudinal analysis of the causal dilemma between goal orientation, self-regulation and cognitive processing strategies in higher education. *Studies in Educational Evaluation, 39*, 4–13.

De Feyter, T., Caers, R., Vigna, C., & Berings, D. (2012). Unraveling the impact of the big five personality traits on academic performance: The moderating and mediating effects of self-efficacy and academic motivation. *Learning and Individual Differences, 22,* 439–448.

de Raad, B., & Schouwenburg, H.C. (1996). Personality in learning and education: A review. *European Journal of Personality, 10*, 303–336.

Deci, E.L., & Ryan, R.M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268.

Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting students drop out: A case study. In T. Barnes, M.C. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the Second International Conference on Educational Data Mining*, Cordoba, Spain, 41–50.

DiBenedetto, M.K., & Bembenutty, H. (2013). Within the pipeline, self-regulated learning, self-efficacy, and socialization among college students in science courses. *Learning and Individual Differences, 23,* 218–224.

Diseth, Á. (2011). Self-efficacy, goal orientations and learning strategies as mediators between preceding and subsequent academic achievement. *Learning and Individual Differences, 21*, 191–195.

Dixon, S., & Brereton, R. (2009). Comparison of the performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure. *Chemometrics and Intelligent Laboratory Systems, 95*, 1–17.

Dollinger, S.J., Matyja, A.M., & Huber, J.L. (2008). Which factors best account for academic success: Those which college students can control or those they cannot? *Journal of Research in Personality, 42*, 872–885.

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics: Understandings and expectations from the stakeholders. *Second International Conference on Learning Analytics and Knowledge*, ACM, Vancouver, British Columbia, Canada.

Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences, 36*, 1907–1920.

Dweck, C.S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040–1048.

Dweck, C.S., & Leggett, E.L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256–273.

Eccles, J.S., & Wigfield, A. (2002). Motivation beliefs, values and goals. *Annual Review of Psychology, 53*, 109–132.

Entwhistle, N. (2005). Contrasting Perspectives in Learning. In *The Experience of Learning*, chapter 1. Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment. Retrieved from http://www.tla.ed.ac.uk/resources/EoL.html

Entwistle, N., & Entwistle, D. (1970). The relationship between personality, study methods and academic performance. *British Journal of Educational Psychology, 40*, 132–143.

Eppler, M.A., & Harju, B.L. (1997). Achievement motivation goals in relation to academic performance in traditional and non-traditional college students. *Research in Higher Education, 38*(5), 557–573.

Eysenck, H. (1994). *Test Your IQ*. London: Thorsons.

Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual Differences, 34*, 1225–1243.

Flanagan, D.P., & McGrew, K.S. (1998). Interpreting intelligence tests from contemporary gf-gc theory: Joint confirmatory factor analysis of the WJ-R and KAIT in a non-white sample. *Journal of School Psychology, 36*(2), 151–182.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *American Psychological Association, 101*(2), 171–191.

Freedman, D. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics, 12*(2), 101–128.

Frey, M.C., & Detterman, D.K. (2003). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science, 15*(6), 373–378.

Furnham, A., & Medhurst, S. (1995). Personality correlates of academic seminar behaviour: A study of four instruments. *Personality and Individual Differences, 19*(2), 197–208.

Furnham, A., & Zhang, J. (2006). The relationship between psychometric and self-estimated intelligence, creativity, personality, and academic achievement. *Imagination, Cognition and Personality, 25*(2), 119–145.

Goldberg, L.R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4*(1), 26–42.

Goldberg, L.R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26–34.

Goodman, S., Jaffer, T., Keresztesi, M., Mamdani, F., Mokgatle, D., Musariri, M., Pires, J., & Schlechter, A. (2011). An investigation of the relationship between students' motivation and academic performance as mediated by effort. *Psychological Journal of South Africa, 41*(3), 373–385.

Gray, E.K., & Watson, D. (2002). General and specific traits of personality and their relationship to sleep and academic performance. *Journal of Personality, 70*(2), 177–206.

Gray, G., McGuinness, C., & Owende, P. (2013). An investigation of psychometric measures for modelling academic performance in tertiary education. *Sixth International Conference on Educational Data Mining,* Memphis, Tennessee, 240–243.

Groth, R. (2000). *Data mining: Building competitive advantage.* Upper Saddle River, NJ: Prentice Hall

Gu, R., Shen, F., & Huang, Y. (2013). A parallel computing platform for training large scale neural networks. *IEEE International Conference on Big Data,* Santa Clara, California, USA, 376–384.

Gupta, P., & Toshniwal, D. (2011). Performance comparison of rule based classification algorithms. *International Journal of Computer Science and Informatics*, *1*(2), 37–42.

Han, J., & Kamber, M. (2006). *Data mining concepts and techniques*. Burlington, MA: Morgan Kaufmann.

HEA (2013). *Higher education key facts and figures 2011–12*. Retrieved from http://www.hea.ie/en/Publications

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Education Research, 58*, 47–77.

Herzog, S. (2006). Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression. *New Directions for Institutional Research, 131*, 17–33.

Jensen, A.R. (1981). *Straight talk about mental tests*. Glencoe, IL: The Free Press (MacMillan).

John, O.P., Naumann, L.P., & Soto, C.J. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114–158). New York, NY: Guildford Press.

Jonsson, A., Johns, J., Mehranian, H., Arroyo, I., Woolf, B., Barto, A., Fisher, D., & Mahadevan, S. (2005). Evaluating the feasibility of learning student models from data: Technical report, American Association for Artificial Intelligence.

Kappe, R., & van der Flier, H. (2010). Using multiple and specific criteria to assess the predictive validity of the big five personality factors on academic performance. *Journal of Research in Personality, 44*, 142–145.

Kaufman, J.C., Agars, M.D., & Lopez-Wagner, M.C. (2008). The role of personality and motivation in predicting early college academic success in non-traditional students at a Hispanic-serving institution. *Learning and Individual Differences, 18*, 492–496.

Klassen, R.M., Krawchuk, L.L., & Rajani, S. (2008). Academic procrastination of undergraduates: Low self-efficacy to self-regulate predicts higher levels of procrastination. *Contemporary Educational Psychology, 33*, 915–931.

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Epistemology, pedagogy, assessment and learning analytics. *Third Conference on Learning Analytics and Knowledge* (LAK 2013). Leuven, Belgium.

Kobrin, J., Patterson, B.F., Shaw, E.J., Mattern, K.D., & Barbuti, S.M. (2008). Validity of the SAT for predicting first year college grade point average, Research Report. *College Board New York*, 2008–2005.

Komarraju, M., Karau, S.J., & Schmeck, R.R. (2009). Role of the big five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences, 19*, 47–52.

Komarraju, M., Karau, S.J., Schmeck, R.R., & Avdic, A. (2011). The big five personality traits, learning styles, and academic achievement. *Personality and Individual Differences, 51*, 472–477.

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences, 25,* 67–75.

Komarraju, M., Ramsey, A., & Rinella, V. (2013). Cognitive and non-cognitive predictors of college readiness and performance: Role academic discipline. *Learning and Individual Differences, 24,* 103–109.

Kuncel, N.R., Hezlett, S.A., & Ones, D.S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1), 148–161.

Larose, D.T. (2005). *Discovering knowledge in data: An introduction to data mining*. Hoboken, NJ: John Wiley and Sons, Inc.

Lauría, E.J.M., Baron, J.D., & Devireddy, M. (2012). Mining academic data to improve college student retention: An open source perspective. *Second International Conference on Learning Analytics and Knowledge* (LAK 2012). ACM, Vancouver, British Columbia, Canada.

Lauría, E.J.M., Moody, E.W., Jayaprakash, S.M., Jonnalagadda, N., & Baron, J.D. (2013). Open academic analytics initiative: Initial research findings. *Third Conference on Learning Analytics and Knowledge* (LAK 2013). ACM, Leuven, Belgium.

Lent, R.W., Brown, S.D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behaviour, 45*, 79–122.

Liang, S., Liu, Y., Wang, C., & Jian, L. (2009). A CUDA-based parallel implementation of k-nearest neighbor algorithm. *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery.* China, 53–60.

López, M.I., Luna, J.M., Romero, C., & Ventura, S. (2012). Classification via clustering for predicting final marks based on student participation in forums. *Proceedings of the Fifth International Conference on Educational Data Mining,* Chania, Greece, 148–151.

Luo, D., Ding, C., & Huang, H. (2012). Parallelization with multiplicative algorithms for big data mining. *IEEE 12th International Conference on Data Mining,* Brussels, Belgium, 489–498.

Marton, F., & Säljö, R. (2005). Approaches to Learning. In *The Experience of Learning*, chapter 3. Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment. Retrieved from http://www.tla.ed.ac.uk/resources/EoL.html

Matarazzo, J.D., & Goldstein, S.G. (1972). The intellectual caliber of medical students. *Journal of Medical Education, 47*(2), 10.

Mayo, M., & Mitrovic, A. (2001). Optimising ITS behaviour with Bayesian networks and decision theory. *International Journal of Artificial Intelligence in Education, 12*, 124–153.

McCrae, R.R., & Costa, P.T.J. (1991). The NEO personality inventory: Using the five-factor model in counseling. *Journal of Counseling and Development, 69*(4), 367–372.

McGrew, K.S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1–10.

Merceron, A., & Yacef, K. (2005). Educational data mining: A case study. *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED),* Amsterdam, 467–474.

Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer, G., & Punch, W.F. (2003). Predicting student performance: An application of data mining methods with educational web-based system lon-capa. *Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conference,* Boulder, Colorado, USA.

Mislevy, R.J., Behrens, J.T., & Dicerbo, K.E. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining, 4*(1), 11–48.

Mooney, O., Patterson, V., O'Connor, M., & Chantler, A. (2010). A study of progression in higher education: A report by the higher education authority, Technical report, Higher Education Authority, Ireland.

Moran, M.A., & Crowley, M.J. (1979). The leaving certificate and first year university performance. *Journal of Statistical and Social Enquiry in Ireland, 24*(1), 231–266.

Munzert, A.W. (1980). *Test Your IQ*. New York, NY: MacMillan.

Naderi, H., Abdullah, H.T., Sharir, J., & Kumar, V. (2009). Creativity, age and gender as predictors of academic achievement among undergraduate students. *Journal of American Science, 5*(5), 101–112.

Nasiriyan, A., Azar, H.K., Noruzy, A., & Dalvand, M.R. (2011). A model of self-efficacy, task value, achievement goals, effort and mathematics achievement. *International Journal of Academic Research, 3*(2), 612–618.

Nelson B., Nugent R., & Rupp A. (2012). On instructional utility, statistical methodology, and the added value of ECD: Lessons learned from the special issue. *Journal of Educational Data Mining, 4*(1), 227–233.

Ng, A.Y., & Jordon, M.I. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and Naïve Bayes. *Advances in Neural Information Processing Systems (NIPS)*, *14*, 841–848.

Ning, H.K., & Downing, K. (2010). The reciprocal relationship between motivation and self-regulation: A longitudinal study on academic performance. *Learning and Individual Differences, 20*, 682–686.

Nisbet, R., Elder, J., & Miner, G. (2009). *Statistical analysis and data mining applications*. Waltham, MA: Academic Press.

O'Connor, M.C., & Paunonen, S.V. (2007). Big five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*, 971–990.

OECD (2013). Education at a glance 2013. Retrieved from http://www.oecd.org/edu/eag2013%20(eng)--FINAL%2020%20June%202013.pdf

Otero, J., & Sánchez, L. (2005). Induction of descriptive fuzzy classifiers with the logitboost algorithm. *Soft Computing, 10*, 825–835.

Pardos, Z.A., Baker, R.S.J.d., Gowda, S.M., & Heffernan, N.T. (2011). The sum is greater than the parts: Ensembling models of student knowledge in educational software. *SIGKDD Explorations, 13(2)*, 37–44.

Pintrich, P., & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal and Educational Psychology, 82*, 33–40.

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1991). A manual for the use of the motivated strategies for learning questionnaire, Technical Report 91-B-004, The Regents of the University of Michigan.

Poropat, A.E. (2009). A meta-analysis of the five-factor model or personality and academic performance. *Psychological Bulletin, 135*(2), 322–338.

Prekopcsák, Z., Makrai, G., Kenk, T., & Gáspár-Papanek, C. (2011). Radoop: Analysing big data with rapidminer and hadoop. *Second Rapidminer Community Meeting and Conference* (RCOMM), Dublin, Ireland.

Prinsloo, P., Slade, S., & Galpin, F. (2012). Learning analytics: Challenges, paradoxes and opportunities for mega open distance learning institutions. *Second International Conference on Learning Analytics and Knowledge*. ACM, Vancouver, British Columbia, Canada.

Quinlan, J.R. (1986). Simplifying decision trees, AI Memo 930, Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Robbins, S.B., Lauver, K., Le, H., Davis, D., & Langley, R. (2004). Do psychosocial and study skill factors predict college outcomes? A meta analysis. *Psychological Bulletin, 130*(2), 261–288.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*, 135–146.

Romero, C., Ventura, S., Espejo, P.G., & Hervás, C. (2008). Data mining algorithms to classify students. *Proceedings of the First International Conference on Educational Data Mining,* Montreal, Canada, 8–17.

Ryan, R.M., & Deci, E.L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67.

Sachin, R.B., & Vijay, M.S. (2012). A survey and future vision of data mining in educational field. *Second International Conference on Advanced Computing & Communication Technologies (ACCT),* Rohtak, India, 96–100.

Sargent, D. (2001). Comparison of artificial neural networks with other statistical approaches. *Conference on Prognostic Factors and Staging Cancer Management, 91*(8), 1636–1642.

Schmitt, N., Oswald, F.L., Pleskac, T., Sinha, R., & Zorzie, M. (2009). Prediction of four-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology, 94*(6), 1479–1497.

Schunk, D.H. (2005). Commentary on self-regulation in school contexts. *Learning and Instruction, 15*, 173–177.

Shute, V., & Ventura, M. (2013). Stealth assessment: Measuring and supporting learning in video games. The John D. and Catherine T. MacArthur Foundation reports on digital media and learning.

Siemens, G. (2012). Learning analytics: Envisioning a research discipline and a domain of practice. *Proceedings of the Second International Conference on Learning Analytics and Knowledge,* Vancouver, Canada, 4–8.

Siemens, G., & Baker, R.S.J.d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the Second International Conference on Learning Analytics and Knowledge,* Vancouver, Canada*,* 252–254.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30–23.

Sins, P.H.M., van Joolingen, W.R., Savelsbergh, E.R., & van Hout-Wolters, B. (2008). Motivation and performance within a collaborative computer-based modeling task: Relations between students'

achievement goal orientation, self-efficacy, cognitive processing, and achievement. *Contemporary Educational Psychology, 33*, 58–77.

Snelgrove, S. (2004). Approaches to learning of student nurses. *Nurse Education Today, 24*, 605–614.

Spearman, C. (1904). General intelligence, objectively determined and measured. *Journal of Psychology, 15*, 201–293.

Spearman, C. (1927). *The abilities of man*. London: MacMillan.

Srivastava, S. (2010). Measuring the big five personality factors. Retrieved from http://pages.uoregon.edu/sanjay/bigfive.html

Steel, P., & Konig, C.J. (2006). Integrating theories of motivation. *Academy of Management Review, 31*(4), 889–913.

Sternberg, R. (1999). Intelligence as developing expertise. *Contemporary Educational Psychology, 24*, 359 – 375.

Sundre, D.L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology, 29*, 6–26.

Swanberg, A.B., & Martinsen, Ø.L. (2010). Personality, approaches to learning and achievement. *Educational Psychology, 30*(1), 75–88.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Boston, MA: Pearson Education.

Tang, M., Chen, X., Hu, W., & Yu, W. (2012). Generation of a probabilistic fuzzy rule base by learning from examples. *Information Sciences, 271,* 21–30.

Tempelaar, D.T., Cuypers, H., van de Vrie, E., Heck, A., & van der Kooij, H. (2013). Formative assessment and learning analytics. In E.D. Dan Suthers, Katrien Verbert, E.D. Xavier Ochoa Dan Suthers, Katrien Verbert, D. Xavier Ochoa Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK '13). ACM, New York, NY.

Tishman, S., Jay, E., & Perkins D.N. (1993). Teaching thinking disposition: From transmission to enculturation. *Theory into Practice, 32*, 147–153.

Trapmann, S., Hell, B., Hirn, J.-O.W., & Schuler, H. (2007). Meta-analysis of the relationship between the big five and academic success at university. *Zeitschrift fur Psychologie, 215*(2), 132–151.

Tupes, E.C., & Cristal, R.E. (1961). Recurrent personality factors based on trait ratings (Report No AD 267 778), United States Air Force, Lackland, Texas.

Vancouver, J.B., & Kendall, L.N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology, 91*(5), 1146–1153.

Volet, S.E. (1996). Cognitive and affective variables in academic learning: The significance of direction and effort in students' goals. *Learning and Instruction, 7*(3), 235–254.

Wolters, C.A. (1998). Self-regulated learning and college students' regulation of motivation. *American Psychological Association, 90*(2), 224–235.

Yu, H.-F., Lo, H.-Y., Hsieh, H.-P., Lou, J.-K., McKenzie, T.G., Chou, J.-W., Chung, P.-H., Ho, C.-H., Chang, C.-F., Wei, Y.-H., Weng, J.-Y., Yan, E.-S., Chang, C.-W., Kuo, T.-T., Lo, Y.-C., Chang, P.T., Po, C., Wang,

C.-Y., Huang, Y.-H., Hung, C.-W., Ruan, Y.-X., Lin, Y.-S., Lin, S.-d., Lin, H.-T., & Lin, C.-J. (2010). Feature engineering and classifier ensemble for KDD cup 2010. *JMLR Workshop and Conference Proceedings, 1*, 1–16.

Yusuf, M. (2011). The impact of self-efficacy, achievement motivation, and self-regulated learning strategies on students' academic achievement. *Procedia: Social and Behavioural Sciences, 15*, 2623–2626.

Zimmerman, B.J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3–17.

Zimmerman, B.J., & Kitsantas, A. (2005). Homework practices and academic achievement: The mediating role of self-efficacy and perceived responsibility beliefs. *Contemporary Educational Psychology, 30*, 397–417.

Zuffianó, A., Alessandri, G., Gerbino, M., Kanacri, B.P.L., Di Giunta, L., Milioni, M., & Caprara, G.V. (2013). Academic achievement: The unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits and self-esteem. *Learning and Individual Differences, 23*, 158–162.

JOURNAL OF LEARNING ANALYTICS

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes

Zachary A. Pardos
University of California, Berkeley, USA

Ryan S.J.d. Baker, Maria O.C.Z. San Pedro
Columbia University Teachers College, USA

Sujith M. Gowda, Supreeth M. Gowda
Worcester Polytechnic Institute, USA

pardos@berkeley.edu

ABSTRACT: In this paper, we investigate the correspondence between student affect and behavioural engagement in a web-based tutoring platform throughout the school year and learning outcomes at the end of the year on a high-stakes mathematics exam in a manner that is both longitudinal and fine-grained. Affect and behaviour detectors are used to estimate student affective states and behaviour based on post-hoc analysis of tutor log-data. For every student action in the tutor, the detectors give us an estimated probability that the student is in a state of boredom, engaged concentration, confusion, or frustration, and estimates of the probability that the student is exhibiting off-task or gaming behaviours. We used data from the ASSISTments math tutoring system and found that boredom during problem solving is negatively correlated with performance, as expected; however, boredom is positively correlated with performance when exhibited during scaffolded tutoring. A similar pattern is unexpectedly seen for confusion. Engaged concentration and, surprisingly, frustration are both associated with positive learning outcomes. In a second analysis, we build a unified model that predicts student standardized examination scores from a combination of student affect, disengaged behaviour, and performance within the learning system. This model achieves high overall correlation to standardized exam score, showing that these types of features can effectively infer longer-term learning outcomes.

KEYWORDS: Learning analytics, affect, confusion, boredom, high-stakes tests, tutoring, automated detectors, prediction, data mining

1 INTRODUCTION

In recent years, researchers have increasingly investigated the relationship between fine-grained details of student usage of tutoring systems and performance on high-stakes examinations (cf. Feng, Heffernan, & Koedinger, 2009; Pardos, Wang, & Trivedi, 2012). Understanding how different student behaviours

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

correspond to student outcomes can help us to understand the larger implications of student choices that might seem only momentary. This information can be useful both in terms of advancing theory on meta-cognition and engagement (cf. Aleven, McLaren, Roll, & Koedinger, 2004; Baker, Gowda, & Corbett, 2011), and to provide actionable information for teachers about factors potentially influencing their students' learning outcomes (Arnold, 2010). Within this paper, we analyze the relationships between a student's affect, engagement, and their outcomes. Several studies have indicated that affect and behavioural engagement can lead to differences in learning (Craig, Graesser, Sullins, & Gholson, 2004; Pekrun, Goetz, Titz, & Perry, 2002; Rodrigo et al., 2009; Baker, 2007; Cocea, Hershkovitz, & Baker, 2009); however, past research on these relationships has been limited by the use of observational or survey methods, which are either coarse-grained, or can only be applied over brief periods (year-long field observations are possible, but prohibitively expensive to conduct for large numbers of students). Longitudinal approaches have been used to predict college attendance (San Pedro, Baker, Gowda, & Heffernan, 2013), suggesting that a similar approach may be feasible to predict long-term learning outcomes. Within this paper, we use automated detectors of affect and behavioural engagement that can be applied to every student action in an entire year's log file data to analyze this question, asking how predictive a student's affect and engagement, throughout the school year, is of his or her end-of-year high-stakes test outcome. Specifically, we investigate overall relationships between affect/engagement and learning, and dig deeper to ask if there are some contexts where a particular affect is constructive and others where it is not. We also compare the overall predictiveness of affect and engagement relative to student performance in the learning system. We investigate these questions in the context of two school years of student learning within the ASSISTments tutoring system (Feng et al., 2009), involving over a thousand students.

1.1 The Tutor and the Test

ASSISTments is a web-based tutoring platform, primarily for 7^{th}–12^{th} grade mathematics. Within ASSISTments, shown in Figure 1, students complete mathematics problems and are formatively *assessed* — providing detailed information on their knowledge to their teachers — while being *assisted* with scaffolding, help, and feedback. Items in ASSISTments are designed to correspond to the skills and concepts taught in relevant state standardized examinations. Figure 1 shows how, after the student answers the original question incorrectly, the system provides scaffolding that breaks the problem down into steps. Hints are provided at each step and the student can ask for a bottom-out hint that eventually tells the answer. Students in the data sets studied within this paper used ASSISTments in classroom computer lab sessions targeted towards preparation for the standardized state test, during school hours. While teachers had the ability to assign students questions of a particular skill, the most popular problem set within the data set that will be analyzed in this paper was one that randomly sampled 8^{th} grade math test prep questions from the system. Because of this, students sometimes received questions with skills they had not encountered in class yet. One data set, which was used to develop models of student affect, represented a few days of software usage. The other data set, used to study

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

the relationship between student affect and learning outcomes, represents an entire year of data of students using the ASSISTments system.

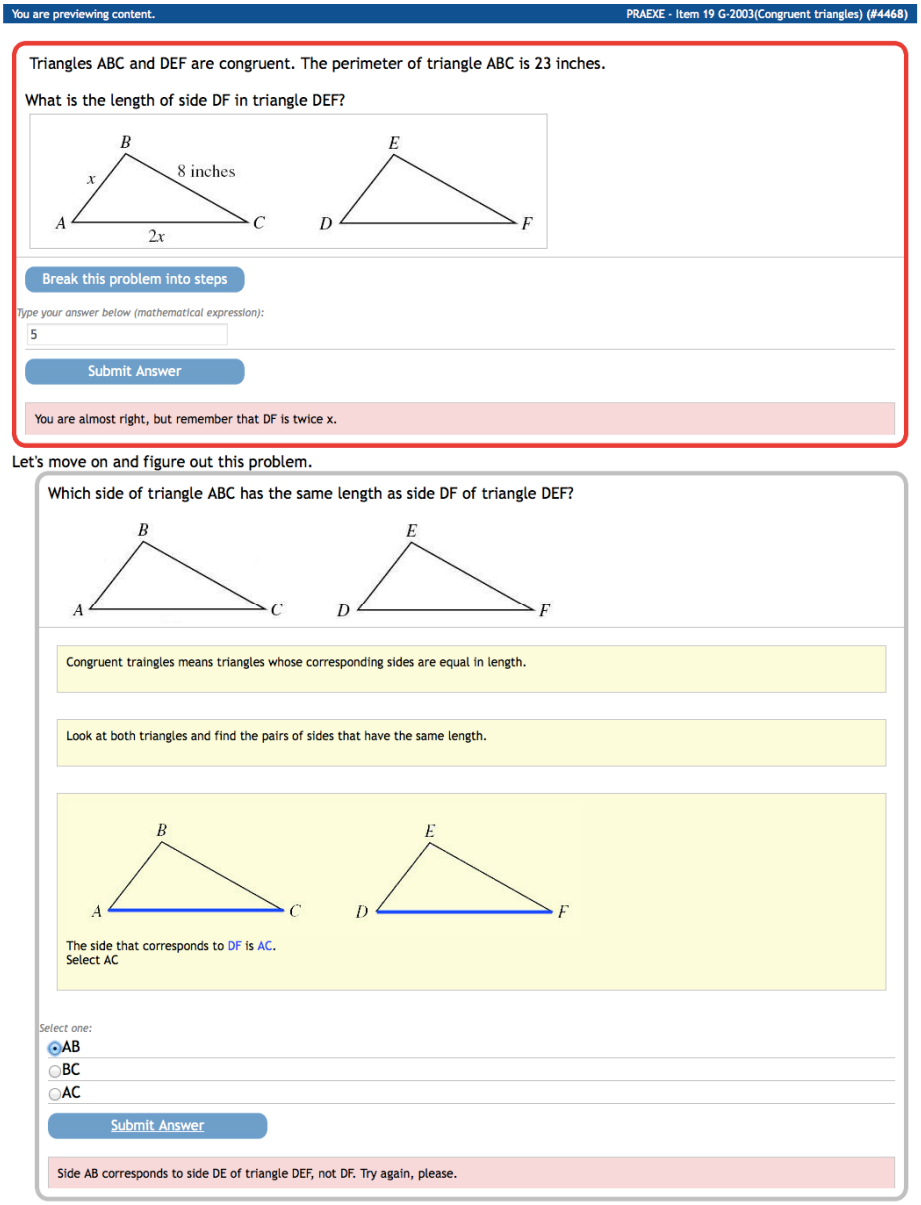

Figure 1: An example of an ASSISTments item where the student answers incorrectly and receives scaffolding help

Near the end of their school year, students took the MCAS (Massachusetts Comprehensive Assessment System) state standardized test. We collected scores for the math portion of the test. Raw scores range from 0 to 54 and are later scaled by the state after all tests are in. The scaling maps four categories;

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

Failing, Needs Improvement, Proficient, and Advanced. Students must score above Failing to graduate high school and an Advanced score earns them an automatic state college scholarship.

2 METHODOLOGY

In this section, we will describe both the methodology for employing the automatic affect detectors to our data set and the methodology for conducting the correlation analysis.

2.1 Affect and Behaviour Detection

In order to assess student affect and behaviour across contexts, we adopt a two-stage process: first labelling student affect and behaviour for a small but reasonably representative sample with field observations (cf. Baker, D'Mello, Rodrigo, & Graesser, 2010), and then using those labels to create automated detectors that can be applied to log files at scale. The detectors are created by synchronizing log files generated by the ASSISTments system with field observations conducted at the same time. To enhance scalability, only log data is used as the basis of the detectors; physical sensors can enhance detector goodness (cf. Conati & Maclaren, 2009; D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008), but reduce the applicability of the resultant models to existing log files. The detectors are constructed using log data from student actions within the software occurring at the same time as or before the observations, making our detectors usable for automated interventions, as well as for the type of discovery with models analysis conducted in this paper. Our process for developing sensor-free affect and behaviour detectors for ASSISTments replicates a process that has been successful for developing affect detectors for a different intelligent tutor, Cognitive Tutor Algebra (Baker et al., 2012).

2.1.1 Data Collection

Two sets of data from ASSISTments were used in this study.

The first data set was used to develop the automated detectors of affect. This data set was composed of field observations of affect and behaviour conducted over several days in an urban middle school in central Massachusetts, sampled from a diverse population of 229 students. Within this school, 40% of students were Hispanic, 14% were African-American, 4% were Asian-American, and 39% were Caucasian. In this school, per capita income was significantly lower than the state average. Information from these observations and the corresponding interaction logs was used to develop and validate the affect detectors discussed below.

The second data set was used to conduct analyses of the relationships between affect and learning. This data set was composed of action log files distilled from a diverse population (racially and socio-economically) of 1,393 students that came from middle schools in the same city in central Massachusetts, in 2004–2005 and 2005–2006 (these years were chosen due to the availability of standardized examination data). In 2004–2005, 629 students used the software and in 2005–2006, the number rose to 764 students. This data set involved students using the software for two hours, twice a

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

week for the entire school year. As this data set represented whole-year usage of the software, 810,000 student actions (entering an answer or requesting help) were represented in the data. The affect models were applied to this larger data set.

2.1.2 Affect and Behaviour Observations

Student affect and behavioural engagement was coded by a pair of expert field observers as students used ASSISTments in 2010. An observation protocol developed for coding affect during the use of educational software (cf. Baker et al., 2010) was implemented using field observation synchronization software (Baker et al., 2012) developed for Google Android handheld devices. Each observation lasted up to twenty seconds, with elapsed observation time so far displayed by the hand-held observation software. If affect or behaviour was labelled before twenty seconds elapsed, the coder moved to the next observation. Each observation was conducted using side-glances, to reduce observer effects. To increase tractability of both coding and eventual analysis, if two distinct affective states were seen during a single observation, only the first state observed was coded. Any affect or behaviour of a student other than the student currently being observed was not coded. The observers based their judgment of a student's affect or behaviour on the student's work context, actions, utterances, facial expressions, body language, and interactions with teachers or fellow students. These are, broadly, the same types of information used in previous methods for coding affect (e.g., Bartel & Saavedra, 2000), and in line with Planalp, DeFrancisco, and Rutherford's (1996) descriptive research on how humans generally identify affect using multiple cues in concert for maximum accuracy rather than attempting to select individual cues. Affect and behaviour coding was conducted on a handheld app previously designed for this purpose (Baker et al., 2012). Student affect or behaviour was coded according to the following set of categories: boredom, frustration, engaged concentration, confusion, off-task behaviour, gaming, and other (comprising any affective or behaviour state not represented by the other categories). These categories were chosen due to past evidence that they are relatively common and are either associated with learning or hypothesized to be associated with learning (cf. Aleven et al., 2004; Baker, 2007; Baker et al., 2010; Baker et al., 2012; Cocea et al., 2009; Craig et al., 2004; Lee, Rodrigo, Baker, Sugay, & Coronel, 2011; Lehman, D'Mello, & Graesser, 2012; Rodrigo et al., 2009). The affective categories were defined for coding according to the definitions in Baker et al. (2010), and the behaviour categories were defined according to the definitions in Baker (2007) and Baker et al. (2010).

At the beginning of data collection, an inter-rater reliability session was conducted, where the two coders coded the same student at the same time, across 51 different coding instances across multiple students. With reference to the categories of affect studied in this paper, inter-rater reliability achieved Cohen's Kappa of 0.72, indicating agreement 72% better than chance. For categories of behaviour, inter-rater reliability achieved Cohen's Kappa of 0.86, agreement 86% better than chance. This level of agreement is substantially higher than the level of agreement typically seen for video coding of affect (D'Mello et al., 2008; Sayette, Cohn, Wertz, Perrott, & Parrott, 2001). After this session, the observers coded students separately, for a total of 3,075 observation codes.

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

Within the observations, not counting observations marked as "?" (which represents cases where coding was impossible due to uncertainty, behaviour outside the coding scheme, a student leaving the room, impossible positioning, or other factors), boredom was observed 21.7% of the time, frustration was observed 5.4% of the time, engaged concentration 65.0% of the time, and confusion was observed 7.9% of the time. In terms of behaviour, off-task behaviour was observed 21.9% of the time, and gaming was observed 1.5% of the time. This distribution of affect and behaviour corresponds to previous studies, where engaged concentration is the most prevalent affect in a classroom environment (Baker et al., 2010; Baker et al., 2012; Sabourin, Mott, & Lester, 2011).

2.1.3 ASSISTments Interaction Logs

During observations, both the handheld devices and the educational software logging server were synchronized to the same internet timeserver, using the same field observation data-collection software as was used in Baker et al. (2012). This enabled us to determine which student actions within the software were occurring when the field observations occurred. Interactions with the software during the twenty seconds prior to data entry by the observer were aggregated as a clip, and data features were distilled.

The original log files consisted of data on every student attempt to respond (and whether it was correct), and requests for hint and scaffolding, as well as the context and time taken for each of these actions. In turn, 43 features were distilled from each action (Table 1), including features distilled for detecting other constructs in ASSISTments (cf. Baker, Goldstein, & Heffernan, 2011), and features developed for detecting student behaviour and affect in Cognitive Tutors (cf. Baker, 2007; Baker et al., 2012). Many of the distilled features pertained to the student's past actions, such as how many attempts the student had previously made on this problem step, how many previous actions for this skill or problem step involved help requests, how many incorrect actions the student had made on this problem step, and so on. To aggregate individual student actions into twenty-second clips, the sum, minimum, maximum, and average values were calculated across actions for each clip. This relatively simple approach to summarizing features was used due to its success in similar problems in other learning systems (cf. Baker et al., 2012). Thus, for the creation of affect and behaviour models, a total of 172 features were used.

2.1.4 Creation of Affect and Behaviour Models

A detector for each affective state or behaviour was developed separately, comparing that affective state to all other affective states (e.g., "bored" was compared to "not bored," "frustrated" was compared to "not frustrated," "engaged concentration" was compared to "not engaged concentration," and "confused" was compared to "not confused"), or comparing that behaviour to all other behaviours (e.g., "off-task" was compared to "not off-task" and "gaming" was compared to "not gaming"). Each detector was evaluated using 5-fold cross-validation at the student-level (e.g., detectors are trained on four groups of students and tested on a fifth group of students). By cross-validating at this level, we increase confidence that detectors will be accurate for new groups of students. Further, in this student-

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

level cross-validation, students were stratified into fold assignments based on their training labels. This guarantees that each fold has a representative number of observations of the majority and minority class. In addition, for unbalanced classes, re-sampling was used on the training sets to make the class frequency more equal for detector development (but detector goodness was validated on a data set that was not re-sampled, to ensure model validity for data with natural distributions). We attempted to fit sensor-free affect detectors using eight common classification algorithms, including J48 decision trees, step regression, JRip, Naive Bayes, K*, and REP-Trees. These algorithms were chosen as a sample of the space of potential algorithms, which can represent data with different patterns, but each of which is relatively conservative and not highly prone to over-fitting.

Table 1: The 43 features generated for affect detection.

The min., max., and avg. were also calculated, totaling 173 features

Total problems attempted in the tutor so far	Problem is original not a scaffolding problem
Bottom-out hint is used	Number of last 8 problems that used the bottom-out hint
Total number of 2 wrong answers in a row across all the problems	Percent of all past problems that were correct on this KC
Answer is correct	Wrong answer after hint
Problem ends with automatic scaffolding	Response is chosen from a list of answers (multiple choice, etc).
Problem ends with scaffolding	Response is filled in (no list of answers available)
First response is a help request	Problem is a scaffolding problem
First response is a help request — scaffolding	Second to last hint is used — indicates a hint that gives considerable detail but is not quite bottom-out
Number of last 5 first responses that included a help request	Long pause after wrong answer
Number of last 5 first responses that were wrong	Long pause after correct answer
Number of last 8 first responses that included a help request	Long pause after help or bug message
Number of last 8 first responses that were wrong	Long pauses after 2 consecutive wrong answers
First response time taken on scaffolding problems	Time since the current KC was last seen
Total first response practice opportunities on this skill so far	Time spent on the current step
First response working during school hours (between 7:00 am and 3:00 pm)	Total first responses attempted in the tutor so far
Time spent on help was under 10 seconds	Total first responses wrong attempts in the tutor so far
Time spent on help was under 1 second	Percent of all past problems that were wrong on this KC
Time spent on help was under 2 seconds	Total first response practice opportunities on this KC so far
Time spent on help was under 5 seconds	Total first response scaffolding opportunities for this KC so far
Immediate help request – help on first response and time spent was under 2 seconds	Total first response time spent on this KC across all problems
Action is a hint response	Total time spent on this KC across all problems divided by percent correct for the same KC
Total number of hints requested so far	

Feature selection for machine learning algorithms was conducted using forward selection with stepwise regression. With this technique, the feature that most improves model goodness is added to the list of features of the model until no more features that improve model goodness can be added (Table 1). During feature selection, cross-validated kappa on the original (non-re-sampled) data set was used as

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

the goodness metric. Prior to feature selection, all features with cross-validated kappa equal to or below zero in a single-feature model were omitted from further consideration, as a check on over-fitting.

The affect and behaviour detectors' performance was evaluated on their ability to predict the presence or absence of each affective state or behaviour in a clip. Detectors were evaluated using A' (Hanley & McNeil, 1980), Cohen's Kappa (Cohen, 1960), and F-measure (Van Rijsbergen, 1974) goodness metrics. The A' metric (equivalent to the area under the ROC curve) is the probability that the model will be able to discriminate a randomly chosen positive case from a randomly chosen negative case. An A' value of 0.5 for a model indicates chance-level performance, and 1.0 performing perfectly. Cohen's Kappa assesses the degree to which the model is better than chance at identifying the affective state or behaviour in a clip. A Kappa of 0 indicates chance-level performance, while a Kappa of 1 indicates perfect performance. A Kappa of 0.45 is equivalent to a detector that is 45% better than chance at identifying affect or behaviour. The F-measure of the F1-score measures the model's accuracy, computing for the weighted average of the model's precision and recall where the best F1 score is 1 and the worst score is 0.

All of the affect and behaviour detectors performed better than chance (Table 2). Detector goodness was somewhat lower than had been previously seen for Cognitive Tutor Algebra (cf. Baker et al., 2012), but better than had been seen in other published models inferring student affect in an intelligent tutoring system solely from log files (where average Kappa ranged from below zero to 0.19 when fully stringent validation was used) (Baker et al., 2012; Conati & Maclaren, 2009; D'Mello et al., 2008; Sabourin et al., 2011). The best detector of engaged concentration involved the K* algorithm, achieving an A' of 0.678, a Kappa of 0.358, and an F-measure of 0.687. The best boredom detector was found using the JRip algorithm, achieving an A' of 0.632, a Kappa of 0.229, and an F-measure of 0.632. The best frustration detector achieved an A' of 0.682, a Kappa of 0.324, and an F-measure of 0.677, using the Naive Bayes algorithm. The best confusion detector used the J48 algorithm, having an A' of 0.736, a Kappa of 0.274, and an F-measure of 0.667. The best detector of off-task behaviour was found using the REP-Tree algorithm, with an A' value of 0.819, a Kappa of 0.506, and an F-measure of 0.693. The best gaming detector involved the K* algorithm, having an A' value of 0.802, a Kappa of 0.370, and an F-measure of 0.687. These levels of detector goodness indicate models that are clearly informative, though there is still considerable room for improvement.

Detector features for boredom include the total number of actions, the total time spent on the last action before the clip and the first action after the clip, and the student's history of help requests and correct answers. For example, students were deemed bored when they spent over 83 seconds inactive immediately before or after the observation (lengthy pauses are also an excellent predictor of off-task behaviour (cf. Baker, 2007), a behaviour thought to be associated with boredom). Students were also deemed bored when they worked on the same problem during the entire observation but did not provide any correct answers either during the observation or immediately afterwards (a serious and

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

actively working student will generally obtain some correct answers in ASSISTments, as increasingly easy scaffolding is given when students make errors).

Table 2: Performances of affect and behaviour models

Affect	Algorithm	A'	Kappa	F-Measure
Boredom	JRip	0.632	0.229	0.632
Frustration	Naive Bayes	0.682	0.324	0.677
Engaged Concentration	K*	0.678	0.358	0.687
Confusion	J48	0.736	0.274	0.667
Off-Task	REP-Tree	0.819	0.506	0.693
Gaming	K*	0.802	0.370	0.750

The detector's features for frustration involve the percent occurrence of incorrect answers on a skill in the past, the largest hint count in that clip, the average correct actions in that clip, the largest number of scaffolding for a problem in that clip, the total number of past help request for that clip, the total number of actions that were second to the last hint for that clip, the largest number of consecutive errors in that clip, and least sum of right actions in that clip. The resulting model showed that students that had a low average of correct actions were frustrated.

Features used in the engaged concentration detector included the number of correct answers during the clip, the proportion of actions where the student took over 80 seconds to respond, whether the student followed scaffolding with a hint request, whether the student received scaffolding on the first attempt in a problem, and how many of the student's previous five actions involved the same problem. The model was created using the K* algorithm, which is an instance-based classifier. Instance-based classifiers predict group membership based on similarities to specific cases in the training set, rather than general rules, enabling them to identify constructs that can manifest in several distinct ways. For example, one group of students in engaged concentration repeatedly answered correctly in less than 80 seconds. Another group of students in engaged concentration answered incorrectly on their first attempt at a problem but then spent considerable time making their first response to the scaffolding question they received.

For confusion, detector features included the total number of consecutive incorrect actions for that clip, number of hints used for that clip, number of correct actions in the clip, total number of past incorrect actions for a skill in that clip, correct actions that took time to answer, actions for a skill that the student

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

got incorrect previously but then took time to answer. The resulting model was fairly complex, but one relationship leveraged in the model is that students who commit consecutive errors in a row for a skill are deemed confused. Students were seen as confused when they had committed a number of incorrect actions in the past for a skill and then took a long time to answer the current clip.

The off-task detector included the total number of attempts made for a skill in that clip, time taken by a student to answer, whether a student had a correct action for that clip, average number of scaffold requests in that clip, and total number of incorrect actions in the past in the clip. The resulting model was also complex, but one relationship shows that if there were few attempts for a problem, and it took the student a long time to answer, then the student is exhibiting off-task behaviour.

The features for the gaming detector included the use of a bottom-out hint in the clip, the number of hint usages for that clip, the average hint counts for a skill in that clip, the total number of actions for that clip that were answered incorrectly, and the occurrence of scaffolding in that clip. The resulting model for gaming, like engaged concentration, used the K* algorithm. Hence, similarities that resulted in the group of gaming students included those that usually used bottom-out hints, scaffolding, and hints.

2.2 Application of Models to Broader Data Set

Once the detectors of student affect and behavioural engagement were developed, they were applied to a broader data set consisting of two school years of student usage of the ASSISTments system by Worcester middle schools, 2004–2005 and 2005–2006. As discussed above, these schools represented a diverse sample of students in terms of both ethnicity and socio-economic status. This data set included 1,393 students and around 810,000 student actions within the learning software. The same features as discussed above were distilled for these data sets. Using these detectors, we were able to predict student affect and behaviour for each student action within the ASSISTments system.

2.2.1 Correlation Analysis

In order to correlate students' affect estimates with their raw state test scores, we first had to summarize their affect during the year, calculating one number per affective state per student. For each affective state, we calculated the mean of the predicted probabilities for that state during performance on each skill in the system. This list of means for each skill was then averaged to produce a summarized overall proportion of affect for the student. This averaging gives equal weighting of affect for each skill. This procedure was used because the MCAS test, which we are correlating to, consists of a random selection of skills. The weighting prevents a more frequently studied skill from having an influence on the students summarized affect that is disproportionate to its representation on the test.

Table 3 shows example affect data for calculating the summary of the bored affective state for one student. To calculate the degree of boredom during the year for the student in Table 2, the following calculation would be used:

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

$$Tricia's\ P(Bored) = \frac{\left[\frac{(0.20 + 0.50 + 0.50)}{3} + \frac{(0.90 + 0.70)}{2}\right]}{2} = 0.60$$

Table 3: Example student affect data set to be summarized

Student	Skill	Probability of Bored	Is Original?
Tricia	Subtraction	0.20	Yes
Tricia	Subtraction	0.50	No
Tricia	Subtraction	0.50	No
Tricia	Addition	0.90	Yes
Tricia	Addition	0.70	Yes

We also calculate the summarized affect for each student for original and scaffold questions separately. In ASSISTments, scaffold questions are given when a student asks for help or answers an original question (main question) incorrectly. The scaffolding often consists of several sub-questions and students know that they will be required to go through the scaffolding if a question is answered incorrectly; therefore, we wanted to allow for the possibility of observing affect differently during original questions than scaffolds.

2.2.2 Correlation Results

After summarizing the estimates of each student's affect, we used Pearson's correlation to observe the correspondence between their affect and their end-of-year state test score. The results below show the correlation of affect to test score for the two years of data. We report separately on the affect experienced by students while answering original questions and the affect while answering scaffold questions, as the patterns of affect were substantially different in these two cases. Across tests, the high sample size resulted in most correlations being statistically significant (using the standard t-test for correlation coefficients, two-tailed).

The strongest positive correlation, as shown in Table 4, was for engaged concentration on original questions. For 2004–2005, r = 0.45, t(624) = 12.56, two-tailed p < 0.01. For 2005–2006, r = 0.26, t(760) = 7.36, two-tailed p < 0.01. This finding is unsurprising, and maps to previous results showing a positive relationship between this affective state and learning (cf. Craig et al., 2004; Rodrigo et al., 2009). Even on scaffolding items, this relationship remained positive. For 2004–2005, r = 0.21, t(624) = 5.36, two-tailed p < 0.01. For 2005–2006, r = 0.09, t(760) = 2.56, two-tailed p = 0.01.

Boredom on original questions was negatively associated with learning outcomes, again matching previous research (cf. Craig et al., 2004; Pekrun et al., 2002; Rodrigo et al., 2009). For 2004–2005, r = -0.12, t(624) = -3.00, two-tailed p < 0.01. For 2005–2006, r = -0.28, t(760) = -8.03, two-tailed p < 0.01. However, boredom on scaffolding questions was associated with better learning. For 2004–2005, r =

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

0.32, t(624) = 8.46, two-tailed p < 0.01. For 2005–2006, r = 0.27, t(760) = 7.69, two-tailed p < 0.01. In interpreting this finding, it is worth considering why a student would become bored on a scaffolding question. One possibility is that the student knew the skill in the original question, but was careless (cf. San Pedro, Baker, & Rodrigo, 2011), which would explain these positive correlations. Another possibility is that high scoring students may know most of the skills involved with an original problem but not enough to answer correctly. When they are forced into the scaffolding, which breaks the main problem into individual skill sub-questions, they become bored because they are being made to work on simpler questions to which they already know the answers.

Table 4: Correlation of student affect to their raw state-test score.
Statistically significant results (p < 0.05) are given in boldface; results where p < 0.01 are also italicized.

Correlation	**ORIGINAL**		**SCAFFOLD**	
AFFECT	**'04–'05**	**'05–'06**	**'04–'05**	**'05–'06**
Boredom	***-0.11930***	***-0.27977***	***0.32082***	***0.26884***
Engaged Concentration	***0.44923***	***0.25794***	***0.20988***	**0.09238**
Confusion	***-0.16538***	**-0.08912**	***0.37370***	***0.23457***
Frustration	***0.30524***	***0.20376***	***0.26182***	***0.22418***
Off-Task	***0.14820***	-0.00662	**0.16985**	**-0.10793**
Gaming	***-0.43083***	***-0.30125***	***-0.32933***	***-0.24688***

Confusion had a similar pattern to boredom, with weak negative associations for original questions. For 2004–2005, r = -0.17, t(624) = -4.19, two-tailed p < 0.01. For 2005–2006, r = -0.09, t(760) = -2.47, two-tailed p = 0.01. By contrast, positive associations were found for scaffolding questions. For 2004–2005, r = 0.37, t(624) = 10.06, two-tailed p < 0.01. For 2005–2006, r = 0.23, t(760) = 6.65, two-tailed p < 0.01. Recent work has suggested that confusion impacts learning differently, depending on whether it is resolved (Lee et al., 2011), and that in some situations, confusion can be beneficial for learning (Lehman et al., 2012). The finding here accords with those papers, suggesting that confusion can be positive if it occurs on items designed to resolve that confusion.

Frustration had a positive correlation to learning, both for original items and scaffolding items. For original items, for 2004–2005, r = 0.31, t(624) = 8.01, two-tailed p < 0.01. For 2005–2006, r = 0.20, t(760) = 5.74, two-tailed p < 0.01. For scaffolding items, for 2004–2005, r = 0.26, t(624) = 6.78, two-tailed p < 0.01. For 2005–2006, r = 0.22, t(760) = 6.34, two-tailed p < 0.01. This finding is unexpected. Past research has suggested little relationship between frustration and learning (Craig et al., 2004; Rodrigo et al., 2009), contrary to hypotheses of a negative correlation. One possibility is that frustration in ASSISTments shows up in teacher reports in terms of negative performance, and that these students

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

receive greater support from their teachers. Clearly, it will be valuable to follow up and study this unexpected result further.

Gaming the system had a negative correlation with learning outcomes. For original items, for 2004–2005, $r = -0.43$, $t(624) = -11.92$, two-tailed $p < 0.01$. For 2005–2006, $r = -0.30$, $t(760) = -8.71$, two-tailed $p < 0.01$. For scaffolding items, for 2004–2005, $r = -0.33$, $t(624) = -11.92$, two-tailed $p < 0.01$. For 2005–2006, $r = -0.25$, $t(760) = -8.71$, two-tailed $p < 0.01$. These findings match previous evidence that gaming is associated with poorer learning (Aleven et al., 2004; Cocea et al., 2009).

The relationship between off-task behaviour and learning was unstable between the two school years, and weak in all cases. It varied between positive and negative, between the years. For original items, for 2004–2005, $r = 0.15$, $t(624) = 3.74$, two-tailed $p < 0.01$. For 2005–2006, $r = -0.01$, $t(760) = -0.18$, two-tailed $p = 0.86$. For scaffolding items, for 2004–2005, $r = -0.17$, $t(624) = 4.31$, two-tailed $p < 0.01$. For 2005–2006, $r = -0.11$, $t(760) = -2.99$, two-tailed $p < 0.01$. It is not clear why the relationships between off-task behaviour and learning were inconsistent between the two school years.

3 AFFECT BY TEST PROFICIENCY CATEGORY

Within this section, we ask if, based on the results above (as well as prior research), successful students are mostly in a state of engaged concentration. Are unsuccessful students mostly gaming the system? To answer these questions we plot the affective state estimates by test proficiency category to reveal the dominant affective states with respect to test outcomes.

Figure 2 plots the state test proficiency category against the average estimate of affect on original questions for all students in that proficiency category. This is an average of the same probability estimates calculated in section 2.2.1. Note that these are the summarized affect estimates and therefore do not necessarily add up to one. Non-summarized estimates may also not add up to one because separate classifiers were used for each affect detector. While a multi-nominal classifier would guarantee a summing to one of predictions for each clip, it would not guarantee a more accurate prediction overall, particularly for underrepresented classes. In this analysis, we applied a second step of offset correction to the affect predictions that was applied in the original test of classifiers (San Pedro et al., 2013). This correction provides a more accurate scaling of the affect summaries but does not change the correlations from the first report of these results (e.g., Pardos, Baker, San Pedro, Gowda, & Gowda, 2013).

We can observe from Figure 2 that the top affective state on original questions among failing students was concentration followed by frustration and boredom. The margin between concentration and frustration narrows as proficiency increases until there are nearly equal parts of the two among students scoring in the Advanced category. For Advanced students, a category that earns them a college scholarship, frustration is unexpectedly tied for the most probable affective state. The position of

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

frustration, on the other hand, is somewhat surprising. It raises the question of whether students react with frustration or boredom in response to material they find too easy.

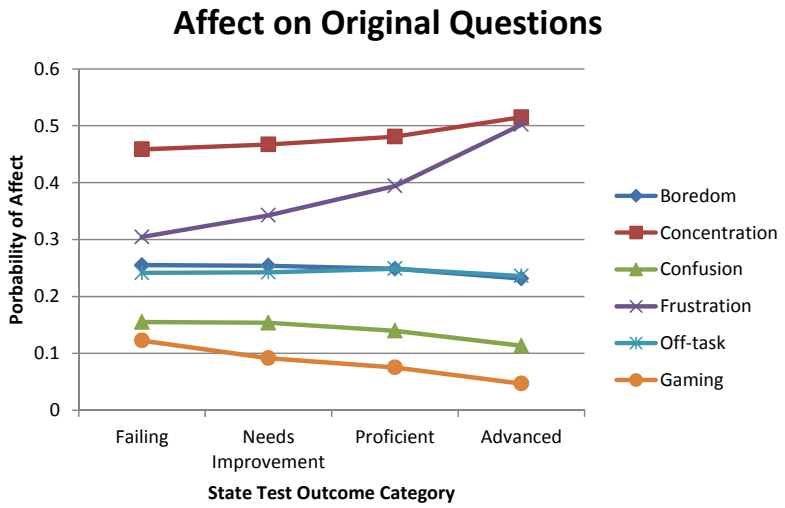

Figure 2: Probability of affect on original questions by test score category (average of both years' data)

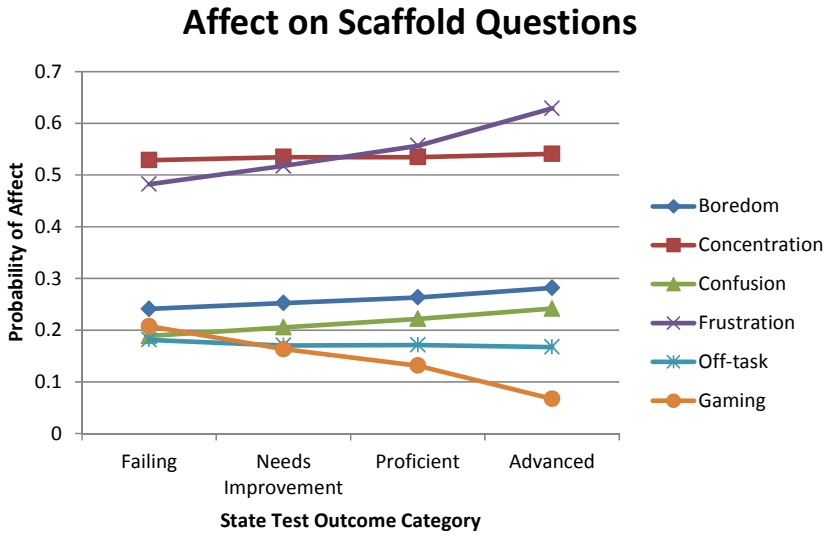

Figure 3: Probability of affect on scaffolds by test score category (average of both years' data)

The breakdown of affective state estimation on scaffold questions, shown in Figure 3, shows similarities to Figure 2 with frustration, engaged concentration, and boredom being the most probable affective states. One difference is that frustration becomes the most prominent affect, instead of concentration, in the proficient and advanced categories, and engaged concentration and boredom show little to no difference in probability between each another. On original questions, the interesting interaction was

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

that engaged concentration and frustration increased in probability over boredom with higher scoring students. On scaffolds, the interesting interaction is among gaming, off-task behaviour, and confusion. Among failing students, gaming is strongest, followed by off-task behaviour, and then by confusion. As the proficiency level increases, off-task behaviour and confusion become more probable as gaming becomes less common. There are equal parts of these three states at the proficient level much as there were equal parts of frustration, engaged concentration, and boredom at the proficient level for original questions. The takeaway for teachers here may be that gaming is generally undesirable, but confusion is not entirely problematic — successful students experience confusion on scaffolding items (perhaps because they are engaging with the material rather than disengaging by gaming the system).

Curiously, once again, highly successful students become frustrated more often on scaffolding items than less successful students. It may be, in these cases, that students become annoyed and then frustrated at receiving scaffolding after making a mistake; or it may be that they are frustrated with themselves when they do not succeed. Higher levels of frustration may reflect a higher level of student emotional investment or pride in mastering the knowledge required to answer the problem. Since the problem sets used by students in these years of the tutor gave a random sampling of 8^{th} grade skills, it is conceivable that this random ordering was a significant source of reasonable frustration for high and low proficiency students alike.

There is an observable difference in the magnitudes of affect estimates on original questions and scaffold questions. Table 5 quantifies this difference by calculating the estimate on scaffolds subtracted by the estimate on originals for each proficiency category. The average of these values across categories is shown in Table 5 along with the standard deviation among the four categories. If the shape of the trend line curve stays the same but is offset from Figure 1 to Figure 2 uniformly across categories, this will result in an average difference but zero standard deviation. A high standard deviation indicates that the change in affect between scaffolds and originals is not of uniform magnitude across categories.

Table 5 shows that students are more likely to be frustrated in scaffolding than when answering original questions. Frustration increases by 0.1543 on average, the highest of the affective states. This increase is fairly uniform across proficiency categories with a standard deviation of only 0.0142. The estimates of confusion, concentration, and boredom increase in the scaffolds but to a far lesser degree than frustration. Gaming and off-task behaviour estimates decrease in scaffolding. The change in these estimates was uniform across proficiency categories, indicated by the low standard deviation. The states with the highest standard deviation (shown in Table 6), although still low, were confusion, boredom, and gaming. The increase in confusion on scaffolds was greater as the proficiency level increased, with failing students showing a 0.0205 increase and advanced students showing a 0.1111 increase. A similar, lower magnitude, trend was observed for boredom. A decrease in gaming was observed with increasing magnitude as proficiency level increased. Boredom and confusion change from being negatively correlated with proficiency on original questions to being positively correlated with proficiency in

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

scaffolds, as shown in Table 4. With this kind of change, we would expect to see a variance in the change in estimates across proficiencies for these states.

Table 5: Scaffold estimate subtracted by original affect estimate and standard deviation across proficiency categories

Affect	Std.	Avg.
Frustration	0.0142	0.1543
Confusion	0.0404	0.0566
Concentration	0.0165	0.0365
Boredom	0.0301	0.0333
Gaming	0.0262	-0.0286
Off-task	0.0067	-0.0778

Table 6: Difference between scaffold and original affect estimates with the highest standard deviation across the proficiency categories

Affect	Failing	Needs Imp.	Proficient	Advanced	Std.
Confusion	0.0205	0.0323	0.0626	0.1111	0.0404
Boredom	0.0037	0.0183	0.0376	0.0735	0.0301
Gaming	-0.0008	-0.0191	-0.0313	-0.0631	0.0262

4 PREDICTION

In previous sections, we have trained affect and behavioural engagement detectors and correlated these constructs with end-of-year outcomes. In this section, we investigate how well student outcomes can be predicted by affect and behaviour as compared to student performance. Prior work has shown that student usage choices while receiving tutoring in ASSISTments can predict as much of the variance in students' end-of-year state test scores as student performance can on items designed to assess test-related knowledge (Feng et al., 2009), a result replicated in Ritter, Joshi, Fancsali, and Nixon (2013). It may also be worth trying to understand the role that affect and behaviour play in predicting student learning outcomes, in the form of end-of-year standardized examinations.

4.1 Methodology

In this section, we predict student performance on the standardized state math exam, the Massachusetts Comprehensive Assessment System, from three potential sets of features: an affect/behaviour feature set, a performance set, and a combined set. Each of these feature sets was

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

compared to a baseline model with the average test score in the training set. A detailed description of each set follows:

Table 7: Description of the four feature sets used in our prediction

Feature set	# features	Feature description
Affect/ Behaviour	12	Six summative affect/behavioural engagement measures for the student on original questions and six on scaffolds
Performance	2	Percent correct of the student on original questions and scaffolds
Both	14	Combination of affect/behavioural and performance features
Baseline	N/A	Average test score in training set

By comparing the affect/behaviour and performance feature sets to predict test scores, we can determine which has more predictive power. Using the combined feature set can tell us if the two sets are capturing the same variance or novel variance between them. The baseline measure gives us a simple prediction heuristic to compare to, the average test score for all students. These feature sets are described in Table 7. The four sets are generated for all students in both of years of data.

We use a five-fold cross-validation for each year of data separately and then a train/test hold out where the training set is the previous year's data. In both validation experiments, we use standard linear regression to learn coefficients for each feature in the feature set that maximizes fit to the target variable of test score. Mean absolute error is used as the error metric and statistical significance between predictions is tested on the absolute errors with a two-tailed paired t-test.

4.2 Prediction Results

In this section, we present the results of predicting the end-of-year scores based on features generated from tutor data collected from students during the school year. The baseline calculates the average test score in the training set and uses that prediction for every student in the test set. We use the different feature sets to compare the predictive power of each. Different data sets are used to observe whether predictive performance of the sets is consistent across years. Finally, a validation is conducted using data from one year as the test set and data from the previous year as the training set in order to test longitudinal model consistency.

Table 8 shows predictive performance results in terms of Mean Absolute Error (MAE). Overall, predictive performance of regression on the combination of the 12 *affect* features and the 2 *performance* features was better than either feature set alone. Specifically, the *both* model was best for all comparisons,

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

although the difference was only marginally significant between *performance* and *both* in the '04–'05 data set ($p = 0.084$). The difference was statistically significant in the '04–'05 ($p < 0.001$) and the different year hold out set ($p = 0.035$).

Table 8: Prediction results for the four feature sets on the two data sets and a one-year training / next year test holdout.

Mean Absolute Error / Pearson Correlation		Validation data set/holdout		
		'04–'05 5-Fold CV	'05–'06 5-Fold CV	'04–'05 train '05–'06 test
Feature set	Aff/Eng	6.48 / 0.736	7.41 / 0.650	8.56 / 0.587
	Performance	6.24 / 0.753	7.56 / 0.693	7.87 / 0.692
	Both	6.08 / 0.765	6.20 / 0.762	7.67 / 0.694
	Baseline	10.15 / NA	10.29 / NA	10.67 / NA

Overall, the *baseline* model was worst for all comparisons, significantly at the $p < 0.001$ level in all cases. The difference between *affect/behaviour* and *performance* was not stable. It was only significant in the case where the '04–'05 model was used on the '05–'06 data, where *performance* performed better than *affect/behaviour.* In the other comparisons, this difference was not significant, $p = 0.129$ in the '04–'05 data set and $p = 0.515$ in the '05–'06 data set.

Overall, then, it can be argued that *affect/behaviour* and *performance* are each good predictors of the state test. Furthermore, a combined feature set generally performs better than either of the feature sets alone. This suggests that while affect and performance provide similar predictive ability, they capture significantly different variance.

The third data set, using the '04–'05 data as training and '05–'06 data as testing, served as a validation that more closely fit how the detectors and prediction might be used in a real-world scenario, where scores of other students within a year cannot be used to train prediction within the same year but instead are used to train a model applied to the next year. With this validation, *affect/behaviour* features performed 9% less accurately than *performance* features but the combination of features resulted in a statistically significant improvement. The overall model, combining both *affect/behaviour* and *performance* features, trained on the combined '04–'05 and '05–'06 data set, is shown in Table 9.

5 CONCLUSION

In this paper, we evaluate the relationship between affect and behavioural engagement in a tutoring system over the course of a year, to performance on an end-of-year high-stakes test. Differentiating affect/behavioural engagement on original problems versus scaffolding help problems elicited

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

interesting results, in terms of boredom and confusion. Students who were bored or confused while answering the main problems, tended to do poorly on the test; however, boredom and confusion on scaffolding problems were associated with positive performance on the test. Gaming the system was, as expected, associated with poorer learning, while off-task behaviour was not consistently associated with poorer learning. One unexpected finding was a positive relationship between frustration and learning, which should be investigated further. These findings are clearly not yet conclusive, representing just a single online learning environment; but the methodological step that they represent — enabling analysis of affect that is both longitudinal and fine-grained, in the service of understanding the relationships between affect and learning — is a potentially valuable step. The data set produced through the application of these detectors is amenable to considerable further analysis of the ways that the context of affect influences learning. This will be a productive and valuable area for future work.

Overall, we find that a model integrating across multiple measures of affect and behavioural engagement can effectively predict student performance in the high-stakes exam. Such a model performs even better if measures of performance are also considered. As such, we can infer not just which affective states matter, but make an integrated prediction of how successful a student will be on a standardized examination.

Table 9: Features of the stepwise regression model in the order they were added to the model. An "(o)" denotes "on originals" and an "s" denotes "on scaffolds."

#	Feature description	Coefficient
1*	Gaming (o)	-8.27
2	Percent correct (o)	52.09
3	Confusion (o)	12.81
4	Frustration (s)	10.94
5	Concentration (o)	-65.74
6	Concentration (s)	48.36
7	Bored (o)	-48.61
8	Bored (s)	61.11
9	Off-task (s)	-35.13
10	Off-task (o)	13.28

Overall, these findings may be useful in the design of reporting on student behaviour and affect for teachers using digital learning and assessment platforms. When reporting on student boredom and confusion, it will be important to report context as well. For example, it may be useful to recommend interventions to teachers if a student is bored or confused on original questions, but not if these

* Gaming (o) was the first feature added to the model; however, it was removed from the model in the last step of the regression.

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

affective states occur during scaffolding. We see this work as leading in the direction of better support for teachers on intervening based on student affect. Real time integration of affect detection into a teacher's tutor dashboard along with an expanded understanding of the conditions that can make an affective state constructive or not, could greatly assist a teacher in signalling when to intervene in a crowded classroom.

ACKNOWLEDGMENTS

We would like to thank Neil Heffernan for sharing the ASSISTments data with us and access to ASSISTments classes; Adam Nakama, Adam Goldstein, and Sue Donas, for their participation and support in the original data collection; Lisa Rossi for copy editing assistance; support from the Bill and Melinda Gates Foundation, award #OPP1048577; and the NSF, award #DRL-1031398.

REFERENCES

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2004). Toward tutoring help seeking: Applying cognitive modeling to meta-cognitive skills. In J.C. Lester, R.M. Vicario, and F. Paraguaçu (Eds.), *Proceedings of Seventh International Conference on Intelligent Tutoring Systems, ITS 2004*, 30 August–3 September, Maceió, Alagoas, Brazil, 227–239.

Arnold, K.E. (2010). Signals: Applying academic analytics. *Educause Quarterly, 33*, 1.

Baker, R.S.J.d. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In *Proceedings of ACM CHI 2007: Computer-Human Interaction*, 28 April–3 May, San Jose, California, USA, 1059–1068.

Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., & Graesser, A.C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223–241.

Baker, R.S.J.d., Goldstein, A.B., & Heffernan, N.T. (2011). Detecting learning moment-by-moment. *International Journal of Artificial Intelligence in Education, 21*(1–2), 5–25.

Baker, R.S.J.d., Gowda, S., & Corbett, A.T. (2011). Towards predicting future transfer of learning. In *Proceedings of 15th International Conference on Artificial Intelligence in Education*, 28 June–2 July, Auckland, New Zealand, 23–30.

Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., & Rossi, L. (2012). Towards sensor-free affect detection in cognitive tutor algebra. In *Proceedings of the 5th International Conference on Educational Data Mining*, 19–21 June, Chania, Greece, 126–133.

Bartel, C.A., & Saavedra, R. (2000). The collective construction of work group moods. *Administrative Science Quarterly, 45*(2), 197–231.

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

Cocea, M., Hershkovitz, A., & Baker, R.S.J.d. (2009). The impact of off-task and gaming behaviors on learning: Immediate or aggregate? In *Proceedings of the 14th International Conference on Artificial Intelligence in Education*, 6–10 July, Brighton, UK, 507–514.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction, 19*(3), 267–303.

Craig, S.D., Graesser, A.C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media, 29*, 241–250.

D'Mello, S.K., Craig, S.D., Witherspoon, A.W., McDaniel, B.T., & Graesser, A.C. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction, 18*(1–2), 45–80.

Feng, M., Heffernan, N.T., & Koedinger, K.R. (2009). Addressing the assessment challenge in an intelligent tutoring system that tutors as it assesses. *Journal of User Modeling and User-Adapted Interaction, 19*, 243–266.

Hanley, J., & McNeil, B. (1980). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29–36.

Lee, D.M., Rodrigo, M.M., Baker, R.S.J.d., Sugay, J., & Coronel, A. (2011). Exploring the relationship between novice programmer confusion and achievement. In *Proceedings of the 4th Biannual International Conference on Affective Computing and Intelligent Interaction*, 9–12 October, Memphis, Tennessee, USA, 175–184.

Lehman, B., D'Mello, S.K., & Graesser, A.C. (2012). Confusion and complex learning during interactions with computer learning environments. *The Internet and Higher Education, 15*(3), 184–194.

Pardos, Z.A., Baker, R.S.J.d., San Pedro, M.O.C.Z., Gowda, S.M., & Gowda, S.M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge*, 8–12 April, Leuven, Belgium, 117–124.

Pardos, Z.A., Wang, Q.Y., & Trivedi, S. (2012). The real world significance of performance prediction. In *Proceedings of the 5th International Conference on Educational Data Mining,* 19–21 June, Chania, Greece, 192–195.

Pekrun, R., Goetz, T., Titz, W., & Perry, R.P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of quantitative and qualitative research. *Educational Psychologist, 37*, 91–106.

Planalp, S., DeFrancisco, V.L., & Rutherford, D. (1996). Varieties of cues to emotion in naturally occurring settings. *Cognition and Emotion, 10*(2), 137–153.

Ritter, S., Joshi, A., Fancsali, S.E., & Nixon, T. (2013). Predicting standardized test scores from cognitive tutor interactions. In *Proceedings of the 6th International Conference on Educational Data Mining*, 6–9 July, Memphis, Tennessee, USA, 169–176.

(2014). Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, *1*(1), 107–128.

Rodrigo, M.M.T., Baker, R.S., Jadud, M.C., Amarra, A.C.M., Dy, T., Espejo-Lahoz, M.B.V., Lim, S.A. L., Pascua, S.A.M.S., Sugay, J.O., & Tabanao, E.S. (2009). Affective and behavioral predictors of novice programmer achievement. In *Proceedings of the 14th ACM-SIGCSE Annual Conference on Innovation and Technology in Computer Science Education*, 5–8 March, Atlanta, GA, USA, 156–160.

Sabourin, J., Mott, B., & Lester, J. (2011). Modeling learner affect with theoretically grounded dynamic bayesian networks. In *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction*, 9–12 October, Memphis, Tennessee, USA, 286–295.

San Pedro, M.O.Z., Baker, R.S.J.d., Gowda, S.M., & Heffernan, N.T. (2013). Towards an understanding of affect and knowledge from student interaction with an intelligent tutoring system. In *Proceedings of the 16th International Conference on Artificial Intelligence and Education*, 9–13 July, Exeter, UK, 41–50.

San Pedro, M.O.C., Baker, R., & Rodrigo, M.M. (2011). Detecting carelessness through contextual estimation of slip probabilities among students using an intelligent tutor for mathematics. In *Proceedings of 15th International Conference on Artificial Intelligence in Education*, 28 June –2 July, Auckland, New Zealand, 304–311.

Sayette, M.A., Cohn, J.F., Wertz, J.M., Perrott, M.A., & Parrott, D.J. (2001). A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *Journal of Nonverbal Behavior, 25*(3), 167–185.

Van Rijsbergen, C.J. (1974). Foundation of evaluation. *Journal of Documentation, 30*, 365–373.

JOURNAL OF LEARNING ANALYTICS

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report

John P. Buerck and Srikanth P. Mudigonda
Saint Louis University, USA
buerckjp@slu.edu

ABSTRACT: Academic analytics and learning analytics have been increasingly adopted by academic institutions of higher learning for improving student performance and retention. While several studies have reported the implementation details and the successes of specific analytics initiatives, relatively fewer studies exist in literature that describe the possible constraints that can preclude an academic or learning analytics initiative from succeeding fully, meeting the criteria of success as defined by the stakeholders affected by such initiatives. Our aim in this article is to describe the constraints that precluded a successful completion of our analytics initiative and how we re-envisioned our approach and scope to achieve our primary goals while operating within the constraints and tools associated with our academic environment.

KEYWORDS: Learning analytics, academic analytics, learning management system, student retention

1 INTRODUCTION

The purpose of this article is to describe the our experience with our initiative in applying analytics to improve student performance and student retention, and situate this in the experiences and best practices reported in the literature pertaining to how academic and Learning Analytics initiatives have been undertaken by various academic institutions of higher learning.

Campbell, DeBlouis, and Oblinger explain that “Analytics marries large data sets, statistical techniques, and predictive modeling” and say that analytics “could be thought of as the practice of mining institutional data to produce ‘actionable intelligence’” (2007, p. 42). When the context is an academic institution, and the goals are student success and retention, the application of analytics at an institutional level to achieve the above-stated goals is termed as “Academic Analytics” (AA) (Long & Siemens, 2011). When the overarching goals of student success and retention are used as a focus within individual courses to improve student performance in individual courses, by measuring, collecting, analyzing, and reporting data related to student learning, participatory, and performance-related behaviours, then the particular analytics approach is termed “Learning Analytics” (LA) (Long & Siemens, 2011). Thus, LA can be seen as an important component of an educational institution’s AA initiative and that successful implementation of LA is crucial to the success of the institution’s AA initiative.

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

According to Greller and Drachsler (2012), there are six dimensions related to an LA initiative that need to be addressed well in order for the LA initiative to be successful. They are: (1) stakeholders, consisting of various groups of individuals who will be affected by the LA initiative, such as students, instructors, administrators, staff and other personnel; (2) objectives, consisting of the specific stakeholders' goals that the LA initiative is intended to address; (3) data, derived from the actions and activities undertaken by students, instructors, administrators, and other personnel involved in the operations of the academic institution; (4) instruments, consisting of the theories related to the behaviours of various actors in an educational setting and how they affect the outcome(s) of interest, technologies consisting of hardware and software, including the analysis algorithms, reporting and visualization tools and formats, etc.; (5) external constraints, such as conventions, norms, and legal requirements pertaining to data privacy (e.g., the US *Family Educational Rights and Privacy Act*, FERPA), research ethics, time period for which the LA initiative needs to exist before its outputs can be seen as beneficial to the stakeholders; (6) internal limitations, such as the competencies of various stakeholders to participate effectively and take advantage of the tools and technologies made available to them through the LA initiative.

Reflecting on our LA-related efforts thus far, and viewing them through the perspective of the six dimensions provided by Greller and Drachsler (2012), we were able to understand the reasons why our analytics initiative, which we originally envisioned as an AA initiative, did not meet our expectations, and how our revised version of the LA has been able to achieve moderate levels of success. In the next section, we will use the perspective provided by Greller and Drachsler (2012) to provide two sets of analyses: (1) the shortcomings of our original AA initiative; (2) a revised version of the initiative, which can now be described as an LA initiative due to a revision of the goals and scope. The revised version has achieved moderate success.

2 OUR ACADEMIC ANALYTICS AND LEARNING ANALYTICS INITIATIVES

2.1 Background

A survey of extant literature related to academic and Learning Analytics revealed that academic institutions of higher learning have successfully deployed analytics at various levels in the institutional hierarchy successfully. For instance, the president of Arizona State University made analytics a central component in university-wide change in focus to improve student performance and retention, while also launching newer academic departments and programs and increasing the university's societal impact, successfully (Crow, 2012). Baylor University is cited as one of the pioneers in using analytics in support of student recruitment and admissions (Campbell et al., 2007). The University of Alabama, Sinclair Community College, and Northern Arizona University have used analytics in improving student retention by identifying under-performing students early and making necessary interventions to ensure that they improve their academic performance and graduate (Campbell et al., 2007). Another "large-scale success" of an Academic Analytics initiative is Purdue's Signals system, which is being used successfully across

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

several courses in various departments and academic programs, and has been converted successfully into a product that can be procured and implemented by other academic institutions as well (Arnold, 2010; Tally, 2010; Norris & Baer, 2013).

Given the above-cited Academic Analytics initiatives that have been successful, our team decided to use a top-down approach for planning the various phases of the initiative and implementing them using an approach familiar to the initiative managers in the IT support services department. The impetus for our AA initiative came from higher-level administration officials in the IT support services and university administration departments, and two faculty members, who, in addition to teaching and researching, had responsibilities in department-level administration.

The higher-level and the department-level administrators realized that a cohesive, data-driven approach was needed to identifying students at risk of performing poorly and consequently either falling behind in their coursework or possibly leaving the university. By identifying such students and by providing them with necessary help and attention, it was expected that their academic performance and consequently their successful completion of their academic programs, could be achieved. From the IT support services side, the top-level officials realized that the large amounts of data being recorded in the learning management system (LMS), Blackboard, which the university uses, alongside the data related to student demographics and extra- and co-curricular activities, could be analyzed using tools related to Big Data and Analytics. Due to the convergence of these two sets of goals, an Academic Analytics initiative was sponsored by one high-ranking administrator and received support from high-ranking officials in IT support services. A team comprising IT personnel with expertise in data management, stewardship, administration of LMS and various student databases, an initiative manager, and two academic faculty members (the researchers/authors of this article) was formed.

According to the original initiative charter, a pilot version of the initiative would involve the collation of data of beginning-level students in the faculty members' department (computer information systems) drawn from their activities and performance recorded in the LMS, along with their demographic and extra-curricular activity data drawn from several disparate databases. These data would be cleaned, converted into a format amenable to data analysis after de-identification of individual student data (to ensure adherence to FERPA and other privacy-related requirements), and then models would be generated for creating profiles of students in an approach akin to that used in the Purdue Signals' initiative (Arnold & Pistilli, 2012).

3 ANALYSIS OF THE FIRST ATTEMPT USING THE GRELLER AND DRACHSLER PERSPECTIVE

Our initial work in the implementation of Academic Analytics at our university did not meet our expectations for several reasons. The various initiative phases were either not initialized on time, or

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

those that were initialized experienced significant delays. We will now present an analysis for our approach by structuring it using the six-dimensional perspective provided by Greller and Drachsler (2012).

3.1 Stakeholders

Having a shared understanding of the initiative's goals and scope among various stakeholders affected by it, in particular among the initiative's sponsors and implementers, is a key determinant of an analytics initiative's success (Norris & Baer, 2013; Crow, 2012). While we began with a common understanding of the scope of the initiative and its expected schedule, over the course of the pilot phase, deviations from the expected schedule led to a reallocation of priorities by various stakeholders, due to the constraints under which they were operating. This modification resulted in further deviations from the expected schedule, culminating in a disbanding of the initiative team and termination of the initiative in its original form.

3.2 Objectives

The primary objective of the pilot phase of the initiative was to develop a comprehensive infrastructure that included automated procedures for obtaining data from multiple data sources (student demographics and extra-curricular activities databases, LMS databases), removing personally identifiable data, and converting the data into a format amenable to analysis. The next step was to create statistical models able to profile students based on their background data, their curricular data (behavioural and performance data), and their extra-curricular data, e.g., their participation in various campus activities, their residency, and so on using an approach similar to that used in the Purdue Signals initiative (Arnold & Pistilli, 2012). The generation of actionable information presented in a form easily understood by students, their advisors, the faculty, and high-level administrators was the goal. In hindsight, and in comparison to the initiative schedules reported from Academic Analytics initiatives of other institutions, we realized that our schedule did not account for the amount of work required for completing all the tasks while requiring access to the personnel with the necessary skills and expertise needed for completing all the tasks.

3.3 Data

The data in individual databases was stored in tables whose naming conventions were unique to each vendor. Making sense of the metadata and determining which tables in which databases contained potentially meaningful and useful data was a stupendous task requiring a significant amount of time and collaboration among various departments including IT support services and the office of student records. Additionally, ensuring compliance to various privacy-related guidelines and laws was also important. The significance of these factors, and the amount of time, effort, and human expertise required to address the data-related issues cannot be understated. However, for a team undertaking an Academic Analytics

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

initiative for the first time, the magnitude of the resources required were not readily apparent at the start of the initiative.

3.4 Instruments

Personnel who have hands-on experience with multiple vendors' database conventions, in addition to the specific naming and design conventions followed by the departments in charge of data and information technologies at an academic institution will play a crucial role in determining optimal approaches for preparing datasets ready for analyses. While the researchers were well conversant in using statistical methods for analyzing data and producing profiles of students, the initiative was hamstrung by a lack of availability of data management experts who could devote the amount of time necessary to produce the datasets in a form that the researchers could use on an ongoing basis. The researchers' attempts to make sense of the metadata combined with obtaining the necessary clearances to access the necessary data were also constrained due to their teaching and departmental administration tasks. An additional hurdle was the difficulty associated with the inter-weaving of data from multiple sources to use in the data analyses. Thus, the few analyses that were run produced results that were not surprising. For instance, in analyzing student data from a course in Computer Ethics, we found that students' access of various learning materials and of the grade book were positively associated with their final score. While this finding is consistent with the expectation that students who consistently obtain the learning materials and look at their performance and feedback are more studious and consequently more likely to perform well, it does not provide any insight into the thought patterns of those who are not performing well. Or as Strader and Thille put it, "students' knowledge state is a blackbox to the instructor" (2012, p. 205). As such, the model that resulted from analyzing the small sample dataset did not produce any actionable information.

3.5 External Constraints

There were relatively few external obstacles faced by our team since our existing university policies ensured compliance with all regulatory requirements related to student privacy. Additionally, by obtaining approval for an Institutional Review Board, we ensured that enough safeguards were in place that current students would not be adversely affected by our research outcomes and that the uncertainty associated with the "pay-off" of the initiative for current students in the future would be acceptable.

3.6 Internal Constraints

The biggest constraints encountered have been internal. We lacked enough personnel with the necessary expertise to produce the needed datasets in a timely manner. Additionally, several of the existing experts did not have the cross-domain expertise needed to create an automated process that produced the necessary datasets to be analyzed. Furthermore, the specific technology-related and

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

policy-related mechanisms for producing meaningful reports that can be used to take the actions necessary for improving student participation, performance, and thereby retention, required analysis and development of tools and policies. These, too, required time and human expertise that were in short supply. Thus, the original schedule for implementation of all aspects of the pilot project could not be met.

Subsequent to the termination of the initiative in its original form, the researchers redefined the scope of the initiative, with a goal of achieving a few noticeable results at the departmental level in the form of improved student retention and performance. In the next section, the new version of the initiative will be analyzed using, once again, the six dimensional perspective provided by Greller and Drachsler (2012).

4 PROJECT REDUX: A REVISION OF SCOPE AND APPROACH

Having analyzed our first attempt at incorporating analytics at an institutional level, which can best be described as an Academic Analytics initiative (Long & Siemens, 2011), we determined that the chances of using analytics successfully within our department is possible if we could define the scope of the initiative by explicitly taking into account the various constraints associated with time, expertise, availability of technologies, and data that constitute the primary components of an analytics initiative. Fortuitously, during the time when we were planning the next attempt at initiating an analytics initiative, the LMS underwent a version upgrade. In the new version of Blackboard (version 9.1, Service Pack 11), a "Retention Center" functionality[1] is available. The functionality provided by the Retention Center allows an instructor or a course-builder to set alerts that are triggered when a student's performance level, activity level (accessing the course), or involvement level (participation in various discussion fora, completion of various evaluative components) fall below certain preset threshold values. This functionality can be used by an instructor, a course-builder, or a teaching assistant to view, via a dashboard, the overall participation and performance of all the students in a course and take the necessary intervention steps. Details of how we are leveraging this functionality, and using it in conjunction with a few other changes to our LMS-based course websites and departmental policies will be explained using the six dimensions.

4.1 Stakeholders

The stakeholders now are the students, the department chair, the assistant chair, the advising staff, the instructors, and the dean. Prior to the upgrade to the new version of the LMS, instructors had to monitor each aspect of student activity and performance levels manually and send an early warning to their academic advisors who would then work with the student, in conjunction with the instructor and the department chair, if necessary, to help the student improve his or her performance in the course. The old version of the early warning system was seen as tedious; because of this perception, not all faculty

[1] http://goo.gl/nopcTe

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

members used it often. The new version of the early warning system, which is integrated into the LMS via the Retention Center dashboard (a) reduces the number of steps that an instructor has to take to identify students who have fallen below the set thresholds for involvement and performance in the course, and (b) simplifies the process of sending an early warning. In their report on the determinants of success of the Purdue Signals system, Tanes, Arnold, King, and Remnet (2011) remark that reluctance on the part of instructors using the system stemmed from a lack of understanding of how the system could benefit them and not knowing the best practices related to its use. Based on this finding, and on feedback from faculty members not using the older early warning process optimally, we decided to implement the Retention Center-based early warning system in a select few courses, with the aim of determining which approaches seem to work best. The expected consequences would be a reduction in the amount of overhead associated with identifying students who need intervention and the provision of necessary information and feedback in a timely manner. Simultaneously, we have informed other faculty members about the availability of this system and our pilot initiative, and are conferring with them to obtain their input, and thus their buy-in, on how best to use the system to benefit our students, while being cognizant of any additional workload this might create.

4.2 Objectives

Our first objective was to identify the procedures that an instructor, the departmental administrators (chair and assistant chair), and the advising staff have to follow to ensure that intervention activities that encourage help-seeking and involvement on the part of students are optimized. This objective is consistent with the determination of similar procedures in the Purdue Signals system (Tanes et al., 2011). Our second objective was to improve communication with students about their participation and performance in the course and guide them to helpful resources (e.g., online and in-person tutoring and writing services).

Quicker and more informative communication with students is achieved by periodic monitoring by the instructor of the Retention Center dashboard to learn about student performance and then contacting students, via the Retention Center, regarding specific areas where they are lagging. Where needed, instructors get in touch with advisors to determine if an intervention requires a more personal approach (e.g., help with issues related to financial aid, overall course load, and so on) is needed. In our university, this is a continuation of existing departmental procedures. The provision of detailed feedback on each evaluative component has been implemented via rubrics that include detailed descriptions of criteria used in the evaluation of each component and comments on the extent to which a student's work addresses each criterion. Based on the anecdotal evidence, in the form of verbal feedback from the students, it appears that using component-specific rubrics and the Retention Center are meeting the intended objectives. More specifically, quantitative data will be obtained via end-of-term evaluations by students of various aspects of the courses.

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

Complementary to the above-described objectives, we are redesigning the navigational and organizational scheme of our course websites on the LMS by following the guidelines specified by Quality Matters.[2] Based on the feedback we received from students during previous academic terms, we believe that by using a standard design template across all courses on the LMS, the cognitive burden experienced by students while navigating a course website is minimized when all the courses have the same navigational scheme and organization of course materials into various sections and folders. The changes to website organization will be implemented in the next academic term.

By grounding our LA initiative in best practices reported in the literature, using well-designed, consistent online course websites, in conjunction with timely intervention actions that encourage help-seeking and involvement by our students, we expect to see improved student performance and retention. The "success" of our initiative will be determined by: (a) our mastery of the functionality and associated options provided by the Retention Center; (b) the development of a set of operational procedures, based on our observations in using the Retention Center, in determining the optimal actions that will lead to an increase in student success and retention; (c) an actual increase in student success and retention, determined by comparing the performance and number of drops[3] in courses where the Retention Center is currently used, with the performance and number of drops in the same courses offered during the same time period in previous academic years; (d) a survey of students who have interacted with course websites built using the new set of guidelines to determine their perceptions of ease of use. When the approach from this pilot, defined to span the current academic year, is applied to other courses offered through our department, we intend to collect more data to determine student perceptions on using multiple course websites that have the same "look and feel" in terms of navigational elements and organization of course materials into different folders and sections.

4.3 Data

As stated previously, the lack of actionable information, derived from data available in a form amenable to analysis, was one of the primary factors affecting the success of our previous Academic Analytics initiative. In the new iteration of our initiative, the information needed to take the necessary interventions is processed and made available in an intuitive format by the LMS itself with no additional work, other than accessing the Retention Center, by those who need the information. Thus, in the current form of our initiative, all the data issues that constrained us previously have been obviated. At the end of the current academic term, we will analyze the feedback provided by students in the two introductory courses in computer information systems that are implementing the LA activities. These data will be used in conjunction with the notes we are keeping on the various actions taken by the instructors in the courses, along with the feedback provided by the instructors on their use of the Retention Center and the rubrics for communicating with the students regarding their participatory

[2] https://www.qualitymatters.org/rubric

[3] A drop is said to occur in a course when a student unregisters from the course.

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

behaviours and performance in the courses. Additionally, we will also compare the data from the current term related to student retention and performance in these courses, and compare it with data from the same courses in previous academic terms to determine whether there have been any changes in the retention and performance of students in these courses. Based on our analysis, we will determine the changes, if any, needed to our current operational procedures and expand the LA activities to a selected set of additional courses.

4.4 Instruments

In our new LA initiative, the instruments necessary for processing the data and viewing the information in a useful form are embedded within the LMS. Additional tools that aid in communication and providing feedback — for example email and video/telephone conferencing — have been in use for a long time, so no additional cognitive or time-related constraints are present.

4.5 External Constraints

As in the previous version of our analytics initiative, currently no external constraints affect our initiative-related activities.

4.6 Internal Constraints

In the current version of our analytics initiative, constraints on technical expertise, which affected our previous pilot project, are absent. While learning to use the Retention Center functionality in an effective manner requires some additional time and effort, it is not imposing too big a burden, so we expect to continue with the initiative.

5 CONCLUSIONS AND FUTURE DIRECTIONS

As with any new project in an educational enterprise, this project encountered a few challenges. These challenges were, however, easy to overcome, as the goal of the project is to develop processes that help students overall to graduate successfully. In this section, we highlight the key challenges that were encountered and are being addressed by the team.

We learned from our two attempts at using analytics in an academic setting for improving student performance and retention that a top-down approach to Academic Analytics may not always work. A bottom-up approach, based on careful consideration of the functionality available in existing tools, can work when it is augmented by complementary communication and support procedures.

Our first attempt at Academic Analytics was guided by a schedule that proved untenable because we had not fully taken into account the diversity and complexity of activities to be completed, and the skills and

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

knowledge required of the team assigned to complete them. In our second attempt, our focus has been on student participation and performance in individual courses, by relying on the "Retention Center" functionality available in the LMS used in conjunction with communication and operational procedures. Consistent with findings reported in the literature (e.g., Tanes et al., 2011), these procedures are directed towards increasing students' help-seeking and participatory behaviours, with an expected improvement in their performance in coursework and consequently in the successful completion of their academic programs of study.

Based on our experiences, we believe that academic institutions intending to undertake AA and LA initiatives need not start with an all-out approach requiring the deployment of tools and analytical procedures subsumed under the Big Data paradigm (see e.g., Barton & Court, 2012). Rather, they can use a multi-phased approach where they undertake small LA initiatives centred around one or a few courses, utilize existing LMS-based tools to determine, given the current constraints and technology-based affordances, the optimal set of communication, intervention, and help-providing procedures so that student performance and retention are maximized. Then, based on what they have learned through this experience, they can determine the next steps needed for scaling their initiative to encompass more courses, and eventually determine a path for a more complete implementation of LA, and eventually AA across their academic institution.

As for our own LA initiative, in the immediate future, we expect to expand the LA activities to a few other courses in the computer information systems department after making any changes deemed necessary based on an evaluation of the data[4] collected from the courses where the LA activities have been introduced during the current term.

REFERENCES

Arnold, K.E. (2010). Signals: Applying academic analytics. *Educause Review Online.* Retrieved from http://www.educause.edu/ero/article/signals-applying-academic-analytics

Arnold, K.E., & Pistilli, M.D. (2012) Course signals at Purdue: Using learning analytics to increase student success, *Proceedings of the 2nd International Conference on Learning Analytics*, 29 April–2 May 2012, Vancouver, British Columbia, Canada.

Barton, D., & Court, D. (2012). Making advanced analytics work for you. *Harvard Business Review*, *90*(10), 78–83.

Campbell, J.P., DeBlouis, P.B., & Oblinger, D.G. (2007). Academic analytics: A new tool for a new era. *Educause Review Online, 42*(4), 40–57.

Crow, M.M. (2012). *"No more excuses": Michael M. Crow on analytics*. Retrieved from http://www.educause.edu/ero/article/no-more-excuses-michael-m-crow-analytics

[4] See description in the Data subsection that precedes this section.

(2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. *Journal of Learning Analytics*, *1*(1), 129–139.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society, 15*(3), 42–57.

Long, P.D., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review Online.* Retrieved from http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education

Norris, D.M., & Baer, L.L. (2013). Building organizational capacity for analytics. *Educause Report*. Retrieved from http://www.educause.edu/library/resources/building-organizational-capacity-analytics

Tally, S. (2010). Purdue-developed student intervention tech available nationally. *Purdue University News Service*. Retrieved from http://www.purdue.edu/newsroom/general/2010/101011CampbellCourseSigna.html

Strader, R., & Thille, C. (2012). The open learning initiative: Enacting instruction online. In D.G. Oblinger (Ed.), *Game Changers: Education and Information Technologies* (pp. 201–213). Educause.

Tanes, Z., Arnold, K.E., King, A.S., & Remnet, M.A. (2011). Using signals for appropriate feedback: Perceptions and practices. *Computers and Education, 57*(4), 2414–2422.

JOURNAL OF LEARNING ANALYTICS
SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH
(2014). Contemporary Privacy Theory Contributions to Learning Analytics. *Journal of Learning Analytics*, *1* (1), 140–149.

Contemporary Privacy Theory Contributions to Learning Analytics

Jennifer Heath
University of Wollongong
Australia
jheath@uow.edu.au

ABSTRACT: With the continued adoption of learning analytics in higher education institutions, vast volumes of data are generated and "big data" related issues, including privacy, emerge. Privacy is an ill-defined concept and subject to various interpretations and perspectives, including those of philosophers, lawyers, and information systems specialists. This paper provides an overview of privacy and considers the potential contribution contemporary privacy theories can make to learning analytics. Conclusions reflect on the suitability of these theories towards the advancement of learning analytics and future research considers the importance of hearing the student voice in this space.

KEYWORDS: Learning analytics, privacy theory, privacy

1 INTRODUCTION

The anticipated benefits of learning analytics in higher education are well documented and frequently focus on data issues and technical matters associated with system development and implementation. As Willis, Campbell, and Pistilli recently discussed:

> Big data and analytics, which marries large data sets, statistical techniques and predictive modelling [to mine] institutional data to produce 'actionable intelligence' present big questions to those of us in higher education. (2013)

This paper focuses on the privacy aspects of learning analytics deployment as a component of the ethical dimension of learning analytics. This paper is written from the position that having the technical capability to conduct a particular learning analytics task does not automatically mean that the task should be performed. As the discussion here will outline, there are many facets to privacy and some of the older concepts may not adequately serve the needs of learning analytics stakeholders, including academics, institutions, technology providers, and — most importantly — students.

2 PRIVACY, WITH A FOCUS ON INFORMATION PRIVACY

Philosopher Herman Tavani provides an insightful phrase that is a useful starting point for considering privacy matters: "Privacy is a concept that is neither clearly understood nor easily defined" (Tavani, 1999, p. 11). Publications concerning privacy matters in Western culture have been provided across multiple disciplines, including law and philosophy (Cohen, 2000; Fried, 1968; Rachels, 1975; Warren & Brandeis, 1890; Westin, 1967) and those with a focus on information

privacy (Floridi, 2005, 2006; Kang, 1998; J. Moor, 2000, 2005; J.H. Moor, 1997; Nissenbaum, 2010; Shoemaker, 2009; H. Tavani, 2007; Tavani & Moor, 2001; H.T. Tavani, 2007). Figure 1 presents a very simplified overview of privacy in order to provide some insight into the foundation concepts of privacy across four broad areas.

The first area uses a concise conceptualization of privacy from Culver et al. (Culver, Moor, Duerfeldt, Kapp, & Sullivan, 1994) where they argue that a person can be said to have privacy if, in a given situation or context, he or she is offered protection from *intrusion*, *interference,* and *information access* by others. This conceptualization of privacy is similar to that raised by Warren and Brandeis (1890) in their seminal paper on the rights of an individual to be left alone and free from intrusion and interference. The second area describes two broad classifications that also assist in the conceptualization of privacy: being normative and descriptive privacy. In a normatively private situation, individuals are protected by cultural norms such as formal laws or informal policies. Normatively private situations often include zones or contexts where normative protection is needed, for example, a patient in consultation with a clinician or a client in discussions with a lawyer. Descriptive privacy results in situations where individuals can expect privacy by natural means such as physical barriers. There are additional suggestions of dichotomies of "personal" versus "public" and zones of privacy (Gerstein, 1984) with privacy expectations varying according to the classification of the information type.

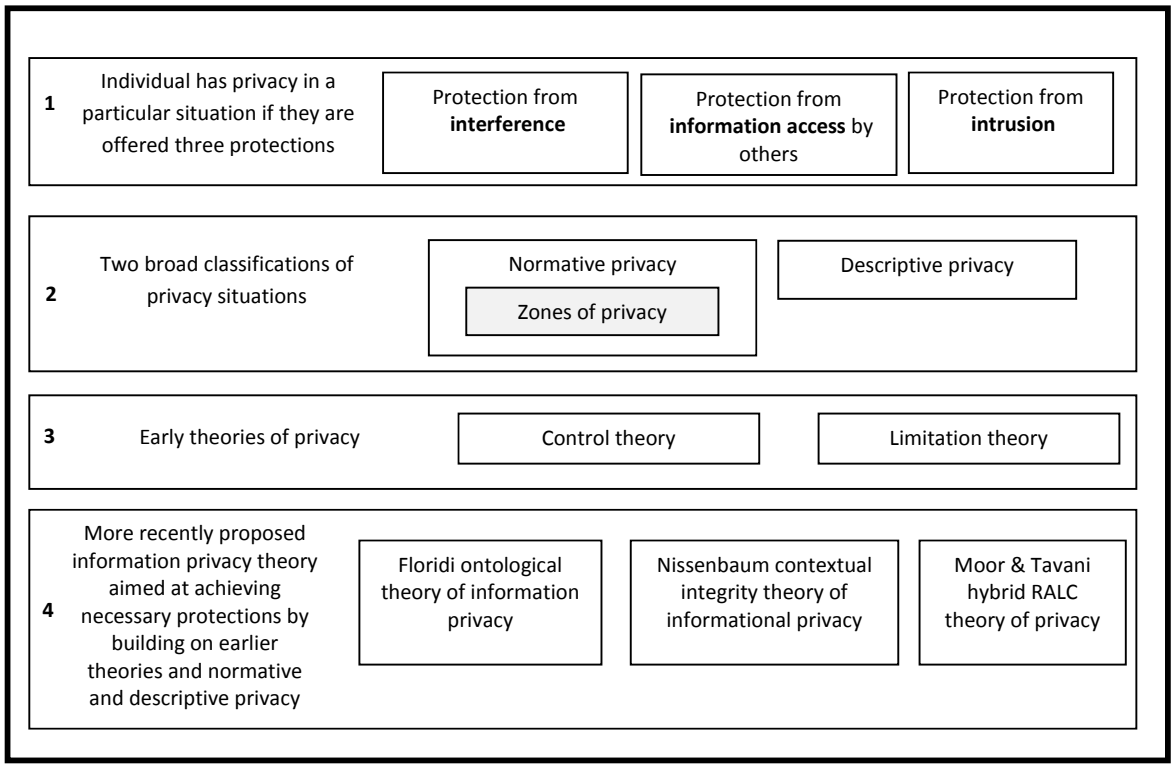

Figure 1: Broad overview of privacy

The third area, in Figure 1, presents two early theories of privacy that focus on allowing individuals control over their personal information, or limitations on the persons who could gain access to personal information. Debate regarding privacy has swung between arguments for and against a particular approach with the limitation theory and control theory dominating. Some publications in the learning analytics domain refer to privacy matters and take a single perspective from Figure 1 as the guide. For example, Petersen and Worona use a control theory interpretation of privacy to suggest how privacy affects learning analytics: "Privacy relates to the ability of individuals to control information about themselves" (2006, p. 16).

The final area includes three contemporary theories of informational privacy that move beyond the control theory versus limitation theory debate and offer a more holistic approach to privacy where the context (or *infosphere* in Floridi's work) emerges as a very important component of privacy theory. Luciano Floridi (2005) proposed an ontological theory of informational privacy based on information ethics. In a follow-up publication, Floridi (2006) provides a concise summary of his theory:

> To summarise: given a certain amount of personal information available in (a region of) the infosphere *I*, the lower the ontological friction in *I*, the higher the accessibility of personal information about the agents embedded in *I*, the smaller the informational gap among them, and the lower the level of informational privacy implementable about each of them. Put simply, informational privacy is a function of the ontological friction in the infosphere. (Floridi, 2006, p. 110)

Applying Floridi's privacy theory to the real world would be possible; however, the more tangible, less esoteric nature of Nissenbaum's, Tavani's, and Moor's theories provide a useful bridge to the "real world" of learning analytics.

3 CONTEMPORARY PRIVACY THEORIES AND APPLICATION TO LEARNING ANALYTICS

Contemporary privacy theories proposed by Nissenbaum (2010), Tavani (H. Tavani, 2007; Tavani & Moor, 2001; H.T. Tavani, 2007) and Moor (J. Moor, 2005; J.H. Moor, 1997) have been developed with the intention of applying them to diverse contexts, such as the rich learning analytics environment.

3.1 Nissenbaum

Nissenbaum moves the privacy debate beyond "control" or "limitation" theory, stating:

> Common usage suggests that intuitions behind both the constraint and control conceptions are sound: namely, that control over information about oneself is an important dimension of privacy, but so is the degree of access that others have to this information, irrespective of who is in control... In my view, the effect of these challenges, coupled with persuasive arguments, is not to prove that one or the other of these approaches is correct, but that both capture essential aspects of

> privacy that we seem to care about. A non-arbitrary resolution of this disagreement is not possible. (Nissenbaum, 2010, p. 71)

Nissenbaum proposes "contextual integrity" as an alternative conception of information privacy. Her approach is comprehensive with a goal of providing a decision heuristic to guide the evaluation of information privacy, which has potential benefits for learning analytics environments.

> … a right to privacy is neither a right to secrecy nor a right to control but a right to appropriate flow of personal information … Privacy may still be posited as an important human right or value worth protecting through law and other means, but what this amounts to is contextual integrity and what this amounts to varies from context to context. (Nissenbaum, 2010, p. 127)

Nissenbaum proposes informational norms that govern activities in contexts that she refers to as "context-relative informational norms." These norms are characterized by four key parameters: (1) contexts, (2) actors, (3) attributes, and (4) transmission principles. Nissenbaum provides a comprehensive definition of contexts as "structured social settings characterized by canonical activities, roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purposes)" (2010, p. 132). In the learning analytics context, the broader higher education environment canonical activities, roles, relationships, and so on are apparent. The learning analytics environment is not a static context. One context in which the student may engage is with the learning management system (LMS) where exchanges take place between academic staff and students engaged in learning. Frequently student involvement is mandatory in this context. The internal values (goals, ends, purposes) of student engagement with the LMS relate to providing a stimulating learning environment and effective management of student engagement. As students engage with online activities, data is generated as a by-product of this activity, including patterns of questions posed and answered (Buckingham Shum & Ferguson, 2012). This data does not inform measures related to learning outcomes as collected by assessment items, but does provide valuable insight into student learning engagement. Where does this sit with the internal values of the LMS context?

Another relevant context is associated with application, admission, and administration of the student journey into and through the Institution. Again, this context sees students required to provide personal information in order to progress through their administrative matters. The internal values (goals, ends, purposes) in this context tend to focus on the efficient management of student administration matters.

The above two contexts operate with different internal values, however learning analytics frequently merges the data from the two contexts into consolidated datasets ready for analysis where yet another set of internal values are encountered, the precise nature of which is emerging and under discussion at many Institutions. The definition of learning analytics provided by Ferguson, as reported by Ellis (2013), is a reasonable indicator of the goals, ends, and purposes of learning analytics: "the measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimising, learning and the environment in which it occurs." Given the mandatory requirement for students to engage in the above two contexts, is it

adequate to assume that they provide tacit agreement with the goals, ends, and purposes of this new context?

Second, Nissenbaum identifies three types of actors: (1) senders of information (2) recipients of information, and (3) information subjects. Table 1 provides an overview of key actors in the learning analytics domain. An "X" indicates typical higher education stakeholders and the actor roles they adopt in learning analytics contexts.

Table 1: *Privacy in Context*: Actors (Learning Analytics environment)

Actors	Individual students	Collaborative groups of students	Academic staff – subject coordinators	Academic staff – tutors, facilitators etc	Information Technology professionals	University Administrators, business analyst, planners
Senders of information	X	X	X	X	X	
Recipients of information	X	X	X	X	X	X
Information subjects	X	X	X	X		

It is interesting to note that academic staff may also be considered information subjects as the details of their engagement with students, including assessment marking and comments, may be information of interest to other actors. Diaz et al. (Diaz, Golas, & Gautsch, 2010) suggest that academics are more concerned about privacy than students are. The surveillance dimensions related to academic staff are interesting and require further research.

About attributes (information types) Nissenbaum says, "Analysis of attributes in contextual integrity is more nuanced than the private/public dichotomy of information. Informational norms render certain attributes appropriate or inappropriate under certain conditions and attributes co-evolve with contexts" (p. 132). In the learning analytics domain, attributes are diverse and constantly evolving, hence Nissenbaum's recognition that they co-evolve with contexts is an important aspect of her privacy theory. This pairing re-enforces the dynamic nature of the always-evolving environment and the need for privacy theory to keep pace. A static, rigid approach to privacy is inadequate in the learning analytics (and many other) technology-enabled activities.

The idea that privacy implies a limitation of access by others is similar to Nissenbaum's concept of an informational norm. In Nissenbaum's theory, diminishment of access is just one way that information flow may be governed. She defines transmission principles as "a constraint on the flow of information from party to party in a context" and the "terms and conditions under which such transfers should occur" (p. 132). In the two contexts described above, student administration and learning-management system context, the students enter into a tacit agreement regarding the transmission principles. For example, by submitting an assignment in a learning management system, students are allowing the information to flow from themselves to the academic staff member who will provide feedback on the assessment item. The flow of information is from the student to the responsible academic and back again. The terms and conditions under which this

information flows is an information norm when engaged in a learning context facilitated by a learning management system.

The transmission principles related to the flow of information from individual students to information technology professionals or University administrators is not necessarily an information norm. Terms and conditions under which these transfers should occur begin to assist in unpacking the complex ethics and privacy issues surrounding learning analytics.

Transmission principles regarding the provision of a student's personal demographic data in the student application, admission, and administration context do not necessarily apply in any other context. If a student agrees to the flow of student equity-related data to support admission processes, he or she is not necessarily agreeing to the same terms and conditions of information flow in another context, such as secondary use of data for learning analytics activities.

The two following scenarios, using the above-described four key parameters, illustrate the variations that can occur across learning analytics initiatives and hence the variation in the breadth and depth of privacy matters:

> **Scenario #1**: Analytics visualization for student use. Colour-coded indicators displayed to individual students to provide clear visualization of their progress in completion of assessments within a subject.
>
> **Context**: Provision of personalized information to individual students during semester using the academic data for specific subjects typically available in learning management systems.
>
> **Actors:** Sender of Information is the teaching academic, Recipient of Information is the individual student, and the Subject of Information is the individual student.
>
> **Attributes:** Assessment data (e.g., name, due date, learning outcomes), Student assessment progress data (e.g., date submitted, assessment mark, days overdue, learning outcome achievement).
>
> **Transmission Principles**: Data flow terms and conditions: Tacit within the teaching-learning relationship that exists between academic and student. No change in context affects this scenario as the data is generated and used within the one context, which is the learning within a particular subject instance.
>
> **Scenario #2**: "At risk" student modelling and associated interventions. Informed by predictive analytics modelling that includes diverse datasets from multiple university transaction processing systems, including student demographics, admission pathway, engagement with support services (including student well-being and academic type support services), attendance records from labs, tutorials, and myriad of data captured when student engages with the learning management system.

Context: Broad use of transaction processing information generated as student engages with mandatory and optional administrative and support services across the university environment.

Actors: Senders of Information are the custodians of diverse information systems, Recipients of Information are the academic or professional staff responsible for using "at risk" predictive models and initiating the interventions for individual students; Subjects of Information are the individual students.

Attributes: The data comprises all the "electronic breadcrumbs" left by students as their higher education journey moves from application to admissions, enrolments, and engagement across the institution.

Transmission Principles: Data flow terms and conditions from the original context where information systems gathered student data — say at the application stage where potential students provide demographic data — are quite different from the data flow terms and conditions that must be considered in the context involving building "at risk" models and encouraging staff to intervene with students. As the data subjects, the students should, ideally, have influence over the data flow terms and conditions, including options to remove themselves from this modelling and intervention scenario.

The descriptions above are a first step in unpacking the potential for Nissenbaum's contemporary privacy theory to assist in the analysis of privacy scenarios in the learning analytics domain.

3.2 Tavani and Moor

The work of Tavani and Moor also assists in navigating through this grey area of privacy and learning analytics. Through a series of individual and jointly authored publications, Tavani (Tavani, 1999; H. Tavani, 2007; Tavani & Moor, 2001; H.T. Tavani, 2007) and Moor (J. Moor, 2000, 2005; J.H. Moor, 1997) proposed a hybrid privacy theory that seeks to move beyond early privacy theories. The result is an identification of the fundamental, essential components necessary in a privacy theory. One outcome of their research is a tripartite model to describe a sufficient theory of privacy that they suggest must include three core aspects: (1) concept of privacy, (2) justification of privacy, and (3) management of privacy:

> A good theory of privacy has at least three components: an account of the concept of privacy, an account of the justification for privacy, and an account of the management of privacy. This tripartite structure of the theory of privacy is important to keep in mind because each part of the theory performs a different function. To give an account of one of the parts is not to give an account of the others. (Tavani & Moor, 2001, p. 6)

Moor and Tavani tackled the fundamental, important matter of developing a privacy theory rather than devising particular justifications or recommendations for the management of privacy suitable

for particular contexts. This approach thus creates a privacy theory foundation that will keep pace with technological innovations and supports the future directions of learning analytics activities. The resulting theory can be effective in a wide range of contexts with sufficient provision to respond to constantly developing technologies that could bring insufficient conceptualizations of privacy undone.

In a practical scenario, the "concept" and "justification" of privacy related to learning analytics can be addressed through careful consideration and clear articulation in learning analytics governance policy. An institution may choose to adopt a particular philosophy to underpin consideration of these privacy matters. The concept of privacy in the learning analytics domain broadly encompasses protection from intrusion and data gathering by actors who are not the subject of the information (i.e., individual students). The justifications of privacy are often "rights" based with the rights of both students and academic teaching staff to be considered within the learning analytics domain. The "concept" and "justification" of privacy are reasonably stable in the higher education learning analytics domain. Each of these aspects of privacy should be addressed in the learning analytics governance policies of higher education institutions, ideally before adopting learning analytics strategies. The scope of "management" of privacy in learning analytics scenarios includes the combination of technologies, policies, and procedures designed to address learning-analytics privacy requirements. With the emergence of new technologies, the "management" aspect of learning analytics privacy will be more volatile than the comparatively stable "concept" and "justification" aspects.

4 FUTURE CHALLENGES

Another recent publication proposes six principles for an ethical framework for learning analytics (Slade & Prinsloo, 2013) and three of these principles intersect with contemporary privacy theory, specifically the following: P2, Students as agents; P3, Student identity and performance are temporal dynamic constructs; and P5, Transparency. *P2, Students as agents,* encourages the view that students are collaborators in learning analytics and should be involved in decisions regarding use of their data. This is similar to the recognition of Actors (Information subjects) and Transmission Principles (Nissenbaum, 2010) and management of privacy expressed through choice, consent, and correction (Tavani & Moor, 2001). *P3, Student identity and performance are temporal dynamic constructs,* and *P5, Transparency,* speaks to the management aspects of privacy. Future challenges surround the effective integration of proposed ethical frameworks with valuable theoretical foundations, such as those proposed in contemporary privacy theories.

The rapidly evolving field of student co-creation of material (Diaz et al., 2010) presents fresh challenges as learning materials shift from being distributed by static online tools to the involvement of third-party providers (Rotenberg & Barnes, 2013). For example, how do third-party providers sit with the actors depicted in Table 1? What transmission principles guide the flow of data to third-party providers? The outcomes of recent research (Drachsler & Greller, 2012) indicate stakeholder concerns regarding intellectual property. How, then, does this rest with co-creation of materials? Drachsler and Greller (2012) described privacy as a "soft barrier" to learning analytics and investigated the opinion of education practitioners and researchers (n=156) in relation to privacy. It

is not clear if a shared definition of privacy was provided to participants and, as has been highlighted in this paper, there are diverse perspectives and interpretations of privacy. The four questions covered are as follows: breaches of privacy and intrusion in personal affairs; ethical principles around sex, political and religious beliefs, and ethnic origin; ownership rights and intellectual property (IP); and freedom of expression. Focussing on the first question, 65.8% of respondents indicated an expectation that learning analytics would affect privacy and personal affairs. Results from question three indicate that 60.1% of respondents indicate that ownership and IP would be affected by learning analytics. Respondent opinion was less clear regarding questions two and four.

This "soft barrier" study did not engage with students but focussed on the opinions of educators and researchers. The contemporary privacy theories considered here clearly recognize the importance of data subjects in determining appropriate privacy solutions. In the learning analytics domain, students are, as illustrated in Table 1, important actors and their voices need to be heard. Therefore, future research must engage with students in order to hear their expectations and concerns about privacy matters regarding advancing learning analytics.

5 CONCLUSION

Contemporary privacy theories can make a valuable contribution to learning analytics by providing clearly articulated, comprehensive conceptualizations of privacy. The theories explored here provide guideposts for considering the privacy dimensions of scenarios from the visualization of assessment progress data by individual students to the far more complex and ethically challenging example of "at risk" student predictive modelling and interventions. Future research must include consideration of the student voice to inform learning analytics ethics and privacy debates, as these voices have largely been silent.

REFERENCES

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society, 15*(3), 3–26.

Cohen, J. (2000). Examined lives: Informational privacy and the subject as object. *Stanford Law Review, 52*, 1373–1437.

Culver, C., Moor, J., Duerfeldt, W., Kapp, M., & Sullivan, M. (1994). Privacy. *Professional Ethics 3 & 4*, 3–25.

Diaz, V., Golas, J., & Gautsch, S. (2010). Privacy considerations in cloud-based teaching and learning environments. *Educause* (November), 2–10.

Drachsler, H., & Greller, W. (2012). *Confidence in learning analytics*. Paper presented at the LAK12: Second International Conference on Learning Analytics & Knowledge, Vancouver, Canada.

Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology, 44*(4), 662–664.

Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology, 7*(4), 185–200.

Floridi, L. (2006). Four challenges for a theory of informational privacy. *Ethics and Information Technology, 8*, 109–119.

Fried, C. (1968). Privacy: A moral analysis. *Yale Law Journal, 77*(1), 475–493.

Gerstein, R. (1984). Intimacy and privacy. In F. Shoeman (Ed.), *Philosophical dimensions of privacy: An anthology* (pp. 265–271). Cambridge: Cambridge University Press.

Kang, J. (1998). Information privacy in cyberspace transactions. *Stanford Law Review, 50*(4), 1193–1294.

Moor, J. (2000). Towards a theory of privacy for the information age. In R. M. Baird, R. Ramsower, & S. Rosenbaum (Eds.), *Cyberethics: Moral, social, and legal issues in the computer age* (pp. 200–212). Amherst, NY: Prometheus Books.

Moor, J. (2005). Why we need better ethics for emerging technologies. *Ethics and Information Technology, 7*, 111–119.

Moor, J.H. (1997). Towards a theory of privacy in the information age. *SIGCAS Computers and Society, 27*(3), 27–32. doi: http://doi.acm.org/10.1145/270858.270866

Nissenbaum, H. (2010). *Privacy in Context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.

Petersen, R., & Worona, S. (2006). Security & privacy: An overview. *Educause* (September/October), 16–17.

Rachels, J. (1975). Why privacy is important. *Philosophy and Public Affairs, 4*(4), 323–333.

Rotenberg, M., & Barnes, K. (2013). Amassing student data and dissipating privacy rights. *Educause* (January/February), 56–57.

Shoemaker, D. (2009). Self-exposure and exposure of the self: Informational privacy and the presentation of identity. *Ethics and Information Technology, 12*(1), 3–15.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1509–1528.

Tavani, H. (1999). Privacy online. *SIGCAS Computers and Society, 29*(4), 11–19. doi: http://doi.acm.org/10.1145/572199.572203

Tavani, H. (2007). *Ethics & technology: Ethical issues in an age of information and communication technology* (2nd ed.). Hoboken, NJ: John Wiley & Sons Inc.

Tavani, H., & Moor, J. (2001). Privacy protection, control of information, and privacy-enhancing technologies. *SIGCAS Computers and Society, 31*(1), 6–11. doi: http://doi.acm.org/10.1145/572277.572278

Tavani, H.T. (2007). Philosophical theories of privacy: Implications for an adequate online privacy policy. *Metaphilosophy, 38*(1).

Warren, S., & Brandeis, L. (1890). The right to privacy (the implicit made explicit). In F.D. Schoeman (Ed.), *Philosophical dimensions of privacy: An anthology* (pp. 193–220). Cambridge, MA: Harvard Law Review.

Westin, A. (1967). *Privacy and freedom*. New York: Atheneum.

Willis, J.E., Campbell, J.P., & Pistilli, M.D. (2013). Ethics, big data, and analytics: A model for application. *Educause Review Online*. Retrieved from http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application