Han, F., & Ellis, R. A. (2021). Predicting Students' Academic Performance by Their Online Learning Patterns in a Blended Course: To What Extent Is a Theory-driven Approach and a Data-driven Approach Consistent? *Educational Technology & Society*, *24* (1), 191-204.

Predicting Students' Academic Performance by Their Online Learning Patterns in a Blended Course: To What Extent Is a Theory-driven Approach and a Data-driven Approach Consistent?

Feifei Han[1, 2*] and Robert A. Ellis[1]
[1]Office of Pro-Vice-Chancellor (Arts, Education and Law), Griffith University, Australia // [2]Griffith Institute for Educational Research, Griffith University, Australia // feifei.han@griffith.edu.au // r.ellis@griffith.edu.au
[*]Corresponding author

ABSTRACT: One of the major objectives of precision education is to improve prediction of educational outcome. This study combined theory-driven and data-driven approaches to address the limitations of current practice of predicting learning outcomes only using a single approach. The study identified the online learning patterns by using students' self-reported approaches and perceptions of online learning and by using the observational digital traces of the sequences of their online learning events in a blended course. The study examined predictions of the academic performance using the online learning patterns generated by the two approaches separately. It also investigated the extent to which the online learning patterns identified by the two approaches were associated with each other. The theory-driven approach adopted a hierarchical cluster analysis using the self-reported data and found a 'deep' and a 'surface' online learning patterns, which were related to differences in the academic performance. The data-driven approach used an agglomerative sequence clustering and detected four patterns of online learning, which not only differed by quantity (number of learning events), but also differed by quality (the proportions of types of learning events). A one-way ANOVA revealed that the online learning pattern which had the most learning events, and was characterized by high proportions of viewing course contents and of performing problem-solving exercises, had the highest academic performance. A cross-tabulation revealed significant association between the self-reported and observational online learning patterns, demonstrating consistency of the evidence by a theory-driven and a data-driven approach and triangulating the results of the two approaches.

Keywords: Online learning patterns, Academic performance, Theory-driven approaches, Data-driven approaches, Blended course designs

1. Introduction

Blended course design, which is "a systematic combination of co-present (face-to-face) interactions and technologically-mediated interactions between students, teachers and learning resources" (Bliuc, Goodyear, & Ellis, 2007, p. 234), has been increasingly adopted in the higher education sector worldwide. As blended courses require students to move back and forth across face-to-face and online contexts (Ellis & Goodyear, 2019), their learning experiences are related to an increasing number of core elements; their cognitions (e.g., conceptions, approaches, and perceptions, Trigwell & Prosser, 2020), their social interactions in learning (Hadwin & Oshige, 2011), the spaces in which they learn (Ellis & Goodyear, 2016), and the different devices they use for learning (Laurillard, 2013). As a result, student learning experiences are becoming more and more complex in their structure. Consequently it requires research methods that move beyond approaches that do not routinely investigate the combined contribution of participants and the things they use to outcomes, such as academic performance (López-Pérez, López-Pérez, & Rodríguez-Ariza, 2011; Wu, Tennyson, & Hsia, 2010). The recent precision education initiative has offered some alternative approaches to framing how we evaluate learning in universities (Hart, 2016, Williamson, 2019, Yang, 2019).

Precision education is based on the philosophy of the Precision Medicine Initiative launched by the former US present Barack Obama after his 2015 State of the Union Address (The White House, 2015). The Precision Medicine Initiative aimed to revolutionize the medical treatment of diseases by transiting away from the one-size-fits-all approach of medical research and practice to the personalized approach, which takes into account of individual differences in genetics, environments, and personal choices (Collins & Varmus, 2015). Rather than producing unique treatments for specific patients, precision medicine emphasizes individual variability to improve the diagnosis, prediction, treatment, and prevention of disease. Underpinned by the same principles, Lu et al. (2018) defines the objectives of precision education as "the improvement of diagnosis, prediction, treatment, and prevention of learning outcome" (p. 221). To achieve these aims, educational researchers have been increasingly using big data analytics, applying artificial intelligence and machine learning, and implementing advanced data mining techniques and complex algorithms (known as data-driven approaches) to identify at-risk students early, and to provide an prediction of students' academic performance by their learning behaviors; so that targeted intervention strategies can be planned in order to prevent drop-out and learning failure (Cook, Kilgus, & Burns, 2018; Tsai, Chen, Shiao, Ciou, & Wu, 2020). Another important focus of precision education lies in the personalized education for enhancing student learning experiences (Lemons et al., 2018; Rojas-López & García-Peñalvo, 2019; Wilson & Ismaili, 2019). To contribute to the development of how a precision education perspective can improve our understanding of university student experiences of learning, this research will focus on improving prediction of students' learning outcomes.

Traditionally, research into student learning experiences and academic performance in higher education has largely adopted theory-driven approaches, which test hypotheses derived from theories in educational psychology, learning sciences, and research in pedagogy and curriculum (Trigwell & Prosser, 2020). Studies using such approaches primarily employ self-reported instruments and data to understand the relations between the processes of learning (e.g., approaches to, and perceptions of, learning) and the product of learning (e.g., the academic performance) (Ellis & Goodyear, 2016). With the advancement of educational data mining techniques and collection of rich learning analytic data, data-driven approaches have gradually gained popularity and a growing research area of learning analytics has emerged to provide a more objective picture of student learning (Baker & Siemens, 2014).

Both theory-driven and data-driven approaches, however, have limitations: the theory-driven approaches are criticized for being subjective and lacking accuracy in self-assessment by subjects (Siemens, 2013); whereas the data-driven approaches are often fragmented from educational theories and rely purely on empiricism, which limit the insights they can offer for directing pedagogical reforms, guiding learning design, and enhancing student learning experiences and outcomes (Buckingham Shum & Crick, 2012). Theoretically speaking, it is useful to investigate in what ways the associations between students' learning processes and academic outcomes revealed by the two approaches are convergent or divergent. To address this issue, the study first investigated the assessments of students' academic performance by using their online learning patterns adopting a theory-driven or a data-driven approach separately. It then examined the extent to which the online learning patterns found in the two approaches are consistent with each other.

From a methodological point of view, the study combined methods used in theory-driven and data-driven approaches. Such combination not only provides complementary information of what students reported and what they actually did for online learning, but also enables the results obtained from each approach to be triangulated. It has the strengths of offering information regarding the intents of students' learning and objective evidence of their learning behaviors to address the limitations of adopting either a theory-driven or a data-centric approach (Reimann, Markauskaite, & Bannert, 2014).

In the context of a theory-driven approach, we applied Student Approaches to Learning (SAL) framework (Trigwell & Prosser, 2020) to investigate students' approaches to using online learning technologies and their perceptions of the online learning environment in the blended course through a self-reported Likert-scale questionnaire to demonstrate their self-reported online learning patterns. For the data-driven approach, we used the digital-trace data of the sequences of the online learning events produced by the learning analytic functions in the learning management system (LMS) to show students' online learning patterns by observation (Jovanović, Gašević, Pardo, Dawson, & Mirriahi, 2017). The following part will review the relevant SAL research and learning analytics research.

2. Literature review

2.1. Related SAL research

SAL is a well-established research framework into student learning in higher education (Trigwell & Prosser, 2020). It focuses on identifying various factors in the learning processes that are able to explain differences in academic performance (Biggs, Kember, & Leung, 2001). SAL research mostly uses self-reported questionnaires to examine variations of the ways how students go about their learning (i.e., approaches to learning) and how they perceive the learning environment and teaching context (i.e., perceptions of learning and teaching) (Ramsden, 2003). The approaches and perceptions in SAL research are considered being responses to different learning and teaching contexts rather than a personality trait (Diseth, 2003). Hence, the same individual may report adopting different approaches and having different perceptions from one environment to another (Biggs & Tang, 2011). In addition, individuals in the same learning context can report contrasting experiences, despite studying the same learning tasks and experiencing the same teaching team.

A key insight from SAL research is that consistent and logical relations have been found between students' approaches, perceptions, and their academic performance (Asikainen & Gijbels, 2017; Entwistle, 2009). Students adopting surface approaches to learning, which are characterized by mechanistic procedures, producing formulaic responses, and being not engaged with the conceptions in learning, tend to perceive poor teaching quality, high workload, and irrelevant assessments. In contrast, students who adopt deep approaches, which involve using meaningful strategies to investigate the subject matter and pursue in-depth understanding of the key concepts and theories, are likely to perceive teaching as fostering independence and clear goal-focused, workload and assessment tasks as appropriate (Lizzio, Wilson, & Simons, 2002; Wilson & Fowler, 2005).

In blended course designs, deep approaches to learning and to using online learning technologies are also found to be logically related to students' appreciation of the online design and valuing of online contributions, and relatively higher course marks; whereas surface approaches to learning and to using online learning technologies are significantly associated with perceptions of unbalanced online learning workload, and a disconnection between face-to-face and online parts of the course, and relatively lower academic learning outcomes (Ellis & Bliuc, 2016; Ellis & Bliuc, 2019; Han & Ellis, 2020a). In this study, a self-reported questionnaire from SAL research is used to examine students' online learning patterns, including their approaches to using online learning technologies and perceptions of the online learning environment in the blended course.

2.2. Related learning analytics research

Learning analytics research has been established at the intersection of learning sciences, computer science, psychology, and education (Gašević, Dawson, & Siemens, 2015). It focuses on how large-scale data derived from technologies can be used to increase the understanding and improvement of the quality of student learning experiences and outcomes (Siemens & Gašević, 2012). The large volume of digital-trace data records what students and teachers do when they interact with a variety of learning technologies. When combining with various students' demographic information, the big learning analytic data are processed by advanced data mining techniques and sophisticated algorithms so that they can be used to: (1) address challenging problems in education, such as identifying at-risk students to minimise course attrition (Krumm, Waddington, Teasley, & Lonn, 2014), increasing program retention (Dawson, Jovanović, Gašević, & Pardo, 2017), and monitoring students' affect (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014); (2) provide empirical evidence to support decision-making, like in the area of academic success prediction (Romero, López, Luna, & Ventura, 2013), education policy reforms (Ferguson et al., 2016), and career advice (Bettinger & Baker, 2013); and (3) improve learning and teaching quality, including assisting learning design (Tempelaar, Rienties, & Giesbers, 2015), identifying learning strategies (Chen, Resendes, Chai, & Hong, 2017), facilitating online discussions (Joksimović, Gašević, Kovanović, Riecke, & Hatala, 2015) and collaboration (Kaendler, Wiedmann, Rummel, & Spada, 2015), and providing personalised learning feedback (Pardo, Jovanović, Dawson, Gašević, & Mirriahi, 2019).

2.3. Combining theory-driven and data-driven approaches

In considering the drawbacks of relying solely on theory-driven or data-driven approaches, some researchers have proposed to combine theory-driven and data-driven approaches in explaining students' learning outcomes (Lockyer, Heathcote, & Dawson, 2013). A combination of theory-driven and data-driven approaches may not only increase the power of detection of learning behaviors and prediction of learning outcomes, but may also be used as a way of triangulation to see if the evidence from different approaches can achieve consistency (Rodríguez-Triana, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2015).

Adopting a combined approach, studies have used different sources of data to predict and assess students' learning outcomes. Some studies reported that student learning processes collected from different data sources contributed uniquely to the learning outcome and increased the predictive power (Han & Ellis, 2017a). For instance, Pardo, Han, and Ellis (2017) reported that adding the frequency of interactions with the online learning resources by observation explained an additional 25% of variance in students' course marks than using students' reported use of self-regulated learning strategies alone. Ellis, Han, and Pardo (2017) also found a similar result that inclusion of students' engagement with various online learning activities in the regression model increased as high as 25% of variance explained for students' course marks than merely using students' reported approaches to study.

However, research findings are not always consistent with regard to whether the self-reported and observational data of student learning complementarily explain students' academic performance or overlap. Some studies found either an indirect contribution (Pardo, Han, & Ellis, 2016) or non-significant contribution of self-reported data (Tempelaar et al., 2015) to learning outcomes after adding observational data. Using a path analysis, Pardo et al. (2016) showed that students' reported positive self-regulated learning strategies only indirectly predicted academic performance via online activity participation recorded by digital traces. Tempelaar et al. (2015) conducted a large-scale study, which included 151 online learning modules involving 11,256 students. Their regression analyses showed that students' reported satisfaction with the online learning modules became a non-significant predictor after entering the observational data of time spent on online learning, which explained 11% of variance of their module retention.

Noting these inconsistencies, researchers have attempted to triangulate evidence offered by theory-driven and data-driven approaches. Han and Ellis (2017b) reported logical associations between students' self-reported perceptions of the course learning environment and the recorded counts of students' use of online learning tools in the LMS. Positive perceptions were found to be related to higher counts of tool use; as well as positive associations between negative perceptions and lower counts of tool use, suggesting a consistency of the learning experiences obtained from the self-reported and observational data.

Using more complex observational data rather than frequency counts as in Han and Ellis (2017b), and Han, Pardo, and Ellis (2020) identified a self-reported "understanding" and a "reproducing" learning orientation by students' approaches to learning in face-to-face and online contexts and perceptions of the blended learning environment. They also identified four different online learning orientations by using the observational sequences of the online study states. The results showed that students who reported an "understanding" learning orientation were involved in more online study states with high volume of formative learning, indicating a level of consistency between learning orientations by self-reported and observational data. One limitation of this study is the mismatch between the self-reported data and the observational data: while the self-reported questionnaire measured the learning in both face-to-face and online parts in the course; the observational data only recorded the online learning. Such mismatch may affect the results and needs to be addressed in future research.

2.4. Research purposes and research questions

Combining theory-driven and data-driven approaches, the current study had three research purposes. The first two research purposes concerned with separate examinations of the self-reported and observational online learning patterns, and the relations between these patterns and the academic performance. While the self-reported online learning patterns were measured by students' reporting on approaches to online learning technologies and perceptions of the online learning environment; the observational online learning patterns were assessed by digital traces of sequences of the online learning events. The third research purpose was to investigate the level of consistency between the self-reported and observational online learning patterns. By combining theory-driven and data-driven approaches, the study aimed to improve the current practice of predicting learning outcomes in order to contribute to the development of precision education.

To be more specific, the study addressed three research questions:

- What are students' self-reported online learning patterns and their relations to the academic performance?
- What are students' observational online learning patterns and their relations to the academic performance?
- To what extent are the online and observational learning patterns consistent with each other?

3. Method

3.1. The participants and the course design

The research was conducted in an Australian metropolitan university with 314 freshmen studying a compulsory engineering course. The course lasted 13 weeks, and was designed with face-to-face and online components. The face-to-face component included weekly lectures (2 hours), weekly tutorials (2 hours), and weekly laboratory sessions (3 hours). The focus of the lectures was in-depth explanations of the key concepts, which were further expanded and discussed in tutorials. The tutorials also included activities of how to apply theoretical principles to tackling practical issues through demonstration of problem-solving tasks. In laboratory sessions, students were given opportunities to gain hand-on skills through projects, such building an electronic circuit, or configuring a computer system. The online component, which was held in a bespoke LMS, required students to engage with the online learning activities both before and after classes through three types of online learning activities, namely: (1) view course contents; (2) doing theory-related exercises; and (3) performing problem-solving exercises. Before classes, students were asked to familiarize themselves with the lectures, tutorials, and laboratory contents, such as key concepts and laboratory procedures through reading and/or watching videos (i.e., view course contents). After classes, there were quizzes to test students' understanding of terminologies, such as Moore's law, System Verilog, and Flynn's taxonomy (i.e., doing theory-related exercises). There were also mini case studies for students to solve practical problem sequences by applying theories, such as digital circuit design improvement, control circuit reset, and pipelined processor implementation (i.e., performing problem-solving exercises).

3.2. Data and the instruments

3.2.1. Self-reported data collected by a questionnaire

The self-report data was collected using a 5-point Likert-scale questionnaire, which consisted of five scales: two scales assessed approaches to using online learning technologies; and the other three scales evaluated perceptions of the online learning environment in the blended course, including perceptions of the integrated learning environment, perceptions of online contributions, and perceptions of online workload. These scales were developed adopting SAL framework (Prosser & Trigwell, 1999) and used and validated in previous SAL studies (Ellis & Bliuc, 2016; Han & Ellis, 2020b). The description and reliability of each scale accompanied by a sample item are provided in Table 1.

Table 1. The details of the questionnaire

Scale	Description	α	Sample
Deep approach to using online learning technologies (6 items)	Using online learning technologies in a meaningful way, such as assisting forming key inquiry questions in learning, deepening concepts in the course, and developing essential skills	.75	*I find I use the learning technologies in this course to further my research into a topic.*
Surface approach to using online learning technologies (5 items)	Using online learning technologies in formulaic and mechanistic ways, such as fulfilling course requirements and downloading documents	.75	*I only use the online learning technologies in this course to fulfil course requirements.*
Perceptions of the integrated learning environment (9 items)	Perceptions of levels of integration of the online learning in the course	.89	*The ideas we reviewed online helped with the assessment of the course.*
Perceptions of online contributions (6 items)	Perceptions of the value of online contributions by other students in the course	.87	*The online contributions from other students helped develop my understanding of particular topics.*
Perceptions of online workload (6 items)	Perceptions of the online workload in relation to the whole course	.77	*The workload for the online activities was too heavy. (negatively worded item)*

3.2.2. Observational digital-trace data collected by the learning analytic functions in the LMS

The observational digital-trace data was collected by the learning analytic functions in LMS, which recorded sequences of timestamped learning events involving three learning activities: (1) viewing course contents of printed and video materials; (2) doing theory-related exercises; and (3) performing problem-solving exercises.

3.2.3. Data of academic performance

The academic performance data was students' course marks, which were comprised by scores for class preparation (25%), the laboratory project (25%), and the close-book examination (50%). The examination assessed students' understanding of the key theories through 20 multiple-choice questions (1.5 marks each) and their abilities to apply theories in solving practical questions through four open questions (5 marks each). The range of the course marks were from 21.25 to 98.75 (M = 67.40, SD = 14.62).

3.3. Ethical consideration and data collection

We obtained the formal approval and strictly followed the ethical requirements stipulated by the institution's ethics committee. The volunteer students signed written consent forms and agreed to answer the questionnaire, permitted the extraction of the digital-trace data of their online learning, and allowed access to their course marks. The different types of the data were matched and then anonymized so that students' identification was revealed.

3.4. Data analysis methods

To answer the first research question, a hierarchical cluster analysis using the mean scores of the five scales of students' approaches to, and perceptions of, online learning was conducted to identify self-reported online learning patterns. Based on the cluster membership, one-way ANOVAs were performed to examine the relations between self-reported online learning patterns and academic performance. To facilitate interpretation, the mean scores of the scales were transformed into z-scores (M = 0, SD = 1) in the analyses. To provide an answer to the second research question, we performed an agglomerative sequence clustering analyses using the timestamped sequences of the online learning events to investigate the observational evidence of the online learning patterns. Then we applied a one-way ANOVA to see the extent to which students' academic performance differed by the observational online learning patterns. For the third research question, we run a cross-tabulation analysis using the self-reported and observational clusters generated from the above analyses.

4. Results

4.1. Self-reported online learning patterns and academic performance

A hierarchical cluster analysis using the five scales produced two clusters of students: cluster 1 had 95 students and cluster 2 had 219 students. As shown by one-way ANOVAs, the two clusters of students differed significantly on all the five scales: deep approach to using online learning technologies: $F(1, 312) = 88.43$, $p < .01$, $\eta^2 = .22$; surface approach to using online learning technologies: $F(1, 312) = 163.25$, $p < .01$, $\eta^2=.34$), perceptions of the integrated learning environment: $F(1, 312) = 119.53$, $p < .01$, $\eta^2 = .28$; perceptions of online contributions: $F(1, 312) = 8.59$ $p < .01$, $\eta^2 = .03$; and perceptions of online workload: $F(1, 312) = 115.60$, $p < .01$, $\eta^2 = .27$). The two clusters of students also differed on the course marks: ($F(1, 312) = 16.46$, $p < .01$, $\eta^2 = .05$); and the scores of each assessment task, class preparation: ($F(1, 312) = 4.11$, $p < .05$, $\eta^2 = .01$); laboratory project: ($F(1, 312) = 4.15$, $p < .05$, $\eta^2 = .01$); close-book examination: ($F(1, 312) = 14.92$, $p < .01$, $\eta^2 = .05$).

Table 2. Self-reported online learning patterns and academic performance

Variables	Deep (N = 95)		Surface (N = 219)		F	p	η^2
	M	SD	M	SD			
Deep approaches to online learning technologies	0.72	0.73	-0.31	0.96	88.43	.00	.22
Surface approaches to online learning technologies	-0.88	0.64	0.40	0.88	163.25	.00	.34
Perceptions of the integrated learning environment	0.80	0.62	-0.35	0.94	119.53	.00	.28
Perceptions of online contributions	0.24	0.95	-0.12	1.02	8.59	.00	.03
Perceptions of online workload	0.80	0.88	-0.34	0.86	115.60	.00	.27
Course marks	72.37	14.15	67.41	14.62	16.46	.00	.05
Class preparation	21.58	2.83	20.80	3.31	4.11	.04	.01
Laboratory project	21.48	6.17	19.84	6.71	4.15	.04	.01
Close-book examination	29.22	11.22	23.90	11.21	14.92	.00	.05

The M values in Table 2 suggested that cluster 1 students self-reported a higher score on deep approaches to using online learning technologies; positive perceptions of the integrated learning environment, of online contributions, and of online workload. These features of approaches and perceptions suggested that cluster 1 students had a "deep" learning pattern. In contrast, cluster 2 students reported higher scores on surface approaches to using online learning technologies, and had negative ratings on perceptions of the integrated learning environment, of online contributions, and of online workload. The learning of the cluster 2 students had characteristics of "surface" pattern of learning. The scores of assessment tasks achieved by students with the "deep" learning pattern were significantly higher than those with the "surface" learning pattern.

4.2. Observational online learning patterns and academic performance

The agglomerative hierarchical sequence clustering using the timestamped online learning events involving the three types of learning activities produced four observational online learning patterns, which are visualized in Figure 1.

- pattern 1 (N = 161): had most learning events (M learning events = 62), amongst which viewing course contents occupied highest proportion, followed by problem-solving exercises, and theory-related exercises accounted for the lowest proportion.
- pattern 2 (N = 64): had second most learning events (M learning events = 27), of which there were relatively balanced learning events of viewing course contents and doing theory-related exercises, with performing problem-solving exercises being lowest.
- pattern 3 (N = 27): had least learning events (M learning events = 13), of which there was predominantly doing theory-related exercises, with very low proportion of performing problem-solving exercises.
- pattern 4 (N = 62): had the second least learning events (M learning events = 18), of which there were high proportions of doing theory-related exercises, followed by viewing course contents, and performing problem-solving exercises had the lowest proportion.

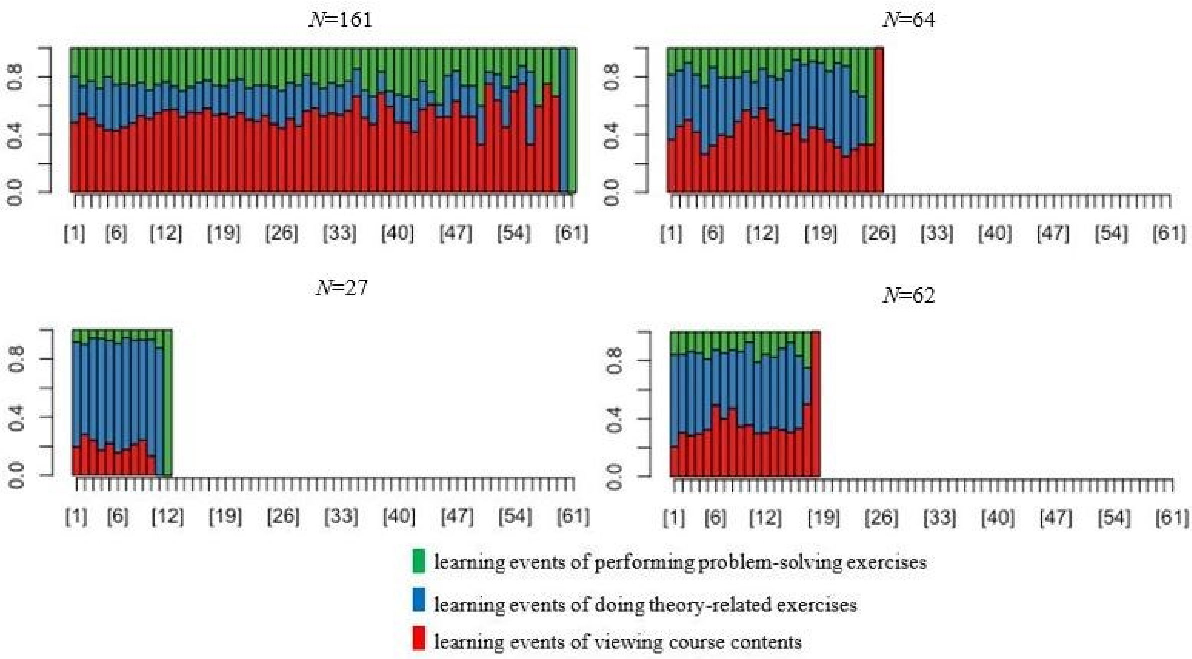

Figure 1. The four observational online learning patterns

The one-way ANOVAs showed that students' course marks (F(3, 310) = 34.24, $p < .01$, $\eta^2 = .25$) and scores on each assessment task (class preparation: (F(3, 310) = 29.50, $p < .01$, $\eta^2 = .22$); laboratory project: (F(3, 310) = 12.96, $p < .05$, $\eta^2 = .11$); and close-book examination: (F(3, 310) = 15.87, $p < .01$, $\eta^2 = .13$) significantly differed by patterns. The post-hoc analyses in Table 3 were summarized (course marks: pattern1 > pattern2 > pattern3 = pattern4; class preparation: pattern1 > pattern2 > pattern3 > pattern4; laboratory project and close-book examination: pattern1 = pattern2 > pattern3 = pattern4).

Table 3. Post-hoc analyses of students' academic performance by patterns

	Course marks			*p* values for pairwise comparisons		
	N	*M*	*SD*	pattern 2	pattern 3	pattern 4
Pattern 1	161	73.13	12.99			
Pattern 2	64	68.39	13.22	.01		
Pattern 3	27	52.95	15.14	.00	.00	
Pattern 4	62	57.85	10.14	.01	.00	.10
	Class preparation			*p* values for pairwise comparisons		
	N	*M*	*SD*	pattern 2	pattern 3	pattern 4
Pattern 1	161	22.22	2.36			
Pattern 2	64	21.08	2.90	.01		
Pattern 3	27	17.61	3.78	.00	.00	
Pattern 4	62	19.41	3.34	.00	.00	.01
	Laboratory project			*p* values for pairwise comparisons		
	N	*M*	*SD*	pattern 2	pattern 3	pattern 4
Pattern 1	161	21.74	5.13			
Pattern 2	64	21.64	4.46	.92		
Pattern 3	27	17.41	7.26	.00	.00	
Pattern 4	62	16.62	9.25	.01	.00	.59
	Close-book examination			*p* values for pairwise comparisons		
	N	*M*	*SD*	pattern 2	pattern 3	pattern 4
Pattern 1	161	28.81	11.55			
Pattern 2	64	26.09	10.69	.09		
Pattern 3	27	17.78	11.02	.00	.00	
Pattern 4	62	19.72	8.04	.01	.00	.43

4.3. Association between self-reported and observational online learning patterns

The results of the cross-tabulation analysis revealed a significant association between the self-reported and observational learning patterns ($\chi^2(3) = 7.95$, $p < .05$). Table 4 shows that amongst the students categorized in observational online learning pattern 1, the proportion of students who reported a "deep" online learning pattern (61.1%) was significantly higher than that of students who reported a 'surface' online learning pattern (47.0%). Of the students in observational online learning pattern 3, the reversed pattern was observed: the proportion of students who reported a "surface" online learning pattern (11.0%) was significantly higher than that of those reporting a "deep" pattern (3.2%).

Table 4. Association between self-reported and observational online learning patterns

Patterns	Count & % within self-reported patterns	Deep	Surface	Total
Pattern 1	Count	58	103	161
	%	61.1%$_a$	47.0%$_b$	51.3%
Pattern 2	Count	18	46	64
	%	18.9%$_a$	21.0%$_a$	20.4%
Pattern 3	Count	3	24	27
	%	3.2%$_a$	11.0%$_b$	10.8%
Pattern 4	Count	16	46	62
	%	16.8%$_a$	21.0%$_a$	19.27%
Total	Count	95	219	314
	%	100.0%	100.0%	100.0%

Note. Different subscript letters denote the proportions differ significantly at $p < .05$.

5. Discussion and practical implications

An important aim of precision education is to provide tailored instructional interventions to address students' problematic learning behaviors in order to enhance their learning achievement (Cook et al., 2018; Tsai et al., 2020). Before effective intervention plans can be carried out, it is essential to know what students' problematic learning behaviors might entail and how they are related to learning outcomes. Understanding the structure of problematic student behaviours and why they do the things they do is one of the two foci of the current study.

Our results show that the different learning patterns identified by the self-reported approaches to, and perceptions of, online learning *and* the differences in the sequences of digital traces of the online learning events recorded in the LMS both are consistent with variations of the academic achievement as indicated by the final marks in their course.

Before discussing the results, it is worthwhile noting the limitations of the study. While the patterns of the observational online learning behaviors detected by the agglomerative sequence clustering considered the proportions of types of learning events as well as the total number of learning events, it did not provide detailed account of frequencies of each type of learning activities. Future studies could apply other statistical methods to make fine-grained analyses of the frequencies of different learning activities. Furthermore, due to ethical issues, it was not practical to obtain item-by-item score of students' close-book examination, hence, the internal reliability of the examination could not be calculated. Notwithstanding these limitations, the outcomes of this study offer some interesting insights into a possible way of improving prediction of students' learning outcomes from a precision education perspective.

In this study, the theory-driven approach found that students who self-reported deep approaches to using online learning technologies as well as positive perceptions of the online learning environment also obtained relatively higher achievement in the course; whereas those who reported using surface approaches and having negative perceptions were more likely to receive lower course marks. The two contrasting patterns of students' online learning were similar to the two contrasting patterns found in the research conducted in the traditional classroom learning context, which also found logical associations between deep approaches to learning, positive perceptions of teaching quality, and quality learning outcomes on the one hand; and surface approaches, perceptions of inappropriate workload and assessment, and poorer learning outcomes on the other (Lizzio et al., 2002; Wilson & Fowler, 2005). The results corroborated with the findings of the studies in blended course designs, which also distinguished between students' learning with contrasting approaches in both face-to-face and online parts, perceptions of integration between face-to-face and online learning, and the academic achievement (Ellis & Bliuc, 2019; Han & Ellis, 2020a). However, different to previous studies in blended contexts, the current study only focused on the approaches and perceptions of the online part in order to compare the observational data, which also concerned with the engagement with the online learning activities only. The comparison between the self-reported and observational data in our study improved the research design by addressing the limitation of a mismatch in Han's et al. (2020) research, in which the self-reported data focused on the whole course experience, whereas the observational data was only about online part of learning.

The self-reported findings show that as high as 70% of the students in the course did not approach learning in a meaningful and considered way. The questionnaire not only identified the students whose learning requires extra support, but also provided evidence to teachers to better understand which specific aspects of the learning experience should be improved so that the intervention strategies are more likely to produce a benefit. Specifically, the less desirable aspects of the learning experience in this study include adopting surface approaches to using online learning technologies, perceiving a fragmentation of the online learning experience in relation to the course, not valuing the online contributions of their peers, as well as considering the online workload to be high. Teachers can help improve these elements early in the course so that students may have more desirable learning experiences later on in the course. The teaching team can invite those whose learning was oriented towards a deep understanding of the subject matter to share their strategies and ideas. There are many useful topics for the students to share: such as how and why they used online learning technologies in a meaningful way to facilitate their learning, how these approaches helped them achieve the learning objectives in the course, what they decided to contribute online and what they found more useful to share in class, how they felt about and learnt from others' online contributions, how they managed their online workload and what proportion of time they allocated to their online learning activities in the overall course experience. In addition, the teachers can help all their students by providing more explicit explanations at the beginning of the course as to how online resources, quizzes, and activities are linked with learning outcomes to reduce the likelihood that students experience the online activities being unrelated to the learning objectives.

With the goal of discovering evidence to improve the students' experience of learning, it is equally valuable to employ a data-driven approach to find out the patterns of what, and how much, the students interacted with their online learning activities through investigating the digital traces left in the LMS. The results demonstrated that students who were involved in most learning events with higher proportions of tackling practical problems (pattern 1) tended to achieve relatively more highly compared to students, who participated the least in the learning events (pattern 4). The latter group also had a higher proportion of exercises focusing on just testing the understanding of theories and obtained the lowest course marks. The possible reasons as to why students in the observational online learning pattern 1 outperformed the students in the other patterns could be their active participation online, as reflected in the quantity of the completed online learning events. It could also be that the problem-solving exercises they favoured provided them opportunities to link theoretical concepts with practical problems, enabling a deeper understanding of the key theories and strengthening their understanding between the subject of learning and its applications in context. In addition, the practice of a higher proportion of the problem-solving type of online learning activities by the students adopting pattern 1 were congruent with the major assessment in the course, accounting for 50% of the course marks. The examination not only tested students' theoretical understanding but also their ability to solve practical problems by applying the theories. In contrast, the students in pattern 4 not only had insufficient practice, but the theory-related exercises they prioritized to work with lacked sophistication to allow them to extend theories into practice.

Difference in the depth of engagement with the online learning activities was also reflected by differences in the self-reported approaches and perceptions. The significant association between the self-reported and observational online learning patterns suggest that what students reported they did in the learning was consistent and coherent with what the observed data suggested they did. This outcome provides a type of triangulated evidence for both approaches (Knight, Buckingham Shum, & Littleton, 2014). The positive association found in the current study are also in line with a trend of associations identified in related research (Han & Ellis, 2017b; Han et al., 2020). This study has added to the quality of the research design by matching the self-reported and observational data so that both emphasised how the online part of learning experience in the blended course design. Interestingly, the results of our study are only partially consistent with the study by Gašević, Jovanović, Pardo, and Dawson (2017), which found significant link between self-reported and observational learning approaches only for the deep aspect but not for the surface aspect. Clearly more studies are needed before more conclusive evidence can be drawn.

The results from the data-drive approach suggest to us that teachers can monitor both the quantity and the quality of students' engagement with the online learning activities through using the learning analytics functions built in the LMS. The early detection of such information can hint to teachers the levels of engagement and the appropriateness of strategies the students approach their online learning in order to implement the class-level and/or personalized intervention strategies. For the whole class, the teaching staff can encourage online participation by selecting some online activities as part of discussions in the class to make these as essential activities for students to prepare beforehand. Teachers can also explain the purpose of each type of the online learning activities and their relations to help students make wise decision as to how much time and effort to be allocated for each one. Personalized intervention strategies could be also arranged to support students with different problems in learning. For example, the dashboard of students' online participation together with the class average online participation rate could be sent to those at the bottom level of online engagement to reminder them of catching up. Personalized strategic plans for how to deal with the online activities in an appropriate way can also be tailored and delivered to those lacking good strategies. It is hoped that through those timely interventions, students can make adjustments and improve their learning behaviors, lowering the potential risk of dropping out or failing the course.

6. Conclusion and implications for precision education

Research in precision education in an era of big data often solely relies on the techniques of advanced statistical learning, deep learning, and sophisticated data mining to achieve the "volume" principle in the four V's of the big data analytics (i.e., volume, velocity, variety, and veracity; IBM, 2018). This often results in the disintegration between the quantitative numbers and the established educational theories. Hence, the figures and the models generated in datacentric approach in precision education may severely limit the power and value in the abilities of guiding practice in learning and teaching or translating the numbers into meaningful interpretations for stakeholders in policy-making processes (Chan, 2019). Therefore, within current international trends in educational research, it is timely to combine theory-driven and data-driven approaches to advance the applications of big data analytics in precision education, so that more attention can be paid to the 'variety' principle to include multiple data sources, data collection methods, and data processing techniques in a single study (Topps & Cullen, 2019).

Our study is an initial attempt to demonstrate how combining theory-driven and data-driven approaches can be used to improve research on predicting students' academic achievement, which is one of the major objectives in precision education. Our study predicted learning outcomes by using students' perceived online learning experience through self-reports and their actual online learning behaviors observed by the digital-trace data in LMS. The strengths of such combined approaches lie in multiple ways. First, it has an advantage of offering richer information in the way of predicting students' learning over using a single approach, with each approach supplementing the other. While the observational digital-trace data are able to provide objective evidence as to what students actually do in their learning (Fincham, Gašević, Jovanović, & Pardo, 2019), they do not, however, have capacity to reflect students' intents and perceptions behind the ways they learn as in the self-reported studies (Asikainen & Gijbels, 2017; Gerritsen-van Leeuwenkamp, Joosten-ten Brinke, & Kester, 2019). Second, combined approaches can serve as a triangulation to check the validity of the results derived from either a theory-driven or data-driven approach. The significant and logical association found between the learning patterns of the two types of data in our study demonstrate complementarity and some degree of consistency and coherence of the two approaches. Third, the multiple data analysis methods used in the combined approaches also strengthen the analytical power of the analyses. Hence, the combined approaches have potential to transfer into other similar investigations, which tackle the complex issues of contemporary student experiences of learning, involving interactions not only between students and other individuals (e.g., peers, teachers, tutors, and laboratory facilitators), but also between students and things (e.g., tools, resources, and learning spaces). All the merits of combining theory-driven and data-driven approaches point out its future applications to advance research in precision education. While the current study demonstrates how combined approaches are used to improve current practice of predicting students' learning outcomes using either a theory-driven or data-driven perspective, future studies may extend this methodology to fulfil other objectives of precision education, such as diagnosing learning problems, personalized learning interventions, and preventing learning failure.

Acknowledgement

The authors are pleased to acknowledge the financial support of the Australian Research Council through grant DP150104163.

References

Asikainen, H., & Gijbels, D. (2017). Do students develop towards more deep approaches to learning during studies? A Systematic review on the development of students' deep and surface approaches to learning in higher education. *Educational Psychology Review, 29*(2), 205-234.

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In R. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 253-272). New York, NY: Cambridge University Press.

Bettinger, E., & Baker, R. (2013). The Effects of student coaching: An Evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis, 36*(1), 3-19.

Biggs, J., Kember, D., & Leung, D. (2001). The Revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133-149.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4th ed.). London, UK: McGraw-Hill.

Bliuc, A.-M., Goodyear, P., & Ellis, R. A. (2007). Research focus and methodological choices in studies into students' experiences of blended learning in higher education. *The Internet & Higher Education, 10*(4), 231-244.

Buckingham Shum, S., & Crick, R. (2012). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 92-101). New York, NY: ACM.

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. Lance & R. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 309-336). New York, NY: Routledge.

Chen, B., Resendes, M., Chai, C., & Hong, H. (2017). Two tales of time: Uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments, 25*(2), 162-175.

Collins, F., & Varmus, H. (2015). A New initiative on precision medicine. *New England Journal of Medicine, 372*(9), 793-795.

Cook, C., Kilgus, S., & Burns, M. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology, 66*, 4-10.

Dawson, S., Jovanović, J., Gašević, D., & Pardo, A. (2017). From prediction to impact: Evaluation of a learning analytics retention program. In X. Ochoa & I. Molenaar (Eds.), *Proceedings of the International Conference on Learning Analytics and Knowledge* (pp. 474-478). Vancouver, Canada: ACM.

Diseth, Å. (2003). Personality and approaches to learning as predictors of academic achievement. *European Journal of Personality, 17*(2), 143-155.

Ellis, R. A., & Bliuc, A.-M. (2016). An Exploration into first year university students' approaches to inquiry and online learning technologies in blended environments. *British Journal of Educational Technology, 47*(5), 970-980.

Ellis, R. A., & Bliuc, A.-M. (2019). Exploring new elements of the student approaches to learning framework: The Role of online Han, F., & Ellis, R. A. (2019). Identifying consistent patterns of quality learning discussions in blended learning. *The Internet & Higher Education, 40*, 12-19.

Ellis, R. A., & Goodyear, P. (2016). Models of learning space: Integrating research on space, place and learning in higher education. *Review of Education, 4*(2), 149-191.

Ellis, R. A., & Goodyear, P. (2019). *The Education ecology of universities: Integrating learning, strategy and the academy*. New York, NY: Routledge.

Ellis, R. A., Han, F., & Pardo, A. (2017). Improving insights from learning analytics – The Value of combining observed and self-report data on university student learning. *Educational Technology & Society, 20*(3), 158-169.

Entwistle, N. (2009). *Teaching for understanding at university: Deep approaches and distinctive ways of thinking.* Eastbourne, UK: Palgrave Macmillan.

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittlemeier, J., Rienties, B., Ullmann, T., & Vuorikari, R. (2016). Research evidence on the use of learning analytics: Implications for education policy. Seville, Spain: Joint Research Centre. doi:10.2791/955210

Fincham, E., Gašević, D., Jovanović, J., & Pardo, A. (2019). From study tactics to learning strategies: An Analytical method for extracting interpretable representations. *IEEE Transactions on Learning Technologies, 12*(1), 59-72.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64-75.

Gašević, D., Jovanović, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics, 4*(2), 113-128.

Gerritsen-van Leeuwenkamp, K., Joosten-ten Brinke, D., & Kester, L. (2019). Students' perceptions of assessment quality related to their learning approaches and learning outcomes. *Studies in Educational Evaluation, 63*, 72-82.

Hadwin, A., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*(2), 240-264.

Han, F., & Ellis, R. A. (2017a). Explaining learning achievement in student experience of blended learning: What can a sociomaterial perspective contribute? In H. Partridge, K. Davis, & J. Thomas. (Eds.), *Proceedings of the 34th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (pp. 89-94). Toowoomba, Australia: University of Southern Queensland.

Han, F., & Ellis, R. A. (2017b). Variations in coherence and engagement in students' experience of blended learning. In H. Partridge, K. Davis, & J. Thomas. (Eds.), *Proceedings of the 34th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (pp. 268-275). Toowoomba, Australia: University of Southern Queensland.

Han, F., & Ellis, R. A. (2020a). Personalised learning networks in the university blended learning context. *Comunicar, 62*(1), 19-30.

Han, F., & Ellis, R. A. (2020b). Initial development and validation of the Perceptions of the Blended Learning Environment Questionnaire. *Journal of Psychoeducational Assessment, 38*(2) 168-181.

Han, F., Pardo, A., & Ellis, R. A. (2020). Students' self-report and observed learning orientations in blended university course design: How are they related to each other and to academic performance? *Journal of Computer Assisted Learning, 36*(6), 969-980.

Hart, S. (2016). Precision education initiative: Moving toward personalized education. *Mind, Brain, and Education, 10*(4), 209-211.

IBM. (2018). *The Four v's of big data. IBM Big Data and Analytics Hub.* Retrieved from https://www.ibmbigdatahub.com/infographic/four-vs-big-data

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning, 31*(6), 638-654.

Jovanović, J., Gašević, D., Pardo, A., Dawson, S., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet & Higher Education, 23*, 74-85.

Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2015). Teacher competencies for the implementation of collaborative learning in the classroom: A Framework and research review. *Educational Psychology Review, 27*(3), 505-536.

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, assessment, pedagogy: Where blearning meets analytics in the middle space. *Journal of Learning Analytics, 1*(2), 23-47.

Krumm, A., Waddington, R., Teasley, S., & Lonn, S. (2014). A Learning management system-based early warning system for academic advising in undergraduate engineering. In J. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 103-119). New York, NY: Springer.

Laurillard, D. (2013). *Rethinking university teaching: A Conversational framework for the effective use of learning technologies.* London, UK: Routledge.

Lemons, C., King, S., Davidson, K., Puranik, C., Al Otaiba, S., & Fidler, D. (2018). Personalized reading intervention for children with Down syndrome. *Journal of School Psychology, 66*, 67-84.

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education, 27*, 27-52.

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

López-Pérez, M., López-Pérez, M., & Rodríguez-Ariza, L. (2011). Blended learning in higher education: Students' perceptions and their relation to outcomes. *Computers & Education, 56*(3), 818-826.

Lu, O., Huang, A., Huang, J., Lin, A., Ogata, H., & Yang, S.-J.-H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology & Society, 21*(2), 220-232.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology, 45*(3), 487-501.

Pardo, A., Han, F., & Ellis, R. A. (2016). Exploring the relation between self-regulation, online activities, and academic performance: A Case study. In *Proceedings of the 6th International Learning Analytics and Knowledge Conference* (pp. 422-429). Edinburgh, UK: ACM Press.

Pardo, A., Han, F., & Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies, 10*(1), 82-92.

Pardo, A., Jovanović, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology, 50*(1), 128-138.

Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching. The Experience in higher education.* London, UK: McGraw-Hill.

Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London, UK: Routledge.

Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528-540.

Rodríguez-Triana, M., Martínez-Monés, A., Asensio-Pérez, J., & Dimitriadis, Y. (2015). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology, 46*(2), 330-343.

Rojas-López, A., & García-Peñalvo, F. (2019). Personalized education for a programming course in higher education. In M. Sein-Echaluce, Á. Fidalgo-Blanco, & F. García-Peñalvo (Eds.), *Innovative trends in flipped teaching and adaptive learning* (pp. 203-227). Hershey, PA: IGI.

Romero, C., López, M.-I., Luna, J.-M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education, 68*, 458-472.

Siemens, G. (2013). Learning analytics: The Emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400.

Siemens, G., & Gašević, D. (2012). Guest editorial: Learning and knowledge analytics. *Educational Technology & Society, 15*(3), 1-2.

Tempelaar, D., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157-167.

The White House (2015). *Fact sheet: President Obama's precision medicine initiative*. Retrieved from http://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative

Topps, D., & Cullen, M. (2019). Scalable use of big data analytics in small group learning. *MedEdPublish*. doi:10.15694/mep.2019.000138.1

Trigwell, K. & Prosser, M. (2020). *Exploring university teaching and learning: Experience and context.* London, UK: Palgrave Macmillan.

Tsai, S., Chen, C., Shiao, Y., Ciou, J., & Wu, T. (2020). Precision education with statistical learning and deep learning: A Case study in Taiwan. *International Journal of Educational Technology in Higher Education, 17*, 1-13.

Williamson, B. (2019). Digital policy sociology: Software and science in data-intensive precision education. *Critical Studies in Education,* 1-17. doi:10.1080/17508487.2019.1691030

Wilson, K., & Fowler, J. (2005). Assessing the impact of learning environments on students' approaches to learning: Comparing conventional and action learning designs. *Assessment and Evaluation in Higher Education, 30*, 85-99.

Wilson, M., & Ismaili, P. (2019). Toward maximizing the student experience and value proposition through precision education. *Business Education Innovation Journal, 11*(2) 119-124.

Wu, J.-H., Tennyson, R., & Hsia, T.-L. (2010). A Study of student satisfaction in a blended e-learning system environment. *Computers & Education, 55*(1), 155-164.

Yang, S. J. H. (2019). Precision education: New challenges for AI in education [conference keynote]. In *Proceedings of the 27th International Conference on Computers in Education (ICCE)* (pp. XXVII-XXVIII). Kenting, Taiwan: Asia-Pacific Society for Computers in Education (APSCE).