Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.

Chapter # - will be assigned by editors

# EDUCATIONAL THEORIES AND LEARNING ANALYTICS: FROM DATA TO KNOWLEDGE

*The whole is greater than the sum of its parts*

AUTHOR(s) – Jacqueline Wong[1], Martine Baars[1], Björn B. de Koning[1], Tim van der Zee[2], Dan Davis[3], Mohammad Khalil[3], Geert-Jan Houben[3], and Fred Paas[1,4]

[1]Erasmus University Rotterdam, The Netherlands
[2]Leiden University, The Netherlands
[3]University of Bergen, Norway
[4]University of Wollongong, Australia

Abstract: The study of learning is grounded in theories and research. Since learning is complex and not directly observable, it is often inferred by collecting and analysing data based on the things learners do or say. By virtue, theories are developed from the analyses of data collected. With the proliferation of technology, large amounts of data are generated when students learn online. Therefore, researchers not only have data on what students have learned, but they also have data on the actions students take to achieve the desired learning outcomes. These data could help researchers to understand how students learn and the conditions needed for successful learning. In turn, the information can be translated to instructional and learning design to support students. The aim of the chapter is to discuss how learning theories and learning analytics are important components of educational research. To achieve this aim, studies employing learning analytics are qualitatively reviewed to examine which theories have been used and how the theories have been investigated. The results of the review show that self-regulated learning, motivation, and social constructivism theories were used in studies employing learning analytics. However, the studies at present are mostly correlational. Therefore, experimental studies are needed to examine how theory-informed practices can be implemented so that students can be better supported in online learning environments. The chapter concludes by proposing an iterative loop for educational research employing learning analytics in which learning theories guide data collection and analyses. To convert data into knowledge, it is important to recognize what we already know and what we want to examine.

Keywords: learning theories, big data, learning analytics, study success

## 1. INTRODUCTION

"Without theories, people could view research findings as disorganized collections of data, because researchers and practitioners would have no overarching frameworks to which the data could be linked"

Schunk (2012, p.10)

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



At all levels of education, the widespread use of new technologies such as interactive learning environments, learning management systems (LMS), intelligent tutoring systems (ITS), and online learning provides access to large amounts of student data (e.g., user interaction with online course content; Gašević, Dawson, & Siemens, 2015). Despite being a rich source of information, student data automatically collected in online learning environments, is typically not transformed into useful information for teaching and learning (Greller & Drachsler, 2012) and is used poorly across the educational domain (Dawson, Gašević, Siemens, and Joksimovic, 2014). In starting to transform large amounts of student data into useful information for learning, educational researchers recently have taken an interest in learning analytics approaches (Knight, & Buckingham Shum, 2017).

Although learning analytics is an evolving discipline, it draws on research, methods, and techniques from multiple established disciplines such as data mining, information visualization, psychology, and educational sciences (Gašević et al., 2015). Learning analytics is commonly defined as “the measurement, collection, analysis, and reporting of data about the learners and their contexts for the purposes of understanding and optimizing learning and the environment in which it occurs” (Siemens & Long, 2011, p.34). Trace data, also known as audit trails, log files, and event traces, are captured in online environments as students study the learning materials (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007). By utilizing learning analytics to examine the trace data, patterns related to learning processes can be identified to deepen our understanding of how students learn and add to the development of learning theories. In turn, this will help guide the design of instructional materials to support and enhance learning.

Given that understanding learning is a highly complex issue (Phillips, 2014), many learning theories have been developed over the last century based on different views of what learning is (Murphy & Knight, 2016). Learning theories are important not only because they can help to explain the phenomenon of learning, but also because design principles for learning environments, materials and tasks can be derived from the theories (Ertmer & Newby, 1993). Moreover, learning theories can help to convert information from learning analytics into actionable knowledge for instructional and learning design.

Importantly, as expressed by Ifenthaler (2017), a synergistic relationship between instructional design and learning analytics exists. On one hand, instructional designers can better evaluate the learning environment, materials, and tasks by processing data about the learners and their complex interactions within the learning environment using learning analytics approaches. On the other hand, learning analytics require theories and principles on instructional design to guide the transformation of the information obtained from the data into useful knowledge for instructional design. Consistent with Ifenthaler’s (2017) view, this chapter emphasizes the importance of taking learning theories into account when employing learning analytics in studies to support study success.

The aim of this chapter is to discuss how learning theories and learning analytics could be integrated in educational research since the whole is greater than the sum of its parts. We will first discuss the definition of learning and the role of learning theories. Then, a qualitative analysis of studies employing learning analytics to examine the current role of learning theories in research using learning analytics will be presented. The fourth section discusses the studies

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



reviewed and proposes an iterative educational research loop to integrate both educational theories and learning analytics.

## 2. UNDERSTANDING LEARNING

Building strong connections with the learning sciences was listed as one of the future directions of learning analytics by Ferguson (2012). The author reasoned that a good understanding of how learning occurs, how learning can be supported, and how student characteristics influence learning are needed if the goal of learning analytics is to understand and optimize learning. To understand the "how" of learning, one has to first define what learning is. Alexander, Schallert, and Reynolds (2009) proposed that learning can be defined as "a multidimensional process that results in a relatively enduring change in a person or persons, and consequently how that person or persons will perceive the world and reciprocally respond to its affordances physically, psychologically, and socially. The process of learning has as its foundation the systemic, dynamic, and interactive relation between the nature of the learner and the object of the learning as ecologically situated in a given time and place as well as over time" (p.186). This definition encapsulates the many perspectives of learning that were derived from the evolution of learning theories.

### 2.1 Evolution of learning theories

Based on a recent review of papers published in Review of Educational Research (RER) journal over the last century, Murphy and Knight (2016) found that learning sciences have been guided by three predominant theoretical lenses: behavioral, cognitive, and contextual. The authors used the word 'lenses' to analogously refer to the theories that researchers use. Just like how a certain lens may be more suitable for taking pictures in one situation than another, one learning theory may be more suitable for understanding learning in one environment than another. At the beginning of the 20th century, learning was viewed as a change in behavior (for an overview of learning theories, see Ormrod, 1999). Using the behavioral lens (e.g., Skinner, 1977), researchers focused on the responses of individuals to the environment and the ways to condition the desired responses. Several theories, such as classical conditioning and drive reduction theory, emerged from the behavioral viewpoint. In the middle of the 20th century, the cognitive lens (e.g., Ausubel, 1969) was used, viewing learning as a change in the mind of an individual. The focus was on understanding the mental processes that influence the processing and storing of information in the mind. Multiple theories, such as information processing theory and cognitive constructivism, developed under the cognitive lens. Although behavioral and cognitive lenses explained changes in one's behavior and mind, researchers were missing theories to explain social factors that influence learning that occurred in groups. The contextual lens arose to fill this gap. Under the contextual lens (e.g., Vygotsky, 1978), learning was viewed as contextually bound and a result of social interactions. Theories that developed from the contextual lens included social constructivism and social learning theory.

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



Murphy and Knight (2016) concluded that the shift in theoretical lens occurs when findings from new studies cannot be explained by the existing lens. However, a shift in theoretical lens does not invalidate the prior lens. Instead, each theoretical lens offers researchers the filter to focus on different areas of learning. More importantly, multiple theories can coexist and be simultaneously used to guide instructional practice. Therefore, it is at the discretion of learning scientists and learning analysts to recognize these nuanced perspectives of learning provided by the different lenses and apply learning theories based on the learning materials, learning condition, learning task, and learner characteristics.

## 3. ROLE OF EDUCATIONAL THEORIES IN LEARNING ANALYTICS

Given that learning theories evolved to accommodate new findings from studies, one might question if there is a need for learning theories. There is no doubt that a learning theory has to be built upon collective findings from studies (Alexander, 2006). Yet, without a theory to begin with, researchers will not know what to look out for. This conundrum of not knowing what to look for is magnified in research utilizing learning analytics since studies conducted in online learning environments usually involve the collection of immense amounts of data. Therefore, a good theory is needed to guide researchers (Alexander, 2006). Using the theoretical lens of a learning theory, researchers will be better positioned to formulate their research questions, make hypotheses about what learning outcome to expect, make decisions on the research methods, and finally, make interpretations of the results derived from learning analytics approaches (Murphy & Knight, 2016).

Since one of the aims of learning analytics is to advance educational research and practice, it is of interest to take a look at how well learning theories are being referred to or investigated in studies employing learning analytics to support study success. Na and Tasir (2017) found mixed effects of the use of learning analytics interventions to support students' success. However, it is not clear whether the learning analytics interventions in the studies reviewed were based on specific learning theories or whether any learning theories were mentioned in the studies. Gaining insight into this is important to aid our understanding of how learning analytics can affect study success. Therefore, the current study extends the Na and Tasir study by investigating whether studies employing learning analytics to support study success take into account learning theories and if so, to what extent the learning theories are guiding the studies. The main research question addressed in our review is:

Which learning theories have been used in the studies examining learning analytics approaches to support study success?

### 3.1 Research methodology

The review methodology consisted of four sequential steps qualifying it as a systematic qualitative review: a) literature search based on keywords to identify relevant papers, b)

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



assessment of search results to select a set of primary studies, c) categorising and integration of the results, and d) reporting the findings (Gikandi, Morrow, & Davis, 2011).

The aim of the first step was to identify published papers examining study success using learning analytics. Given that learning analytics has been applied to examine success in different domains and at various levels of education, broad search terms (i.e., study success, student success, and achievement) were used to capture all forms of success and achievement related to study and student. The search terms “learning analytics” AND “stud* success” OR “achievement” were used to search for papers indexed in the databases of Scopus (http://www.scopus.com) and Web of Science (http://www.webofknowledge.com/wos) in December 2017. These two databases were chosen because of their multidisciplinary indexing of articles across journals and conferences. We only included papers published in journals and conferences over the last seven years starting from 2011 when the first learning analytics and knowledge conference proceeding was published. After removing duplicates, 164 papers that were published in 79 distinct journals (46) and conference proceedings (33) remained.

The second step was to select a set of primary studies. Given the aim of the study was to qualitatively review the role of learning theories in studies employing learning analytics, impact factors were used to identify papers that were published in top five journals and conferences. We ranked the scientific influence of the 46 journals based on impact factors obtained from Scimago Journal and Country Rank (SJR; SCImago, 2007) and Journal Citation Reports (JCR). The two impact factors were taken into account as SJR is built on Scopus database while JCR is built on Web of Science database. We ranked the conferences using H-index obtained from Google Scholar Metrics since conferences were not ranked by SJR or JCR. Table 1 shows the distribution of papers published across the top five journals and conferences according to the SJR, JCR, and H-index. This selection process resulted in a set of 27 papers published in six journals and five conferences.

Table 1. *Number of papers selected based on five highest-ranked journals according to the journal titles in alphabetical order*

| **Publications** | **Number of papers** | **SJR** | **JCR** |
|---|---|---|---|
| **Journal titles** | | | |
| Computers and Education | 6 | 2.61 | 3.82 |
| International Journal of Computer-Supported Collaborative Learning | 6 | 1.47 | 3.47 |
| Computers in Human Behavior | 1 | 1.60 | 3.44 |
| Internet and Higher Education | 4 | 2.83 | 4.24 |
| Journal of Computer Assisted Learning | 1 | 1.65 | 1.25 |
| Soft Computing | 1 | .75 | 2.47 |
| **Conference titles** | | **H-index** | |
| Americas Conference on Information Systems (AMCIS) | 1 | 22 | |
| ACM Conference on International Computing Education Research (ICER) | 2 | 19 | |
| Conference on User Modeling, Adaptation and Personalization (UMAP) | 1 | 21 | |

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



| | | |
|---|---|---|
| IEEE Global Engineering Education Conference (EDUCON) | 2 | 19 |
| International Learning Analytics & Knowledge Conference (LAK) | 2 | 32 |

The 27 papers went through a second selection process based on the study type (i.e., experimental, correlational, student survey only, and conceptual/review). We selected only empirical papers (i.e., experimental and correlational studies) for the review, specifically papers that used learning analytics approaches to analyze trace data obtained from the online learning environments. This allowed us to examine whether the studies referred to learning theories when employing learning analytics approaches to analyze the trace data. We refer to the definition of learning analytics as "the measurement, collection, analysis, and reporting of data about the learners and their contexts for the purposes of understanding and optimizing learning and the environment in which it occurs" (Siemens & Long, 2011, p.34). Therefore, we selected studies that collected data about the learner in online learning environment. During this selection process, papers that used student surveys only (Atif, Bilgin, & Richards, 2015; Tan, Yang, Koh, & Jonathan, 2016; Zhuhadar, Yang, & Lytras, 2013), reviews (Tlili, Essalmi, Jemni, & Chen, 2016), and conceptual papers (Kim & Moon, 2017; Wise and Schwarz, 2017, Yassine, Kadry, & Sicilia, 2016) were removed. This resulted in a final set of 20 empirical papers involving the analysis of trace data using learning analytics approaches.

In the third step, the 20 papers were read in detail and categorised according the learning theories mentioned in the papers. For each paper, further information on the learning environment investigated, the learning analytics techniques/application applied, and the types of data collected were extracted from the papers. Finally, the findings of the papers were integrated and qualitatively reviewed based on learning theories mentioned in the papers to answer the research question.

## 3.2 Results and discussion

Among the set of 20 papers, there were only two (quasi)experimental papers (i.e., Rowe et al., 2017; Tabuenca et al., 2015) comparing different treatment conditions. Tabuenca et al. (2015) compared the effects of delivering notifications between a fixed and a random schedule to support self-regulated learning while Rowe et al. (2017) compared the use of in-game measures of implicit science knowledge either as a bridge or as a supplement to teaching activities to enhance learning. The rest of the 18 papers were correlational studies.

### 3.2.1 Learning theories and learning analytics applications

After categorising the papers, 16 studies were found to mention theories related to learning while the other four studies did not. Table 2 shows a summary of the learning theories mentioned in the 16 studies, the learning environments in which the studies were deployed, the learning analytics

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



approaches, and the types of data that were collected. Most studies tended to be situated within self-regulated learning *(n* = 6), followed by motivation (*n* = 2), and social constructivism (*n* = 2). Another six individual studies used other concepts related to learning (i.e., learner effort, feedback, deep learning, engagement, implicit knowledge, and a combination of concepts).

#### 3.2.1.1 Self-regulated learning

Self-regulated learning (SRL) was the most employed theory related to learning in the selected studies. Models of SRL characterize self-regulated learners as students who actively use and adjust their learning strategies to achieve their learning goals (Bos & Brand-Gruwel, 2016; Kizilcec, Perez-Sanagustin, & Maldonado, 2017). There were six studies (i.e., Bos & Brand-Gruwel, 2016; Jovanović, Gašević, Dawson, Pardo, & Mirriahi, 2017; Kizilcec et al., 2017; Siadaty, Gašević, & Hatala, 2016; Tabuenca, Kalz, Drachsler, & Specht, 2015; You, 2016) which examined the use of learning analytics albeit in different learning environments (e.g., MOOCs and LMS). You (2016) used hierarchical regression analyses to identify events from data generated in Learning Management Systems (LMS) to predict course achievement in e-learning courses. The results showed that students who accessed the content videos within the instructor-scheduled time and watched the full-length of the video was the strongest predictor of course achievement, followed by number of late submissions, number of course log-ins, and whether the course information was downloaded.

Instead of predictive modelling, Jovanović et al. (2017) employed an exploratory learning sequence analysis to compare learning sequences of high performers and low performers in a flipped classroom. Low performers mostly focused on summative assessments that counted towards their final course scores, while high performers engaged with all the activities (i.e., formative assessment, summative assessments, reading materials, and videos) evenly. Using agglomerative hierarchical clustering based on Ward’s method, the authors identified five student profiles (i.e., intensive, highly strategic, strategic, selective, and highly selective) based on the activities that students chose to engage in (e.g., focus on summative assessment or focus on course video). While the learning analytics approach helped to detect and describe differences in students’ learning behavior, it could not provide reasons as to why students’ behavior differed.

To be able to explain differences in students’ behaviors, Kizilcec et al. (2017) correlated student behavioral data with student self-reports about their learning approach. The authors examined the relationship between SRL survey data, student interactions with course contents in MOOC, and personal goal attainment. The results showed that students’ self-reported level of SRL was related to their intentions in completing the course. Students who scored higher on goal setting and strategic planning were more likely to attain their goals, while students who reported more help-seeking were less likely to attain their goals. In general, students with higher self-reported use of SRL strategies spent more time revisiting assessments. Based on the results, the authors suggested MOOC instructors to guide students in goal setting and strategic planning activities.

Instead of analysing temporal learning sequences, Bos and Brand-Gruwel (2016) chose a more direct method of counting the number of times an activity was done and the time spent on the activities in the learning environment. Similar to Kizilcec et al.'s (2017) study, Bos and Brand-Gruwel (2016) combined SRL survey data with data generated in a LMS platform. In the study, students were first clustered based on their scores on the administered SRL surveys. The analysis resulted in three clusters: i) students who reported lack of regulation when external regulation is absent, ii) students who reported use of self-regulation strategies when external regulation is absent, and iii) students without a clear regulation strategy. The results showed that although students in the three clusters used the online resources to a similar extent (e.g., number of videos watched), they benefited differently from the use of the same resources. Frequencies of login and time spent in the LMS alone were found to be poor predictors of students' performance. This is not surprising given that the duration measured may not be the actual time students spent processing information on the page in an online environment.

Two studies were found to examine interventions that support SRL. Siadaty et al. (2016) examined the relationship between students' perceived usefulness of the interventions and actual use of SRL interventions. Seven SRL scaffolds were embedded in a technologically-enhanced learning environment: i) usage information, ii) social context of the workplace, iii) progress towards goal attainment, iv) peer-recommended learning goal, v) system-recommended competencies, vi) system-recommended learning path, and vii) learning resources students own or have shared with the organization. The authors predefined activities in the online environment to measure SRL processes. For example, rating a learning path in the online environment is a measurement of self-evaluation as a SRL process. The analysis of students' activities in the online environment showed that i) frequencies of planning activities were related to looking at usage information, social context of workplace, and system recommended competencies and learning path, ii) frequencies related to performance phase were related to information about social context of the workplace and learning resources they own or have shared with the organization, and iii) frequencies related to reflection phase were related to competences of goals. The findings suggested that providing information on social context of the workplace had the highest impact on processes of SRL. The authors concluded that recommender system technology should be integrated in modern workplace environments to support SRL. Although this study showed that recommender system technology enhances SRL on the whole, it is not clear which factors in particular (e.g., system-recommended competencies or system-recommended learning path) influenced SRL. Moreover, a recommender system might increase students' reliance on the recommendations instead of their own regulation of learning.

In another experimental intervention study by Tabuenca et al. (2015), a within-subjects design was used to examine the effect of a mobile tool for tracking and monitoring study time on SRL. At different time points in the study, students received notifications containing tips for time management that were either generic or based on learning analytics at random time or on a fixed schedule. Students reported an increase in perceptions of time management and planning skills after the notification intervention. Students specifically preferred notifications sent early in the day with learning analytics information about their personal time-management and behavior. Activities in the time-logs showed that students were more active at certain time periods and on certain days, and there were more records of study time whenever notifications were sent.

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



However, students who had more time-logs did not score higher in the final exam than students who had less time-logs.

The six discussed studies exemplify the complexity of examining SRL in an online environment. SRL processes consist of a broad range of learning strategies such as time management, goal setting, and planning. The studies used different learning analytics approaches to examine the trace data. Trace data can be examined by aggregating an action in terms of frequencies and time spent on the online materials (e.g., Bos & Brand-Gruwel, 2016), action in context such as submitting an assignment on time (e.g., You, 2016), transitions of activities (e.g., Kizilcec et al., 2017), and learning sequences (e.g., Jovanović et al., 2017). The learning analytics approaches provide insights into what students do in the online environment that might relate to SRL. However, trace data alone are insufficient to explain students' behavior. Among the selected studies, four studies attempted to shed more light on this by relating trace data to self-report data. The combination of trace data and self-reports enables a deeper understanding on the relationship between SRL and students' behavior. For example, students who reported higher levels of SRL also spent more time revisiting assessments (Kizilcec et al., 2017). It should be noted that these studies involved primarily correlational analyses, so causality cannot be inferred from these studies. Therefore, there is a need for more experimental studies such as the Tabuenca et al.'s (2015) study. Together, the selected studies suggest that SRL is a promising area in which learning theories and learning analytics converge. The fact that SRL turned out to be the most investigated learning theory in learning analytics research is understandable given that SRL has been shown to be crucial to academic success in online learning environments (Broadbent & Poon, 2015).

#### 3.2.1.2 Motivation

Two studies (i.e., Barba, Kennedy, & Ainley, 2016; Lonn, Aguilar, & Teasley, 2015) examined motivation, each with a different theoretical approach. Barba et al. (2016) examined the impact of general motivation (i.e., individual interest, mastery-approach, utility value beliefs) and state-level motivation (i.e., situational interest). Motivation in this study was defined as systems of beliefs that can be activated by contextual and personal factors. Using structural equation modelling, they investigated the relationship between motivation, participation, and study success in MOOCs. The different types of motivation were measured by surveys whereas participation in MOOC activities was measured by the number of videos viewed and the number of quizzes attempted. The results showed that students who reported a mastery-approach towards learning attempted more quizzes. Students' report of higher situational interest was related to larger number of videos watched. The strongest predictor of final grades in the MOOCs was the number of quizzes attempted followed by situational interest. These results suggest that it is important for MOOC designers to focus on supporting situational interest.

The study by Lonn et al. (2015) focused on achievement goal theory to measure the effects of a learning analytics intervention in a summer bridge programme. Achievement goal theory was used to conceptualize students' two types of motivation orientation: mastery goals focus on the development of personal competencies while performance goals focus on showing competence compared to others. The intervention in Lonn et al.'s (2015) study consisted of an early alert

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



system that tracked students' progress to identify whether they were at-risk. Student advisors in the course could then look at the information provided by the early alert system and act accordingly. Results of the study showed that the mastery-approach decreased over time, suggesting that the learning analytics intervention is negatively correlated to mastery-approach. Therefore, the study suggested that this learning analytics intervention should be implemented with caution as it may have a negative influence on student motivation.

Both discussed studies used surveys to measure motivation instead of predefining student activities in the log data as proxies of motivation (as was for example done in the SRL study by Siadaty et al., 2016). This could be due to the fact that motivation is a cognitive process related to goal-directed behavior (Schunk, 2012). The two studies exemplify the important relationship between learning theories and learning analytics. Barba et al. (2016) linked student motivation to participation, providing insights to how motivation can be manifested in learning behaviors. This suggests that learning analytics can help to quantify learning behaviors to deepen our understanding of motivation –what behaviors are related to motivation. Lonn et al.'s (2015) study showed that learning analytics interventions can affect motivation. This suggests that learning theories can help guide the implementation of learning analytics interventions –how can motivation be supported to enhance study success.

#### 3.2.1.3 Social constructivism

Two studies (i.e., Carter & Hundhausen, 2016; Joksimović, Gašević, Loughin, Kovanović, & Hatala, 2015) were categorized under the theoretical framework of social constructivism. As discussed in Section 2, social constructivism can be viewed from a contextual lens. Under this view, learning does not only occur only within the learner but is contextualized and dependent on the environment. These studies examined the interactions in online learning environments and related the interactions to theory of social constructivism. Carter and Hundhausen (2016) examined peer interactions using trace data generated in a programming environment where students could pose and answer questions. The results showed that students who asked a question, received a suggestion, and acknowledge the suggestion were more likely to make progress in the course and achieve better final grades.

Joksimović et al. (2015) not only examined student-student interaction but also interaction between student and instructor, student and content, and student and system in an online course. The analytical approach involved identifying the interactions, classifying them into interaction types, calculating the frequency and time spent on each interaction type, and statistically analysing the relationship between interaction types and final grades. The results showed that student-system interactions were positively related to final grades while student-content interactions were negatively related to final grades. Also, student-instructor interactions were negatively correlated to final grades in core courses only. Based on these results, the authors suggested the different courses (i.e., core, elective and foundational courses) require different forms of interactions to support the learning process.

The discussed studies demonstrate that using learning analytics enables researchers to examine the effect of actual interactions instead of relying on only perceived interactions. The results from the two studies showed that interactions such as student-student interactions (Carter &

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



Hundhausen, 2016) or student-system interactions (Joksimović et al., 2015) can differentially affect grades. Future studies can build on these two studies to further compare different properties of interactions (e.g., asynchronous, synchronous, virtual, augmented). In addition, learning analytics can also be used to help students monitor their interactions. To conclude, there is a reciprocal relationship between learning analytics and social constructivism. Learning analytics provide evidence for learning from a social constructivist perspective while social constructivism helps to make sense of interaction data provided by learning analytics.

#### 3.2.1.4 Studies using specific learning concepts

In this section, other specific learning concepts mentioned in individual papers are discussed. What stands out is that the extent to which the learning theories were discussed in the studies as well as the moment at which they were introduced varied. Most studies introduced the learning theories at the beginning but failed to link the patterns or clusters obtained back to the learning theories. In some studies, certain concepts related to learning were mentioned although no clear learning theories were stated.

Zhao, Davis, Chen, Lofi, Hauff, and Houben (2017) investigated the link between assessment and learner effort within a MOOC. Educational researchers suggest that learner effort should be distributed evenly across topics and course weeks. This appears to be related to the concept of distributed practice (Dunlosky & Rawson, 2015). Results of the study showed that MOOC students behaved differently after meeting the minimum passing requirement. Some students reduced their engagement with videos and quizzes after passing, suggesting that students who passed did not necessarily have complete mastery of all course content. The authors concluded that differences in post-passing behaviors may be related to students' motivation for taking the course. However, student motivation was not actually measured in this study.

The role of feedback is mentioned in Sedrakyan, Snoeck, and De Weerdt's (2014) study. Feedback can be linked to several learning theories depending on the focus of the feedback (Thurlings, Vermeulen, Bastiaens, & Stijnen, 2013). Feedback is also viewed as an important component of self-regulated learning (Butler & Winne, 1995). Sedrakyan et al. (2014) examined whether quality of work can be predicted by differences in students learning patterns. Based on a three-dimensional analysis (i.e., hierarchical, modelling, and time-trend), the results showed that the quality of work can be predicted by students' learning pattern. This suggested that instructors can identify poor performing students and provide process-oriented feedback during the task to enhance their quality of work. The potential of feedback to support learning is proposed but not investigated in the study.

Romero-Zaldivar, Pardo, Burgos, and Kloos (2012) employed the concept of deep learning (Webb, 1997) to evaluate the effectiveness of a virtual appliance where students interact with the tools from a pre-installed application on the computer. Based on the assumption of deep learning, learning is enhanced when students have high level of interactions with the learning tools. Predictive modelling based on the frequency and time spent with the tools in the learning environment showed that students' final grades can be predicted by the use of two out of the six tools available. However, the authors did not relate the activities back to the concept of deep learning.

Likewise, predictive modelling was used in Junco and Chen's (2015) study in which theory of engagement was mentioned. The authors gave a brief background on the theory on engagement by Astin (1984) which suggested that amount of learning is related to the amount of time and effort that students invest. Course outcomes were predicted based on the usage data generated from a digital textbook. The results showed that time spent reading was significantly related to course grades. Also, students in the top 10th percentile used more highlights than students in the lower 90th percentile. The study did not further examine the texts that were highlighted, as such, it is not clear how students were using the highlights to support their reading.

Rowe et al. (2017) examined the assessment of implicit science knowledge in digital games. Implicit knowledge is defined as what learners are able to do given their existing understanding. In-game measures of implicit learning were first developed using educational data mining technique. The digital games were then used either used as a bridge for science class, as an extra activity outside of class, or not used at all in an experimental study. Using hierarchical linear models, the results showed that the in-game measures of implicit knowledge correlated to external measures of learning (i.e., post-assessment). Moreover, students did better in the course when teachers use information about students' implicit knowledge for explicit teaching.

Kim, Park, Yoon, and Jo (2016) constructed proxy variables in an asynchronous online discussion environment to measure various concepts related to learning: active participation in the course, engagement with discussion topics, consistent effort and awareness, and interaction. Psychological and behavioral characteristics of high performing students were then identified for each concept. For instance, psychological and behavioral characteristics of consistent effort and awareness were responsibility, punctuality, time management, and intrinsic motivation. These characteristics were further operationalized by proxy variables that can be measured by the log file data such as interval regularity of visit to the online environment, total time spent, number of LMS visits, number of discussion board visits, and number of posts. To evaluate how well the proxy variables were able to predict good and poor performers, the authors used random forest technique to develop the prediction model. The results indicated that, using the proxy variables, the prediction model was highly accurate. The authors suggested that for whole-class discussions, students can be encouraged to reply to others and be supported to work towards more in-depth discussion. For team-based discussion, the authors suggested employing support for cognitive engagement at the beginning and sustain engagement throughout the course.

The studies mentioned above suggest that learning analytics have the potential to provide information on various learning-related concepts. Learning analytics add value to educational research through the collection of different sources of data (e.g., trace data) and measuring and analysing the data in ways that can be related to learning theories (e.g., predictive models and clustering). However, for learning analytics to achieve the potential of providing deeper insights to learning, it is important to first clearly determine which learning theories are being investigated so that decisions can be made on which data to be collected and which analytical method to be used.

Table 2. *Learning theories identified from the selected papers*

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



| **Learning theories used** | **Authors** | **Learning environment** | **LA technique / application** | **Trace data collected** | **Performance-related measures** |
|---|---|---|---|---|---|
| Self-regulated Learning (SRL) | Bos & Brand-Gruwel (2016) | LMS, Blended course | - Clustering<br>- Multiple regression analysis | Time spent viewing recorded lectures, number of formative assessment complete and score on the formative assessment, time spent using the LMS and number of clicks in the LMS (e.g., announcements, video files, viewing grades) | Mid-course and final course assessment, self-reported inventory of learning styles (ILS) |
| | Jovanović et al., (2017) | Flipped course with learner dashboard | - Learning sequence analysis<br>- Clustering | Number of correctly and incorrectly solved summative and formative assessment items, number of solutions requested, number of videos played, number of access to content, dashboard, and schedule | Midterm and final exam scores |
| | Kizilcec at al., (2017) | MOOCs | - Logistic regression models<br>- Transition graphs | Number of transitions from one interaction state type (e.g., begin a video to complete a video) and time spent on each type of learning material, number of learning materials interacted | Course goals (i.e., earning a course certificate, complete all assessments, and complete all lectures), self-reported self-regulation of learning |
| | Siadaty et al. (2016) | Learn-B environment | - Trace-based methodology | Number of actions performed by students in the learning environment (e.g., clicking on different competencies, choose an available learning path, rate a learning path) | Perceived usefulness of the features provided in the learning environment |
| | Tabuenca et al. (2015) | Online course with support from mobile application | - SQL queries to examine the distribution of study time | Students log their study time on the mobile application which in turn visualizes the summary of their recording that shows time spent per assignment | Course grades, self-reported self-regulation of learning |
| | You (2016) | LMS, e-learning | - Hierarchical regression | Time spent viewing the instructional videos, number of course logins, number of late submission, students' reply to instructor's post, fulfilment of attendance, Number of posting in the discussion board | Midterm and final exam scores |
| Motivation (achievement goal) | Lonn et al. (2015) | Summer bridge program | - Multiple linear regression | An early warning system that assigned students one of the three statuses (i.e., encourage, explore, engage) based on the points students earned on their coursework, difference between the course average, and number of LMS logins | Course grades, pre- and post-measures of self-reported achievement goals (i.e., mastery- and performance- approach, and performance-avoidance orientation) |

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



| | | | | | |
|---|---|---|---|---|---|
| Motivation (mastery, value beliefs, individual interest, and situational interest | Barba et al. (2016) | MOOC | - Structural equation modelling | Number of clicks on videos and number of quiz attempts | Final grade |
| Socio-constructivism (interaction types: student-content, student-instructor, student-student) | Joksimović et al. (2015) | LMS | - Hierarchical linear mixed models using restricted maximum likelihood (REML) estimate | Number of and time spent on four types of interaction (i.e., student-student, student-content, student-teacher, student-system) | Final course grades |
| Social learning theory | Carter & Hundhausen (2016) | Social Programming Environment | - Chi-squared test | Number of interaction types (i.e., post, reply, receive a suggestion), topic of post, progress in the course | Average grade for programming assignment and final course grade |
| Learner effort (distributed practice) | Zhao et al. (2017) | MOOC | - k-means clustering | Time spent watching videos and quiz score | Eligibility to earn a course certificate |
| Feedback (Process-oriented) | Sedrakyan et al. (2014) | Conceptual modelling environment (JMermaid) | - Process model discovery and dotted chart analysis | Event log of students' group work during the modelling process (i.e., create, edit, delete, redo, and copy). | Scores on the group project's final solution |
| Deep Learning | Romero-Zaldivar et al. (2012) | Virtual appliance | - Multiple regression<br>- Prediction | Time spent in the learning environment, number of times action was performed (i.e., write a command, open a webpage, open a file with an Editor, and using the C compiler, memory profiler, and C debugger, time spent performing each action. | Final grades |
| Engagement | Junco & Clem (2015) | Digital textbooks | - Hierarchical linear regression | Number of reading days, number of reading sessions, time spent reading, number of pages read, number of highlights, number of bookmarks, number of notes | Final course grades |
| Implicit Knowledge | Rowe et al. (2017) | Computer game | - Approach Map for network clustering | Implicit knowledge measured by in game behavior involving specific strategic moves | Pre-post assessment improvement |
| Combination of concepts (active participation, engagement, consistent effort and awareness, interaction) | Kim et al. (2016) | LMS in blended course | - Random forest technique to create a prediction model | Time spent on LMS, number of LMS visits, number of discussion board visits, number of posts, post length, interval between LMS visits, interval between discussion board visits, number of replies received by a student, | Final course grades |

number of replies generated by a student

---

### 3.2.2 Absence of learning theories

Out of the 20 empirical studies that used correlational and experimental design, 16 studies were found to mention certain learning theories or concepts related to learning. The four studies that did not mention any learning theories were mainly focused on using exploratory approaches to identify student behaviors predictive of academic achievement. Studies by Brooks, Erickson, Greer, and Gutwin (2014) and Liu and d'Aquin (2017) used clustering methods to identify groups of learners that were most likely to be successful. The third study by Carter, Hundhausen, and Adesope (2015) argued that theories in learning research lacked the ability to predict "student performance that are dynamic, robust, and continuously updated throughout a course". Therefore, they proposed a normalized programming state model that explained how removing compilations errors from a program is related to better achievement. Finally, Marbouti, Diefes-Dux, and Madhavan (2016) compared seven prediction methods to evaluate the models' accuracy in identifying at-risk students: i) logistic regression, i) support vector machine, iii) decision tree, iv) multi-layer perceptron, v) naives bayes classifier, vi) k-nearest neighbour, and vii) ensemble model. The accuracy of the models depends on the performance data collected which can be affected by quality and reliability of the grading. This suggests that there is no one prediction method that is the most accurate. Together, while the studies using various learning analytics methodologies without mentioning learning theories do provide insights into factors influencing student success, we argue that more direct links with learning theories would help to advance the conversation from 'what are the factors that influence learning?' to 'how and why do these factors influence learning?'.

## 4. CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

The aim of the current review was to investigate which theories have been used in studies employing learning analytics to support study success. We searched for studies in two major databases and selected 20 empirical papers for the final review. Based on the studies reviewed, self-regulated learning (SRL) appears to be widely referenced in studies employing learning analytics (i.e., Bos & Brand-Gruwel, 2016; Jovanović et al., 2017; Kizilcec et al., 2017; Siadaty et al., 2016; Tabuenca et al., 2015; You, 2016). There are also two studies related to theories about motivation (i.e., Barba et al., 2016; Lonn et al., 2015) and two studies related to theories on social constructivism (i.e., Carter & Hundhausen, 2016; Joksimović et al., 2015). There are

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



several single studies on different concepts related to learning such as learner effort (i.e., Zhao et al.,2017), feedback (i.e., Sedrakyan et al., 2014), deep learning (i.e., Romero-Zaldivar et al., 2012), engagement (i.e., Junco & Clem, 2015), and implicit knowledge (i.e., Rowe et al., 2017). Kim et al.'s (2016) study is the only exception that examined multiple concepts related to learning (i.e., active participation, engagement, consistent effort and awareness, interaction).

All of these studies are examples of how learning theories are used in studies that employed learning analytics to examine student behaviors in online learning environments. We observed that, at present, learning theories have been used in studies employing learning analytics in two ways. First, learning theories help to guide decisions on the types of data to be collected and the learning analytics approaches to take. From the studies, it is noted that similar data points (e.g., time spent on an activity) can be used as proxies related to different learning theories (e.g., SRL and engagement). Therefore, learning theories play an important role in explaining the concept of learning that is being measured. For example, researchers examining SRL may focus on learning sequences (e.g., Jovanović et al., 2017), while researchers taking the perspectives of socio-constructivism may focus on students' interactions with instructors and other students. Second, learning theories help researchers to explain why students might behave in certain ways and why behaving in certain ways might lead to study success. For example, students who are better at SRL, are more inclined to revisit assessments, and hence, more likely to be successful learners (Kizilcec et al., 2017).

Although this chapter has identified several learning theories mentioned in studies employing learning analytics approaches to support study success, a trend that we observed is that learning theories are often briefly mentioned or introduced at the beginning of the articles but rarely circled back to contextualise the results with the learning theory mentioned (e.g., Romero-Zaldivar et al., 2012). While the first part (introducing the theory) is certainly a step in the right direction, we contend that a robust, thorough employment of learning theory in learning analytics should use the results obtained from the various analyses to make direct inferences about the applicability of the theory on the learning behavior observed (and also, perhaps, the method applied, as learning analytics borrows from a very wide variety of methodologies). As learning analytics is a young, blossoming, interdisciplinary field, it is comprised of researchers from a plethora of other fields, each bringing with them various levels of expertise in different topics. And, as is often the case in interdisciplinary research, knowledge from some fields will inevitable be more prominent than others. For example, a large part of learning analytics research comes from computer science departments (Dawson et al. 2014). To move forward within the learning analytics field, it is imperative that learning analytics researchers, regardless of their base discipline, go beyond a surface-level understanding of the learning theory or theories they are employing. Instead of having it merely as a framing at the beginning of a paper, the learning theories should be integral to the research narrative and provide explanations at every stage about how the theory informed each decision along the way.

Learning theories play an important role in transforming results obtained from learning analytics into insights about learning. While learning analytics can help to identify patterns of student behaviors and add new understanding to the field of educational research, it alone does not provide explanations for underlying mechanism. The analysis of trace data in Jovanović et

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



al.'s (2017) study helped to detect series of student actions corresponding to the unfolding of learning strategies used by the students, yet the results fall short in explaining what underlying factors could have accounted for the differences in the use of learning strategies between different groups of students. In accordance with learning theory related to self-regulated learning in Zimmerman's model (Zimmerman & Campillo, 2003), the use of learning strategies is preceded by self-motivational beliefs and processes of task analysis. By adopting Zimmerman's model in their study, Jovanović et al.'s (2017) could examine whether motivational beliefs influence students' use of learning strategies manifested in the different series of student actions. When using learning theories, researchers should recognize that a theory may have a number of constructs, for instance motivational beliefs can include self-efficacy beliefs, goal orientation, and task interest. Therefore, discussions among researchers are needed to discern learning theories that may align better with learning analytics. The potential of learning analytics can only be realized when the nuances of learning theories are aligned with the nuances of the data.

Another trend that we observed was the considerable overlap in the analytical techniques found in several studies. For instance, regression was mostly used as the analytical method in the first stage followed by clustering in the second stage (Bos & Brand-Gruwel, 2016; You, 2016; Lonn et al., 2015; Romero-Zaldi var et al., 2012; and Junco & Clem, 2015). There were also studies that explore novel analytic approaches such as trace-based methodology (Siadaty et al., 2016) and process model discovery (Sedrakyan et al., 2014). The multiple analytic approaches used in the studies demonstrate the ability of learning analytics to deep dive into rich data sources of log files, discussion forums, time spent on tasks, and number of interactions to extrapolate learning as a holistic and social process based on students' behaviors. However, as noted by Gašević at al. (2015), the interpretation of students' behaviors can change depending on the understanding of the students' internal conditions (e.g., cognitive load, self-efficacy, achievement-goal orientation, and interest) as well as external conditions (e.g., instructional design, and previous experience with using the tool). Therefore, future studies should include multiple sources of data that can be derived from learning theories (e.g., prior knowledge, self-report of motivation) to supplement the analysis of student data generated in the online environments.

We propose an iterative loop as illustrated in Figure 1 to guide future educational research employing learning analytics. The iterative loop starts with a theory of learning (learning theory 1.0) that is used to examine how students learn in a learning environment. This is followed by theory-guided data collection so that a predefined set of data is collected. Subsequently, theory-guided selection of learning analytics methods is used to analyze the data. The analysis based on learning analytics can either provide evidence to support the hypotheses derived from learning theory 1.0 or suggest how the theory can be developed (learning theory 1.n). The process is iterative until the findings fit a theory. Rowe et al.'s (2017) study is an example of a study which already fits well with what we proposed. Based on theory of implicit knowledge, data were collected in a digital game environment to detect student actions related to implicit knowledge based on learning analytics approaches. Hypotheses were derived to examine whether students whose teachers used the digital game to assess implicit knowledge as a bridge in class would perform better than students whose teachers use the digital game as a supplementary activity and students whose teachers did not use the digital game at all. New data are collected in the digital game environment along with course grade to understand how assessing implicit knowledge can support teacher, and ultimately enhance learning.

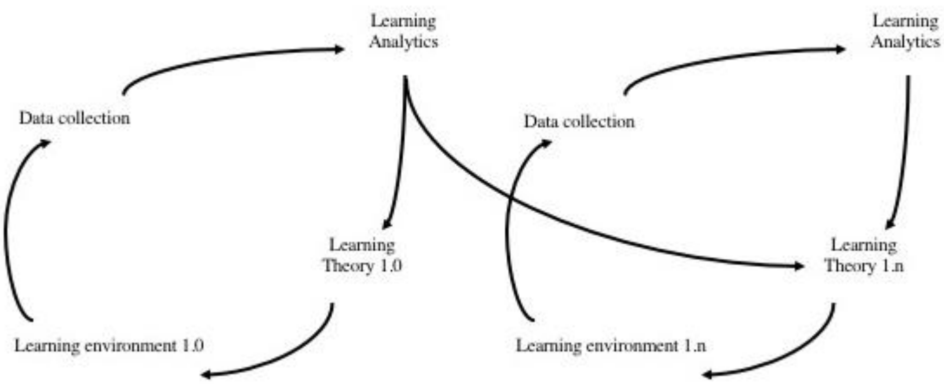


*Figure 1.* Propose Iterative Loop in which learning theory is integral to study employing learning analytics

Beside the iterative loop, we also suggest three ways in which learning theories can and should be used. First, learning theories can guide decisions on which research questions to investigate or not to investigate. By keeping abreast of the development of learning theories, future studies employing learning analytics can focus on research questions that are not yet answered instead of running the risk of claiming new discoveries that are perhaps long-established findings. For example, in digital learning environments it is typically easy to collect data about students' levels of activity, which is commonly found to be a great predictor of study success (e.g., You, 2016). This mirrors the finding that in higher education, class attendance is one of the strongest predictors of study success (Credé, Roch, & Kieszczynka, 2010). Second, learning theories can guide the operationalization of research questions into testable hypotheses, which is a critical step in designing an empirical test. Knowledge from educational research helps to sidestep collection of problematic or inappropriate variables. For example, researchers might be tempted to rely on students' evaluations of online courses and educational technologies to infer about better or more effective approaches. However, student evaluations of courses and/or teachers are only minimally related to learning outcomes, and should not be used as a proxy of learning (Clayson, 2009).

Finally, learning theories can guide the design and evaluation of tools and interventions. In the learning analytics literature, dashboards and other educational technologies are a popular subject of research. Learning theories provide highly relevant frameworks to guide the process of creating as well as evaluating dashboards and other educational technologies. For example, the added, or possibly detrimental, value of visualizations in dashboards can and should be empirically assessed, for example by using Cognitive Load Theory (Sweller, 2011) and the Cognitive Affective Theory of Multimedia Learning (Mayer, 2011). Similarly, these large fields of research are invaluable to design and create dashboards and other tools based on decades of relevant empirical research.

In conclusion, the current study shows that learning theories are often mentioned without much depth in the studies employing learning analytics. While learning analyst may be proficient with analytical approaches, they may be less familiar with the nuances of learning. Similarly, learning scientist may be apt at recognising the nuances of learning but not equipped with skills to perform the analytics using trace data. Therefore, the study of learning can benefit from the joint effort of learning scientists and learning analysts in conducting research that integrate learning theories and learning analytics. This will help to achieve an understanding of learning of which the whole is greater than the sum of its parts.

## ACKNOWLEDGEMENTS

This research is supported by Leiden, Delft, Erasmus Center for Education and Learning.

## REFERENCES

Alexander, P. A. (2006). Evolution of a learning theory: A case study. *Educational Psychologist*, *41*(4), 257-264.

Alexander, P. A., Schallert, D. L., & Reynolds, R. E. (2009). What is learning anyway? A topographical perspective considered. *Educational Psychologist*, *44*(3), 176-192.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267-270.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of college student personnel*, *25*(4), 297-308.

*Atif, A., Bilgin, A., & Richards, D. (2015). Student preferences and attitudes to the use of early alerts. Paper presented at the Twenty-first Americas Conference on Information Systems.

Ausubel, D. P. (1969). A cognitive theory of school learning. *Psychology in the Schools*, *6*(4), 331-335.

*Barba, P. D., Kennedy, G. E., & Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning*, *32*(3), 218-231.

*Bos, N., & Brand-Gruwel, S. (2016). Student differences in regulation strategies and their use of learning resources: implications for educational design. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 344-353.

*Brooks, C., Erickson, G., Greer, J., & Gutwin, C. (2014). Modelling and quantifying the behaviours of students in lecture capture environments. *Computers & Education*, *75*, 282-292.

*Carter, A. S., Hundhausen, C. D., & Adesope, O. (2015). The normalized programming state model: Predicting student performance in computing courses based on programming behavior. In *Proceedings of the eleventh annual International Conference on International Computing Education Research*, 141-150.

*Carter, A. S., & Hundhausen, C. D. (2016) With a Little Help From My Friends: An Empirical Study of the Interplay of Students' Social Activities, Programming Activities, and Course Success. In *Proceedings of the 2016 ACM Conference on International Computing Education Research,* 201-209.

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, *31*(1), 16-30.

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research*, *80*(2), 272-295.

Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current state and future trends: A citation network analysis of the learning analytics field. In *Proceedings of the 4th International Conference on Learning Analytics and Knowledge,* 231-240.

Dunlosky, J., & Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: Tips for classroom use and for guiding students' learning. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 72.

Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance improvement quarterly*, *6*(4), 50-72.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5-6), 304-317.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64-71.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, *15*(3), 42-57.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, *57*(4), 2333-2351.

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2-3), 107-124.

Harasim, L. (2012). Learning Theory and Online Technologies (First ed.). New York: Taylor and Francis Group.

Heward, W. L. (2005). Reasons applied behavior analysis is good for education and why those reasons have been insufficient. Focus on Behavior Analysis in Education: Achievements, Challenges, and Opportunities, 316-348.

Ifenthaler, D. (2017). Learning analytics design. In L. Lin & M. Spector (Eds.), *The sciences of learning and instructional design: Constructive articulation between communities* (pp. 202–211). New York, NY: Routledge.

*Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. *Computers & Education*, *87*, 204-217.

*Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*, 74-85.

*Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data. *The Internet and Higher Education*, *27*, 54-63.

*Kim, K., & Moon, N. (2017). A model for collecting and analyzing action data in a learning process based on activity theory. *Soft Computing*, 1-11.

*Kim, D., Park, Y., Yoon, M., & Jo, I. H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *The Internet and Higher Education*, *30*, 30-43.

*Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, *104*, 18-33.

Knight, S. & Buckingham Shum, S. (2017). Theory and learning analytics. In C., Lang, G., Siemens, A. F., Wise, & D., Gašević (Eds.), *The handbook of learning analytics (*pp. 17–22). Society for Learning Analytics Research (SoLAR): Alberta, Canada.

*Liu, S., & d'Aquin, M. (2017). Unsupervised learning for understanding student achievement in a distance learning setting. In *Global Engineering Education Conference (EDUCON), 2017 IEEE*, 1373-1377.

*Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*, *47*, 90-97.

*Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, *103*, 1-15.

Murphy, P. K., & Knight, S. L. (2016). Exploring a Century of Advancements in the Science of Learning. *Review of Research in Education, 40*(1), 402-456.

Na, K. S., & Tasir, Z. (2017). A Systematic Review of Learning Analytics Intervention Contributing to Student Success in Online Learning. In *International Conference on Learning and Teaching in Computing and Engineering (LaTICE) 2017,* 62-68.

Ormrod, J. E. (1999). *Human learning* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Phillips, D. C. (2014). Research in the hard sciences, and in very hard "softer" domains. *Educational Researcher*, *43*(1), 9-11.

Schunk, D. H. (2012). *Learning theories: An educational perspective* (6th. ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

SCImago. (2007). SJR — SCImago Journal & Country Rank. Retrieved July 21, 2015, from http://www.scimagojr.com

*Sedrakyan, G., Snoeck, M., & De Weerdt, J. (2014). Process mining analysis of conceptual modeling behavior of novices–empirical study using JMermaid modeling and experimental logging environment. *Computers in Human Behavior*, *41*, 486-503.

*Siadaty, M., Gašević, D., & Hatala, M. (2016). Associations between technological scaffolding and micro-level processes of self-regulated learning: A workplace study. *Computers in Human Behavior*, *55*, 1007-1019.

Siemens, G. & Long, P. (2011). Penetrating the fog: analytics in learning and education. *Educause Review*, *48*, 5, 31–40.

Skinner, B. F. (1977). Why I am not a cognitive psychologist. *Behaviorism*, *5*(2), 1-10.

Smith, V. C., Lange, A., & Huston, D. R. (2012). Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks*, *16*(3), 51-61.

Sweller, J. (2011). Cognitive load theory. In Psychology of learning and motivation (Vol. 55, pp. 37-76). Academic Press.

*Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education*, *89*, 53-74.

*Tan, J. P. L., Yang, S., Koh, E., & Jonathan, C. (2016). Fostering 21st century literacies through a collaborative critical reading and learning analytics environment: user-perceived benefits and problematics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 430-434.

*Tlili, A., Essalmi, F., Jemni, M., & Chen, N. S. (2016). Role of personality in computer based learning. *Computers in Human Behavior*, *64*, 805-813.

Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2013). Understanding feedback: A learning theory perspective. *Educational Research Review*, *9*, 1-15.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

*Romero-Zaldivar, V. A., Pardo, A., Burgos, D., & Kloos, C. D. (2012). Monitoring student progress using virtual appliances: A case study. *Computers & Education*, *58*(4), 1058-1067.

*Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., ... & Edwards, T. (2017). Assessing implicit science learning in digital games. *Computers in Human Behavior*, *76*, 617-630.

Webb, G. (1997). Deconstructing deep and surface: Towards a critique of phenomenography. *Higher Education*, *33*(2), 195-212.

*Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, *12*(4), 423-467.

Wise, A., Zhao, Y., & Hausknecht, S. (2014). Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics*, *1*(2), 48-71.

*Yassine, S., Kadry, S., & Sicilia, M. A. (2016). A framework for learning analytics in moodle for assessing course outcomes. In *Global Engineering Education Conference (EDUCON), 2016 IEEE*, 261-266.

*You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education*, *29*, 23-30.

*Zhao, Y., Davis, D., Chen, G., Lofi, C., Hauff, C., & Houben, G. J. (2017). Certificate Achievement Unlocked: How Does MOOC Learners' Behaviour Change?. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* , 83-88.

*Zhuhadar, L., Yang, R., & Lytras, M. D. (2013). The impact of Social Multimedia Systems on cyberlearners. *Computers in Human Behavior*, *29*(2), 378-385.

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. In J.E. Davidson and R.J. Sternberg (Eds.), The psychology of problem solving (pp. 233-262). Cambridge, UK: Cambridge University Press.

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



## AUTHOR INFORMATION

Include the following information for each contributing author:

**Name**: Jacqueline Wong

**Institutional affiliation:** Dept. of Psychology, Education and Child Studies, Erasmus University Rotterdam, The Netherlands.

**Institutional address:** Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

**Complete mailing address**: Mail: P.O. Box 1738 | 3000 DR Rotterdam | The Netherlands

**Telephone number:** +31104089588

**Email address**: wong@essb.eur.nl

**Short biographical sketch:** Jacqueline Wong is a PhD candidate at the Department of Psychology, Education and Child Studies of Erasmus University Rotterdam. Her research focuses on motivation and self-regulated learning in open online learning environments. She examines the influence of student characteristic and the effect of learning supports on student success in Massive Open Online Courses (MOOCs).

**Name**: Martine Baars

**Institutional affiliation:** Dept. of Psychology, Education and Child Studies, Erasmus University Rotterdam, The Netherlands.

**Institutional address:** Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

**Complete mailing address**: Mail: P.O. Box 1738 | 3000 DR Rotterdam | The Netherlands

**Telephone number:** +31104088669

**Email address**: baars@essb.eur.nl

**Short biographical sketch:** Martine Baars is an assistant professor of Educational Psychology at the Erasmus University Rotterdam. Her research is focused on instructional strategies to improve self-regulated learning. She investigates what cues are used for self-regulated learning.

Also, she explores the role of cognitive load, task complexity, and motivation in self-regulated learning.

**Name**: Bjorn B. De Koning

**Institutional affiliation:** Dept. of Psychology, Education and Child Studies, Erasmus University Rotterdam, The Netherlands.

**Institutional address:** Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

**Complete mailing address**: Mail: P.O. Box 1738 | 3000 DR Rotterdam | The Netherlands

**Telephone number:** +31104089620

**Email address**: b.b.dekoning@essb.eur.nl

**Short biographical sketch:** Björn B. de Koning is an assistant professor of Educational Psychology in the Department of Psychology, Education, and Child Studies at Erasmus University Rotterdam, Rotterdam, the Netherlands. De Koning's research concentrates on designing effective instructional support for learners and using process-oriented measures such as eye-tracking to uncover the learning process. His research projects focus on optimizing learning from dynamic visualizations, designing effective multimedia instruction, using cognitive strategies to foster reading comprehension, supporting mental model construction in mathematical word problem solving, and understanding the influence of collaboration on learning processes and performance in online and offline learning environments. These projects involve learners from primary education to higher education. His work is characterized by a multidisciplinary approach that aims to provide insights for improving educational practice by integrating cognitive science, developmental and educational psychology research perspectives.

**Name:** Tim van der Zee

**Institutional affiliation:** Leiden University Graduate School of Teaching (ICLON), Leiden University, The Netherlands

**Institutional address:** Rapenburg 70, 2311 EZ Leiden

**Complete mailing address:** Leiden University, PO Box 9500, 2300 RA Leiden

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



**Telephone number:** +31715275004

**Email address:** t.van.der.zee@iclon.leidenuniv.nl]

**Short biographical sketch:** Tim van der Zee has a MSc in Psychology and is currently a PhD student at the Leiden University Graduate School of Teaching (ICLON) in the Netherlands. In his research, he focuses on understanding and improving the educational quality of open online courses such as MOOCs (Massive Open Online Courses).

**Name:** Dan Davis

**Institutional affiliation:** Delft University of Technology

**Institutional address:** Van Mourik Broekmanweg 6, 2628 XE Delft, South Holland, The Netherlands

**Complete mailing address:** EEMCS, Web Information Systems, P.O. Box 5031, 2600 GA Delft

**Telephone number:** +31152787484

**Email address:** m.f.d.khalil@tudelft.nl

**Website:** http://mohdkhalil.wordpress.com

**Short biographical sketch:** Dan Davis is a PhD candidate at Delft University of Technology. He develops methods to gain a deeper understanding about how the design of online learning environments affects learner success and engagement, often by designing, developing, and evaluating instructional interventions at scale.

**Name:** Mohammad Khalil

**Institutional affiliation:** Delft University of Technology

**Institutional address:** Van Mourik Broekmanweg 6, 2628 XE Delft, South Holland, The Netherlands

**Complete mailing address:** EEMCS, Web Information Systems, P.O. Box 5031, 2600 GA Delft

**Telephone number:** +31645779321

**Email address:** m.f.d.khalil@tudelft.nl

Pre-print; to cite use: Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M., ... & Paas, F. (2019). Educational Theories and Learning Analytics: From Data to Knowledge. In Utilizing Learning Analytics to Support Study Success (pp. 3-25). Springer, Cham.



**Website:** http://mohdkhalil.wordpress.com

**Short biographical sketch:** Mohammad Khalil is a postdoctoral candidate at Delft University of Technology, funded by the Leiden-Delft-Erasmus Centre for Education and Learning (LDECEL) consortium. Mohammad has a doctoral degree in computer science from Graz University of Technology. His PhD dissertation was about learning analytics in Massive Open Online Courses (MOOCs). At the moment, his research is strongly related to MOOCs, online learning, and learning analytics. For publications as well as further research activities, visit his website: http://mohdkhalil.wordpress.com.

**Name:** Geert-Jan Houben

**Institutional affiliation:** Delft University of Technology

**Institutional address:** Van Mourik Broekmanweg 6, 2628 XE Delft, South Holland, The Netherlands

**Complete mailing address:** EEMCS, Web Information Systems, P.O. Box 5031, 2600 GA Delft

**Telephone number:** +31 15 27 87486

**Email address:** g.j.p.m.houben@tudelft.nl

**Website (optional):** http://www.wis.ewi.tudelft.nl/houben

**Short biographical sketch:** Geert-Jan Houben is full professor of Web Information Systems at the Software Technology department at Delft University of Technology (TU Delft). His main research interests are in Web Engineering, Web Science, and User Modeling, Adaptation and Personalization. He is managing editor of the Journal of Web Engineering (JWE), editorial board member for the Journal of Web Science (JWS), the International Journal of Web Science (IJWS), User Modeling and User-Adapted Interaction (UMUAI), and ACM Transactions on the Web (ACM TWEB).

In Delft, he is scientific director of Delft Data Science (DDS), TU Delft's coordinating initiative in the field of Data Science, holding the KIVI-chair Big Data Science, leading TU Delft's research program on Open & Online Education in TU Delft Extension School, and principal investigator in AMS, Amsterdam Institute for Advanced Metropolitan Solutions. He is currently serving as Director of Education at the faculty of Electrical Engineering, Mathematics and Computer Science at TU Delft.

**Name**: Fred Paas

**Institutional affiliation:** Dept. of Psychology, Education and Child Studies, Erasmus University Rotterdam, The Netherlands.

**Institutional address**: Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

**Complete mailing address:** Mail: P.O. Box 1738 | 3000 DR Rotterdam | The Netherlands

**Telephone number**: +31104082705

**Email address**: paas@essb.eur.nl

**Short biographical sketch:** Fred Paas is Professor of Educational Psychology at Erasmus University Rotterdam in the Netherlands, and Professorial Fellow at the University of Wollongong in Australia. Since 1990 he has been using the theoretical framework of cognitive load theory to investigate the instructional control of cognitive load in the training of complex cognitive tasks. He has (co)authored more than 250 SSCI listed journal articles, which have generated more than 28.000 citations. He is editor-in-chief of the journal *Educational Psychology Review*, and on the editorial board of several other renowned journals, such as the *Journal of Educational Psychology*. He is a fellow of the American Educational Research Association. In 2016, he was recognized as the world's most productive author in the five best journals in the field of Educational Psychology for the period 2009-2014