Intelligent Learning Analytics Dashboards: Automated Drill-Down Recommendations to Support Teacher Data Exploration

Hassan Khosravi[1], Shiva Shabaninejad[2], Aneesha Bakharia[3], Shazia Sadiq[4], Marta Indulska[5], Dragan Gašević[6]

Abstract

Learning analytics dashboards commonly visualize data about students with the aim of helping students and educators understand and make informed decisions about the learning process. To assist with making sense of complex and multidimensional data, many learning analytics systems and dashboards have relied strongly on AI algorithms based on predictive analytics. While predictive models have been successful in many domains, there is an increasing realization of the inadequacies of using predictive models in decision-making tasks that affect individuals without human oversight. In this paper, we employ a suite of state-of-the-art algorithms, from the online analytics processing, data mining, and process mining domains, to present an alternative human-in-the-loop AI method to enable educators to identify, explore, and use appropriate interventions for subpopulations of students with the highest deviation in performance or learning process compared to the rest of the class. We demonstrate an application of our proposed approach in an existing learning analytics dashboard (LAD) and explore the recommended drill-downs in a course with 875 students. The demonstration provides an example of the recommendations from real course data and shows how recommendations can lead the user to interesting insights. Furthermore, we demonstrate how our approach can be employed to develop intelligent LADs.

Notes for Practice

- The increase in the volume and complexity of student data makes it challenging for instructors to use conventional learning analytics dashboards (LADs) to make sense of students' learning processes and to decide on pedagogical actions to enhance learning.

- Instructors may be better supported in such decision making by using predictive learning analytics, which can identify at-risk students and automate the process of providing pedagogical interventions. However, use of predictive models comes with growing concerns about the fairness, accountability, transparency, and ethics (FATE) of such models.

- The approach presented in this paper recommends to instructors which students should be considered in more detail due to the highest deviation in performance or learning process from their classmates. The approach responds to concerns related to FATE by refraining from automatically labelling students; instead, it guides instructors to make data-informed decisions via recommendations.

- Instructors can gain interesting insights when drilling down into student data in LADs on the basis of the recommendations. However, a risk of inequity still exists. Rigorous evaluation of our proposed approach is required before it can be integrated into existing dashboards.

Corresponding author [1] *Email: h.khosravi@uq.edu.au Address: Institute for Teaching and Learning Innovation, The University of Queensland, St Lucia, Brisbane, QLD4072, Australia. ORCID: 0000-0001-8664-6117*
[2] *Address: Institute for Teaching and Learning Innovation, The University of Queensland, St Lucia, Brisbane, QLD4072, Australia. ORCID: 0000-0003-4007-6600*
[3] *Address: Institute for Teaching and Learning Innovation, The University of Queensland, St Lucia, Brisbane, QLD4072, Australia. ORCID:*

0000-0002-5802-1280
[4] *Address: School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, Brisbane, QLD4072, Australia. ORCID: 0000-0001-6739-4145*
[5] *Address: Business School, The University of Queensland, St Lucia, Brisbane, QLD4072, Australia. ORCID: 0000-0002-2156-4097*
[6] *Address: Faculty of Information Technology, Monash University, Melbourne, VIC3800, Australia ORCID: 0000-0001-9265-1908*

1. Introduction

Learning analytics aims to analyze data on students for the purpose of understanding and improving student learning, taking into consideration the diverse needs of the student population (Siemens & Gasevic, 2012). Within learning analytics, the development and evaluation of learning analytics dashboards (LADs) have received significant attention (Matcha, Ahmad Uzir, Gasevic, & Pardo, 2019; Jivet, Scheffel, Specht, & Drachsler, 2018; Schwendimann et al., 2017; Verbert, Ochoa, Croon, Dourado, & Laet, 2020). As the volume, velocity, and variety of student data increase, making sense of multidimensional data using conventional LADs and identifying actionable insights (e.g., identifying students in need of assistance) become more challenging.

Many learning analytics systems and dashboards have relied strongly on predictive analytics to automatically identify students in need of assistance (Hu, Lo, & Shih, 2014; Marbouti, Diefes-Dux, & Madhavan, 2016; Ahadi, Lister, Haapala, & Vihavainen, 2015). As an example, Course Signals (Arnold & Pistilli, 2012) used predictive models to label students' risk status using traffic light signals. However, with the wide adoption of predictive models to support automatic decision making, there are increasing concerns about using predictive models without human oversight in decision-making tasks that affect individuals (Hajian, Bonchi, & Castillo, 2016). Many factors, such as identifying important variables, dealing with poor-quality or imbalanced data, determining the appropriate algorithm and model for the problem at hand, hyperparameter tuning, and knowing when to retrain the algorithm with new data, may bias or reduce the accuracy of the result of a predictive model. As a consequence, the development of fair, accountable, and transparent AI systems that rely on instructor judgment has been recognized as an important line of research in learning analytics (Putnam & Conati, 2019; Verbert et al., 2020; Buckingham Shum & Luckin, 2019; Khosravi, Sadiq, & Gasevic, 2020). In this context, recent LADs have been moving away from using predictive analytics and toward prescriptive models that focus on pedagogical interventions (Jivet et al., 2018; Park & Jo, 2019).

Exploratory data-analytic techniques are a well-known alternative to predictive modelling techniques in the analysis of large datasets. Exploratory techniques allow users to navigate a dataset and make their own judgment in understanding and interpreting the results. In particular, using the online analytical processing (OLAP) drill-down operation enables users to meaningfully zoom in to the data at a more granular level (Gray et al., 1996). OLAPs have been successfully adopted and utilized across many domains, such as finance, telecom, health, and retail (Ain, Vaia, DeLone, & Waheed, 2019). In the context of LADs, drill-down operations would be performed by progressively adding filters to student attributes (e.g., "program = Computer Science" and "videos viewed < 5") to explore the performance or learning process of subpopulations of students that satisfy the drill-down criteria. The filtering feature can be beneficial in acquiring knowledge about the performance and learning process of subpopulations as well as seeking possible correlations between students' features and their learning progress. Manual drill-down operations work particularly well for curiosity-driven explorations. Wise and Jung (2019, p. 53) found that "instructors did not always come to analytics use with specific questions, but rather with general areas of curiosity." In our study, as later discussed in Section 4.2, examples of curiosity-driven explorations included queries such as "how did international students or female students perform compared to other students." Despite their strengths and benefits in data exploration, the use of manual drill-downs is associated with some shortcomings, including not knowing what attributes to use, failing to find insightful results, and encountering a drill-down fallacy "where incomplete insights result from potentially confounding factors not explored along a drill-down" (Lee, Dev, Hu, Elmeleegy, & Parameswaran, 2019, p. 187). These challenges hamper the most effective use of data, especially by users without a formal background in data analysis.

In light of these challenges, the overarching aim of this paper is to contribute to the understanding of how human-in-the-loop AI methods in the form of guided exploratory approaches can be employed to develop successful LADs. In particular, it aims to address the challenges of manual drill-downs in LADs to provide a viable alternative to predictive modelling approaches. To this end, the paper presents a new vision for intelligent LADs that uses human-in-the-loop AI to support data-driven exploratory analysis. We employ a suite of state-of-the-art algorithms from different domains, such as online analytics processing, data mining, and process mining, to present an approach called automated insightful drill-down (AID). AID consists of two algorithms that use drill-down trees and a scoring function to search the space across all possible drill-downs to recommend insightful drill-downs. We define the notion of an insightful drill-down as a set of filtering rules that identify a subpopulation of

students that deviate from the rest of the class based on performance- and learning process–based metrics, as introduced in Shabaninejad, Khosravi, Indulska, Bakharia, and Isaias (2020) and Shabaninejad, Khosravi, Leemans, Sadiq, and Indulska (2020). For performance-based divergence, the Kullback–Leibler (KL) divergence function (Kullback & Leibler, 1951) is used. For process-based divergence, a process mining technique called Earth Movers' Stochastic Conformance Checking (Leemans, Syring, & van der Aalst, 2019) is used to examine students' learning process considering three aspects of these learning processes: performed learning activities, activity frequency, and activity sequence. Utilizing the learning process, rather than focusing on aggregated engagement metrics, which is the common approach in LADs (Schwendimann et al., 2017), is increasingly recognized as essential to understanding and optimizing learning (Trcka & Pechenizkiy, 2009; Matcha, Gasevic, Uzir, Jovanovic, & Pardo, 2019).

We present a practical application of our approach in an existing LAD called Course Insights, which provides users with manual drill-down functionality. First, we explore how this drill-down functionality was used in a manual unaided manner by 71 teaching staff in their course at an Australian university. We then take the case of one such course, with 875 students in higher education with high demographic and educational diversity, and explore the possible insights that can be derived from the performance-based and learning process–based drill-downs recommended by AID. Our findings suggest that manual drill-downs without guidance can be overwhelming and tend to result in simple drill-downs to answer curiosity-driven questions, while AID provides the opportunity for instructors to utilize data-driven drill-downs alongside curiosity-driven drill-downs.

2. Background

2.1 LADs

Schwendimann et al. (2017) define a LAD as a "display that aggregates different indicators about learner(s), learning process(es) and/or learning context(s) into one or multiple visualisations." Learning analytics dashboards have provided a means to assemble and display visualizations back to stakeholders after data has been pre-processed, thereby facilitating access for a broader range of stakeholders. Early LADs, such as LeMO (Fortenbacher et al., 2013) and GLASS (Leony, Pardo, de la Fuente Valentín, de Castro, & Kloos, 2012), focused on visualizing clickstream data from popular learning management systems such as Blackboard and Moodle. LOCO-Analyst is a stand-out example of an innovative approach to LADs, released in 2006, five years before the first LAK conference (Jovanovic, Gasevic, Brooks, Devedzic, & Hatala, 2007). LOCO-Analyst provided course content access summaries at a granular level, including visualizations for assessment and forum participation as well as information on individual students (Ali, Hatala, Gasevic, & Jovanovic, 2012). Other LAD examples are the Student Activity Meter (SAM) visualization tool (Govaerts, Verbert, Duval, & Pardo, 2012), developed to increase the awareness of time spent online using learning resources; Course Signals (Arnold & Pistilli, 2012), which predicts and visualizes student success using course grades, time on task, and past performance; and the Loop tool (Bakharia et al., 2016), which seeks to link learning analytics with learning design. All early attempts at LADs suffered from the "clicks-to-construct" issue (Knight & Buckingham Shum, 2017), often failing to deliver high-level insights or an understanding of the learning process occurring from discrete trace data. More recently, there has been a move away from only providing LADs to instructors and including visualizations to support study advisors in tools such as LISSA (Charleer, Moere, Klerkx, Verbert, & Laet, 2018) and the Student Explorer (Lonn, Aguilar, & Teasley, 2015). An increasing emphasis has also been placed on involving stakeholders (i.e., instructors, tutors, learning designers, and management decision makers) as partners in a co-design process (Wise & Jung, 2019; Echeverría et al., 2018; Ahn, Campos, Hays, & Digiacomo, 2019). Data sources for LADs have moved beyond analyzing online trace data captured by a learning management system to include richer audiovisual multimodal data (Ez-zaouia & Lavoué, 2017).

In the last five years, LADs have rapidly evolved to support interventions and the provision of feedback at scale. Tools such as OnTask (Pardo et al., 2018) and Student Relationship Engagement System (SRES) (Liu, Bartimote-Aufflick, Pardo, & Bridgeman, 2017) provide visualizations and also include the ability to filter student subpopulations (using engagement and assessment attributes) and communicate targeted feedback to students (via either email or SMS). Both OnTask and SRES are examples of LADs that have been developed using co-design techniques (Dollinger, Liu, Arthars, & Lodge, 2019). Jovanovic, Dawson, Joksimovic, and Siemens (2020), however, argue that the student attributes made available to instructors have a low level of abstraction (e.g., activity access counts, number of forum posts, overall assessment grades, and even responses to individual questions in a quiz) and only allow instructors to provide simplistic process-oriented feedback. Jovanovic and colleagues (2020) go on to articulate the need to support actionable intelligence within LADs and suggest that "feedback provision systems can benefit from information-rich, interpretable insights into learning behaviour patterns." Verbert and colleagues (2020) outline a research agenda for LADs proposing reusable design patterns linking visualizations and outcomes, the use of rigorous evaluation methodologies, development of visual analytic tools, and the need for responsible learning analytics. Of particular relevance to this paper is the call for a new LAD development paradigm using visual analytics tools that

are "explainable, configurable, interactive, integrated and personalised." While the importance of human-in-the-loop presence is stressed to prevent algorithmic bias and ensure fairness within predictive algorithms, the bias that human instructors and stakeholders bring to analysis is under-researched within learning analytics.

The AID approach proposed in this paper seeks to make a significant contribution to flipping from the predictive analytic paradigm to one of recommendation, allowing stakeholders to analyze and make decisions, and also to removing the bias involved in manual subpopulation selection.

2.2 OLAP Drill-Downs

OLAP is a multidimensional data analysis approach that supports the decision-making process by enabling rapid analysis of large collections of historical data (Chaudhuri & Dayal, 1997). It assists analysts, managers, and executives in decision making by providing insight into a system's performance through interactive access to various aggregated data views (Queiroz-Sousa & Salgado, 2020). To analyze a dataset, OLAP includes roll-up (increasing the level of aggregation) and drill-down (decreasing the level of aggregation) along one or more dimension hierarchies, as well as slice-and-dice (selection and projection) and pivot (re-orienting the multidimensional view of data) operations (Chaudhuri & Dayal, 1997).

Operations such as drill-down and roll-up can be used by an analyst for more granular data interpretation and exploration in a system (Caron & Daniels, 2004); however, navigating a highly multidimensional dataset is time-consuming, and making sense of it is error prone. We use a student-based example to illustrate these challenges in the context of LADs. Let us assume that three attributes of students, Residential Status, Program, and Assessment Score, are captured. Considering Assessment Score as the target feature and that Residential Status and Program each have two possible values in our dataset, eight drill-down actions can be performed. Figure 1, which is inspired by the example used by Lee and colleagues (2019), shows all possible drill-downs (labelled B to I).

Each chart represents the distributions of assessment scores of the subpopulation versus the rest of the class population. Each bar in the charts represents the percentage of students with a score value of high, mid, or low. Two drill-down criteria (marked by blue arrows) identify two subpopulations (highlighted with blue squares)—D (students with Program = "Tourism") and I (students with Program = "Engineering" and Residential Status = "Domestic")—with a significant difference in their assessment results compared with the rest of the class population. Considering the presented example, three main challenges with a manual drill-down analysis may be encountered, as discussed below.

1. *Too many drill-down choices.* To find interesting insights into subpopulations of students, the user has to determine what features to use for filtering. As the number of attributes and attribute values increases, the number of possible drill-down actions increases exponentially.

2. *Lack of insightful results.* A large number of possible drill-down criteria can lead to results that may not be insightful. As a consequence, insightful drill-downs may be missed via a manual drill-down process. Referring to the example in Figure 1, many of the drill-downs, such as those presented by subpopulations B, C, and H, may not seem insightful because the filtered population has similar assessment results to the entire population. In contrast, the two blue-highlighted subpopulations D and I have significantly different assessment results than the rest of the class population.

3. *Drill-down fallacies.* An incomplete drill-down analysis may produce a class of errors known as *drill-down fallacy* (Lee et al., 2019). Such a fallacy occurs when incorrect reasoning for a deviation found in the dataset is attributed to a smaller subpopulation while "in fact it is a more general phenomenon" (Lee et al., 2019, p. 186). For instance, as shown in Figure 1, if the user takes the drill-down path marked by the red arrows, they could conclude that the high percentage of Low scores and the absence of High scores (shown in chart G) are associated with international students who study tourism. However, if students are filtered first to tourism (marked with a green checkmark), the user could infer that this distribution is associated with the tourism subpopulation in general, instead of international students who study tourism specifically.

Several attempts to address the above challenges have been made. Sarawagi (2001) focuses on finding what is informative in visited parts of the data to model the users' expected values in the unvisited parts of the data. Lee and colleagues (2019) focus on accelerating drill-down via visualization and preventing users from drill-down fallacies. Joglekar, Garcia-Molina, and Parameswaran (2019) introduce a smart drill-down operator that finds drill-down rules that result in frequent patterns. Many of the proposed methods for discovering insightful drill-downs focus on detecting anomalies in small data portions (e.g., Sarawagi (2000, 2001); Lee et al. (2019)), while some focus on identifying interesting differences in larger data subsets (e.g., Joglekar et al. (2019)). In the AID approach, we let LAD users request drill-down recommendations at the level of granularity they are interested in and, thus, reduce drill-down choices without affecting user autonomy.
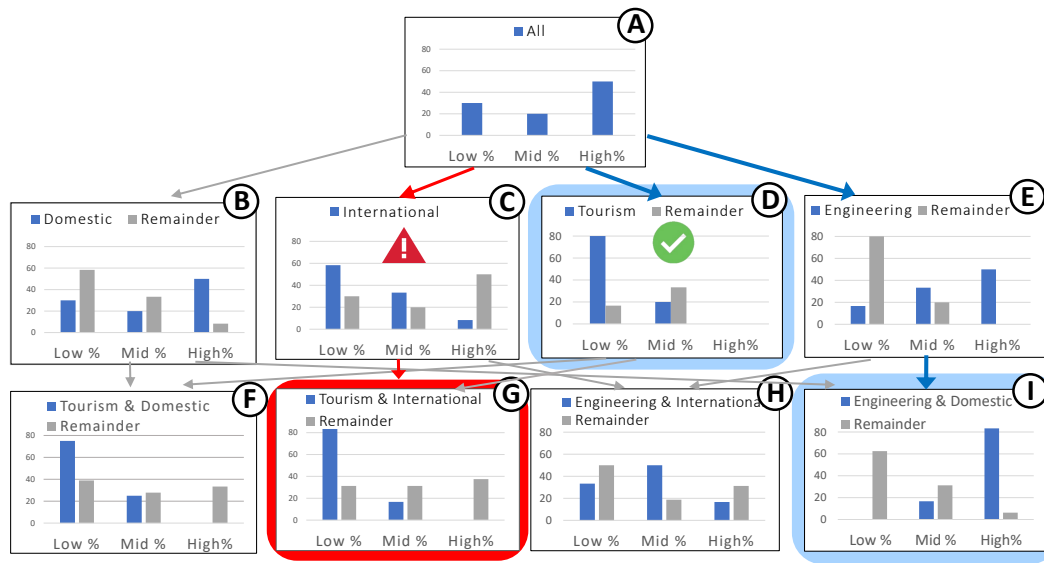
Figure 1. Drill-down and fallacy illustration.

2.3 Educational Process Mining

Process mining aims to extract process-related knowledge from event logs to discover, monitor, and improve processes (van der Aalst, 2016). Process mining techniques have been used in the educational context in recent years mainly to visualize, evaluate, and improve educational and learning processes (Bogarín, Cerezo, & Romero, 2018), for instance, to visualize student learning behaviour (Cairns, Gueni, Assu, Joubert, & Khelifa, 2015); to investigate a correlation between student learning behaviour and assessment outcomes (Mukala, Buijs, Leemans, & van der Aalst, 2015); to study whether student behaviour corresponds to a learning model (van der Aalst, Guo, & Gorissen, 2013); to detect bottlenecks in educational processes (Pechenizkiy, Trcka, Vasilyeva, van der Aalst, & Bra, 2009); and to identify patterns in educational processes (Dekker, Pechenizkiy, & Vleeshouwers, 2009), learning strategies (Matcha, Gasevic, et al., 2019), and time management (Uzir, Gasevic, Matcha, Jovanovic, & Pardo, 2020).

In the educational process context, several attempts have been made to compare student subpopulations. For instance, van der Aalst and colleagues (2013) explore a comparative process analysis within a process cube by employing a process alignment method to compare the process of students that passed a course with that of the students who failed it. Saint, Gasevic, Matcha, Uzir, and Pardo (2020) use an approach to process mining based on first-order Markov models in combination with expectation-maximization clustering to identify learning tactics followed by students in online learning activities. Visual properties of some of the process mining tools (e.g., `pMineR` or `bupaR` libraries in R) have been used to compare process models of different student subpopulations. Prior works (Bogarín et al., 2018) indicate that process mining techniques in education have been utilized through general-purpose process mining tools (e.g., PRoM, Disco) and are yet to be integrated into LADs that are available to frontline users in education (Saint, Gasevic, & Pardo, 2018).

The work presented in this paper extends our previous works, which have focused on employing AI and process mining techniques to develop intelligent LADs to support instructors with data-driven exploratory analysis of students' learning (Shabaninejad, Khosravi, Indulska, et al., 2020; Shabaninejad, Khosravi, Leemans, et al., 2020; Leemans et al., 2020). Shabaninejad, Khosravi, Indulska, and colleagues (2020) present an approach that can provide instructors with meaningful and efficient ways to gain insight into subsets of students with the highest deviation in an attribute (e.g., performance) compared to the overall class and to recommend subsets of students. Shabaninejad, Khosravi, Leemans, and colleagues (2020) present an approach that uses a process mining lens to examine learning process differences of students and to identify and recommend subsets of students with the highest difference in learning processes compared to the rest of the class. Leemans and colleagues (2020) introduce the notion of *cohort identification* as a method of eliciting features from trace attributes, measuring the stochastic distance between cohorts defined by sets of these features, and presenting to users this landscape of sets of features and their influence on process behaviour. In this paper, we present how cohort identification can be incorporated into LADs and demonstrate different use cases of how it can provide easy-to-interpret actionable insights into students' learning processes.

3. AID

In this section, we introduce our proposed approach, which consists of two underlying algorithms for recommending insightful performance-based and process-based drill-down criteria in LADs. In Subsection 3.1, we define our problem statement and introduce formal notation. We present an overview of our approach in Subsection 3.2. Furthermore, we present our proposed algorithms in Subsection 3.3 and Subsection 3.4. Lastly, in Subsubsection 3.3.1 we present an illustrative example for each of the two algorithms based on the datasets introduced in Subsection 2.2.

3.1 Notation and Problem Statement

Assume that a LAD has access to an event log L that captures a collection of traces $T = \{t_1, \ldots, t_N\}$, each representing a student. A trace t_i has a unique identifier (e.g., a student ID); a set of features $F = f_1, \ldots, f_M$, where $f_{im} = v$ presents v being assigned to feature f_i for user s_i; and a sequence of events $E_i = \langle e_{i1}, \ldots, e_{iL_i} \rangle$ representing the learning path taken by student s_i, where the trace length L_i can vary for each student. Each event e_{iL_i} has a timestamp and a label representing the learning activity. Let a rule r express a condition on a feature (e.g., "Program" = "Computer Science"). For a feature with numerical values, the corresponding rule value can be a range instead of a single value (e.g., "Age" > 25). The drill-down criterion σ is defined as the conjunction of a set of rules (e.g., "Program" = "Computer Science" $\wedge$ "Age" > 25). A drill-down criterion σ is said to cover a student s_n if all of the rules in σ match the corresponding values for s_n. Consequently, applying σ to L leads to the selection of a set of students $S' \in S$ such that σ covers each $s_n \in S'$. We define the *coverage* of a drill-down criterion C_σ as $\frac{|S'|}{|S|}$, which is the fraction of students S covered in the resulting subpopulation S'.

Problem statement   Using this notation, our problem of finding performance-based and process-based insights can be formalized as follows:

- *Performance-based insights.* Given an event log L, a set of features $F' \subseteq F$, a performance-based feature $\hat{f} \in F$, a constant $0 \leq \alpha \leq 1$, and a constant k, find a set of drill-down criteria $\Sigma = \{\sigma_1, \ldots, \sigma_K\}$ that uses a set of features in F' such that each criterion σ_k (1) has a larger coverage than α (i.e., $C_{\sigma_k} > \alpha$) and (2) maximizes between-subpopulation performance variability between S' and the remaining students $S \setminus S'$ based on $\hat{f}$.

- *Process-based insights.* Given an event log L, a set of features $F' \subseteq F$, a constant $0 \leq \alpha \leq 1$, and a constant k, find a set of drill-down criteria $\Sigma = \{\sigma_1, \ldots, \sigma_K\}$ that uses a set of features in F' such that each criterion σ_k (1) has a larger coverage than α (i.e., $C_{\sigma_k} > \alpha$) and (2) maximizes between-subpopulation learning process variability between S' and the remaining students $S \setminus S'$ in terms of events, relative frequency of each different learning path, and the order using the sequence of events captured in events E.

The rest of this section presents our approach for finding drill-down actions that result in performance-based and process-based learning insights.

3.2 Overview

Figure 2 provides an overview of the interaction between instructors and our proposed approach within LADs. The interaction is an iterative process involving the generation of recommendations by the algorithm and the exploration of the recommended drill-downs by the instructor. In each iteration, instructors may target different subsets of data and repeat the process.

Figure 3 shows a high level of how recommendations are produced. Using the event log as the input, the steps are as follows:

1. Elicit all the features from the event log.

2. Build a cohort drill-down tree from all possible subsets of the features.

3. For each node in the tree, compute the distance value by measuring the difference between the learning processes of the cohort and those of the rest of the students.

4. Convert the cohorts (i.e., tree nodes) with the largest distance value into recommendations.
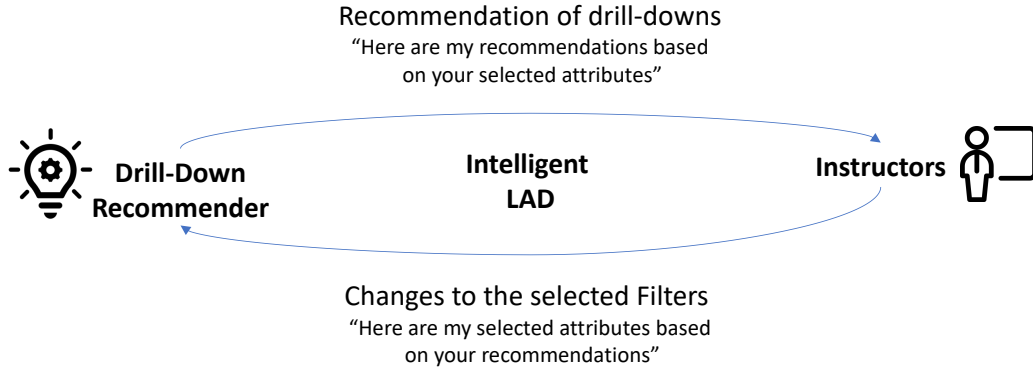
Recommendation of drill-downs
"Here are my recommendations based
on your selected attributes"

Drill-Down
Recommender

Intelligent
LAD

Instructors

Changes to the selected Filters
"Here are my selected attributes based
on your recommendations"

Figure 2. An overview of the role of the drill-down recommender in the process of instructors' data exploration in the context of intelligent LADs.

Learning system's
event log

Elicit features

Build
drill-down tree

Compute
subpopulation's
distance
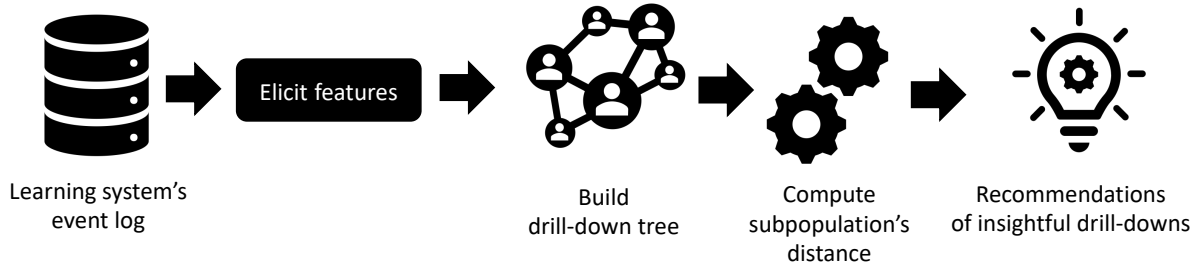
Recommendations
of insightful drill-downs

Figure 3. Overview of the approach: students' features are elicited and a drill-down tree of all cohorts produced, where each cohort shares a set of features (e.g., residential status: international), the distance between each cohort and the rest of the students is computed, and drill-downs with the highest distance values are chosen to produce recommendations.

3.3 Performance-Based AID

We present our approach for recommending performance-based drill-downs by first providing a high-level overview of the underlying algorithm and then describing the automatic drill-down process using an example.

Algorithm 1 provides the high-level pseudocode of our proposed approach for recommendation of performance-based drill-downs. It takes five parameters as input: the event log L; the features F'; the target performance-based features $\hat{f}$; the minimum coverage α; and the number of drill-down criteria to be recommended, k. The output of the algorithm is a set of top k scored drill-down criteria. The algorithm consists of three main components, as described below.

Create drill-down tree To provide educators with performance-based drilling-down recommendations, the algorithm examines all of the cohorts by first extracting all possible drill-down actions from the selected students' features. The *TransformLogToFeature* function transforms the log into a tabular dataset D that captures data on selected feature F', target feature $\hat{f}$, and aggregated level features from events on each student. The *BuildTree* function takes the dataset D as input and returns a drill-down tree. The function obtains all the values of each feature in D and generates a tree-like collection of nodes T, where each node represents a splitting rule r for one feature. Each path in the tree consists of a set of feature-value pairs.

Score nodes and prune the tree The algorithm evaluates the insighfulness of each drill-down action by measuring the difference between each cohort's targeted performance feature and that of the rest of the students. The tree embodies all possible drill-down paths, of which not all will necessarily result in a subpopulation with the required minimum size (i.e., α). *PerformancePruneAndScore* traverses the tree recursively to examine all of the possible drill-down actions. *ObtainSubData* takes each node, which is a feature-value pair, and dataset D as input and filters D to obtain a sub-dataset *subCohortData* containing only the data of the subpopulation S. The subpopulation's size is checked for the covered fraction of the student population; if the fraction is smaller than α, then *childNode* and its underlying sub-tree are pruned. Otherwise dataset D is filtered to obtain the data on S', which is captured in *remainderCohortData*. The function *ComputeDeviance* takes data on the two subpopulations *subCohortData* and *remainderCohortData* and the target feature $\hat{f}$ as input and scores the

---

Algorithm 1 Finding a set of k performance-based drill-down criteria

---

function MAIN(Log L, Features F', Target Feature $\hat{f}$, Minimal Coverage α, k)
    $D \leftarrow TransformLogToFeature(L)$ ▷ Transform log to a dataset
    $T \leftarrow BuildTree(D)$ ▷ Create drill-down tree
    PERFORMANCEPRUNEANDSCORE($D, T, \hat{f}, \alpha$)
    $topK \leftarrow TopDeviation(T, k)$ ▷ Sort and return the top K drill-down criteria
    return $NodeToDrillDown(topK)$
end function
function PERFORMANCEPRUNEANDSCORE(dataset D, Node $parentNode$, Feature $\hat{f}$, Minimal Coverage α) ▷ Score nodes based on $\hat{f}$ and prune the tree
    for $childNode \in parentNode.children$ **do**
        $subCohortData \leftarrow ObtainSubData(childNode, D)$
        if $(CohortSize(subCohortData)|/CohortSize(D) \leq \alpha$ **then**   ▷ Pruned because coverage $\leq \alpha$
            remove $childNode$
        else   ▷ Compute score for current child and recurse on child
            $remainderCohortData \leftarrow ObtainRemainderData(D, subCohortData)$
            $childNode.deviance \leftarrow ComputeDeviance(subCohortData, remainderCohortData, \hat{f})$
            PERFORMANCEPRUNEANDSCORE($D, childNode, \hat{f}, \alpha$)
        end if
    end for
end function

---

path by measuring the distance between S and S' using KL divergence, defined as $d = \sum_{x \in X} S'(x) \log(\frac{S'(x)}{S(x)})$, where d is the distance between the distribution of the values in S and S' and X is a set that consists of the values in the domain of $\hat{f}$. *PerformancePruneAndScore* will then be called to explore the sub-tree with root *childNode* for other potential candidates.

Sort and return the top *K* drill-down criteria   The algorithm returns the cohorts with the most different targeted performance features. *TopDeviation* takes the scored drill-down tree T and k as input and returns the k highest scored paths. Our algorithm then converts the chosen nodes to a set of drill-down criteria Σ, each annotated with distance score, and returns them as a recommendation to users.

3.3.1 Illustrative Example

We illustrate our approach using an event log with a small set of six students, $k = 1$, and $\alpha = 0.2$. We explain how the algorithm is used to find the insightful drill-down criteria (namely the criteria that identify a subpopulation with the highest deviation in terms of performance or behaviour) for the event log given in Figure 4 with students $\{S1 \cdot S6\}$ and the feature set $\{$Residential Status, Assessment$\}$ as F'. Our example course has learning activities of $\{$Lecture 1, Lecture 2, Quiz A, Lecture 3, Lecture 4, Quiz B, and Lecture final$\}$, which were made available to students weekly in the mentioned order. The trace of triggered learning events by each student is shown in Figure 4(a). Each event is represented by an activity label and the timestamp. The feature set and associated values for each student are given in Figure 4(b). The Activity Engagement attribute is computed by first taking the sum of activities for each user and then using a discretization algorithm to transform the integer output into a discrete counterpart to be used by the drill-down tree.

    The algorithm initially generates the drill-down tree T. Next, the tree is traversed depth first; based on each node's filtering criteria, the dataset is divided into the subpopulations. The nodes covering less than $\alpha = 0.4$ of the student population are pruned. For instance, the node [Residential Status = "Domestic" AND Activity Engagement = "High"] is pruned because only one student (i.e., 0.16 coverage) adheres to this criterion. After pruning, the resulting decision tree includes four candidates, shown in blue and labelled as P1 to P4, that can be mapped to drill-down paths. AID uses a scoring function based on KL divergence to evaluate the "insightfulness" of each of these paths. For instance, path $P1$: [Residential Status = "Domestic"] with the assessment distribution Low: 0.33, Mid: 0.33, High: 0.34, and the distributions in the assessment of the remainder population Low: 0.67, Mid: 0, High: 0.33%, will be scored as $0.33 \log(\frac{0.33}{0.67}) + 0.34 \log(\frac{0.34}{0.001}) + 0.34 \log(\frac{0.34}{0.33}) = 0.73$. As a side-note, 0 is replaced with a small ε (0.001 in our example) in the computation of the score to address dealing with undefined log values. The score of the other drill-down candidates is computed in a similar way. These drill-down paths and their resulting coverage and significance scores are shown in Figure 5(a).

    The path with the highest scores ($P3$) is converted to a set of drill-down criteria and is recommended to the user. Figure 5(b) shows a prototype representing the input and the resulting recommendation, including the drill-down criteria, coverage, and significance. The green upward arrow indicates that the average Assessment Score in the subpopulation is higher than the average Assessment Score in the remaining population. Note that the returned result satisfied both criteria of the problem statement.

#	Time Stamp	Activity	#	Time Stamp	Activity
S1	19-02-02T23:02:32	Lecture 1	S4	19-02-02T23:02:32	Lecture 1
S1	19-04-02T23:02:33	Lecture 2	S4	19-03-02T23:02:33	Lecture 2
S1	19-06-02T23:02:34	Lecture 4	S4	19-04-02T23:02:33	Quiz A
S1	19-08-02T23:02:36	Lecture final	S4	19-05-02T23:02:34	Lecture 3
S1	19-09-02T23:02:35	Quiz A	S4	19-06-02T23:02:34	Lecture 4
S1	19-10-02T23:02:35	Quiz B	S4	19-07-02T23:02:35	Quiz B
			S4	19-09-02T23:02:36	Lecture final
S2	19-02-01T23:02:32	Lecture 1			
S2	19-03-02T23:02:32	Lecture 2	S5	19-01-02T23:02:33	Lecture 1
S2	19-04-02T23:02:33	Lecture 3	S5	19-02-02T23:02:33	Lecture 2
S2	19-06-02T23:02:34	Lecture 4	S5	19-03-02T23:02:33	Quiz A
S2	19-8-02T23:02:36	Lecture final	S5	19-04-02T23:02:33	Lecture 3
S2	19-09-02T23:02:35	Quiz A	S5	19-05-02T23:02:34	Quiz B
S2	19-10-02T23:02:35	Quiz B	S5	19-07-02T23:02:35	Lecture final
S3	19-02-02T23:02:32	Lecture 1	S6	19-02-02T23:02:32	Lecture 1
S3	19-03-02T23:02:33	Lecture 2	S6	19-03-02T23:02:33	Lecture 2
S3	19-04-02T23:02:33	Quiz A	S6	19-04-02T23:02:33	Quiz A
S3	19-05-02T23:02:34	Lecture 3	S6	19-05-02T23:02:34	Lecture 3
S3	19-06-02T23:02:34	Lecture 4	S6	19-06-02T23:02:34	Quiz B
S3	19-07-02T23:02:35	Quiz B	S6	19-08-02T23:02:35	Lecture final
S3	19-09-02T23:02:36	Lecture final			

#	Assessment	Residential Status	Activity Engagement
S1	Low Grade	International	High
S2	Low Grade	International	High
S3	High Grade	International	High
S4	High Grade	Domestic	High
S5	Low Grade	Domestic	Mid
S6	Mid Grade	Domestic	Mid

(a) Students' learning events. **(b)** Students' features.

Figure 4. A sample dataset used for the illustrative example.
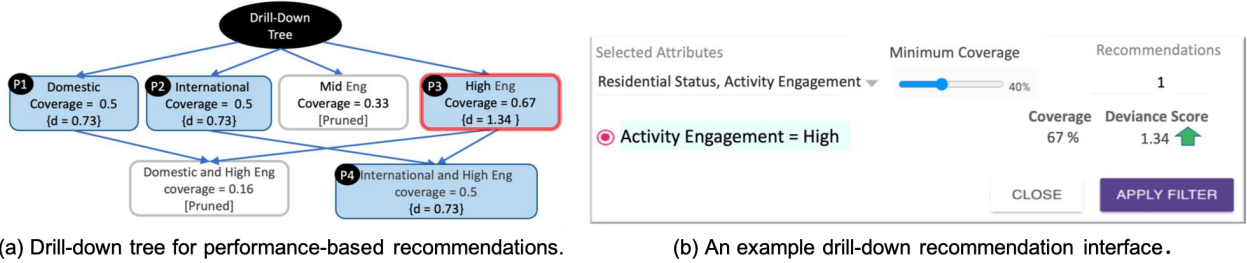
(a) Drill-down tree for performance-based recommendations. (b) An example drill-down recommendation interface.

Figure 5. The drill-down tree and the associated drill-down recommendation for the performance-based illustrative example.

3.4 Process-Based AID

We present our approach for recommending behaviour-based drill-downs by first providing a high-level overview of the underlying algorithms and then describing the automatic drill-down process using an example.

Algorithm 2 provides the high-level pseudocode of our proposed approach for recommendation of behaviour-based drill-downs. It takes four parameters as input: the event log L; the features F'; the minimum coverage α; and the number of drill-down criteria to be recommended, k. The output of the algorithm is a set of the top k scored drill-down criteria represented by Σ. The algorithm consists of three main blocks, as described in the remainder of this subsection.

Create the drill-down tree   To provide educators with process-based drilling-down recommendations, the algorithm examines all the cohorts by first extracting all possible drill-down actions from the selected students' features. The *BuildTree* function takes two parameters as input—the event log L and the list of selected features F'—and returns a drill-down tree. The function obtains all the values of each feature in F' that exist within L and generates a tree-like collection of nodes T, where each node represents a splitting rule r for one feature. Each path in the tree consists of a set of feature-value pairs.

Score nodes and prune the tree   The algorithm evaluates the insightfulness of each drill-down action by measuring the difference between each cohort's learning process and that of the rest of the students. The tree embodies all possible drill-down paths, of which not all will necessarily result in a subpopulation with the required minimum size (i.e., α). *BehaviourPruneAndScore* traverses the tree recursively to examine all the possible drill-down actions. *ObtainSubLog* takes each node, which is a feature/value pair, and the log L as input and filters L to obtain a sub-log *cohortL* containing only the data of the subpopulation. The subpopulation's size is checked for the covered fraction of the student population; if the fraction is smaller than α, then *childNode* and its underlying sub-tree are pruned. Otherwise, log L is filtered to obtain the data on S', which is captured in

remainderL. The function *ComputeDistance* uses a process mining lens to compute the learning process distance between subpopulations S and S'. It takes two subpopulation logs, *cohortL* and *remainderL*, as input and computes the distance between them using Earth Movers' Stochastic Conformance Checking (Leemans et al., 2019). *BehaviourPruneAndScore* will then be called to explore the sub-tree with root *childNode* for other potential candidates.

Sort and return the top *K* drill-down criteria The algorithm returns the cohorts with the most different learning process. *TopDistances* takes the scored drill-down tree T and k as input and returns the k highest-scored paths. The algorithm then converts the chosen nodes to a set of drill-down criteria Σ, each annotated with distance score, and returns them as a recommendation to users.

Algorithm 2 Finding a set of k process-based drill-down criteria.

---

function MAIN(Log L, Features F', Minimal Coverage α, k)
    $T \leftarrow BuildTree(L, F')$ ▷ Create drill-down tree
    BEHAVIOURPRUNEANDSCORE(L, T, α)
    $topK \leftarrow TopDistances(T, k)$ ▷ Sort and return the top K drill-down criteria
    return *NodeToDrillDown(topK)*
end function
function BEHAVIOURPRUNEANDSCORE(Log L, Node *parentNode*, Minimal Coverage α) ▷ Score nodes based on difference in learning process and prune
    for *childNode* $\in$ *parentNode.children* **do**
        cohortSublog $\leftarrow ObtainSublog(childNode, L)$
        if $(CohortSize(cohortSublog)|/CohortSize(L) \leq \alpha$ **then**   ▷ Pruned because coverage $\leq \alpha$
            remove *childNode*
        else   ▷ Compute score for current child and recurse on child
            remainderSublog $\leftarrow ObtainRemainderSublog(L, cohortSublog)$
            childNode.deviance $\leftarrow Computedistance(cohortSublog, remainderSublog)$
            BEHAVIOURPRUNEANDSCORE(L, *childNode*, α)
        end if
    end for
end function

3.4.1 Illustrative Example

In this section, we explain how the algorithm is used to find the insightful process-based drill-down criteria for the data given in Figure 4 with feature set {Residential Status, Assessment} as F', $\alpha = 0.2$, and $k = 1$. The algorithm initially extracts all values of F' that are present in the event log and generates the drill-down tree T. Next, the tree is traversed depth first; based on each node's filtering criteria, the event log is divided into the subpopulation's sub-log and the remaining students' sub-log. The nodes covering less than $\alpha = 0.2$ of the student population are pruned. For instance, the node [Assessment = "Mid Grade"] is pruned because only one student (i.e., 0.16 coverage) adheres to this criterion. As a result, five actionable drill-down paths remain (shown in Figure 6(a)); P1: [Residential Status = "Domestic"], P2: [Residential Status = "International"], P3: [Assessment = "Low Grade"], P4: [Assessment = "High Grade"], and P5: [Assessment = "Low Grade" and Residential Status = "International"]. The algorithm computes the distance between the sub-logs for each drill-down path and annotates each node by the distance d and the coverage. The drill-down path P5, which has the highest difference (57%), is the resulting recommendation. Figure 6(b) shows an example drill-down interface, representing the input and the resulting recommendation, including the drill-down criteria, coverage, and distance.

To understand the difference between the learning process of the subpopulation and that of the rest of the class, we used the well-known process mining tool Disco (Fluxicon, n.d.) to visualize the underlying learning processes of each group. Disco generates a process map in which boxes represent activities, numbers in the boxes represent the frequency of each activity, arrows represent the sequence the activities were performed in (i.e., the control flow), numbers on the arrows represent the frequency with which the two connected activities were performed, and the thickness of the arrows represents relative frequencies. For demonstration purposes, we highlighted in red the activities that were performed in a different order. To compare the two modelled learning processes, we look at the difference between the activities, their frequencies, and their sequence. For instance, Figure 6(c) shows that Lecture 3 was skipped by one of the two students in the subpopulation, while Figure 6(d) shows that the remaining students have done this activity. From a control flow perspective, Quiz A and Quiz B were performed as the last activities by the selected subpopulation, while the remaining students performed these quizzes during the semester.
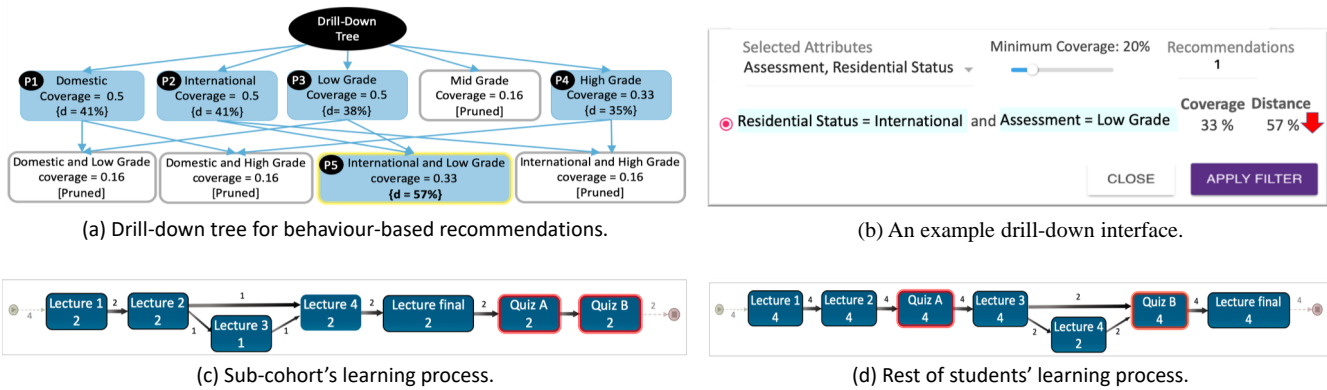
(a) Drill-down tree for behaviour-based recommendations.

(b) An example drill-down interface.

(c) Sub-cohort's learning process.

(d) Rest of students' learning process.

Figure 6. The drill-down tree and the associated drill-down recommendation and visualization of the learning processes for the behaviour-based illustrative example.

4. Practical Application

This section presents an application of the AID approach using an existing LAD called Course Insights, which is equipped with manual drill-down functionality[1]. Subsection 4.1 introduces Course Insights and its main segments. Subsection 4.2 provides insights into how manual drill-down functionality was employed by 71 Course Insights users. Finally, Subsection 4.3 presents results from applying our approach to student data from one of the courses that has previously employed Course Insights. The recommended performance-based and process-based drill-down criteria and how they are presented to users are discussed in this section. In addition, we summarize feedback from the instructor of the course.

4.1 Course Insights

Course Insights is an instructor-facing LAD that is under active development at the University of Queensland and is currently accessible to all courses with a Blackboard presence (i.e., approximately 2,000 in Semester 1, 2020). Course Insights includes student data from numerous university data sources, including the university's student information system (i.e., enrolment and demographic information), trace data from interactions with learning management systems used at the university (i.e., Blackboard and edX Edge), and assessment grades. The processing pipeline for the dashboard uses AWS Data Lake technologies (i.e., AWS Athena) and the Elasticsearch engine with the user interface developed in React and deployed as a serverless lambda function.

The vision for Course Insights has been to "provide actionable insight," which has been achieved by adhering to the following design principles:

1. *Provide analytics across the course life-cycle.* Providing analytics relevant to time periods across the course life-cycle (e.g., orientation week, the first week of the semester, post–census date, post-assessment, and end of course) was seen as fundamental to the success of the Course Insights dashboard and has been achieved by integrating multiple data sources.

2. *Provide filterable and comparative student subpopulation visualizations.* An essential element in instructor decision making is comparative analysis, whereby the instructor is able to compare different student subpopulations. Course Insights includes unique filtering functionality and automatically includes comparative (i.e., multiple series) visualizations to allow the whole student population to be compared with the filtered subpopulation (e.g., students from a specific program and/or demographics). The Course Insights filter allows instructors to define query criteria using field-level attributes from any of the integrated data sources. The query builder is included on all Course Insights dashboards.

3. *Facilitate the provision of whole-class scaffolding and personalized feedback.* Course Insights has been designed to bridge the gap between dashboards that provide aggregate cohort data and feedback tools that allow subpopulations of students to be identified for targeted feedback. The Enrolment, Engagement, and Assessment dashboard screens provide aggregate visualizations to compare the whole class with a filtered subpopulation. The aggregate visualizations enable scaffolding at the whole-class level. An additional Students dashboard is included to enable personalized feedback. The Students dashboard displays students and their attributes from all data sources in a tabular format with the ability to filter students and provide feedback via email.

---

[1]Approval from our Human Research Ethics Committee (#2019002181) of the University of Queensland was received for conducting this study.

The inclusion of a filter builder across all dashboards available in Course Insights has proved useful and allowed instructors with very specific criteria to find subpopulations of students (e.g., find students enrolled in the Bachelor of Computer Science program who are retaking the course and have not engaged with course material before an assessment task). As discussed throughout the paper, manual drill-downs are challenging to use without assistance. The combination of student attributes across a growing list of data sources provides a very large search space, making it almost impossible to apply all filter combinations and analyze the comparative visualization for large distributional differences. Instructors, as reported in Subsection 4.2, have only used very simple demographic filters, introducing bias and perhaps reasoning that a deviation in distribution is only present in a smaller subpopulation when it is in fact present in a larger population (i.e., the drill-down fallacy) (Lee et al., 2019). The inclusion of the AID recommendations within Course Insights can help direct instructors to focus their attention and minimize any bias present in the selection of student subpopulations. Implementation is underway to incorporate AID into Course Insights based on two different approaches. The first approach positions the drill-down recommendations as an enhancing feature as part of the manual filtering process. The second approach utilizes the recommendations to provide actionable periodic (e.g., weekly) insights for instructors to consider and act upon.

The filter recommendation approach   Figure 7(a) demonstrates a current prototype of our vision in terms of including a filter recommendation feature as part of a manual filtering process. This interface enables instructors to utilize manual drill-downs for curiosity-driven explorations while providing recommendations to support data-driven exploration. Figure 7(b) illustrates a potential interface for presenting drill-down recommendations to instructors, including the drill-down criteria, coverage (fraction of students covered by the drill-down), and significance (distance score of the target feature).
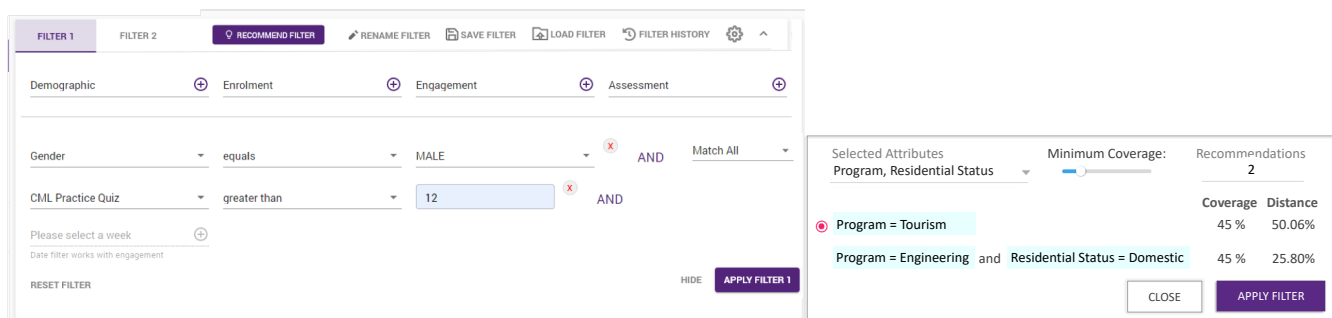
(a) Filtering interface with recommendation support.   **(b)** The drill-down criteria recommendation interface.
Figure 7. The filter recommendation approach illustration.

The periodic insights approach   Figure 8 demonstrates a current prototype of our vision in terms of providing periodic actionable insights for instructors. Each drill-down in the interface provides insight into a subpopulation with a different performance or learning process from the rest of the class. The interface provides three main calls to action: (1) *Explore Further* automatically loads the recommended drill-down in the filter interface to enable further exploration, (2) *View Students* provides access to available data about students that are covered by the drill-down, (3) *Take Action* enables instructors to act upon the provided insights by, for example, contacting the identified subpopulation via email. In the screenshot shown in Figure 8, the first recommended drill-down is an example of a performance-based insight for a subpopulation with the common features of Program = "Computer Science" and Video Engagement = "Low", acknowledging that the subpopulation's score on Quiz 1 was 15% lower than that of the rest of the class. The second recommended drill-down represents an example of a learning process–based insight, reporting on a subpopulation whose learning processes are noticeably different from those of the rest of the class and with an overall engagement level 30% lower than that of the rest of the class. These students have common features of Residential Status = "International" and Assessment = "Low".

4.2 Manual Drill-Downs in Practice

To understand the common challenges with manual drill-downs in LADs, and to study how they are being used, we have analyzed logs from Course Insights. Our investigation was guided by the following questions:

1. How complex are the manual drill-downs applied by users?

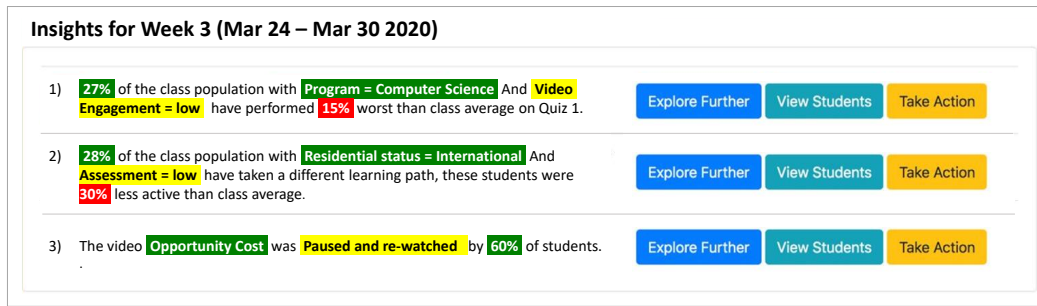2. Which manual drill-downs are commonly applied?

Figure 8. The periodic insights approach implementation.

To answer these questions, we processed 356 manual drill-down actions performed by 71 teaching staff members who were involved in teaching graduate or postgraduate courses during 10 weeks of a semester at the University of Queensland in 2019. The drill-down could be performed on student demographics (i.e., age, gender, language, brand new, residential status), enrolment (i.e, program, selected tutorial session, full-time/part-time, whether or not student is repeating the course), engagement (five video and platform relation actions), and assessment (each assessment score range) features.

Analysis of data related to the first question shows that using a combination of attributes was less common. Overall, 84% of drill-downs were performed on a single attribute and the average number of attributes for each drill-down was < 2 (Enrolment $= 1.35 \pm 0.55$, Engagement $= 1.34 \pm 0.68$, Assessment $= 1.29 \pm 0.59$, and Students $= 1.15 \pm 0.36$). Table 1 reports the 10 most common drill-down actions. These results demonstrate that most of the drill-downs consisted of a single attribute and that only one of the 10 most common drill-downs used a combination of more than two attributes. This result might be explained by the well-established phenomenon called *choice overload* from choice theory (Gourville & Soman, 2005), which suggests that increasing the number of choices given to a user may increase their effort in decision making.

Table 1. Top 10 common manual drill-downs

Single Criterion	Total	Multi Criteria	Total
Age	98	Age & Study Load	22
Program	71	Age & Gender	6
Residential Status	66	Program & Gender	6
Study Load	35	Program & Age	6
Gender	32	Program & Residential Status & Study Load	5

As is evident, teaching staff have commonly used a small subset of the features and rarely drilled down more than one level into the features. This may suggest that the drill-down functionality has been under-used and that many users may require assistance to fully benefit from drill-down operation. Furthermore, as evident by the results of Table 1 and as supported by the findings of Wise and Jung (2019), many of the drill-downs posed by instructors seem to have had a curiosity-driven nature, focusing on the age, program, residential status, study load, or gender of the students.
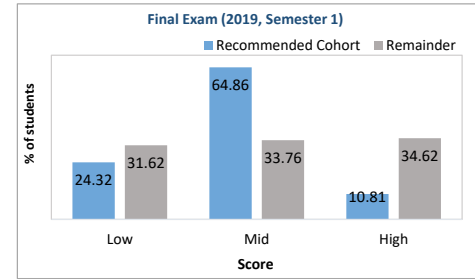
4.3 AID Recommendations in Action

We applied AID to an introductory calculus and linear algebra course offered in 2019 to 875 undergraduate students from 48 programs at the University of Queensland. Within the course, students could do three types of learning activities: (1) accessing course materials—accessing course materials by chapter; (2) participating in formative quizzes—submitting chapter-based practice quizzes (practice quizzes were formative assessments and thus optional); and (3) reviewing summative assessment solutions—accessing chapter-based workbook solutions, released weekly. Workbooks were summative assessments, assigned weekly with a weekly requirement to submit their answer sheets (paper-based submissions). Data from 736 students who completed the course were used for the study. The event logs included a set of attributes {Brand New (Yes 87%, No 13%) (i.e., indicating whether they are first semester students), Gender (Male 69%, Female 31%, Other 0.1%), Program (48 different programs), Residential Status (Domestic 71%, International 29%), Final Exam Score (High 38%, Mid 32%, Low 30%), and the level of engagement in each learning activity (Course Materials 11%, Formative Quiz 42%, Solution of Summative 47%). In addition, as a trace of learning events for each student, the learning activity and the related course chapter were included in the event log. In the remainder of this section, we present the recommendations generated for this course using relatively small ($\alpha = 0.05$), medium ($\alpha = 0.1$), and large ($\alpha = 0.3$) student populations.

4.3.1 Performance-Based Drill-Down Recommendations

As the input for the performance-based drill-down, we selected {Brand New, Gender, Program, Residential Status, and the level of engagement in each of the three learning activities} and {Final Exam Score} as the target feature. Figure **??** presents the recommendations generated for this course. The arrows indicate whether the average result of the final exam in the subpopulation was higher ($\uparrow$) or lower ($\downarrow$) than that of the remainder of the class. As demonstrated in the figure, the use of $\alpha = 0.3$ leads to the generation of drill-down recommendations with a relatively low distance score (average distance = 0.035) that finds relatively large subpopulations (average coverage = 0.325). As expected, a decrease in α leads to an increase in distance score. For example, the use of $\alpha = 0.05$ leads to drill-down recommendations with a higher distance score (average score = 0.1) on smaller subpopulations (average coverage = 0.05). Contrary to our initial assumptions, the average length of a drill-down action is not necessarily a good indicator of its distance score. In the reported results, the average length of drill-downs for $\alpha = 0.05$ is 3, which is higher than that of drill-downs for $\alpha = 0.3$ (average length is 1.5) but is lower than that of $\alpha = 0.1$ (average length is 4). In relation to the actual features used in the recommendations, it is reassuring to see that low engagement with learning activities was chosen in five of the drill-down recommendations that had a lower exam grade than the rest of the class.

α	Recommended Drill-Down Criteria	Coverage	Distance
0.05	(1) [Brand New = "No"] and [Review Solution of Summative = "Low"] and [Access to Course Materials = "Low"]	0.05	0.11 $\downarrow$
	(2) [Program = "Bachelor of Engineering (Honours)"] and [Brand New = "No"]	0.05	0.09 $\downarrow$
0.10	(3) [Gender = "Male"] and [Program = "Bachelor of Engineering (Honours)"] and [Review Solution of Summative = "Low"] and [Submission of Formative Quiz = "Low"] and [Access to Course Materials = "Low"]	0.10	0.06 $\downarrow$
	(4) [Program = "Bachelor of Engineering (Honours)"] and [Review Solution of Summative = "Low"] and [Submission of Formative Quiz = "Low"] and [Access to Course Materials = "Low"]	0.11	0.06 $\downarrow$
0.30	(5) [Review Solution of Summative = "Low"] and [Submission of Formative Quiz = "Low"]	0.34	0.04 $\downarrow$
	(6) [Submission of Formative Quiz = "Mid"]	0.31	0.03 $\uparrow$

(b) Distribution of final exam scores in subpopulation of recommendation (1).

(a) Resulting recommendations generated by AID

Figure 9. Performance of filtered students (by the recommended drill-down) versus the rest of the students.

Visualizing subpopulation performance deviations   To further investigate the insights that can be derived from the recommended drill-downs, we projected that the distribution of final exam scores for the identified subpopulation and the remaining students on a comparative bar graph recommended drill-down (1) (shown in Figure 9(b)). This drill-down results in a subpopulation of Brand New = "No", Review Solution of Summative = "Low", and Access to Course Materials = "Low". According to the AID result, the performance of the subpopulation on the final exam was 11% different from that of the remaining students. As Figure 9(b) illustrates, two major differences were the percentage of medium-scoring students (+31 percentage points) and high-scoring students (-23.81 percentage points). Furthermore, on average, the subpopulation performed 20% worse than the remaining students, which might be explained (possible correlation) by their low engagement with Review Solution of Summative and Access to Course Materials activities.

4.3.2 Process-Based Drill-Down Recommendations

As the input for the process-based drill-down, we used {Brand New, Gender, Program, Residential Status} as the selected features set. Table 2 presents the recommendations generated for this course. It is worth noting that the distance scores resulting from the two algorithms are not comparable because they use different underlying scoring functions. Similar to the case of performance-based recommendations, a decrease in α leads to an increase in the distance score. In the reported results, the use of $\alpha = 0.3$ leads to the generation of drill-down recommendations with a relatively low distance score (average distance = 0.635) that finds relatively large subpopulations (average coverage = 0.51). Decreasing α to 0.05 leads to the generation of drill-down recommendations with a higher distance score (average score = 0.71) on smaller subpopulations (average coverage = 0.05).

Unlike the case of performance-based recommendations, the average length of drill-down actions correlates with the distance score. In the reported results, the average length of drill-downs for $\alpha = 0.05$ is 3.5, which is higher than that of drill-downs for $\alpha = 0.1$ (average length is 3), which is higher than that of $\alpha = 0.3$ (average length is 1.5). In relation to the actual features used in the recommendations, an interesting observation is that the Brand New (i.e., being in the first semester) feature was a common feature in five of the drill-down recommendations, which indicates that the learning behaviour was commonly different in students that were new in the education system of the University of Queensland than that of the rest of the class.

Table 2. Process-based recommendations visualization

α	Recommended Drill-Down Criteria	Coverage	Distance
0.05	(1) [Brand New = "Yes" and Residential Status = "International" and Final Exam = "High" and Gender = "Male"]	0.05	0.72
	(2) [Brand New = "Yes" and Residential Status = "International" and Gender = "Female"]	0.05	0.70
0.1	(3) [Brand New = "Yes" and Residential Status = "International" and Program = "Bachelor of Engineering (Honours)"]	0.10	0.69
	(4) [Brand New = "Yes" and Residential Status = "International" and Gender = "Male"]	0.12	0.68
0.3	(5) [Final Exam = "High"]	0.33	0.64
	(6) [Brand New = "Yes" and Residential Status = "Domestic"]	0.69	0.63

Visualizing deviations of subpopulation learning processes   Here, we demonstrate the insights derived from the recommended drill-down (1) (shown in Table 2). This drill-down results in a subpopulation of Brand New = "Yes" and Residential Status = "International" and Final Exam = "High" and Gender = "Male". According to the AID result, this subpopulation's learning process is 72% different from that of the remaining students. To investigate the difference between the two learning processes, we visualized the underlying process of the subpopulation (shown in Figure 10(a)) and the remaining students (Figure 10(b)). Each box in the map is an activity, which is labelled by the action type and the relevant chapter (e.g., Formative Quiz—Chapter1). We used colour coding to visually emphasize the differences in the three types of learning activities in the process map. In the subpopulation's process, the arrows between the three different types of activities indicate switching between the types of learning tasks. Such switching can indicate that the three types of tasks were performed every week before the next chapter's activities were made available. In contrast, the underlying process of the remaining students shows that the activity types related to Chapters 9 to 18 (highlighted in Figure 10(b)) were mainly performed sequentially, which indicates students performing them at the end of the semester when all tasks were available.

To further investigate our initial findings, we used the Events graph of Disco (Fluxicon, n.d.) to compare the distribution of the events over the semester. Figures 10(c) and 10(d) demonstrate that the subpopulation was more active during the semester than the remaining students. Furthermore, the average number of events per student was 36 in the subpopulation and 25 for the remaining students. To conclude our analysis, the identified subpopulation had a high rate of activities throughout the semester compared to the remaining students. One of the common features of this subpopulation was their high performance on the final exam, which might be associated with their learning process. Some other differences identified by comparing the two process maps are the Formative Quiz of Chapter 8 not being performed by any students in the subpopulation; Solution Review of Chapters 2, 7, 8, and 9 being the highest-rated activities in the subpopulation; and Solution Review of Chapters 1, 2, 6, 7, 8, and 9 being the highest-rated activities by the rest of the class.

4.3.3 Feedback from the Instructor

We presented the reported drill-down recommendations and the process visualizations to the instructor of the course to capture their feedback and comments on the findings. Their feedback can be summarized as follows: (1) While the instructor had access to Course Insights throughout the semester, they rarely used it and generally found it to be overwhelming. They considered the large number of potential drill-down options within the platform to be the main reason that using the platform was overwhelming. (2) Findings of learning process that have led to successful outcomes can be used for positive deviance (Marsh, Schroeder, Dearden, Sternin, & Sternin, 2004) purposes. The instructor indicated that they would like to share Figure 10 as a recommended pattern of successful learning with their students as evidence that consistent engagement with learning activities throughout the semester is related to better outcomes. (3) Providing the ability to receive drill-down recommendations based on a rule (e.g., "Midterm" < 50) would be useful. The instructor indicated that they would like to understand deviations in low-performing and at-risk students to help them pass the course.

In summary, the practical application confirmed the value of automated drill-down recommendations, while also providing us with a number of helpful learnings with which to refine and extend AID, as discussed next.
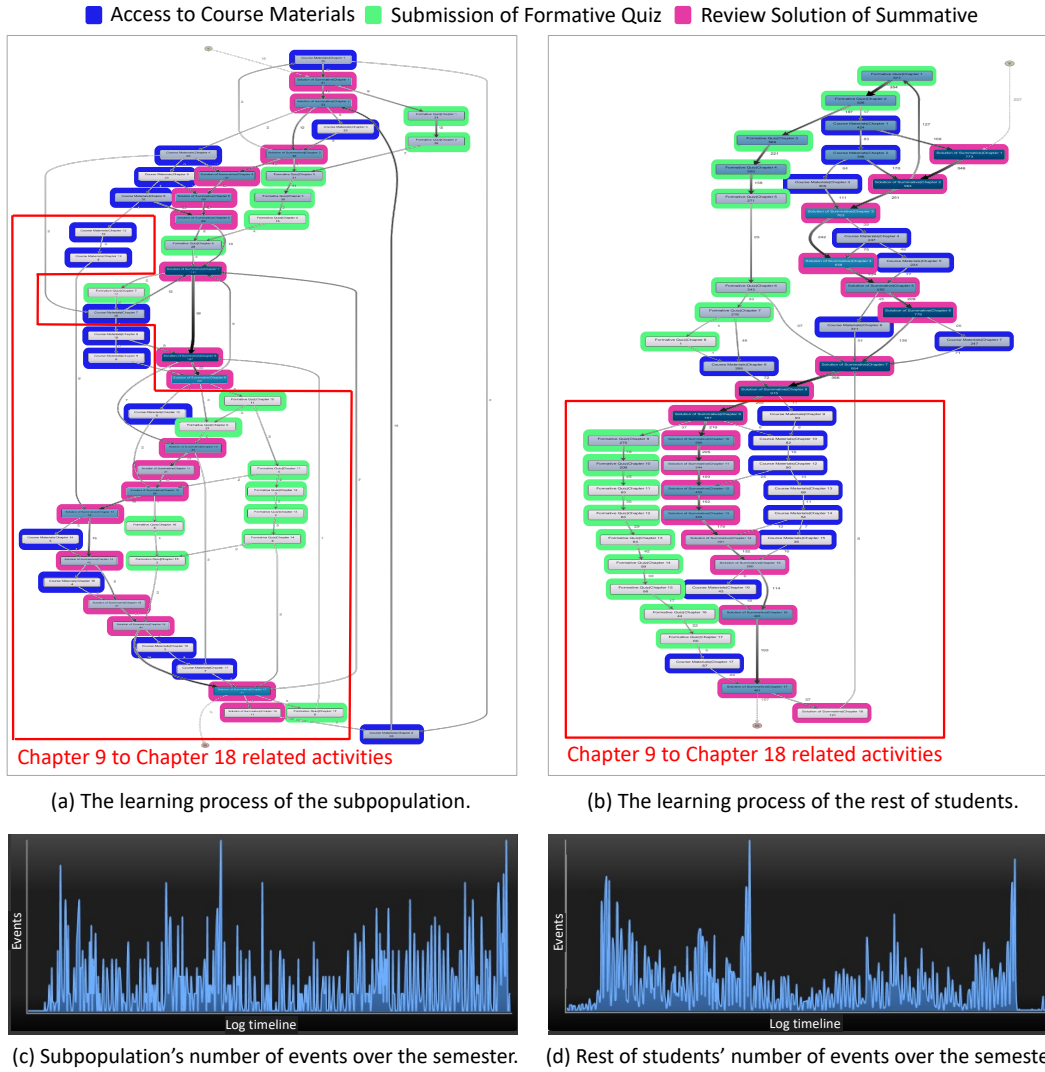
■ Access to Course Materials   ■ Submission of Formative Quiz   ■ Review Solution of Summative

Chapter 9 to Chapter 18 related activities

(a) The learning process of the subpopulation.

Chapter 9 to Chapter 18 related activities

(b) The learning process of the rest of students.

(c) Subpopulation's number of events over the semester.

(d) Rest of students' number of events over the semester.

Figure 10. Learning process of filtered students (by the recommended drill-down) versus the rest of students.

5. Discussion and Conclusions

Our overarching aim is to contribute to the understanding of how human-in-the-loop AI methods in the form of guided exploratory approaches can be employed in the development of successful LADs. In particular, we have developed and presented the AID approach, which identifies subpopulations of students with the highest deviation in performance or learning process compared to the rest of the class. Our practical application of AID indicates that AID can be integrated into LADs to produce intelligent LADs. Such LADs can provide instructors with meaningful and efficient ways to gain insight into subpopulations of students that would otherwise be missed, or misinterpreted, in current manual or predictive approaches. In particular, we note that manual drill-downs without guidance can be overwhelming, and many instructors tend to pose simple drill-downs to answer curiosity-driven questions. Approaches such as AID provide an opportunity for instructors to utilize data-driven drill-downs alongside curiosity-driven drill-downs.

There are several interesting directions to pursue in future work to overcome current limitations and help prepare AID for incorporation into LADs, which are discussed below.

Algorithm updates and interface implementations In Subsection 2.2, we referred to fallacy as one of the existing challenges with a manual drill-down analysis. We have shown how fallacy can occur due to the user failing to explore the entire decision tree and, therefore, missing the attribute that contributes most to increasing the distance score. Our proposed algorithms explore

the entire tree to address this issue. However, fallacy can still occur if a child node of the tree increases scores minimally due to overfitting to the data. Future work aims to explore how we can use regularizers to avoid overfitting, and thus fallacy, affecting the proposed approach.

To date, more than 10 workshops with over 50 participants, who were mostly academics or learning designers, have been facilitated on Course Insights. The workshops have generally focused on presenting the current version of the system and receiving feedback from participants. A few of the sessions focused on co-creation and participatory design, leading to the design of the presented interfaces (Figure 7 and Figure 8) for inclusion of the AID recommendations within Course Insights. The participants had very different levels of digital literacy with diverse expectations from the available features in the system. This makes the task of designing interfaces that are easy to navigate while satisfying the expectations of a diverse audience quite challenging. Nevertheless, improvements to the underlying algorithms and implementation of easy-to-use interfaces for engaging with the recommendations may be two of the most effective ways of making AID more accessible to non-technical users. We provide three examples of the tasks that are in our roadmap: (1) The constant α plays a very important role in the quality of the generated recommendations. A poor choice of α may lead to disappointing results for the user. Our future work should extend AID such that the algorithm itself scans over a range of values for α and recommends promising drill-downs with various coverages. (2) Although AID can generate insightful (high significance score) performance-based or process-based drill-down recommendations instantly, further work is needed to investigate what criteria can be used to determine insightfulness. We hope to partner with instructors and learning designers to uncover additional criteria beyond deviation. (3) AID currently uses a different scoring function for ranking performance-based and process-based recommendations. This makes it challenging for instructors to compare and contrast performance-based and process-based recommendations. Future work should aim to develop strategies such as utilizing a common scoring function to compare the insightfulness of performance-based recommendations against process-based recommendations.

Engaging with the recommendations   The focus of this paper has been on the development of an approach for generating the recommendations; however, an equally important aspect that needs to be investigated is the development of effective methods to help instructors understand and act upon the provided recommendations. Here, we outline three lines of promising future work for assisting instructors in engaging with the recommendations. (1) As mentioned in Section 2.3, existing work on process mining techniques in education has been utilized mostly by educational researchers with a strong technical background in computer science and data science through general-purpose process mining tools (Saint et al., 2018). We hope to partner with instructors and learning designers to develop appropriate visualization method(s) for presenting the results of the drill-down recommendations to frontline users in education. (2) An important component of empowering instructors with insights and returning meaningful learning analytics data to them is to inform learning design decisions (Corrin et al., 2016). A number of frameworks have been proposed for connecting learning analytics and learning design (Bakharia et al., 2016; Lockyer, Heathcote, & Dawson, 2013; Schmitz, van Limbeek, Greller, Sloep, & Drachsler, 2017). We hope to work with instructors and learning designers to operationalize these frameworks into guidelines and practical tips for employing AID toward actionable pedagogical recommendations. (3) Another important component of empowering instructors with insights is to enable them to share these insights with students to promote self-regulation (Roll & Winne, 2015) and to encourage change (i.e., application of positive deviance). However, previous work has demonstrated that inappropriate sharing of information on student performance in comparison with their peers can also have negative impacts and lead to student demotivation (Lonn et al., 2015). Future work should investigate best practices of sharing insights derived from the use of AID in LADs with students to encourage self-regulation and positive deviance without negative impacts.

Evaluation   In this paper, we presented the use case of automated drill-down recommendations in the context of one course. As a first step, future directions include replicating this study with courses across different disciplines to investigate the generalizability of our current findings. The second step of evaluation includes the incorporation of AID into Course Insights to investigate how instructors may utilize recommended data-driven drill-downs alongside manually navigated curiosity-driven drill-downs.

Equity and fairness considerations   There is an increasing awareness in AI about the risk of inequity and unfairness with the knowledge generated through the machine learning algorithms and the way we interpret and operationalize this knowledge for education-oriented applications (Dudík et al., 2020; Holstein, Wortman Vaughan, Daumé III, Dudík, & Wallach, 2019). Because machine learning techniques learn from historical data, there is a danger of attracting extreme attention to historically disadvantaged groups, and some populations may become vulnerable to incorrect predictions or judgments (Rajkomar, Hardt, Howell, Corrado, & Chin, 2018); this also holds for the approaches introduced in this paper. While this paper provides a human-in-the-loop approach, the risk of inequity still exists. For instance, if students from a particular ethnicity are shown to have lower grades, could that be harmful in any way? Could it further bias the instructor and lead to further stereotyping? Other

concerns exist in terms of overreliance on the recommendations for assisting students, which may disadvantage students who were not identified by recommendations. Further investigation is needed to avoid this challenge.

References

Ahadi, A., Lister, R., Haapala, H., & Vihavainen, A. (2015). Exploring machine learning methods to automatically identify students in need of assistance. In *Proceedings of the 11th Annual International Conference on International Computing Education Research* (ICER 2015), 9–13 July 2015, Omaha, NE, USA (pp. 121–130). New York: ACM. doi: https://doi.acm.org/10.1145/2787622.2787717

Ahn, J., Campos, F., Hays, M., & Digiacomo, D. (2019). Designing in context: Reaching beyond usability in learning analytics dashboard design. *Journal of Learning Analytics*, *6*(2), 70–85. doi: https://doi.org/10.18608/jla.2019.62.5

Ain, N., Vaia, G., DeLone, W. H., & Waheed, M. (2019). Two decades of research on business intelligence system adoption, utilization and success—A systematic literature review. *Decision Support Systems*, *125*, 113113. doi: https://doi.org/10.1016/j.dss.2019.113113

Ali, L., Hatala, M., Gasevic, D., & Jovanovic, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, *58*(1), 470–489. doi: https://doi.org/10.1016/j.compedu.2011.08.030

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at purdue: Using learning analytics to increase student success. In S. Dawson, C. Haythornthwaite, S. B. Shum, D. Gasevic, & R. Ferguson (Eds.), *Proceedings of the Second International Conference on Learning Analytics and Knowledge* (LAK 2012), Vancouver, BC, Canada, 29 April–2 May 2012 (pp. 267–270). New York: ACM. doi: https://doi.org/10.1145/2330601.2330666

Bakharia, A., Corrin, L., de Barba, P. G., Kennedy, G. E., Gasevic, D., Mulder, R., . . . Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. In D. Gasevic, G. Lynch, S. Dawson, H. Drachsler, & C. P. Rosé (Eds.), *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (LAK 2016), 25–29 April 2016, Edinburgh, UK (pp. 329–338). New York: ACM. doi: https://doi.org/10.1145/2883851.2883944

Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. *WIREs Data Mining and Knowledge Discovery*, *8*(1), e1230. doi: https://doi.org/10.1002/widm.1230

Buckingham Shum, S., & Luckin, R. (2019). Learning analytics and AI: Politics, pedagogy and practices. *British Journal of Educational Technology*, *50*(6), 2785–2793. doi: https://doi.org/10.1111/bjet.12880

Cairns, A. H., Gueni, B., Assu, J., Joubert, C., & Khelifa, N. (2015). Analyzing and improving educational process models using process mining techniques. In *Proceedings of the Fifth International Conference on Advances in Information Mining Management* (IMMM 2015), 21–26 June 2015, Brussels, Belgium (pp. 17–22).

Caron, E., & Daniels, H. (2004). Automated business diagnosis in the OLAP context. In H. A. Fleuren, D. den Hertog, & P. M. Kort (Eds.), *Operations Research, Proceedings 2004, Selected Papers of the Annual International Conference of the German Operations Research Society (GOR), Jointly Organized with the Netherlands Society for Operations Research* (NGB), 1–3 September 2004, Tilburg, Netherlands (pp. 425–433). Springer. doi: https://doi.org/10.1007/3-540-27679-3_53

Charleer, S., Moere, A. V., Klerkx, J., Verbert, K., & Laet, T. D. (2018). Learning analytics dashboards to support adviser-student dialogue. *IEEE Transactions on Learning Technologies*, *11*(3), 389–399. doi: https://doi.org/10.1109/TLT.2017.2720670

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *SIGMOD Record*, *26*(1), 65–74. doi: https://doi.org/10.1145/248603.248616

Corrin, L., De Barba, P., Lockyear, L., Gašević, D., Williams, D., Dawson, S., . . . Bakharia, A. (2016). *Completing the Loop: Returning Meaningful Learning Analytic Data to Teachers.* Australian Government Office for Learning and Teaching. Retrieved from `https://ltr.edu.au/resources/ID13_3068_Kennedy_Report_2016.pdf`

Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting students drop out: A case study. In T. Barnes, M. C. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the Second International Conference on Educational Data Mining* (EDM 2009), 1–3 July 2009, Cordoba, Spain (pp. 41–50). www.educationaldatamining.org. Retrieved from `https://www.educationaldatamining.org/EDM2009/uploads/proceedings/dekker.pdf`

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. (2019). Working together in learning analytics towards the co-creation of value. *Journal of Learning Analytics*, *6*(2), 10–26. doi: https://doi.org/10.18608/jla.2019.62.2

Dudík, M., Chen, W., Barocas, S., Inchiosa, M., Lewins, N., Oprescu, M., . . . Wallach, H. (2020, September). *Assessing and mitigating unfairness in credit models with the fairlearn toolkit* (Tech. Rep. No. MSR-TR-2020-34). Microsoft. Retrieved from `https://go.ey.com/3kVTOY8`

Echeverría, V., Martínez Maldonado, R., Granda, R., Chiluiza, K., Conati, C., & Buckingham Shum, S. (2018). Driving data storytelling from learning design. In A. Pardo et al. (Eds.), *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge,* (LAK 2018), 7–9 March 2018, Sydney, Australia (pp. 131–140). New York: ACM. doi: https://doi.org/10.1145/3170358.3170380

Ez-zaouia, M., & Lavoué, É. (2017). EMODA: A tutor oriented multimodal and contextual emotional dashboard. In M. Hatala et al. (Eds.), *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (LAK 2017), 13–17 March 2017, Vancouver, BC, Canada (pp. 429–438). New York: ACM. doi: https://doi.org/10.1145/3027385.3027434

Fluxicon. (n.d.). *Disco.* Retrieved from `https://fluxicon.com/`

Fortenbacher, A., Beuster, L., Elkina, M., Kappe, L., Merceron, A., Pursian, A., . . . Wenzlaff, B. (2013). LeMo: A learning analytics application focussing on user path analysis and interactive visualization. In *Proceedings of the IEEE Seventh International Conference on Intelligent Data Acquisition and Advanced Computing Systems,* (IDAACS 2013), 12–14 September 2013, Berlin, Germany (pp. 748–753). IEEE. doi: https://doi.org/10.1109/IDAACS.2013.6663025

Gourville, J. T., & Soman, D. (2005). Overchoice and assortment type: When and why variety backfires. *Marketing Science*, *24*(3), 382–395. doi: https://doi.org/10.1287/mksc.1040.0109

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). The student activity meter for awareness and self-reflection. In J. A. Konstan, E. H. Chi, & K. Höök (Eds.), *Proceedings of the CHI Conference on Human Factors in Computing Systems* (CHI 2012), Extended Abstracts Volume, 5–10 May 2012, Austin, TX, USA (pp. 869–884). ACM. doi: https://doi.org/10.1145/2212776.2212860

Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., . . . Pirahesh, H. (1996). Data Cube: A relational aggregation operator generalizing GROUP-BY, CROSS-TAB, and SUB-TOTALS. In *Proceedings of the 12th International Conference on Data Engineering,* 26 February–1 March 1996, New Orleans, LA, USA (pp. 152–159). doi: https://doi.org/10.1109/ICDE.1996.492099

Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In B. Krishnapuram, M. Shah, A. J. Smola, C. C. Aggarwal, D. Shen, & R. Rastogi (Eds.), *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 13–17 August 2016, San Francisco, CA, USA (pp. 2125–2126). New York: ACM. doi: https://doi.org/10.1145/2939672.2945386

Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudík, M., & Wallach, H. M. (2019). Improving fairness in machine learning systems: What do industry practitioners need? In S. A. Brewster, G. Fitzpatrick, A. L. Cox, & V. Kostakos (Eds.), *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* 4–9 May 2019, Glasgow, UK (pp. 1–16). New York: ACM. doi: https://doi.org/10.1145/3290605.3300830

Hu, Y., Lo, C., & Shih, S. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior*, *36*, 469–478. doi: https://doi.org/10.1016/j.chb.2014.04.002

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In A. Pardo et al. (Eds.), *Proceedings of the Eighth International Conference on Learning Analytics and Knowledge* (LAK 2018), 7–9 March 2018, Sydney, Australia (pp. 31–40). New York: ACM. doi: https://doi.org/10.1145/3170358.3170421

Joglekar, M., Garcia-Molina, H., & Parameswaran, A. G. (2019). Interactive data exploration with smart drill-down. *IEEE Transactions on Knowledge and Data Engineering*, *31*(1), 46–60. doi: https://doi.org/10.1109/TKDE.2017.2685998

Jovanovic, J., Dawson, S., Joksimovic, S., & Siemens, G. (2020). Supporting actionable intelligence: reframing the analysis of observed study strategies. In C. Rensing & H. Drachsler (Eds.), *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 161–170). New York: ACM. doi: https://doi.org/10.1145/3375462.3375474

Jovanovic, J., Gasevic, D., Brooks, C. A., Devedzic, V., & Hatala, M. (2007). LOCO-analyst: A tool for raising teachers' awareness in online learning environments. In E. Duval, R. Klamma, & M. Wolpers (Eds.), *Creating New Learning*

Experiences on a Global Scale, Proceedings of the Second European Conference on Technology Enhanced Learning (EC-TEL 2007), 17–29 September 2007, Crete, Greece (Vol. 4753, pp. 112–126). Springer. doi: https://doi.org/10.1007/978-3-540-75195-3_9

Khosravi, H., Sadiq, S. W., & Gasevic, D. (2020). Development and adoption of an adaptive learning system: Reflections and lessons learned. In J. Zhang, M. Sherriff, S. Heckman, P. A. Cutter, & A. E. Monge (Eds.), *Proceedings of the 51st ACM Technical Symposium on Computer Science Education,* (SIGCSE 2020), 11–14 March 2020, Portland, OR, USA (pp. 58–64). New York: ACM. doi: https://doi.org/10.1145/3328778.3366900

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* (1st ed., pp. 17–22). Alberta, Canada: Society for Learning Analytics Research (SoLAR). doi: https://doi.org/10.18608/hla17.001

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. doi: https://doi.org/10.1214/aoms/1177729694

Lee, D. J. L., Dev, H., Hu, H., Elmeleegy, H., & Parameswaran, A. G. (2019). Avoiding drill-down fallacies with *VisPilot*: assisted exploration of data subsets. In W. Fu, S. Pan, O. Brdiczka, P. Chau, & G. Calvary (Eds.), *Proceedings of the 24th International Conference on Intelligent User Interfaces* (IUI 2019), 17–20 March 2019, Marina del Ray, CA, USA (pp. 186–196). New York: ACM. doi: https://doi.org/10.1145/3301275.3302307

Leemans, S. J. J., Shabaninejad, S., Goel, K., Khosravi, H., Sadiq, S. W., & Wynn, M. T. (2020). Identifying cohorts: Recommending drill-downs based on differences in behaviour for process mining. In G. Dobbie, U. Frank, G. Kappel, S. W. Liddle, & H. C. Mayr (Eds.), *Proceedings of the 39th International Conference on Conceptual Modeling* (ER 2020), 3–6 November 2020, Vienna, Austria (Vol. 12400, pp. 92–102). Springer. doi: https://doi.org/10.1007/978-3-030-62522-1_7

Leemans, S. J. J., Syring, A. F., & van der Aalst, W. M. P. (2019). Earth movers' stochastic conformance checking. In T. T. Hildebrandt, B. F. van Dongen, M. Röglinger, & J. Mendling (Eds.), *Proceedings of the Business Process Management Forum* (BPM Forum 2019), 1–6 September 2019, Vienna, Austria (Vol. 360, pp. 127–143). Springer. doi: https://doi.org/10.1007/978-3-030-26643-1_8

Leony, D., Pardo, A., de la Fuente Valentín, L., de Castro, D. S., & Kloos, C. D. (2012). GLASS: A learning analytics visualization tool. In S. Dawson, C. Haythornthwaite, S. Buckingham Shum, D. Gasevic, & R. Ferguson (Eds.), *Proceedings of the Second International Conference on Learning Analytics and Knowledge,* (LAK 2012), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 162–163). New York: ACM. doi: https://doi.org/10.1145/2330601.2330642

Liu, D., Bartimote-Aufflick, K., Pardo, A., & Bridgeman, A. (2017). Data-driven personalization of student learning support in higher education. In A. Peña-Ayala (Ed.), *Learning Analytics: Fundaments, Applications, and Trends* (Vol. 94, pp. 143–169). doi: https://doi.org/10.1007/978-3-319-52977-6_5

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), 1439–1459. doi: https://doi.org/10.1177/0002764213479367

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. doi: https://doi.org/10.1016/j.chb.2014.07.013

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. doi: https://doi.org/10.1016/j.compedu.2016.09.005

Marsh, D., Schroeder, D., Dearden, K., Sternin, J., & Sternin, M. (2004). The power of positive deviance. *BMJ (Clinical Research Edition)*, 329, 1177–1179. doi: https://doi.org/10.1136/bmj.329.7475.1177

Matcha, W., Ahmad Uzir, N., Gasevic, D., & Pardo, A. (2019). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies*, 13(2), 226–245. doi: https://doi.org/10.1109/TLT.2019.2916802

Matcha, W., Gasevic, D., Uzir, N. A., Jovanovic, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the Ninth International Conference on Learning Analytics and Knowledge* (LAK 2019), 4–8 March 2019, Tempe, AZ, USA (pp. 461–470). New York: ACM. doi: https://doi.org/10.1145/3303772.3303787

Mukala, P., Buijs, J. C. A. M., Leemans, M., & van der Aalst, W. M. P. (2015). Learning analytics on Coursera event data: A process mining approach. In P. Ceravolo & S. Rinderle-Ma (Eds.), *Proceedings of the Fifth International Symposium on Data-Driven Process Discovery and Analysis* (SIMPDA 2015), 9–11 December 2015, Vienna, Austria (Vol. 1527, pp. 18–32). CEUR-WS.org. Retrieved from `http://ceur-ws.org/Vol-1527/paper2.pdf`

Pardo, A., Bartimote, K., Buckingham Shum, S., Dawson, S., Gao, J., Gašević, D., ... Vigentini, L. (2018). OnTask:

Delivering data-informed, personalized learning support actions. *Journal of Learning Analytics*, *5*(3), 235–249. doi: https://doi.org/10.18608/jla.2018.53.15

Park, Y., & Jo, I.-H. (2019). Factors that affect the success of learning analytics dashboards. *Educational Technology Research and Development*, *67*(6), 1547–1571. doi: https://doi.org/10.1007/s11423-019-09693-0

Pechenizkiy, M., Trcka, N., Vasilyeva, E., van der Aalst, W. M. P., & Bra, P. D. (2009). Process mining online assessment data. In T. Barnes, M. C. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the Second International Conference on Educational Data Mining* (EDM 2009), 1–3 July 2009, Cordoba, Spain (pp. 279–288). Retrieved from https://educationaldatamining.org/EDM2009/uploads/proceedings/pechenizkiy.pdf

Putnam, V., & Conati, C. (2019). Exploring the need for explainable artificial intelligence (XAI) in intelligent tutoring systems (ITS). In C. Trattner, D. Parra, & N. Riche (Eds.), *Joint Proceedings of the ACM IUI 2019 Workshops co-located with the 24th ACM Conference on Intelligent User Interfaces* (ACM IUI 2019), 20 March 2019, Los Angeles, CA, USA (Vol. 2327). CEUR-WS.org. Retrieved from http://ceur-ws.org/Vol-2327/IUI19WS-ExSS2019-19.pdf

Queiroz-Sousa, P. O., & Salgado, A. C. (2020). A review on OLAP technologies applied to information networks. *ACM Transactions on Knowledge Discovery from Data*, *14*(1), 8:1–8:25. doi: https://doi.org/10.1145/3370912

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, *169*(12), 866-872. (PMID: 30508424) doi: https://doi.org/10.7326/M18-1990

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12. doi: https://doi.org/10.18608/jla.2015.21.2

Saint, J., Gasevic, D., Matcha, W., Uzir, N. A., & Pardo, A. (2020). Combining analytic methods to unlock sequential and temporal patterns of self-regulated learning. In C. Rensing & H. Drachsler (Eds.), *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 402–411). New York: ACM. doi: https://doi.org/10.1145/3375462.3375487

Saint, J., Gasevic, D., & Pardo, A. (2018). Detecting learning strategies through process mining. In V. Pammer-Schindler, M. Pérez-Sanagustín, H. Drachsler, R. Elferink, & M. Scheffel (Eds.), *Lifelong Technology-Enhanced Learning—Proceedings of the 13th European Conference on Technology Enhanced Learning* (EC-TEL 2018), 3–5 September 2018, Leeds, UK (Vol. 11082, pp. 385–398). Springer. doi: https://doi.org/10.1007/978-3-319-98572-5_29

Sarawagi, S. (2000). User-adaptive exploration of multidimensional data. In A. E. Abbadi et al. (Eds.), *Proceedings of the 26th International Conference on Very Large Data Bases* (VLDB 2000), 10–14 September 2000, Cairo, Egypt (pp. 307–316). Morgan Kaufmann. Retrieved from https://www.vldb.org/conf/2000/P307.pdf

Sarawagi, S. (2001). User-cognizant multidimensional analysis. *The VLDB Journal*, *10*(2-3), 224–239. doi: https://doi.org/10.1007/s007780100046

Schmitz, M., van Limbeek, E., Greller, W., Sloep, P. B., & Drachsler, H. (2017). Opportunities and challenges in using learning analytics in learning design. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Proceedings of Data Driven Approaches in Digital Education—12th European Conference on Technology Enhanced Learning* (EC-TEL 2017), 12–15 September 2017, Tallinn, Estonia (Vol. 10474, pp. 209–223). Springer. doi: https://doi.org/10.1007/978-3-319-66610-5_16

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., . . . Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41. doi: https://doi.org/10.1109/TLT.2016.2599522

Shabaninejad, S., Khosravi, H., Indulska, M., Bakharia, A., & Isaias, P. (2020). Automated insightful drill-down recommendations for learning analytics dashboards. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 41–46). New York: ACM. doi: https://doi.org/10.1145/3375462.3375539

Shabaninejad, S., Khosravi, H., Leemans, S. J. J., Sadiq, S. W., & Indulska, M. (2020). Recommending insightful drill-downs based on learning processes for learning analytics dashboards. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Proceedings of the 21st International Conference on Artificial Intelligence in Education, Part I* (AIED 2020), 6–10 July 2020, Ifrane, Morocco (Vol. 12163, pp. 486–499). Springer. doi: https://doi.org/10.1007/978-3-030-52237-7_39

Siemens, G., & Gasevic, D. (2012). Guest editorial—Learning and knowledge analytics. *Journal of Educational Technology & Society*, *15*(3), 1–2. Retrieved from https://www.jstor.org/stable/jeductechsoci.15.3.1

Trcka, N., & Pechenizkiy, M. (2009). From local patterns to global models: Towards domain driven educational process mining. In *Proceedings of the Ninth International Conference on Intelligent Systems Design and Applications* (ISDA 2009), 30 November–2 December 2009, Pisa, Italy (pp. 1114–1119). IEEE Computer Society. doi: https://doi.org/

10.1109/ISDA.2009.159

Uzir, N. A., Gasevic, D., Matcha, W., Jovanovic, J., & Pardo, A. (2020). Analytics of time management strategies in a flipped classroom. *Journal of Computer Assisted Learning*, *36*(1), 70–88. doi: https://doi.org/10.1111/jcal.12392

van der Aalst, W. M. P. (2016). *Process Mining—Data Science in Action* (2nd ed.). Springer. doi: https://doi.org/10.1007/978-3-662-49851-4

van der Aalst, W. M. P., Guo, S., & Gorissen, P. (2013). Comparative process mining in education: An approach based on process cubes. In P. Ceravolo, R. Accorsi, & P. Cudré-Mauroux (Eds.), *Data-Driven Process Discovery and Analysis—Third IFIP WG 2.6, 2.12 International Symposium* (SIMPDA 2013), 30 August 2013, Riva del Garda, Italy, Revised Selected Papers (Vol. 203, pp. 110–134). Springer. doi: https://doi.org/10.1007/978-3-662-46436-6_6

Verbert, K., Ochoa, X., Croon, R. D., Dourado, R. A., & Laet, T. D. (2020). Learning analytics dashboards: The past, the present and the future. In C. Rensing & H. Drachsler (Eds.), *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020), 23–27 March 2020, Frankfurt, Germany (pp. 35–40). New York: ACM. doi: https://doi.org/10.1145/3375462.3375504

Wise, A. F., & Jung, Y. (2019). Teaching with analytics: Towards a situated model of instructional decision-making. *Journal of Learning Analytics*, *6*(2), 53–69. doi: https://doi.org/10.18608/jla.2019.62.4