

A tale of three cases: Examining accuracy, efficiency, and process differences in diagnosing virtual patient cases

Tenzin Doleck and Amanda Jarrell
McGill University

Eric G. Poitras
University of Utah

Maher Chaouachi and Susanne P. Lajoie
McGill University

Clinical reasoning is a central skill in diagnosing cases. However, diagnosing a clinical case poses several challenges that are inherent to solving multifaceted ill-structured problems. In particular, when solving such problems, the complexity stems from the existence of multiple paths to arriving at the correct solution (Lajoie, 2003). Moreover, the approach one employs in diagnosing a clinical case is in some measure dependent upon the complexity of the case. This leads us to the question: Are there differences in the manner in which novices solve cases with varying levels of complexity in a computer-based learning environment? More specifically, we are interested in understanding and elucidating if there are clinical reasoning differences in regards to accuracy, efficiency, and process across three virtual patient cases of varying difficulty levels. Examining such differences may have implications from both a learner modelling and system enhancement perspective. We close by discussing the implications for practice, limitations of the study and future research directions.

Introduction

Clinical reasoning involves the “application of knowledge and clinical experience towards a clinical presentation to derive a solution” (Noll, Key, & Jensen, 2001, p. 41) and consequently is a crucial skill for medical students and professionals (Delany & Golding, 2014; Norman, 2005) because of the tangible implications for patient outcomes and safety (Levett-Jones et al., 2010). Clinical reasoning is inextricably linked with healthcare quality and outcomes and thus is a focal point in medical education (Ryan & Higgs, 2008). Nendaz and Perrier (2012) note that diagnostic errors are associated with 8% of adverse events in medicine and up to 30% of malpractice claims. The challenge in teaching clinical reasoning is that the process is complex and tacit (Higgs & Jones, 2000), thus making it a difficult task to teach (Delany & Golding, 2014). The most commonly used instructional approach to teach these skills is through clinical rotations (Lee et al., 2010); however, Gigante (2013) argues that “relying on time and experience to develop these skills is insufficient” (p. 1). Moreover, Levett-Jones et al. (2010) note that current teaching and learning approaches are lacking in developing the necessary clinical reasoning skills. Technology can complement clinical internship and classroom teaching by providing additional opportunities for students to develop clinical reasoning skills in an asynchronous and supportive environment.

Clinical reasoning is complex and ill-defined in that there is no singular problem-solving path for arriving at the correct diagnosis; there are multiple routes to diagnosing a problem (Lajoie, 2003). The ill-defined nature of clinical problems makes them more difficult to solve because there are no set procedures or algorithms that will lead to the correct diagnosis. Diagnosis correctness, while crucial, may be an inadequate measure of assessing learners’ development of expertise in clinical reasoning. For example, Eva (2005) cautions that “one should not assume that because a student has provided an accurate diagnosis and/or management plan, he or she fully understands the physiological mechanisms underlying the process” (p. 104). Despite such calls, much less attention has been paid to examining the problem-solving path taken to diagnose a case. To illustrate, consider the actions (for instance, reading about diabetes in the library after ordering a lab test for fasting blood glucose level) taken by a learner in diagnosing a case (see Figure 1); a deeper look into such fine-grained information may reveal meaningful insights about how learners synthesise complex information and perform.

Research investigating clinical reasoning has illustrated the importance of case specificity in medical education (Fitzgerald et al., 1994; van der Vleuten & Swanson, 1990; Wimmers, Splinter, Hancock, & Schmidt, 2006); this phenomenon refers to the variability in performance from case to case in medical problem solving. The prior knowledge that is associated with different diseases and their underlying physiological processes, combined with the prevalence of certain diseases, opportunities to encounter them in practice and gather detailed information, means that cases may vary in their level of difficulty. However, little research has examined the differences in performance variables across differing cases and how this impacts learning and the need to adjust feedback provided to learners.

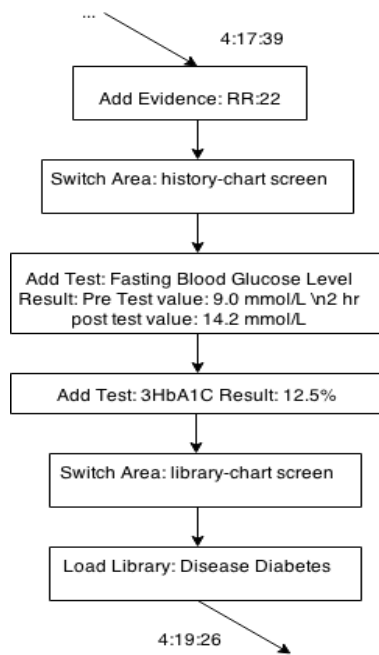


Figure 1. Example of a line of clinical reasoning

Thus, a particularly relevant question is how learners' performance differs in diagnosing different clinical cases; explication of such differences may have implications for understanding how novices diagnose different simulated cases and in turn develop expertise. The present study makes an initial step towards resolving this issue by comparing diagnostic performance and processes across different cases or diseases, but within a similar type of physiological system. Furthermore, this undertaking is linked to the question of how best to support and foster learning and practice of clinical reasoning skills in simulated learning environments. As such, the findings will inform efforts to tailor instruction within BioWorld, a computer-based learning environment (CBLE) for clinical reasoning, to the specific needs of different learners and to recommend challenging cases to learners that are not beyond their level of competency.

The purpose of the present study was twofold: to examine performance differences (measured by accuracy and efficiency) in clinical reasoning across three different virtual patient cases in BioWorld and, to investigate and exemplify whether there are process differences across cases using a data mining approach called process mining. Our hope is that this initial exploration represents a useful step forward in our ongoing efforts to better understand clinical reasoning of simulated cases by novices; specifically, we hope the findings will yield insights into how performance and process differences vary across different cases, and serve as an instructional design roadmap for developing a recommender system to assign cases to novice physicians and for augmenting our current comprehension of various problem-solving trajectories involved in clinical reasoning. As a potential system modification, this system could capture individual differences in prior knowledge and experience to assign cases at the right level of difficulty to better support learners' needs.

The remainder of this paper is organised as follows: The first section delineates the background for our work and discusses the related literature; the second section presents a brief overview of the learning

environment used in the study, namely, BioWorld; the third section outlines the methods; the fourth section presents the analysis and results of performance differences across the three clinical cases; the fifth section outlines the procedure and findings of using process mining for ascertaining the process differences across the three cases; and the final section highlights the results, some limitations, and future extensions of the current study.

Related work

In recent years, the field of education has seen a veritable explosion of the use of technology. Investigations of CBLEs have demonstrated that learning can be supported in a wide range of educational domains, from algebra to history (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995; Beal, Wallis, Arroyo, & Woolf, 2007; Matsuda et al., 2013; Poitras, Lajoie, & Hong, 2011; Vanlehn et al., 2005). This body of work elucidates the salient role that technology plays in learning experiences and outcomes. CBLEs provide an effective means to practise skills in authentic contexts and gain expertise by providing opportunities for deliberate practice and feedback (Lajoie & Azevedo, 2006). Deliberate practice refers to engaging in effortful activities that are designed to enhance one's competence (Ericsson, 2006). Expertise in clinical reasoning can be acquired through the deliberate practice of diagnosing virtual patient cases (Lajoie, 2009). BioWorld (Lajoie, 2009) provides such opportunities for deliberate practice of clinical reasoning by scaffolding novice physicians as they practise with realistic virtual cases. Expert performance models are provided as scaffolds for novices to compare their own performance against an ideal solution for each case. Fine-grained learner-system usage data can be tracked and captured in CBLEs (Baker & Yacef, 2009). Similarly, BioWorld logs learner actions as they practise and gain clinical reasoning skills. In a concerted effort to support learners, CBLEs have been increasingly viewed as a promising tool in training and fostering learning; BioWorld represents one of many such tools to afford rich learning and training opportunities.

One way to provide medical students with opportunities for deliberate practice is virtually through either high- or low-fidelity simulations. Medical simulations have shown promising results for the transfer of processes and skills measured or identified in simulation environments to clinical practice. For example, one longitudinal study showed that medical student clinical problem-solving processes at the beginning of their training is very similar to that of doctors (Neufeld, Norman, Feightner, & Barrows, 1981). Simulated environments allow researchers to map the reasoning or problem-solving process employed by students and experts. These process maps are useful for understanding the similarities and differences in diagnostic processes for different medical cases and across levels of expertise. These process maps can be used to determine the steps taken to arrive at a diagnosis, how the process used for one case compares to the process used for another case and if a specific process is more or less likely to lead to diagnostic accuracy.

Performance on one medical case does not necessarily predict performance on subsequent cases because case specificity leads to different performance outcomes (Fitzgerald et al., 1994; van der Vleuten & Swanson, 1990; Wimmers et al., 2006). In fact, the correlation between diagnostic performances across cases has been reported as low as 0.1 to 0.3 (Norman, Tugwell, Feightner, Muzzin, & Jacoby, 1985). One reason for this lack of performance transfer is that there is substantial variability between and within clinical cases. For example, among 100 patients with pheochromocytoma, a rare tumour on the adrenal gland, 83 patients had functional tumours producing dangerous hormones while 17 patients had silent tumours and produced no hormones. Of the functional tumours, 15% were misdiagnosed leading to tragic consequences (Melicow, 1977). Thus, experience with one case or one type of disease does not necessitate transfer of diagnostic accuracy to other cases.

There are few studies that have investigated problem-solving process in CBLEs or simulation environments. Exploring medical problem solving in a CBLE has several distinct advantages over traditional paper and pencil assessment situations. CBLEs are able to track the moment-to-moment decisions made in order to reach a final diagnosis. Technology can be used to accurately capture processes used in real time, across cases that range in complexity. The CBLE approach allows for a pure assessment of clinical reasoning that is independent of retrospective self-report or memory biases. Studies that have investigated the diagnostic process using CBLEs have been primarily concerned with authentic methods for clinical assessment (Fitzgerald et al., 1994; Schuwirth & Van der Vleuten, 2003). It was found that case specificity hampered the reliability of student performance on the CBLE clinical problems

(Fitzgerald et al., 1994) and in simulated clinical reasoning environments (Schuwirth & Van der Vleuten, 2003). This finding calls for a different approach to evaluating how medical trainees solve clinical cases when cases differ in complexity and medical domain knowledge. One alternative to merely evaluating final performance metrics of medical students is to map their problem-solving process in order to determine the clinical reasoning strategy they used and how this contributes to their performance. This two-pronged approach, presented in this paper, might clarify why some students perform well on some cases and poorly on other cases, despite having similar foundational clinical knowledge.

BioWorld: An intelligent tutoring system for clinical reasoning

BioWorld is a CBLE designed with cognitive tools to help medical students practise clinical reasoning skills in an authentic learning space while receive expert feedback (Lajoie, 2009) as they diagnose simulated patient cases. In practising clinical reasoning in BioWorld, a learner is tasked with diagnosing a simulated patient case. While the learner engages in the act of diagnosis, the system captures fine-grained learner actions. BioWorld consists of four learning spaces (Problem, Chart, Library, and Consult). The Problem space provides the patients' case history, which contains information such as the patients' profile (gender, age, etc.), history and symptoms. In the Chart space, learners can review patients' vital signs and order lab tests to confirm or disconfirm specific diagnosis. The Library and Consult serve as help-seeking tools. In BioWorld, each diagnosis exercise begins with reviewing a patient's case history, which describes the patient symptoms and other relevant details (which can be highlighted and sent to the evidence table). In solving the patient case, learners review the patient summary and formulate a differential diagnosis (with the help of the Hypothesis Manager tool), along with updating their level of confidence in relation to the most likely diagnosis (via the Belief Meter). The problem-solving trajectory involves identifying relevant symptoms, ordering lab tests to confirm or disconfirm specific diagnosis, seeking help (via the Library and Consult tools), and reasoning about the nature of the underlying disease. The final step involves submitting their final diagnosis, sorting and prioritising evidence and writing a final case summary. After the final diagnosis submission, learners can view and compare their solution to that of an expert, providing learners an opportunity to become cognisant of and reflect on differences, if any, of their solution path from the one of an expert.

Method

The data for this study was collected as part of a larger project that investigated the antecedent factors that led to attention allocation towards feedback in the BioWorld environment (Naismith, 2013). For examining accuracy and efficiency, we use the number of evidence matches of each student with the expert solution, total time taken to solve the case, and number of laboratory tests ordered for a more granular exploration of performance differences. More specifically, accuracy is operationalised as the number of evidence matches with the expert solution, while efficiency is defined by the total time taken to solve the case and the number of laboratory tests ordered. For examining process differences, we leverage the recorded learner-system usage data (i.e., the specific actions of learners) for generating process maps for the three cases. The cases used in the study are referred to by the patient names in the simulated cases. As a measure of case difficulty, we used the results from an earlier study to ascertain the difficulty levels of the various patient cases based on accuracy alone (Gauthier & Lajoie, 2014). The anticipated accuracies (represented as percent accuracies) for the three cases, ordered from easiest to the most difficult, were Amy (94%; easy), Susan Taylor (78%; moderate), and Cynthia (33%; difficult). These cases were developed by a content expert and were subsequently tested and validated by two other content experts (Gauthier & Lajoie, 2014). For the purpose of case development and validation, an expert was defined as someone with "prolonged or intense experience through practice and education in a particular field" (as cited in Ericsson, 2006b, p. 3).

Participants

Participants were recruited through advertisements (on classified websites) and newsletters (via email). Participants consisted of 30 volunteer undergraduate students and were compensated \$20 for completing a 2-hour study session. The participants, 28 medical students and 2 dental students, were registered in the same classes at a large north-eastern Canadian University. The sample comprised 11 men (37%) and 19 women (63%), with an average age of 23 ($SD = 2.60$).

Procedure

Participants completed both a demographics questionnaire and the achievement goal questionnaire. This was followed by a training session to enable participants to learn and practise how to navigate and use the BioWorld system. Participants were also given instructions on how to think aloud while engaging in diagnostic reasoning in BioWorld. After the training session, participants were tasked with solving each of the three cases in BioWorld on an individual basis for a total duration of 2 hours. The three endocrinology cases used in this study were referred to by the patient names: Amy, Cynthia, and Susan Taylor. The correct diagnosis for each was diabetes mellitus (type 1), pheochromocytoma and hyperthyroidism respectively, and were classified as easy, medium and difficult cases. The order of the cases was counterbalanced to mitigate practice and fatigue effects. Upon completion of each case, participants completed a retrospective outcome achievement emotions questionnaire.

Measures

Like many CBLEs, the BioWorld system also records user-system interactions in the log files. In the log files there are three types of performance metrics, namely, diagnostic efficacy (e.g., count of matches with experts), efficiency (e.g., number of tests ordered and time to solve the case) and affect (e.g., confidence). Information recorded in the log file included the attempt identifier (participant and case ID), a time stamp, the BioWorld space (e.g., chart), the specific action taken (e.g., add test: Fasting blood glucose level), and details in relation to the action (e.g., Result: Pre Test value: 9.0 mmol/L; Post Test value: 14.2mmol/L). The focus of this study is only on the logs that contain the user actions recorded by the system while the participants solved the three patient cases. Data mining techniques were used to reveal patterns in the log files that could be used to determine learner strategies. Although participants' think-aloud data were collected, the data analysis of the think-alouds was used to answer different research questions and is not the focus of this paper. For the analyses conducted in this paper, the accuracy and performance variables along with learning behaviours were extracted from the log files.

Performance differences: Accuracy and efficiency

In diagnosing a virtual patient case in BioWorld, diagnostic reasoning is assessed during problem-solving using a novice-expert overlay system (Shute & Zapata-Rivera, 2012). The system log tracks a number of key learner activities such as the evidence items highlighted as relevant, the lab tests ordered and the total time taken to solve a case. The goal of the novice-overlay model in BioWorld is to provide a means to automatically compare learners' solution to an expert solution to obtain an explicit representation of their diagnosis steps and to enable learners to become aware of how their own solution differed from an experts' solution

In order to answer the research question, "are there performance differences across cases?", we extracted diagnostic accuracy and efficiency from the log file database. Accuracy is operationalized as the number of evidence matches with the expert solution (the overlay model does an automatic assessment of the learner-submitted evidences in comparison to an experts' list of evidences). Efficiency is defined by the total time taken to solve the case and the number of laboratory tests ordered. These three performance indices were included as dependent variables (i.e., number of correct matches with the expert solution, number of lab tests ordered and time taken to solve the case) and the three different cases (i.e., Amy, Cynthia, Susan Taylor) were included as the independent variable. The descriptive statistics for the dependent variables are presented in Table 1. To test differences across the three cases, MANOVA analysis was performed.

There was no missing data for any of the variables. A box plot analysis reveals that there were four outliers for the Amy case and two outliers for the Susan Taylor case for the number of correct matches with the expert solution, one outlier for the Susan Taylor case and one outlier for the Amy case for number of lab tests ordered, and one outlier for the Cynthia case and one outlier for the Susan Taylor case for time taken to solve the case. Down weighting outliers to the next most extreme value is considered to be a reasonable way to handle outliers (Chatfield, 2003; Lavrakas, 2008). Therefore, the extreme values were replaced using the next most extreme value within the corresponding case. The outlier adjustments, which are necessary to meet the assumptions of the statistical analyses, strengthened the correlations between the dependent variables but did not change the direction or significance of the omnibus result.

Table 1
Descriptive statistics

	Case name	Mean	Standard deviation	N
Number of correct matches with expert	Amy	6.37	2.30	30
	Cynthia	4.23	1.55	30
	Susan	7.07	1.70	30
	Taylor			
	Total	5.89	2.22	90
Number of lab tests ordered	Amy	9.20	6.02	30
	Cynthia	13.27	5.93	30
	Susan	7.33	4.189	30
	Taylor			
	Total	9.93	5.93	90
Time taken to solve the case	Amy	1550.93	768.34	30
	Cynthia	1945.10	1014.49	30
	Susan	1335.20	646.35	30
	Taylor			
	Total	1610.41	853.42	90

The correlations between the dependent variables were checked; the correlations were low to moderate, which is considered appropriate (Tabachnick & Fidell, 2007) (Table 2). All assumptions for a MANOVA were tested. Bartlett's test of sphericity was significant and therefore the DVs are sufficiently correlated. Box's M (35.794) was significant, $p = .001$; thus the assumption of homogeneity of variance was violated. Consequently, Pillai's trace test is reported as it is robust against this violation (Field, 2009).

Table 2
Correlations between dependent variables

Performance	1	2	3
1. Number of correct matches with expert	–		
2. Number of lab tests ordered	-.167 (-.044)	–	
3. How long to solve the case	.015 (.017)	.462* (.384*)	–

Note. *Correlation is significant at the 0.01 level (2-tailed). Values in brackets are values prior to outlier adjustment.

The results from the MANOVA (Table 3) reveal that there is a significant difference in the pattern of means between cases across participant performance indices, $F(6,172) = 8.056, p < .001$.

Table 3
Multivariate tests

	F	Hypothesis df	Error df	Sig.
Pillai's trace test	8.056 (5.623)	6.000 (6.000)	172.000 (172.000)	.000 (.000)

Note. Values in brackets are values prior to outlier adjustment.

To further understand the nature of these differences, a series of ANOVA post-hoc comparisons were conducted. Levene's test of equality of variance suggests that the assumption of homogeneity of variance was met for all the performance variables. The results of each ANOVA were analysed using the Bonferroni adjusted alpha of .017 (.05/3). ANOVA post-hoc comparisons indicate that there was a significant difference in number of matches with the expert solution between cases, $F(2,87)=18.57, p < .001, \eta^2=.30$ and in how many lab tests were ordered, $F(2,87) = 9.31, p < .001, \eta^2 = .18$. However, there were no significant differences in elapsed time across cases, $F(2,87) = 4.23, p = .018, \eta^2 = .09$. These results suggest there were differences in performance accuracy and efficiency across cases. To understand how the cases differed, Tukey's HSD post-hoc comparisons were conducted. The results suggest that participants had significantly more matches with the expert solution for the easier cases, (Amy and Susan Taylor) cases when compared to the most difficult case (Cynthia) ($M = 6.37 SE = .34, M = 7.07 SE = .34$ and $M = 4.23 SE = .34$, respectively). The results also indicate that participants ordered significantly more

lab tests for the Cynthia case when compared to the Amy and Susan Taylor cases ($M = 13.27$ $SE = .99$, $M = 9.2$ $SE = .99$ and $M = 7.33$ $SE = .99$, respectively).

Discussion: Performance differences

The results suggest that there were distinct performance differences across the three cases participants solved in BioWorld, which supports the phenomenon of case specificity in clinical problem solving. To recapitulate, with regards to accuracy, findings suggest that participants had significantly more matches with the expert solution for the easier cases, Amy and Susan Taylor, when compared to the most difficult case, Cynthia. With regards to efficiency, the results indicate that participants ordered significantly more lab tests for the Cynthia case when compared to the Amy and Susan Taylor cases. Taken together, these two results provide evidence of performance differences across patient cases of varying difficulty levels. For the easier cases (Amy and Susan Taylor), participants had more matches with the expert solution and ordered fewer lab tests. For the most difficult case (Cynthia), participants had the least matches with the expert solution and ordered more lab tests. However, there were no significant differences in elapsed time across cases. From a theoretical perspective, the findings suggest the presence of a case-specificity effect in clinical diagnostic reasoning performance in accordance with prior research in this domain.

Process differences: A process mining approach

Learner modelling is often considered a formidable challenge. Previous studies on clinical reasoning have largely focused on diagnosis correctness. Much less is known, however, about the processes involved, such as the problem-solving paths that learners employ in diagnosing a case. The way learners solve a problem is important for the learning process. Given that the process of clinical reasoning is linked to clinical uncertainty and correctness, clearly, there is a need to better understand the way that learners arrive at a solution. Moving beyond diagnosis correctness, examinations of learner actions and behaviours will be beneficial to our understanding of clinical reasoning.

In recent years, educational data mining has been gaining in popularity, as there has been burgeoning evidence showing the utility of various data mining techniques in addressing scores of educational questions (Baker & Yacef, 2009). Advances in data mining have produced new and powerful means for examining learning data. Access to effective data mining techniques have provided strong incentives and galvanised an interest in the educational community to address a wide spectrum of educational questions. Nevertheless, there are some data mining techniques that have not yet seen widespread use in educational contexts. Process mining, a data mining technique that uses “event data to extract process-related information” (Van der Aalst, 2011, p. 1), has been frequently used for investigating and understanding process data in business contexts and has been a staple of business process research. Little research exists, however, to guide efforts to mine clinical reasoning paths of novice learners. This initial examination leverages the value extracted by and success of process mining, as illustrated in the business process research literature, to make embryonic contributions in this regard.

We make a case for the utility of process mining, a promising method for mining process data, in examining usage data from CBLEs to explain latent mechanisms that mediate diagnostic performance and the case-specificity effect found in the initial phase of this study. CBLEs provide affordances that make it possible to capture and track learner behaviours (Baker & Yacef, 2009) that are difficult to do in other learning environments like traditional classrooms. One of the most common ways to track learners’ actions in CBLEs is via log files; data mining methods can then be used to analyse and investigate various questions about learning material and learner outcomes. In this study, we consider clinical reasoning from a process perspective; thus, drawing on this perspective we argue that the learner actions can essentially be viewed in terms of steps or actions in a process. Process mining can be useful towards modelling learners’ problem-solving paths. According to Rozinat (2015a), the “core functionality of process mining is the automated discovery of process maps by interpreting the sequences of activities” (p. 3). We employ the Disco Miner (2015), based on the framework of Fuzzy Miner, to generate process maps from the BioWorld log files.

In process mining, an event log serves as the starting point. Rozinat (2015b) notes that the minimum requirements for an event log include, but are not limited to, Case ID, Activity, and Timestamp. The BioWorld log data meets the minimum requirements for an event log. To mine the data, the necessary

pre-processing steps were conducted: check for missing values, separate the data for the three cases, and correctly map Case ID, Activity and Timestamp data. The BioWorld log file was imported into the Disco tool and a process model (as a process map) was generated for the data. For each of the three cases, we generated a separate map. Thus, the 30 records extracted from the log files for each case were employed for generating the process maps. The Disco tool generates a frequency-based process map, which enables inspection of the process flow between actions; the “process flows ... are automatically reconstructed (“discovered”) based on the sequence and timing of the activities” (Rozinat, 2015b, p. 52) in the log data. Thus, the generated process map illustrates both the order in which the actions have been performed and the relations between the actions (in terms of directionality). In the process map, the triangle symbol and the stop symbol represent the start and end of the process respectively. Each specific action is housed in a box and an arrow marks the process flow between actions. The numbers marked juxtaposed on the arrows (thicker arrows associated with higher frequencies) and in the boxes (different colours associated with different frequencies) are the absolute frequencies for the transitions and the instances of the actions respectively. The Disco tool allows control of the level of detail presented in the generated maps via two slider controls: Activities and Path. The slider values can be set between 0% and 100%: setting the sliders at 0% shows only the most frequent actions and at 100% all the actions are revealed. Setting the two sliders to a low value results in the most frequent activities and paths. To allow both interpretability and focus, we decided to set low values for both the slider values, thus giving the most frequent activities and paths for meaningful interpretation.

Being exploratory, the goal of this preliminary analysis was to present process mining as an approach in knowing and understanding learner behaviours; thus, the intention of the findings is to present a global discussion on the utility of process mining. This exercise in general has import for understanding learner behaviours in clinical reasoning.

Findings

Given the performance differences across the three clinical cases, it is of interest to ascertain whether there are also differences in the actions taken to arrive at a diagnosis. In this section, we present the major themes (highlighting points of convergence and divergence) that emerged from the process maps for the three cases.

Amy

The approach employed in this case (Figure 2) begins with adding evidence, followed by two distinct blocks of actions. The first block revolves around the formulation of hypotheses (hypotheses formulation block), which begins with adding hypothesis and ends with the final prioritisation and submission of summary. After adding hypothesis, the most likely actions include linking evidence and changing hypothesis conviction. The second block includes two sets of behaviours: ordering lab tests and seeking help (test and help block). Here the relationship between lab tests and help-seeking assumes a reciprocal relationship: ordering of lab tests is followed by help-seeking, and after seeking help, learners tend to order lab tests. For the tail end of the process, a similar pattern of actions was seen across all three cases (prioritisation, categorisation and so on); a plausible explanation for this occurrence could be due to the more structured nature of the BioWorld interface.

Susan Taylor

The approach employed in this case (Figure 3) is similar to the Amy case in that the solution begins with adding evidence, followed by two distinct blocks of actions. The first block revolves around the formulation of hypotheses (hypotheses formulation block), which begins with adding hypothesis and ends with the final prioritisation and submission of summary. Unlike the Amy case, here, after adding hypothesis, the follow-on action is changing hypothesis conviction. The second block includes two sets of behaviours: ordering lab tests and seeking help (test and help block). Here, similar to the Amy case, the same reciprocal relationship between lab tests and help-seeking can be seen. As pointed out earlier, for the tail end of the process, a similar pattern of actions was seen across all three cases (prioritisation, categorisation and so on).

Cynthia

Similar to the Amy and Susan Taylor cases, the starting point for the Cynthia case (Figure 4) is also adding evidence. Furthermore, similarly, adding evidence is followed by two loops of actions. In contrast

to the Amy and Susan Taylor cases, however, the two loops assume different compositions compared to the Amy and Cynthia cases. The first loop revolves around the formulation of hypotheses (hypotheses formulation block), which begins with adding hypothesis and ends with the final prioritisation and submission of summary. However, the hypothesis submission, prioritisation and summary submission activities are not unique to this hypotheses formulation block. Here, the aforementioned activities overlap with the second block that includes, but is not limited to, ordering lab tests and seeking help. Moreover, the reciprocal relationship seen between lab tests and help-seeking in the Amy and Susan Taylor cases is absent in the Cynthia case. For the tail end of the process, a similar pattern of actions was seen across all three cases (prioritisation, categorisation and so on).

Overall, the process maps generated for the three clinical cases yielded some interesting insights into process similarities and differences across the three cases. In all three cases, the starting point for the solution is adding evidence. From adding evidence, the two most likely actions are adding hypothesis and ordering lab tests. An interesting behaviour pattern is seen across the three cases: help-seeking is a common subsequent action to ordering lab tests. For the easier cases (Amy and Susan Taylor), there seems to be a reciprocal relationship between ordering lab tests and help-seeking. This is suggestive that either the learners tend to seek help to understand the consequences of the lab test result or they find information in the library that leads them to order new lab tests. However, this particular relationship did not occur in the more difficult Cynthia case; the reason for this is not known and this might be an interesting examination in future studies.

The examination of the process model across different cases suggests important differences in steps taken to obtain a final solution. In the case of Cynthia, stating a hypothesis and ordering a lab test most often preceded library searches. However, in both the Amy and Susan Taylor cases, which are easiest to solve, learners searched the library only after ordering a lab test, but not following the selection of a hypothesis in the hypothesis management panel. Again, this finding is suggestive in terms of the design of feedback provided in the context of BioWorld. This may suggest that the feedback delivered in more complex cases should elaborate on the requisite declarative knowledge in relation to both the underlying physiological processes that characterise a disease and the relevant inferences derived from information gained from lab tests that may explain abnormalities in such processes. However, the cases with the lowest levels of difficulty require feedback that clarifies the meaning of the lab tests, without a detailed elaboration as to the physiological underpinnings as these are more commonly known to learners. As to the interpretation of the later stages of the process model, a similar pattern of activities was found across all three cases (prioritisation, categorisation and so on) due to the more structured nature of the BioWorld interface.

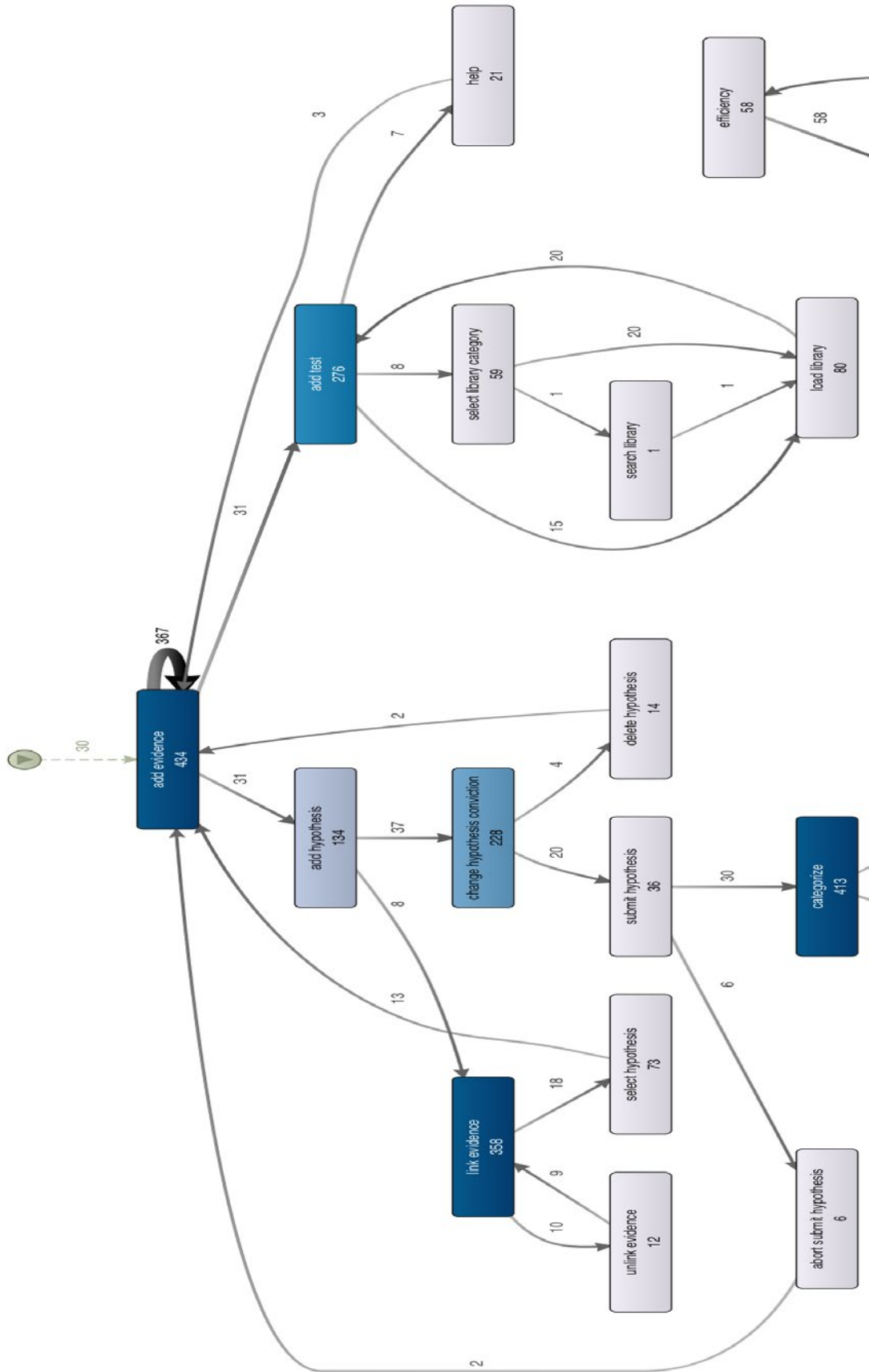


Figure 2. Process model – Amy

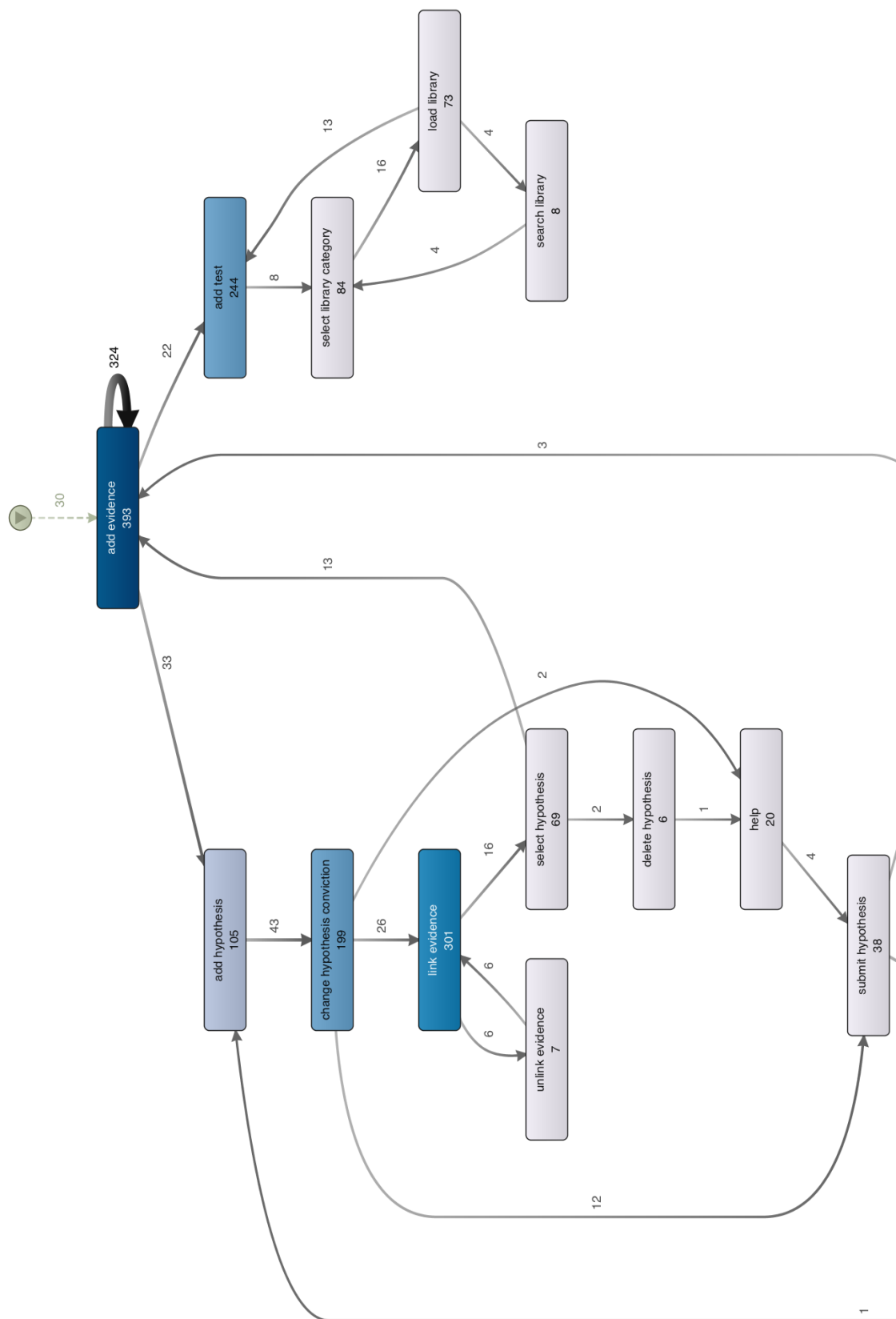


Figure 3. Process model – Susan Taylor

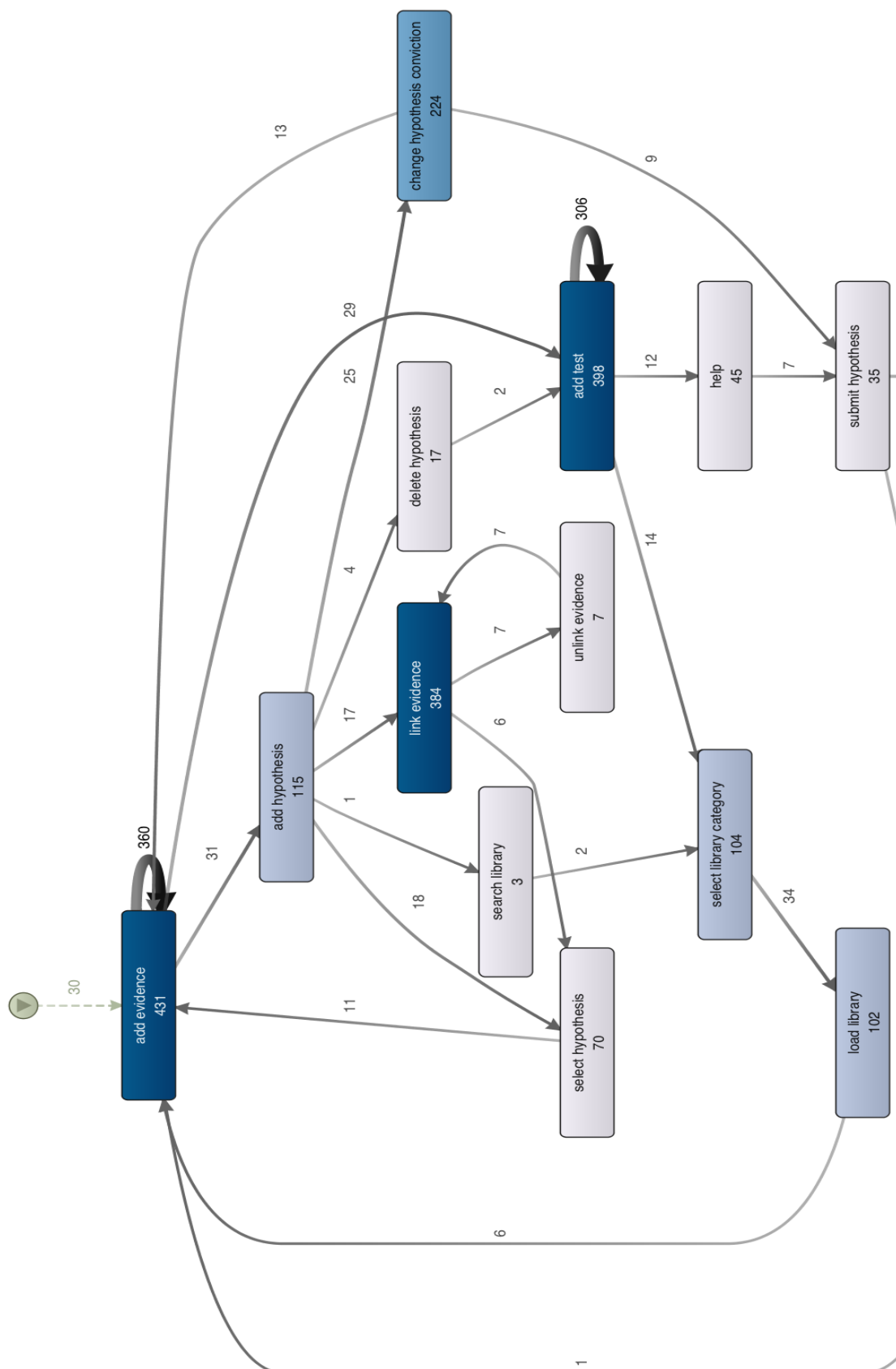


Figure 4. Process model – Cynthia

Discussion: Process differences

The process models generated in this study suggest that the actions taken to solve each case were different. This finding further supports the notion of case specificity in clinical problem solving. The process model for the Cynthia case consisted of more connections to and more cycles between nodes when compared to the process model for the Amy and Susan Taylor cases. This observation illustrates that participants arrived at a final diagnosis through a more convoluted manner for the Cynthia case when compared to the Amy and Susan Taylor cases. The complexity of the process model for the Cynthia case may be a reflection of the difficulty level of the case. In addition, the process models reveals that participants consulted the library when adding hypotheses for only the Cynthia case suggesting that participants consulted the library in order to identify possible diagnoses.

The process mining approach employed in this study represents a novel way of investigating learner behaviours and represents our ongoing efforts to examine and understand learner behaviours so as to provide a more coherent picture about clinical reasoning. The findings substantiate process mining as a useful approach for tracing and uncovering learner behaviours; in particular, the process mining approach identified different learner profiles based on case complexity. Furthermore, our research adds to the growing area of literature on leveraging data mining techniques for improving education by identifying individual differences in problem solving.

Conclusion

Our research supports earlier empirical findings that found case difficulty and case specificity influence clinical reasoning processes and performance (Fitzgerald et al., 1994; van der Vleuten & Swanson, 1990). In diagnosing the three endocrinology clinical cases in BioWorld, our findings suggest that participants had significantly more matches with the expert solution for the easier cases and ordered significantly more lab tests for the most difficult case. However, there were no significant differences in elapsed time across cases. From a theoretical perspective, the findings suggest the presence of a case- effect in clinical diagnostic reasoning performance in accordance with prior research in this domain. Taken together, the findings shed light on the differences across clinical cases and the performance elements involved in diagnosing a clinical case.

From an instructional technology perspective, the results have implications for the design of features and functionalities in CBLEs to support clinical reasoning and medical education in general. Nendaz and Perrier (2012) note, “the majority of cognitive errors are not related to knowledge deficiency but to flaws in data collection, data integration, and data verification that may lead to premature diagnostic closure” (p. 1). Understanding the problem-solving paths of learners can be crucial in getting learners to be cognisant of their errors or prevent such errors entirely. Our research helps identify the steps learners take to arrive at a diagnosis. Viewing and examining diagnostic reasoning as a process through mining the learner-system usage data from a process perspective can lead to meaningful insights. Furthermore, such an approach may also lead to intriguing possibilities for system improvements. According to Shute and Zapata-Rivera (2012), the particular steps encompassed by our mining exercise include the first two steps of effective learner modelling, namely, capturing information about the learner and analysing learner interactions. We determined that the ideal sequence of cases that should be delivered in BioWorld should consider case complexity. In this way, learners would benefit more by beginning to solve the easiest cases and making the transition to the most complex case upon showing mastery of the prerequisite skills. These patterns can be detected through the use of algorithms embedded within the system, thereby showing the benefits of process mining as a means to adaptively sequence content in intelligent tutoring systems in the medical domain. The process mining exercise conducted sets the stage for new ways of uncovering and understanding what is arguably a crucial aspect of learning, learner behaviours in CBLEs. Moreover, our work in this direction contributes to a growing literature shedding light on the use of data mining and learning analytics in educational research.

Taken together, the findings from the performance differences analysis and process modelling lend support to the phenomenon of case specificity in clinical reasoning and extend previous research. The results also provide some support for the view that learners select different clinical reasoning strategies based on case complexity and that problem-solving strategies might influence performance. These results have implications for the design of features and functionalities in CBLEs to support clinical reasoning

and problem solving in general. For example, since solving difficult cases is associated with a more circuitous approach to solving the problem and less effective reasoning strategies, students might require more scaffolding to guide them towards effective strategies when solving difficult problems. One challenge is that case specificity can cause certain cases to be more difficult for some learners and easier for others. Thus, a generalised approach based on problem characteristics alone is not appropriate. To individualise the support provided to each learner, process modelling could be used to evaluate if the actions taken by the learner converge with pre-defined user profiles classified according to strategy effectiveness where effective problem-solving strategies are associated with good performance. Using these templates, the CBLE could respond promptly when users begin to engage in ineffective problem-solving strategies. Thus, an important future direction is to use process modelling techniques to identify common user profiles and categorise them according to effectiveness.

We acknowledge that the work reported here has limitations and scope for improvements. First, like most empirical studies, increasing the sample size can strengthen the study; with an increased sample size, more detailed empirical analyses can be performed. Second, this study was conducted with a single cohort of medical students; further work is needed to examine whether these results generalise across cohorts and levels of education. Third, our analyses were limited to three endocrinology cases; but there is room to expand the study to include a wider array of cases to mitigate issues of generalisability and context specificity. Finally, we limited our analyses to a small number of accuracy and efficiency variables to ease the analysis and interpretation of the results. We plan to conduct an expanded analysis to establish a holistic picture of performance differences across cases in clinical reasoning. Furthermore, we took an exploratory approach in our use of process mining to examine learner behaviours; as a first pass, we only considered the instances of actions and links between actions. Along with the instances, process mining also affords a way to consider the time factor in transitions between actions. For future analyses, we plan to consider the time factor as well.

Several extensions of the present effort deserve to be addressed in future research. Individuals differ in their prior knowledge and their use of such knowledge to solve cases. Those with more prior knowledge have a richer schema to guide their pattern recognition during case resolution and the processes that they use to solve such cases may be more strategic. These individual differences will be most noticeable on the more difficult cases. For starters, it would be interesting to explore how students' perception of case complexity differs based on diagnostic strategy. Another line of extension of the present study could consider and address how novices differ from experts in terms of the ways that the two groups diagnose clinical cases.

These findings shed light on and enrich our understanding of clinical reasoning in simulated cases. The findings from this study will inform our ongoing efforts to improve learner modelling (Doleck, Basnet, Poitras, & Lajoie, 2015) and system design through the design of cognitive and metacognitive tools (Poitras, Lajoie, Doleck, & Jarrell, 2016; Lajoie et al., 2013). Furthermore, the examples and findings offered in this paper represent a template for other researchers and practitioners who are working towards improving learner modelling and the design and modification of learning systems.

References

- Anderson, J., Corbett, A., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*(2), 167–207. doi:10.1207/s15327809jls0402_2
- Baker, R.S.J.d., & Yacef, K. (2009) The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17. Retrieved from <http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/8/2>
- Beal, C. R., Walles, R., Arroyo, I., & Woolf, B. P. (2007). On-line tutoring for math achievement testing: A controlled evaluation. *Journal of Interactive Online Learning, 6* (1), 43–55. Retrieved from <http://www.ncolr.org/jiol/issues/pdf/6.1.4.pdf>
- Chatfield, C. (2003). *The analysis of time series: An introduction*. Boca Raton, FL: Chapman & Hall/CRC.
- Delany, C., & Golding, C. (2014). Teaching clinical reasoning by making thinking visible: An action research project with allied health clinical educators. *BMC Medical Education, 14*(1), 20. doi:10.1186/1472-6920-14-20
- Disco Miner [Computer software]. (2015). Retrieved from <http://fluxicon.com/disco/>

- Doleck, T., Basnet, R., Poitras, E., & Lajoie, S. (2015). Mining learner–system interaction data: Implications for modeling learner behaviors and improving overlay models. *Journal of Computers in Education*, 2(4), 421–447. <http://dx.doi.org/10.1007/s40692-015-0040-3>
- Eva, K. (2005). What every teacher needs to know about clinical reasoning. *Medical Education*, 39(1), 98–106. [doi:10.1111/j.1365-2929.2004.01972.x](https://doi.org/10.1111/j.1365-2929.2004.01972.x)
- Ericsson, K. A. (2006a). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). Cambridge: Cambridge University Press.
- Ericsson, K. A. (2006b). An introduction to *Cambridge handbook of expertise and expert performance*: Its development, organization, and content. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 3–20). Cambridge: Cambridge University Press.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
- Fitzgerald, J., Wolf, F., Davis, W., Barclay, M., Bozynski, M., Chamberlain, K., ... Zelenock, G. B. (1994). A preliminary study of the impact of case specificity on computer-based assessment of medical student clinical performance. *Evaluation & The Health Professions*, 17(3), 307–321. [doi:10.1177/016327879401700304](https://doi.org/10.1177/016327879401700304)
- Gauthier, G., & Lajoie, S.P. (2014). Do expert clinical teachers have a shared understanding of what constitutes a competent reasoning performance in case-based teaching? *Instructional Science*, 42(4), 579–594. [doi:10.1007/s11251-013-9290-5](https://doi.org/10.1007/s11251-013-9290-5)
- Gigante, J. (2013). Teaching clinical reasoning skills to help your learners “get” the diagnosis. *Pediatrics & Therapeutics*, 2(4). [doi:10.4172/2161-0665.1000e121](https://doi.org/10.4172/2161-0665.1000e121)
- Higgs, J., & Jones, M. A. (2000). Clinical reasoning in the health professions. In J. Higgs & M. A. Jones (Eds.), *Clinical reasoning in the health professions* (pp. 3–32). Oxford: Butterworth-Heinemann.
- Lajoie, S. (2003). Transitions and trajectories for studies of expertise. *Educational Researcher*, 32(8), 21–25. [doi:10.3102/0013189x032008021](https://doi.org/10.3102/0013189x032008021)
- Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from Avionics and Medicine. In K. A. Ericsson (Ed.), *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 61–83). Cambridge: Cambridge University Press.
- Lajoie, S. P., & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 803–821). Mahwah, NJ: Erlbaum. .
- Lajoie, S. P., Naismith, L., Hong, Y. J., Poitras, E., Cruz-Panesso, I., Ranellucci, J., Mamane, S., & Wiseman, J. (2013). Technology rich tools to support self-regulated learning and performance in medicine. In R. Azevedo & V. Alevan (Eds.), *International handbook of metacognition and learning technologies* (pp. 229–242). New York, NY: Springer.
- Lajoie, S. P., Poitras, E. G., Doleck, T., & Jarrell, A. (2015). Modeling metacognitive activities in medical problem-solving with BioWorld. In A. Peña-Ayala (Ed.), *Metacognition: Fundamentals, applications, and trends. A profile of the current state-of-the-art* (pp. 323–343). Cham: Springer.
- Lavrakas, P. (2008). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage.
- Lee, A., Joynt, G. M., Lee, A. K., Ho, A. M., Groves, M., Vlantis, A. C., ... Aun, C. S. (2010). Using illness scripts to teach clinical reasoning skills to medical students. *Family Medicine*, 42(4), 255–261. <http://www.stfm.org/fmhub/fm2010/April/Anna255.pdf>
- Levett-Jones, T., Hoffman, K., Dempsey, J., Jeong, S., Noble, D., Norton, C., ... Hickey, N. (2010). The ‘five rights’ of clinical reasoning: An educational model to enhance nursing students’ ability to identify and manage clinically ‘at risk’ patients. *Nurse Education Today*, 30(6), 515–520. [doi:10.1016/j.nedt.2009.10.020](https://doi.org/10.1016/j.nedt.2009.10.020)
- Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., William, W. C., Stylianides, G. J., & Koedinger, K. R. (2013). Cognitive anatomy of tutor learning: Lessons learned with SimStudent. *Journal of Educational Psychology*, 105(4), 1152–1163. [doi:10.1037/a0031955](https://doi.org/10.1037/a0031955)
- Melicow, M. M. (1977). One hundred cases of pheochromocytoma (107 tumors) at the Columbia - Presbyterian Medical Center, 1926–1976. A clinicopathological analysis. *Cancer*, 40(5), 1987–2004. [doi:10.1002/1097-0142\(197711\)40:5<1987::aid-cnrcr2820400502>3.0.co;2-r](https://doi.org/10.1002/1097-0142(197711)40:5<1987::aid-cnrcr2820400502>3.0.co;2-r)
- Naismith, L. (2013). *Examining motivational and emotional influences on medical students’ attention to feedback in a technology-rich environment for learning clinical reasoning* (Unpublished doctoral dissertation). Retrieved from <http://digitool.library.mcgill.ca/R/?func=dbin-jump->

- [full&object_id=117101&local_base=GEN01-MCG02](#)
- Nendaz, M., & Perrier, A. (2012). Diagnostic errors and flaws in clinical reasoning: Mechanisms and prevention in practice. *Swiss Medical Weekly*. doi:10.4414/smw.2012.13706
- Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem-solving by medical students: A cross-sectional and longitudinal analysis. *Medical Education*, 15(5), 315–322. doi:10.1111/j.1365-2923.1981.tb02495.x
- Noll, E., Key, A., & Jensen, G. (2001). Clinical reasoning of an experienced physiotherapist: insight into clinical decision-making regarding low back pain. *Physiotherapy Research International*, 6(1), 40–51. doi:10.1002/pri.212
- Norman, G. (2005). Research in clinical reasoning: Past history and current trends. *Medical Education*, 39(4), 418–427. doi:10.1111/j.1365-2929.2005.02127.x
- Norman, G. R., Tugwell, P., Feightner, J. W., Muzzin, L. J. & Jacoby, L. L. (1985). Knowledge and clinical problem-solving. *Medical Education*, 19(5), 344–356. doi:10.1111/j.1365-2923.1985.tb01336.x
- Poitras, E., Lajoie, S., & Hong, Y. (2011). The design of technology-rich learning environments as metacognitive tools in history education. *Instructional Science*, 40(6), 1033–1061. doi:10.1007/s11251-011-9194-1
- Poitras, E. G., Lajoie, S. P., Doleck, T., & Jarrell, A. (2016). Subgroup discovery with user interaction data: An empirically guided approach to improving intelligent tutoring systems. *Educational Technology and Society*, 19(2), 204–214. http://www.ifets.info/journals/19_2/15.pdf
- Rozinat, A. (2015a). *Disco tour* (pp. 1–13). Retrieved from <http://fluxicon.com/disco/files/Disco-Tour.pdf>
- Rozinat, A. (2015b). *Disco user's guide* (pp. 1–13). Retrieved from <http://fluxicon.com/disco/files/Disco-User-Guide.pdf>
- Ryan, S., & Higgs, J. (2008). Teaching and learning clinical reasoning. In J. Higgs, M. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (pp. 379–387). Amsterdam: Elsevier.
- Schuwirth, L. W., & Van der Vleuten, C. P. (2003). The use of clinical simulations in assessment. *Medical Education*, 37(s1), 65–71. doi:10.1046/j.1365-2923.37.s1.8.x
- Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach, & A. Lesgold (Eds.), *Adaptive technologies for training and education* (pp. 2–27). New York, NY: Cambridge University Press.
- Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics*. Boston, MA: Pearson/Allyn & Bacon.
- Van der Aalst, W. (2011). *Process mining: Discovery, conformance and enhancement of business processes*. Berlin: Springer-Verlag.
- van der Vleuten, C., & Swanson, D. (1990). Assessment of clinical skills with standardized patients: State of the art. *Teaching And Learning In Medicine*, 2(2), 58–76. doi:10.1080/10401339009539432
- Vanlehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., ... Wintersgill, M. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education*, 15(3), 147–204. <http://content.iospress.com/articles/international-journal-of-artificial-intelligence-in-education/jai15-3-02>
- Wimmers, P. F., Splinter, T. A., Hancock, G. R., & Schmidt, H. G. (2006). Clinical competence: General ability or case-specific? *Advances in Health Sciences Education*, 12(3), 299–314. doi:10.1007/s10459-006-9002-x

Corresponding author: Tenzin Doleck, tenzin.doleck@mail.mcgill.ca

Australasian Journal of Educational Technology © 2016.

Please cite as: Doleck, T., Jarrell, A., Poitras, E. G., Chaouachi, M., & Lajoie, S. (2016). A tale of three cases: Examining accuracy, efficiency, and process differences in diagnosing virtual patient cases. *Australasian Journal of Educational Technology*, 36(5), 61-76. <https://doi.org/10.14742/ajet.2759>