



Using Sequential Pattern Mining to Understand How Students Use Guidance While Doing Scientific Calculations

Sjors Verstege¹ · Yingbin Zhang^{2,3} · Peter Wierenga¹ · Luc Paquette² · Julia Diederer¹

Accepted: 18 July 2023
© The Author(s) 2023

Abstract

In natural science education, experiments often lead to the collection of raw data that need to be processed into results by doing calculations. Teaching students how to approach such calculations can be done using digital learning materials that provide guidance. The goal of this study was to investigate students' behaviour regarding the use of guidance while doing scientific multi-step calculations, and to relate this behaviour to learning. Sequential pattern mining was used to i) identify students' behaviour patterns while doing calculations in an online learning environment, ii) study the relation between use of guidance and success on first attempt at submitting a calculated value, iii) study the relation between students' use of guidance and learning gain, and iv) study the relation between students' use of guidance and prior knowledge. Data showed that all students frequently used the guidance provided in the learning task. Moreover, students who used the option to check their intermediate calculations and students who studied worked examples were more likely to successfully complete the calculation on their first attempt than students who did not use this guidance. Guidance in the form of hints was used frequently. However, using the hints did not result in more success at the first attempt. We did not find a relation between learning gain and use of guidance, but we did find a trend that students with a low prior knowledge used more guidance compared to students with a high prior knowledge. The results of this study imply that providing hints and intermediate calculations is of utmost importance for students to independently complete scientific multi-step calculations.

✉ Julia Diederer
julia.diederer@wur.nl

Yingbin Zhang
zyingbin@m.scnu.edu.cn

Luc Paquette
lpaq@illinois.edu

¹ Laboratory of Food Chemistry, Wageningen University, Bornse Weiland 9, 6708 WG Wageningen, The Netherlands

² Department of Curriculum and Instruction, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

³ Institute of Artificial Intelligence in Education, South China Normal University, Guangzhou, China

Keywords Calculations · Guidance · e-learning materials · Sequential pattern mining

1 Introduction

In natural science education, experiments often lead to the collection of raw data, which needs to be processed into results by doing calculations. As such, teaching students how to approach calculations, or in more general terms, teaching students how to solve (calculation) problems, is an important aspect of natural science education. To help students overcome challenges when doing multi-step calculations, they can be provided with an online learning material that includes guidance in the form of procedural information (Van Merriënboer & Kirschner, 2018), such as hints and feedback, so students are able to solve the problems without the help of a teacher. An advantage of the use of online learning materials over traditional instruction is that guidance can be given to many students simultaneously and just-in-time (Diederer et al., 2003). Presenting procedural information just-in-time, when that information is relevant for the task that the student is working on, activates working memory during the task and helps students to construct appropriate cognitive rules in their long-term memory (Hausmann et al., 2013; Van Merriënboer & Kirschner, 2018).

In traditional classrooms, the teacher can observe how students receive their personal guidance, while in online learning materials, this feedback from students is less visible. The shift in guidance from teachers to the online learning material raises questions such as: How do students use the guidance in an online environment, is the use of guidance influenced by their prior knowledge in doing calculations, and does the use of guidance have an effect on learning outcomes?

1.1 Guidance in an Online Environment

An online environment for practising calculations can be designed in such a way that the students are enabled to complete the tasks independently. To ensure that this is indeed possible, guidance needs to be built in (Aleven et al., 2003; Bokhove & Drijvers, 2012; Olsson & Granberg, 2019). For calculations in an online environment the following three design principles are in place with regards to guidance (Verstege et al., 2021): (1) providing formative feedback based on student input (e.g. feedback for several possible answers (Bokhove & Drijvers, 2012), and visual cues on students' progress), (2) providing access to procedural hints that guide the students' thinking process (Koedinger & Aleven, 2007; Roll et al., 2011), and (3) providing the opportunity for students to check their intermediate calculations. These design principles are based on the principle that feedback, especially formative feedback, has a large effect on student learning (Hattie, 2009; Hattie & Timperley, 2007; Neelen & Kirschner, 2020; Van der Kleij, et al., 2015).

There is ample proof that providing guidance in an online environment is indeed effective. A review article published in 2003 described how on-demand help in interactive learning environments, such as hints and hyperlinks to information, substantially improves learning processes and outcomes if students use the help (Aleven et al., 2003). In the past decade, studies have been performed on the effect of automated built-in feedback and guidance in online learning environments. A review article by Anderson et al. (2021) on the effect of automatic feedback in online learning environments on student performance found

that out of the 63 reviewed articles 65% showed a positive effect on student performance, 0% showed a negative effect and 35% showed no evidence of an effect. In this review no specific automatic feedback generation system was selected. As such, the feedback systems discussed in the 63 reviewed articles were designed for diverse domains, and only 3 articles dealt with feedback in mathematics, the domain that is most related to scientific calculations. Olsson and Granberg (2019) investigated the effect of guidance generated by dynamic software on the construction of a mathematical rule. They found that students who were provided with guidance were more successful in constructing the rule (68% of the students) than the students working with the unguided task (33% of the students). On the other hand, students who successfully solved the unguided task had significantly higher grades than students who successfully solved the guided task. In the study of Bokhove and Drijvers (2012), which investigated automatic feedback in a digital learning environment for algebra, it was suggested that addition of buttons for hints and solutions results in fewer attempts. They concluded that “relevant feedback fosters algebra learning by decreasing the number of attempts needed for a task while improving the scores”.

Students in higher education can differ strongly in their prior knowledge and skills related to performing calculations. Students’ prior knowledge was historically found to be the best predictor of subsequent academic success (Zimmerman, 2013). When combining this finding with the expertise reversal effect (Kalyuga, 2007), which states that “students with little prior knowledge benefit from assistance and students with higher prior knowledge benefit from minimal assistance”, we predict that students with little prior knowledge will benefit most from the available guidance in an online learning environment, whereas students with a higher prior knowledge benefit from less guidance.

Brenner et al. (2017) performed a study to investigate the use of assistance in an online science inquiry environment in relation to prior knowledge. Although they expected to find the previously mentioned expertise reversal effect, their data shows that “although prior knowledge plays a part in determining how a student might benefit from assistance, it cannot be relied upon solely to guide what sort of assistance might help most.” A study by Zhou et al. (2021) reports that adult students with less prior knowledge request significantly more guidance in the form of correct answers, resulting in poorer performance when solving similar problems at a later time. Likewise, Zhang et al. (2022) found that when giving students hints, those with more prior knowledge were more likely to use the hints. Hausmann et al. (2013) shows that hints help students to master skills more quickly when they have sufficient understanding of the hint.

Further research is needed to get a better understanding of how students make use of guidance provided in an online learning environment, how this is influenced by their prior knowledge and how this might affect learning. One way to answer these questions is to apply learning analytics to investigate how students make use of the learning materials (e.g. Kew & Tasir, 2022; Law & Liang, 2020; Lockyer et al., 2013; Mangaroska & Gianakos, 2019; Wiley et al., 2020). Learning analytics involves the collection, analysis, and interpretation of (log) data of students’ interactions with an environment, for purposes of understanding and optimizing learning in the environment (Gašević et al., 2015; Greller & Drachsler, 2012; Hernández-Leo et al., 2019; Siemens & Long, 2011).

1.2 Sequential Pattern Mining and its Application in Learning Process Data

Sequential pattern mining (SPM) is a pattern recognition technique that aims at discovering sequential patterns in a dataset containing multiple sequences of items (Agrawal &

Srikant, 1995). It is a common method in the field of learning analytics (Baker, 2010). SPM has been applied to data from a wide range of educational settings, such as Massive Open Online Courses, Intelligent Tutoring Systems, and educational games. Within these contexts, SPM has been used to highlight similarities and differences in engagement and learning across students with regards to self-regulated learning (Kinnebrew et al., 2014; Taub & Azevedo, 2018; Taub et al., 2018), inquiry-based learning (Chen & Wang, 2020; Kang et al., 2017), and collaborative learning (Zheng et al., 2019; Zhu et al., 2019).

Researchers have applied SPM to educational data for a variety of purposes, such as mining learning behaviours, enriching educational theories, and evaluating the efficacy of interventions. Frequent sequential patterns generated by SPM from learning process data, such as event logs, may represent common behaviour patterns across students (Zhou et al., 2010). These behaviour patterns may reveal the order of their activities within a learning environment and provide insights for improving the learning design (Mirzaei & Sahebi, 2019). For example, Kang et al. (2017) applied SPM to gameplay logs in *Alien Rescue*, a serious game for teaching middle school students scientific problem-solving skills, to investigate how students' playing behaviours might vary on different days.

SPM has been used to investigate how interpretation of mined sequential patterns relate to educational theories. For instance, Taub and Azevedo (2018) applied SPM to investigate how self-regulated learning behaviours and emotions influenced learning and gameplay within *Crystal Island*, an educational game that teaches scientific inquiry skills and microbiology. They linked gameplay behaviours, such as hypothesis testing, to metacognitive monitoring strategies aligned with self-regulated learning theories.

Studies have also explored whether sequential learning behaviour patterns could capture the effect of interventions. For example, Wong et al. (2019) designed weekly prompt videos to facilitate students to think about their plan, monitoring, and reflection on learning in a Coursera course. They compared students who watched at least one prompt video (prompt viewers) and those not watching any prompt video (non-viewers). The group of prompt viewers shared more sequential behaviour patterns than non-viewers and tended to watch videos in the order that instructors planned.

1.3 Goal and Research Questions

The goal of this study was to investigate students' behaviour regarding the use of guidance while doing calculations, and to relate this behaviour to learning. The following research questions were formulated:

1. What behaviour patterns can be observed while students are doing calculations in an online environment?
2. Is there a relation between use of guidance and success on first attempt at submitting a calculated value?
3. Is there a relation between students' use of guidance and learning gain?
4. Is there a relation between students' use of guidance and prior knowledge?

2 Method

2.1 Participants

All students participating in the study ($N=81$) were enrolled in a master level course in enzymology (168 study hours) at a university in the Netherlands. All students had previously obtained a bachelor level degree in natural sciences, approximately half of them from the Netherlands (53%), and the other half from universities outside the Netherlands (47%). All students were enrolled in a master programme, either Food Technology (81%) or Biotechnology (19%). The students were 22.9 years old ($SD=1.9$) on average. About 64% of the students were female, while the rest were male.

4. Calculations !

Using the results you obtained, you can calculate the **enzyme activity**, **residual activity**, **inactivation constant** (k_d), **half-lifetime** ($t_{1/2}$) and **activation energy** (E_a). Estimated time for data processing: 40 minutes.

In case you need help in processing your data, several hints are provided:

▶ [Click for hint #1](#)

Fill in the inactivation constants (k_d) and half-lifetimes ($t_{1/2}$) of **α -amylase** in the extract of 4 days germinated ($t=4$) wheat for each temperature.

T (°C)	Inactivation constant (k_d)	Half-lifetime ($t_{1/2}$)
20	<input type="text"/> day ⁻¹	<input type="text"/> days
30	<input type="text"/> day ⁻¹	<input type="text"/> days
40	<input type="text"/> day ⁻¹	<input type="text"/> days
50	<input type="text"/> day ⁻¹	<input type="text"/> days

▶ [Check your intermediate calculations \(optional\)](#)

[Submit Answer](#) [Reset](#)

In case you need help in processing your data, several hints are provided:

▶ [Click for hint #1](#)

Calculate the average of the 3 replicates, and **subtract the (average) blank** (BSA 0.0g/L) value from all the other average values.

▶ [Click for hint #2](#)

Plot the **calibration curve** (BSA concentration on the x-axis, absorbance on the y-axis).

▶ [Click for hint #3](#)

T (°C)	Inactivation constant (k_d)	Half-lifetime ($t_{1/2}$)
20	0.025 ✓ day ⁻¹	28.1 ✓ days
30	0.037 ✓ day ⁻¹	18.9 ✓ days
40	0.101 ✓ day ⁻¹	6.8 ✓ days
50	0.188 ✓ day ⁻¹	37 ✗ days

Feedback

Incorrect.

You can use the **available hints** (above the questions), and **check your intermediate calculations**.

[Submit Answer](#) [Reset](#)

▶ [Check your intermediate calculations \(optional\)](#)

Formula calibration curve: $y = 0.27 \checkmark x + 0 \checkmark$.

Protein concentration **1 day germinated ($t=1$)** (not yet corrected for dilution) = ✗ mg/mL.

Protein concentration **4 days germinated ($t=4$)** (not yet corrected for dilution) = mg/mL.

Feedback

Incorrect.

You can use the **available hints** (above the questions), and **check your intermediate calculations**.

▶ [Please give me the protein concentration for \$t=1\$ \(not yet corrected for dilution\)](#)

Fig. 1 Screenshots of interactive calculation questions. **a** Overview of an interactive calculation question; **b** Procedural hints; **c** Feedback on submitted answers; **d** Example of checking intermediate calculations, with feedback

2.2 Learning Task and Types of Guidance

The learning task used as the context of this study resulted from an educational design research project (Verstege et al., 2021) and was embedded within a larger virtual experiment environment. In the virtual experiment environment students obtain raw data corresponding to chemical methods, and process the raw data into results by means of three types of interactive multi-step calculation questions (Verstege & Diederer, 2019). Each type of calculation question requires students to use a different approach and must be solved multiple times for different datasets. The amount and type of guidance vary between the question types.

Four types of guidance are provided throughout the learning task. First, students can choose to access procedural hints (Fig. 1a). Procedural hints can be opened one at the time, while each hint provides a link to the next hint (3 to 5 hints depending on the difficulty of the calculation). Each hint provides detailed information on the (subsequent) calculation step. All hints together provide the procedural information required to solve the problem. Second, when an answer is submitted, the system will indicate with a green check or a red cross which values are correct, and which values are incorrect (Fig. 1b). Third, students are provided with the option to check their intermediate calculations, which decompose the calculation into smaller parts (Fig. 1c). The system indicates whether each intermediate calculation is correct or incorrect. This functionality can be used to check if the procedure (possibly acquired through the hints) was carried out correctly. When students submitted an incorrect intermediate value, they can choose to request the correct intermediate value from the system. An overview of an interactive calculation question is given in Fig. 1d. Fourth, students can access a worked example.

As mentioned before, the three question types vary in the amount and type of guidance. Type 1 questions are relatively simple: students need one set of calculation steps to process the raw data into the final value. Three hints are available, and one intermediate value can be calculated, but no worked example is provided. Type 2 questions are more difficult: students receive a diverse set of raw data (i.e. calibration values and values for the sample of interest) and need several sets of calculation steps to obtain the final value. Five hints are available and 3 different intermediate values can be calculated, but no worked example is provided. Type 3 questions are most complex: students make several set of calculation steps, need to calculate 2 types of final values for 4 different samples (i.e. different temperatures). There are three hints available, and one intermediate calculation, which asks for 24 values needed to come to the final 8 values. A worked example is available for this question type.

Table 1 Student activities related to the study, with corresponding measurements, measurement output, timing, and duration

Code	Student activity	Measurement	Measurement output	Day*	Duration
A1	Pre-test	Ability to approach calculations	Test score	1	30 min
A2	Calculations	Student behaviour while doing calculations	User log data	1 and 2	4 h
A3	Post-test	Ability to approach calculations	Test score	8	30 min

*Running days relative to the start of the experiment. There were no course activities in between the learning task and the post-test

The design (Verstege & Diederens, 2019) and the effectivity of the learning task were already evaluated in previous studies (Verstege et al., 2021; Verstege et al., 2019) in which it was shown that the virtual experiment environment was perceived as a positive learning experience, contributed to achievement of the intended learning outcomes, and provided sufficient guidance to students to complete the learning task independently.

2.3 Study Design

All student activities related to the study are shown in Table 1. The course activities before the pre-test comprises of lectures and tutorials on diverse enzymes present in food. The calculation that students work on in the virtual experiment environment are not part of the lectures and tutorials. After the post-test students will work on laboratory assignments, applying the calculations they practiced in the virtual experiment environment.

The calculations were embedded in a virtual experiment environment that was implemented in a master level course. Given the pre-requisites for this course, students were expected to have already developed the skill to do scientific calculations to some degree. To measure students' ability to approach calculations, they were asked to complete the pre-test (A1), for which they received a maximum of 30 min. The pre-test consisted of three open questions related to approaching calculations, on which students could score a maximum of 13 points. An example of an open question is: "Describe as detailed as possible all calculation steps you need to do to calculate...". Students voluntarily completed the pre- and post-test (A1 and A3), while they were aware that the anonymized results could be used for research purposes. Most of the students ($N=77$) completed the pre-test. The pre-test scores ($M=5.2$, $SD=2.7$) were used as a measure of students' ability to approach calculations before the learning task.

After the pre-test, students started working in the virtual experiment environment. Most students completed the tasks in the virtual experiment environment in 8 h spread over 2 days. The calculations (A2) were part of the tasks in the virtual experiment environment, for which students needed an average of 4 h to solve. User log data was collected, which is explained in more detail in Sect. 2.4. All students ($N=81$) completed the tasks of the virtual experiment environment (including the calculations) as this was a compulsory course activity. One week after the pre-test, just before the start of the next course activity, students were asked to complete the post-test (A3), for which they received a maximum of 30 min. The post-test was identical to the pre-test and was completed by 79 students. Students were not informed about the post-test, to prevent targeted studying for it. There were no course activities in between doing the calculations and the post-test. The post-test scores ($M=8.5$, $SD=2.0$) were used as a measure of students' ability to approach calculations after the learning task.

2.4 Log Data Pre-Processing

Log data of the students problem-solving behaviours related to completing calculations for any of the three question types were recorded automatically by the virtual experiment environment. The logs were labelled with action codes, which are described in Table 2. To distinguish between code representing a single action and code representing consecutive actions, consecutive actions of the same type were condensed and labelled as one action code by adding a suffix (Kinnebrew et al., 2013). The used suffixes are: '+', which

Table 2 Action codes, including condensed actions, with description and corresponding category

Action code	Description	Category
AC	Student submits a calculated value and is provided with feedback that the value is correct	Attempt
AC+	Multiple consecutive AC	Multiple attempts
AI	Student submits a calculated value and is provided with feedback that the value is incorrect	Attempt
AI+2	Two consecutive AI	Multiple attempts
AI+3+	Three or more consecutive AI	Multiple attempts
GH	Student opens a hint	Guidance
GH+	At least two hints opened consecutively	Multiple guidance
GH+all	All hints opened consecutively	Multiple guidance
GNI	Navigate to the section in which intermediate values can be submitted	Guidance
GI	Student submits a value of an intermediate calculation and is provided with feedback whether the value is correct (GI is always preceded by GNI)	Guidance
GI+2	Two consecutive GI	Multiple guidance
GI+3+	Three or more consecutive GI	Multiple guidance
GW	Student accesses a worked example (only available after mistake in GI)	Guidance
GA	After submitting an incorrect intermediate value, the student requests and receives the correct intermediate value	Guidance
GA+2	Two consecutive GA	Multiple guidance
GA+3+	Three or more consecutive GA	Multiple guidance
IFC	After submitting a correct value (AC), or after correctly submitting all intermediate values (GI) and submitting the final calculated value incorrectly (AI), students click a link in the feedback which provides a pdf with the full calculation	Information
IB	Navigate to background information	Information
NAD	Student acquires an excel file with the raw data required to answer a calculation question	Not categorized
NC	Navigate to an interactive calculation question	Not categorized

means two or more consecutive actions, '+2', means: exactly two actions, '+3+' means: three or more. For hints, '+all' means: all available hints. AC actions did not distinguish between AC+, AC+2 and AC+3+ as the number of AC+3+ was very small, instead only the '+' suffix was used. GW and GNI do not have the suffixes because it was impossible to do multiple of these actions consecutively. No suffixes were used for IFC, NC, IB, NAD as, for these actions, the meaning for a single action and multiple consecutive actions is the same.

To be able to answer our research questions, all action codes were assigned a category based on their purpose within the learning environment. All actions in which students submit a value, i.e. an answer to a calculation question, in the VEE were categorized as 'attempt'. All actions in which students make use of guidance were categorized as 'guidance'. All actions providing students with general information were categorized as 'information', for example when students clicked a link to access a pdf document providing details about the required calculations (IFC). In the case of condensed actions, 'multiple' was added to the categories. The actions NC and NAD were not categorized, as these actions are required by the system in specific situations. They were included in the table and in some results as they do provide meaning. For example, the sequence NAD=>NC=>AC indicates that students acquired the dataset, navigated to an interactive calculation question, and submitted a correct value, without using guidance in the meantime. After the pre-processing, a student's log data contained 224.37 action codes on average ($SD=116.07$).

2.5 Analyses

Table 3 provides an overview the analyses for each research question.

The cSPADE algorithm was applied (Zaki, 2000) to find behaviour patterns of students while doing calculations in the online environment (research question 1). cSPADE is an efficient SPM algorithm that has been used in other educational studies (Jiang et al., 2015; Kang et al., 2017; Wong et al., 2019). We implemented cSPADE via the *arulesSequences* package in R (Buchta et al., 2020). The algorithm computed the support value for each sequential pattern, which represents the proportion of students who used the sequential pattern at least one time. We set the minimum support for discovered frequent to a small value, 0.2, since we wanted to uncover as many sequential patterns as possible. Other values were investigated (including 0.1, 0.3, 0.4, and 0.5), but higher minimum support (e.g. 0.4) ignored many relevant patterns, and lower minimum support (e.g. 0.1) generated excessive meaningless patterns. We specified the maximum gap as 1 so that the algorithm would only consider action sequences that exactly matched a sequential pattern exactly as an instance of this pattern. For example, AI followed by GNI further followed by AC was an instance of AI=>GNI=>AC but not an instance of AI=>AC. For each sequential pattern returned by cSPADE, its average occurrence in a student's log data was computed. cSPADE generated 324 sequential patterns whose support were greater than 0.2. We inspected these patterns and removed meaningless patterns (patterns that do not contain more information than their sub patterns). For example, when students first have to navigate to a calculation question, before they can make an attempt to answer it (e.g. NC=>AC). Such a pattern does not contain more information than its sub pattern without the first element (e.g. AC). After removing the meaningless patterns, 88 sequential patterns remained.

Table 3 Overview of the analyses for each research question

Research question	SPM Algorithm	Input data	Statistical test	Independent variable	Dependent variable	Effect size
1	cSPADE	All log data	–	–	–	–
2	Differential SPM	Log data before the first attempt to questions	Fisher exact test	Correct first attempt versus incorrect first attempt	Whether a sequential pattern existed in one's log data (binary)	Risk ratio
3	Differential SPM	All log data	Kruskal Wallis test Mann-Whitney U test	High, medium, and low learning gain	The occurrences of a sequential pattern in one's log data (continuous)	η^2 r
4	Differential SPM	All log data	Kruskal Wallis test Mann-Whitney U test	High, medium, and low pre-test score	The occurrences of a sequential pattern in one's log data (continuous)	η^2 r

We investigated if there was a relation between use of guidance and early success by comparing differences in use of guidance between students whose first attempt was correct versus students whose first attempt was incorrect (research question 2). The learning task of this study involved three types of calculation questions (as described in Sect. 2.2). Each type of calculation question requires students to use a different approach and has to be done multiple times for different datasets. We examined the difference in use of guidance before students' first attempt for each calculation type. Students' log data were split into three subsets, each containing the actions corresponding to each of the question types. For each subset, a student's first attempt to answer a calculation question and all actions after this first attempt were removed. The average length of the remaining sequences for a correct first attempt was 7.68 ($SD=7.23$), 15.53 ($SD=13.25$), and 17.35 ($SD=10.37$) at question type 1, 2, and 3, respectively. That for an incorrect first attempt was 5.59 ($SD=7.39$), 8.32 ($SD=6.96$), and 14.25 ($SD=20.22$) at question type 1, 2, and 3, respectively.

Differential SPM (Kinnebrew et al., 2013) was used to identify sequential patterns used differentially by students who submit the correct answer at their first attempt versus students who submit an incorrect answer at their first attempt across the three subsets of log data. Differential SPM typically achieves its goal through the following steps:

1. Discover frequent sequential pattern candidates within each group through traditional SPM algorithms, such as cSPADE.
2. Compute the occurrences per sequential pattern per student for all candidate patterns identified in the first step.
3. Apply statistical tests, such as t-tests, to identify candidate patterns with significantly different occurrences between groups.

Because the investigated sequences were short (9.33 actions on average), it was rare that a student executed the same pattern more than once. As such, instead of treating the frequency of occurrences of sequential patterns as continuous, we treated them as binary variables: whether a sequential pattern occurred in a student's action sequence or not. The Fisher exact test was used to examine the differences in the binary occurrences between the groups whose first attempts were correct versus incorrect by comparing their support values (He et al., 2019). The p -values were adjusted using the Benjamini and Yekutieli correction to control the false discovery rate due to multiple comparison (Benjamini & Yekutieli, 2001). The magnitude of the difference in a sequential pattern was characterized by the risk ratio $= (n11 / (n11 + n12)) / (n21 / (n21 + n22))$. $n11$ is the number of students whose first attempt was correct and executed the pattern, $n12$ is those whose first attempt was correct but did not execute the pattern, $n21$ was those whose first attempt was incorrect but executed the pattern, and $n22$ was those whose first attempt was incorrect and did not execute the pattern. In cases where $n21$ was zero, we added 0.5 to $n21$ to avoid an infinite risk ratio.

To investigate the potential relation between students' use of guidance and learning gain (research question 3), we applied differential SPM to identify sequential patterns used differentially by students with high, medium, and low learning gains. Interpreting learning gain values is challenging since they are dependent on students' prior knowledge. In other words, students who get a high score on the pre-test have limited potential to improve towards the post-test. Many studies split their participants into several equally sized groups based on learning gain. This method is appropriate when the group of participants is homogenous and has very similar prior knowledge, so the learning gain is indeed comparable between students. However, we found a large variation on students' prior knowledge in

Table 4 Clusters with corresponding average pre- and post-test scores and average learning gain

Cluster pre-test	# students	Average score (SD)	Average learning gain (SD)
Low	19	1.66 (1.13)	4.84 (2.66)
Medium	37	5.18 (0.97)	3.34 (2.11)
High	21	8.50 (1.13)	1.31 (1.51)
Cluster post-test	# students	Average score (SD)	Average learning gain (SD)
Low	20	5.78 (0.79)	2.38 (2.01)
Medium	24	8.19 (0.79)	3.52 (2.95)
High	35	10.24 (0.71)	3.94 (2.6)

Table 5 The pre- and post-test clusters including the number of students assigned to each of the options, with corresponding label of learning and average learning gain

Cluster pre-test	Cluster post-test	# students	Learning	Average learning gain (SD)
High	High	15	Medium learning	1.73 (1.40)
High	Medium	6	Low learning	0.25 (1.33)
High	Low	0	N/A	N/A
Medium	High	16	High learning	4.94 (1.42)
Medium	Medium	11	Medium learning	3.14 (1.53)
Medium	Low	10	Low learning	1.00 (1.11)
Low	High	3	High learning	7.50 (0.87)
Low	Medium	6	High learning	6.58 (1.74)
Low	Low	9	Medium learning	3.44 (1.57)

our group of participants. To provide as much contextual meaning to the learning gains as possible, the following approach was taken. K-means clustering was used to create two sets of three clusters of students. Based on the pre-test score, students were assigned to cluster low, medium, or high, and based on the post-test score students were once again assigned to cluster low, medium, or high. Table 4 shows the average pre- and post-test scores for each cluster. We also calculated the average learning gain for students in each cluster to illustrate that students with lower pre-test scores were more likely to show larger learning gains.

A student could for example be in cluster medium based on the pre-test score, learn a lot during the learning task, and end up in cluster high based on the post test score. This led to a total of nine possibilities, which were grouped and labelled as low learning, medium learning, or high learning. Students were assigned to low learning ($N=16$) if they dropped one or two clusters (e.g. cluster medium—> low; see Table 5), medium learning ($N=35$) if they remained in the same cluster and high learning ($N=25$) if they increased one or two clusters.

The Kruskal Wallis test was used to examine whether the three learning groups (i.e. low learning, medium learning and high learning) had statistically significant differences in sequential patterns. The effect size measure η^2 was reported. A rule of thumb for interpreting this measure is that η^2 greater than 0.01, 0.06, and 0.14 represents small, medium, and large effects respectively (Cohen, 1988). If the Kruskal Wallis test indicates that a sequential pattern was used differentially by the three groups, Mann-Whitney U tests were conducted on each pair of the three groups to identify between which two there was a difference. The Benjamini and Yekutieli correction was conducted to control the false discovery rate due to multiple comparison. An effect size, Cohen's r , was reported for the Mann-Whitney U test, which ranges between -1 and 1 (Fritz et al., 2012). The same analyses were repeated to examine the three pre-test score groups' differences in sequential patterns (research question 4).

3 Results and Discussion

3.1 Frequent Sequential Patterns Throughout the Learning Task

After analysis of all log data and filtering out meaningless patterns, 88 frequent sequential patterns remained. These remaining patterns were categorized, using the categories enumerated in Table 2, as either (multiple) attempt, (multiple) guidance, information, and the combination of these codes. The 18 sequential patterns with the highest support are shown in Table 6.

Table 6 The 18 sequential patterns (SP) with the highest support, average occurrences, and corresponding category

Code	Sequential pattern	Support	Average occurrences	Category
SP1	AC	1.00	15.52	Attempt
SP2	AI	0.99	6.04	Attempt
SP3	GNI	0.96	8.37	Guidance
SP4	GH+	0.90	2.86	Multiple guidance
SP5	AI+2	0.89	2.11	Multiple attempt
SP6	GH+all	0.87	2.80	Multiple guidance
SP7	GH	0.86	3.13	Guidance
SP8	AI+3+	0.84	2.61	Multiple attempt
SP9	GI	0.81	5.91	Guidance
SP10	AI=>AC	0.80	1.81	Multiple attempt
SP11	NAD=>NC=>AC	0.76	2.30	Attempt, no guidance
SP12	AI=>GNI	0.75	1.63	Attempt=>guidance
SP13	GI+2	0.73	2.14	Multiple guidance
SP14	GI+3+	0.70	2.87	Multiple guidance
SP15	GA	0.67	2.30	Guidance
SP16	AC+	0.65	1.70	Multiple attempt
SP17	NAD=>NC=>AI	0.61	1.03	Attempt, no guidance
SP18	IB=>NC=>GNI	0.58	1.22	Information=>guidance

The results indicate that students use a lot of guidance while working on the calculation questions. Almost all students navigated to the section in which intermediate values can be submitted (SP3). Students submitted many intermediate calculations (SP9, 13, and 14), but most of those were not submitted directly after navigating towards this section. Taking all 88 frequent sequential patterns in consideration, the only match that includes submitting an intermediate value directly after navigating to the just mentioned section, is the sequential pattern $GNI = >GI + 3 +$, with a support of 0.48 and an average occurrence of 0.94. Students navigating to the section in which intermediate values can be submitted without immediately submitting one or more values is indicative of them using a hint (i.e. which intermediate values can be calculated).

From SP4, 6, and 7 it is clear that students use a lot of hints. What stands out is that students were more likely to open two or more hints consecutively (SP4 and 6) than opening only one hint (SP7). Since most calculations are done outside the system (e.g. in Microsoft Excel), and the time factor is not included in this analysis, it is not possible to conclude whether students quickly opened all hints, which could indicate students who are “gaming the system” and trying to have the learning environment provide them with the answer (Baker et al., 2008), or whether they have been following the hint, after which they needed more guidance, and opened the next hint.

More or less the same holds for multiple consecutive incorrect attempts (SP5 and 8). Based on this analysis we cannot tell whether students were guessing the correct value, or whether they really tried to (re)do their calculations after they received feedback on their mistake. SP5, 8, 10, and 16 all show multiple consecutive attempts. This means that students sometimes decide not to use guidance in between attempts. Another situation in which students do not seem to use guidance, is when they make an attempt directly after acquiring the raw data required to do the calculation (SP11 and 17). However, students may have used guidance before acquiring the raw data. It should be noted that given the learning task, students needed to apply a different calculation approach for each question type, but they had to repeat this approach multiple times for different datasets. Given the results, it is plausible that students initially used guidance to find the correct calculation approach, during which they learned, and subsequently required less guidance when solving similar problems.

3.2 Use of Guidance Before First Attempt

Table 7 shows all frequent sequential patterns found before students’ first attempt for the three different question types. For each frequent sequential pattern, the support was given for students whose first attempt was correct and for students whose first attempt was incorrect.

For each question type the support for frequent sequential patterns was in general higher for students whose first attempt was correct versus students whose first attempt was incorrect. In other words, students who use more guidance in the system are more likely to have their first attempt correct. This is in line with the effect found by Bokhove and Drijvers (2012) where adding extra guidance resulted in less attempts. The support for almost all patterns that contain intermediate calculations (i.e. GI , $GI + 2$ or $GI + 3 +$) was significantly higher for students whose first attempt was correct versus students whose first attempt was incorrect (SP9, 13, 14, 19–23). Similarly, the support for frequent sequential patterns that involve opening a worked example, which was

Table 7 Frequent sequential patterns before a first (correct or incorrect) attempt for different question types

Code	Sequential pattern	Question type 1		Question type 2		Question type 3				
		Support	Risk ratio	Support	Risk ratio	Support	Risk ratio			
		Correct (N=38)	Incorrect (N=41)	Correct (N=30)	Incorrect (N=50)	Correct (N=26)	Incorrect (N=53)			
SP3	GNI	0.47	0.20	2.35	0.30	2.10	0.62	0.30	0.30	2.07
SP4	GH +	0.26	0.27	0.96	0.46	0.87	0.42	0.32	0.32	1.31
SP6	GH + all	0.26	0.20	1.30	0.24	1.54	0.42	0.45	0.45	0.93
SP7	GH	0.34	0.15	2.27	0.32	1.88	0.39	0.28	0.28	1.39
SP9	GI	0.34	0.00	28.22***	0.00	27.27**	0.46	0.06	0.06	7.67**
SP13	GI+2	0.26	0.00	21.58**			0.42	0.04	0.04	10.50**
SP14	GI+3+				0.20	20.20*	0.39	0.04	0.04	9.75**
SP15	GA				0.32	32.32***				
SP19	GI => GW						0.39	0.04	0.04	9.75**
SP20	GI+2 => GW						0.23	0.00	0.00	24.61*
SP21	GW => GI						0.31	0.04	0.04	7.75*
SP22	GW => GI+2						0.23	0.00	0.00	24.61*
SP23	GW => GI+3+						0.23	0.04	0.04	5.75
SP24	GNI => IB				0.27	0.08				
SP25	GH => IB				0.23	0.16	0.23	0.23	0.23	1.00
SP26	GH+ => IB				0.33	0.42	0.31	0.28	0.28	1.11
SP27	GH+all => GNI				0.27	0.14	0.27	0.08	0.08	3.38
SP28	GH+all => IB				0.23	0.18	0.12	0.26	0.26	0.46
SP29	IB => NC => GH				0.23	0.18	0.23	0.21	0.21	1.10
SP30	IB => NC => GNI				0.23	0.08	0.23	0.21	0.21	1.10

The sequential patterns for which no support is shown, had a support lower than 0.2, for both a first correct and incorrect attempt

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. p was adjusted by the Benjamini and Yekutieli correction

only available for question type 3 (GW, sequential patterns 19–23), was significantly higher for students whose first attempt was correct than for students whose first attempt was incorrect. This is in line with the positive effects of worked examples discussed by Kirschner et al. (2006). The use of GW was only possible after submitting an incorrect intermediate value, so this result is dependent on the use of GI. These results suggest that decomposing the calculation into smaller parts and providing students with feedback on their intermediate calculations and/or subsequent worked examples, contribute to students' ability to successfully calculate the correct answer on their first attempt. The support for sequential patterns that contain hints (SP4, 6, 7, 25–29) was usually frequent for both groups, and there were no statistically significant differences between students who had a first correct attempt compared to students with a first incorrect attempt. This suggests that the use of hints alone did not help students to directly calculate the correct final value.

3.3 Use of Guidance in Relation To Learning Gain

After splitting the participants into the three learning groups, 127 frequent sequential patterns were found in one or more of the three groups.

No pattern showed a statistically significant difference between the different learning groups after the Benjamini and Yekutieli correction. Before the correction, seven sequential patterns were significant (Table 8). Those patterns were investigated to check whether they might reveal patterns that could be related to learning gains. However, those patterns were considered as not meaningful due to their rarity (based on overall support and average occurrences).

In fact, all the seven sequential patterns in Table 8 show relatively low average number of occurrences and overall support when compared to sequential patterns identified in previous analyses (see Table 6). In general, there were no clear trends differentiating the occurrence of sequential patterns between the different learning groups, and we were not able to draw any meaningful conclusions about how specific patterns related to how students use of guidance might relate to improved learning gains.

Based on the results, we concluded that there are no statistical differences in the use of guidance, when students are categorized based on their learning gain. One possible explanation is that the guidance did support students' learning, but its contribution to learning was similar across students. The differences in learning were not the result of the use of guidance. Another reason may be the expertise reversal effect, which refers to how the effect of instructional support depends on learners' prior knowledge (Kalyuga, 2007). In our study, the cluster with high prior knowledge might be able to learn by solving the calculation without using the guidance, while the cluster of low prior knowledge needed the guidance to solve the calculation to learn. Thus, the use of guidance may have had different effects on learning among prior knowledge clusters. However, students' prior knowledge was heterogeneous within each group of learning gains. For example, the medium learning group was a cluster of students who initially were categorized in the high, medium, or low prior knowledge cluster (see Table 5). Consequently, the use of guidance did not vary across groups with different learning gains. To further investigate this possibility, we also examined whether there are significantly different patterns in relation to students' prior knowledge.

Table 8 Sequential patterns that were used differentially by high, medium, and low learning students (SP31–37) and corresponding effect sizes (before correction)

Code	Sequential pattern	Average occurrences			Overall sup- port (N = 76)	η^2	H	r		
		high learn- ing group (N = 25)	medium learning group (N = 35)	low learn- ing group (N = 16)				Low-medium	Low-high	Medium-high
SP31	AI => GNI => GI => AC	0	0	0.31	<0.20	0.24	19.81 *	0.48	0.46	N/A
SP32	NAD => IB => NC => GH + all	0	0	0.25	<0.20	0.19	15.63 *	0.43	0.41	N/A
SP33	IFC => NC => AC	0.04	0.09	0.38	<0.20	0.10	9.01 *	0.33	0.37	0.04
SP34	GI => GW => GI + 2	0.24	0.06	0	<0.20	0.07	7.46 *	0.14	0.33	0.26
SP35	GI + 2 => GA	0.28	0.26	0.56	0.24	0.07	6.83 *	0.33	0.32	0.01
SP36	NAD => NC => AI + 3 +	0.12	0.69	0.50	0.30	0.07	6.99 *	0.1	0.25	0.34
SP37	GH + all => IB => IB => NC => AC	0.08	0.03	0.25	<0.20	0.06	6.42 *	0.34	0.23	0.12

* $p < 0.05$ before the Benjamini and Yekutieli correction

Table 9 Sequential patterns that were used differentially by students with high, medium, and low prior knowledge, and corresponding effect sizes (before correction)

Code	Sequential pattern	Average occurrences			Overall support (N = 77)	η^2	H	r	
		high prior knowledge group (N = 21)	medium prior knowledge group (N = 37)	low prior knowledge group (N = 19)				Low-medium	Low-high Medium-high
SP4	GH +	2.48	2.51	4.00	0.90	0.08	7.68*	0.36	0.03
SP6	GH + all	2.38	2.24	4.33	0.87	0.06	6.41*	0.33	0.02
SP7	GH	2.71	2.73	4.67	0.86	0.08	7.58*	0.32	0.08
SP14	GI + 3 +	2.76	1.84	5.44	0.70	0.07	7.19*	0.35	0.15
SP18	IB = > NC = > GNI	0.67	1.03	2.39	0.58	0.12	10.8*	0.32	0.49
SP38	AC = > IFC	0.48	1.05	2.17	0.51	0.11	10.2*	0.31	0.49
SP39	IB = > NC = > AI = > AC	0.19	0.65	0.78	0.39	0.06	6.42*	0.09	0.40
SP40	IB = > NC = > GI	0.29	0.51	1.11	0.34	0.08	8.05*	0.27	0.43
SP41	IB = > NC = > GI + 2	0.33	0.08	0.61	0.23	0.12	11.0*	0.44	0.17
SP42	IB = > NC = > GNI = > GI	0.10	0.19	0.72	0.22	0.09	8.41*	0.31	0.40
SP43	GNI = > GH	0.14	0.19	0.50	0.22	0.06	6.58*	0.30	0.34

* $p < 0.05$ before the Benjamini and Yekutieli correction

3.4 Use of Guidance in Relation to Prior Knowledge

After splitting the participants in three groups of high, medium, and low prior knowledge, 142 sequential patterns were found with a support larger than 0.2 in at least one of the three groups. No statistically significant differences were found after the Benjamini and Yekutieli correction, but 29 patterns were significant before the correction and had medium to high effect sizes (η^2 greater or equal to 0.06). Of these 29 patterns, we show the 11 patterns that had an overall support greater than 0.2 (Table 9). These 11 patterns were investigated to identify insights into possible relationships between prior knowledge and use of guidance when performing calculations within the learning environment.

What stands out is that the differential sequential patterns included all types of guidance that could potentially be used. In terms of the prior knowledge groups, the results suggest that there is no difference between use of guidance for students with high or medium prior knowledge. However, the average occurrences of all sequential patterns for students in the low prior knowledge group is higher compared to the other two groups. The trend suggests that students with a low prior knowledge used more guidance than students with a higher prior knowledge, comparable to the results of Zhou et al. (2021). However, additional data should be collected to make sure that these trends can be observed in other datasets and are not false discoveries, given that the three groups had no statistically significant difference in these sequential patterns after the Benjamini and Yekutieli correction.

4 Limitations and Further Research Directions

SPM algorithms require researchers to prespecify some parameters, such as the minimum support. The prespecified parameters are similar to the hyperparameters of machine learning models, but, currently, there is no way to tune the prespecified parameters in SPM as tuning the hyperparameters in machine learning. Thus, the minimum support for this study was chosen based on the trade-off between including more interesting learning behaviour patterns and excluding less meaningful patterns. We acknowledge that the decision might be a little subjective.

Researchers have applied differential SPM in educational data and selected patterns based on statistical significance (Kinnebrew et al., 2013). Using the statistical significance as the selection criterion has limitations. When the sequences are long, many sequential patterns may be discovered (Zhou et al., 2010), and all of them will be further examined via statistical tests. Multiple tests inflate the Type I error rate, and the p -values need to be adjusted (Benjamini & Yekutieli, 2001). However, when the sample size is small, the statistical power may be small (Cohen, 1988). There may be no sequential pattern statistically different between groups of interest after applying the adjustment method, as shown in the results for the third and fourth research questions. Thus, in this study, we selected the patterns based on both significance and the effect size. We found some patterns with large effect sizes, which would be ignored if the selection criterion was purely based on statistical significance. The results suggest further research should consider the interestingness of sequential patterns based on both statistical significance and effect sizes.

Similar studies that applied SPM into educational data have considered the time factor (Dermy & Brun, 2020; Emara et al., 2018). For instance, Emara et al. (2018) added long or short suffixes to reading actions based on whether a read is longer than 3 s. However, the virtual experiment environment recorded the time of action log at the minute level, and

thus, we were unable to compute the duration of each action. As a result, it could be that similar sequential patterns should have been interpreted differently. For example, students who make three consecutive incorrect attempts could be guessing (gaming the system), or they could have tried to do the calculations, but have made a mistake each time.

The studied learning task was part of a larger virtual experiment environment. Since the virtual experiment environment was designed with the purpose to provide students with the best possible learning experience, rather than with the purpose to investigate how students use guidance while doing calculations (which is only a part of the assignment in the virtual experiment environment), the context for this study investigated a sub-set of the larger learning environment. Further studies would benefit from considering how the calculation tasks are integrated in the larger learning task in the virtual experiment environment, and how this provides additional structure in terms of the order in which students have to progress through the learning task. In the current situation, students are free to determine the order in which they do the calculations. Since there are three question types, with each multiple questions, it is hard to control for what students might or might not have previously learned in the other question types.

Last, the study was performed only once within a real educational context. This raises the question whether and how the participants of this study represent the subset of participants that may use this type of technology. As such, there is need for further study to investigate how generalizable the results are and whether different participants would show similar or different patterns.

5 Conclusions

The goal of this study was to investigate students' behaviour regarding the use of guidance while doing calculations, and to relate this behaviour to learning. Using sequential pattern mining, we were able to find several patterns of student behaviour. Students made a lot of use of the guidance provided in the learning task. Students who used the option to check their intermediate calculations were more likely to make a successful first attempt to complete the calculation when compared to students who did not use this guidance. This effect was not found for the use of guidance in the form of hints. Although the hints were frequently used by almost all students, we were not able to quantify the usefulness of the hints in this study. This result suggests that many students may perceive the hints as a useful source of guidance. Guidance in the form of worked examples was found to have a positive contribution to students' success at first attempt, which was also expected based on literature (Clark & Mayer, 2016). Despite the large amount of guidance used by students and its established contribution to students' success when they first submit the answer to a calculation, we were unable to find a relation between learning gain and students' use of guidance. When exploring the relation between students' use of guidance and their prior knowledge, we found a trend that students with a low prior knowledge used more guidance than students with a higher prior knowledge. However, additional research should be done to confirm this result and to provide insights into whether this additional use of guidance leads to improved learning.

When designing the learning task that was the subject of this study, we used, among others, the design principles "provide access to hints (one by one) that guide the students' thinking process" and "Provide the opportunity for students to check their intermediate calculations", to guide students when doing scientific calculations (Verstege et al., 2021).

These design principles were based on our experience with designing digital learning materials (Diederer et al., 2003; Diederer et al., 2006; Van der Kolk et al., 2012; Van der Kolk et al., 2013; Verstege et al., 2021). The results of this study imply that the aforementioned design principles are of utmost importance for students when doing scientific calculations. We advise instructional designers to make use of hints, intermediate calculations, and/or worked examples when designing learning materials that include multi-step calculations. Based on the results of this study, it was not possible to deduct which type of guidance is useful in which situation. Therefore, future research could focus on which combination of these three types of guidance contributes the most to students' ability to independently complete scientific multi-step calculations, and how this is related to the difficulty level of the calculations.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. The authors have no relevant financial or non-financial interests to disclose.

Data availability The data that support the findings of this study are available from the authors upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agrawal, R., & Srikant, R. (1995). *Mining sequential patterns*. Paper presented at the Proceedings of the eleventh international conference on data engineering.
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73, 277–320. <https://doi.org/10.3102/00346543073003277>
- Anderson, P. C., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 2, 100027. <https://doi.org/10.1016/j.caeai.2021.100027>
- Baker, R. (2010). Data mining for education. In B. McGaw, P. Peterson, & E. Baker (Eds.), *International encyclopedia of education* (Vol. 7, pp. 112–118). Elsevier Ltd.
- Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in “gaming the system” behavior in interactive learning environments. *Journal of Interactive Learning Research*, 19, 185–224.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29, 1165–1188.
- Bokhove, C., & Drijvers, P. (2012). Effects of feedback in an online algebra intervention. *Technology, Knowledge and Learning*, 17, 43–59. <https://doi.org/10.1007/s10758-012-9191-8>
- Brenner, D. G., Matlen, B. J., Timms, M. J., Gochyyev, P., Grillo-Hill, A., Luttgen, K., & Varfolomeeva, M. (2017). Modeling student learning behavior patterns in an online science inquiry environment. *Technology, Knowledge and Learning*, 22, 405–425. <https://doi.org/10.1007/s10758-017-9325-0>

- Buchta, C., Hahsler, M., & Diaz, D. (2020). arulesSequences: Mining frequent sequences (version R package version 0.2-24). Retrieved from <https://CRAN.R-project.org/package=arulesSequences>
- Chen, C.-M., & Wang, W.-F. (2020). Mining effective learning behaviors in a web-based inquiry science environment. *Journal of Science Education and Technology*, 29, 519. <https://doi.org/10.1007/s10956-020-09833-9>
- Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.
- Cohen, J. (1988). *The effect size index: d. Statistical power analysis for the behavioral sciences*. Routledge Academic.
- Dermay, O., & Brun, A. (2020). Can we take advantage of time-interval pattern mining to model students activity? In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, 69–80.
- Diederer, J., Gruppen, H., Hartog, R., Moerland, G., & Voragen, A. G. (2003). Design of activating digital learning material for food chemistry education. *Chemistry Education Research and Practice*, 4, 353–371. <https://doi.org/10.1039/B3RP90020G>
- Diederer, J., Gruppen, H., Hartog, R. J., & Voragen, A. G. (2006). Design and evaluation of digital assignments on research experiments within food chemistry. *Journal of Science Education and Technology*, 15, 227. <https://doi.org/10.1007/s10956-006-9010-x>
- Emara, M., Rajendran, R., Biswas, G., Okasha, M., & Elbanna, A. A. (2018). Do students' learning behaviors differ when they collaborate in open-ended learning environments? *Proceedings of the ACM on human-computer interaction* 2: 49. doi: <https://doi.org/10.1145/3274318>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141, 2. <https://doi.org/10.1037/a0024338>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. <https://doi.org/10.1007/s11528-014-0822-x>
- Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15, 42–57.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. <https://doi.org/10.3102/003465430298487>
- Hausmann, R. G. M., Vuong, A., Towle, B., Fraundorf, S. H., Murray, R. C., Connelly, J. (2013). An evaluation of the effectiveness of just-in-time hints. In H. C. Lane, K. Yacef, J. Mostow, P. Pavlik (Eds.), *Artificial intelligence in education. AIED 2013. Lecture Notes in Computer Science* (vol 7926). Springer. doi: https://doi.org/10.1007/978-3-642-39112-5_114
- He, Z., Zhang, S., Gu, F., & Wu, J. (2019). Mining conditional discriminative sequential patterns. *Information Sciences*, 478, 524–539. <https://doi.org/10.1016/j.ins.2018.11.043>
- Hernández-Leo, D., Martínez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J. A., & Rodríguez-Triana, M. J. (2019). Analytics for learning design: A layered framework and tools. *British Journal of Educational Technology*, 50, 139–152. <https://doi.org/10.1111/bjet.12645>
- Jiang, Y., Paquette, L., Baker, R. S., & Clarke-Midura, J. (2015). Comparing novice and experienced students within virtual performance assessments. *Proceedings of the 8th International educational data mining society*, 136–143.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19(4), 509–539. <https://doi.org/10.1007/s10648-007-9054-3>
- Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, 72, 757–770. <https://doi.org/10.1016/j.chb.2016.09.062>
- Kew, S. N., & Tasir, Z. (2022). Learning analytics in online learning environment: A systematic review on the focuses and the types of student-related analytics data. *Technology, Knowledge and Learning*, 27, 405–427. <https://doi.org/10.1007/s10758-021-09541-2>
- Kinnebrew, J. S., Loretz, K. M., & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining*, 5(1), 190–219.
- Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and Learning*, 9, 187–215. <https://doi.org/10.1007/s11409-014-9112-4>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985Sep4102_1

- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19, 239–264. <https://doi.org/10.1007/s10648-007-9049-0>
- Van der Kolk, J., Beldman, G., Hartog, R. J., Moerland, G., & Gruppen, H. (2013). *ExperD: Web-based support for laboratory class workflow design*. Paper presented at the 5th International Conference on Computer Supported Education, Aachen, Germany.
- Law, N., & Liang, L. (2020). A multilevel framework and method for learning analytics integrated learning design. *Journal of Learning Analytics*, 7, 98–117. <https://doi.org/10.18608/jla.2020.73.8>
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57, 1439–1459. <https://doi.org/10.1177/0002764213479367>
- Mangaroska, K., & Giannakos, M. (2019). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, 12, 516–534. <https://doi.org/10.1109/TLT.2018.2868673>
- Mirzaei, M., & Sahebi, S. (2019). Modeling students' behavior using sequential patterns to predict their performance. In: S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, R. Luckin (Eds.), *Artificial Intelligence in Education. AIED 2019. Lecture Notes in Computer Science* (vol. 11626, pp. 350–353). Springer. Doi: https://doi.org/10.1007/978-3-030-23207-8_64
- Neelen, M., & Kirschner, P. A. (2020). *Evidence-informed learning design: Creating training to improve performance*. Kogan Page.
- Olsson, J., & Granberg, C. (2019). Dynamic software, task solving with or without guidelines, and learning outcomes. *Techology, Knowledge and Learning*, 24, 419–436. <https://doi.org/10.1007/s10758-018-9352-5>
- Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, 21, 267–280. <https://doi.org/10.1016/j.learninstruc.2010.07.004>
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46, 30.
- Taub, M., & Azevedo, R. (2018). Using sequence mining to analyze metacognitive monitoring and scientific inquiry based on levels of efficiency and emotions during game-based learning. *Journal of Educational Data Mining*, 10, 1–26. <https://doi.org/10.5281/zenodo.3554711>
- Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., & Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. *Learning and Instruction*, 54, 93–103. <https://doi.org/10.1016/j.learninstruc.2017.08.005>
- Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85, 475–511. <https://doi.org/10.3102/0034654314564881>
- Van der Kolk, J., Beldman, G., Hartog, R. J., & Gruppen, H. (2012). Students using a novel web-based laboratory class support system: A case study in food chemistry education. *Journal of Chemical Education*, 89, 103–108. <https://doi.org/10.1021/ed1005294>
- Van Merriënboer, J. J., & Kirschner, P. A. (2018). *Ten steps to complex learning: A systematic approach to four-component instructional design* (3rd ed.). Routledge.
- Verstege, S., & Diederén, J. (2019). *Virtual experiment environment: A showcase of a preparation tool for laboratory classes*. Paper presented at the EdMedia+ Innovate Learning conference (Amsterdam, The Netherlands).
- Verstege, S., Pijera-Díaz, H. J., Noroozi, O., Biemans, H., & Diederén, J. (2019). Relations between students' perceived levels of self-regulation and their corresponding learning behavior and outcomes in a virtual experiment environment. *Computers in Human Behavior*, 100, 325–334. <https://doi.org/10.1016/j.chb.2019.02.020>
- Verstege, S., Vincken, J.-P., & Diederén, J. (2021). Blueprint to design virtual experiment environments. *Computers and Education Open*, 2, 100039. <https://doi.org/10.1016/j.caeo.2021.100039>
- Wiley, K. J., Dimitriadis, Y., Bradford, A., & Linn, M. C. (2020). *From theory to action: Developing and evaluating learning analytics for learning design*. Paper presented at the proceedings of the tenth international conference on learning analytics & knowledge, Frankfurt, Germany. Doi: <https://doi.org/10.1145/3375462.3375540>
- Wong, J., Khalil, M., Baars, M., de Koning, B. B., & Paas, F. (2019). Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course. *Computers & Education*, 140, 103595. <https://doi.org/10.1016/j.compedu.2019.103595>
- Zaki, M. J. (2000). *Sequence mining in categorical domains: Incorporating constraints*. Paper presented at the Proceedings of the ninth international conference on Information and knowledge management, McLean, Virginia, USA. Doi: <https://doi.org/10.1145/354756.354849>

- Zhang, Y., Paquette, L., Bosch, N., Ocumpaugh, J., Biswas, G., Hutt, S., & Baker, R. S. (2022). The evolution of metacognitive strategy use in an open-ended learning environment: Do prior domain knowledge and motivation play a role? *Contemporary Educational Psychology*, *69*, 102064. <https://doi.org/10.1016/j.cedpsych.2022.102064>
- Zheng, J., Xing, W., & Zhu, G. (2019). Examining sequential patterns of self- and socially shared regulation of STEM learning in a CSCL environment. *Computers & Education*, *136*, 34–48. <https://doi.org/10.1016/j.compedu.2019.03.005>
- Zhou, Y., Andres-Bray, J.M., Hutt, S., Ostrow, K., & Baker, R.S. (2021). A comparison of hints vs. scaffolding in a MOOC with adult learners. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, V. Dimitrova (Eds.), *Artificial Intelligence in Education. AIED 2021. Lecture Notes in Computer Science*, 12749. Springer. doi:https://doi.org/10.1007/978-3-030-78270-2_76
- Zhou, M., Xu, Y., Nesbit, J. C., & Winne, P. H. (2010). Sequential pattern analysis of learning logs: Methodology and applications. *Handbook of Educational Data Mining*, *107*, 107–121.
- Zhu, G., Xing, W., & Popov, V. (2019). Uncovering the sequential patterns in transformative and non-transformative discourse during collaborative inquiry learning. *The Internet and Higher Education*, *41*, 51–61. <https://doi.org/10.1016/j.iheduc.2019.02.001>
- Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist*, *48*, 135–147. <https://doi.org/10.1080/00461520.2013.794676>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.