JOURNAL OF LEARNING ANALYTICS

Volume 11(2), 174–196. https://doi.org/10.18608/jla.2024.8083

How Does a Data-Informed Deliberate Change in Learning Design Impact Students' Self-Regulated Learning Tactics?

Zhongzhou Chen[1], Tom Zhang[2] and Michelle Taub[3]

Abstract

The current study measures the extent to which students' self-regulated learning tactics and learning outcomes change as the result of a deliberate, data-driven improvement in the learning design of mastery-based online learning modules. In the original design, students were required to attempt the assessment once before being allowed to access the learning material. The improved design gave students the choice to skip the first attempt and access the learning material directly. Student learning tactics were measured using a multi-level clustering and process mining algorithm, and a quasi-experiment design was implemented to remove or reduce differences in extraneous factors, including content being covered, time of implementation, and naturally occurring fluctuations in student learning tactics. The analysis suggests that most students who chose to skip the first attempt were effectively self-regulating their learning and were thus successful in learning from the instructional materials. Students who would have failed the first attempt were much more likely to skip it than those who would have passed the first attempt. The new design also resulted in a small improvement in learning outcome and median learning time. The study demonstrates the creation of a closed loop between learning design and learning analytics: first, using learning analytics to inform improvements to the learning design, then assessing the effectiveness and impact of the improvements.

Notes for Practice

- Learning analytics methods are used to evaluate the impact of a deliberate change in learning design on student learning behaviours in an online learning environment.
- A quasi-experiment design is used to account for variations in multiple extraneous factors.
- New analysis methods are introduced to quantify the change in student adoption of different learning tactics.
- When provided with the opportunity, most students engage in effective self-regulation in an online learning environment.
- Data collection and data analysis should be an integral part of learning design.

Keywords: Learning analytics, learning design, process mining, self-regulated learning, mastery-based learning

Submitted: 06/06/2023 — **Accepted:** 21/05/2024 — **Published:** 25/07/2024

Corresponding author [1]Email: zhongzhou.chen@ucf.edu Address: Department of Physics, University of Central Florida, 4111 Libra Dr., Orlando, Florida, 32816-2385, USA. ORCID iD: https://orcid.org/0000-0003-3324-5330
[2]Email: zhang@lifeexpectancy.org Address: Life Expectancy Project, 1439 – 17th Ave, San Francisco, California, 94122, USA.
[3]Email: michelle.taub@ucf.edu Address: Department of Learning Sciences and Educational Research, University of Central Florida, 12494 University Blvd., Orlando, Florida, 32816, USA. ORCID iD: https://orcid.org/0000-0002-6238-1334

1. Introduction

The rapid development of online learning technology empowers today's instructors to implement a wide range of different learning designs in their courses relatively easily (Persico & Pozzi, 2015). Here, we use the term learning design based on the definition of Lockyer et al. (2013) to mean "the sequence of learning tasks, resources, and supports that a teacher constructs for students over part of, or the entire, academic semester" (p. 1411). While modern technology gives the instructor a high level of flexibility to customize the learning design, the average instructor is also responsible for making an increasingly large number of learning design decisions at different scales, from deciding which online platform to use and which pedagogy to implement, to microscopic, detailed decisions such as release and due dates of assignments, number of allowed attempts on assessments, and whether assignments are mandatory or optional. Any of those choices may or may not have a significant impact on student learning experiences, behaviours, and outcomes, yet in many cases, most of those decisions are still largely being made based on instructor experiences and anecdotal evidence (Koedinger et al., 2013).

Over the past decade, there has been a significant increase in efforts to use learning analytics (LA) as a method to inform and evaluate learning design (LD; Holmes et al., 2019; Lockyer et al., 2013; Mangaroska & Giannakos, 2019). As explained in detail in section 1.1, by conducting extensive analysis of student learning behaviours, outcomes, and tactics, those existing studies revealed strong associations between student learning behaviours or learning tactics and different LDs of the learning environment. For example, differences in student learning behaviours have been identified for different LDs implemented in different online courses (Holmes et al., 2019), in different versions of the same course (Lancaster et al., 2020), and with different LDs implemented in different sections of the same course (Fan et al., 2021). Naturally, the next key questions to ask is this one: Can we first implement an improvement to an existing LD—based on learning theory and measurements of student learning—and then measure via LA how student learning changes because of the improvement? This type of targeted, deliberate improvement can be seen as one way to “actively intervene with the learning environment,” which aligns well with the goals of learning analytics research and is relatively infrequent in current LA literature, according to a recent open peer commentary paper (Motz et al., 2023).

Existing studies fell short of establishing such a causal relationship between changes in LD and differences in student learning behaviours or learning tactics because they predominantly analyze existing LDs chosen for different courses or different segments of the same course. Also explained in detail in section 1.1, this type of analysis provides weak control over confounding factors such as differences in content, student population, individual instructor, and instructional methods. In addition, since the different LDs are chosen to fit the content of the course, the differences between them are usually much larger than what the average instructor would consider changing to their existing LD on a course. Building on the results of earlier studies, the current study seeks to answer this question by following a four-step process:

1. Based on an existing LD, collect data and analyze student learning behaviours and outcomes.
2. Propose a theory-driven improvement to the LD based on existing LA; implement the new LD in a classroom setting.
3. Compare student learning behaviours and outcomes under both the new LD and the existing LD and utilize a quasi-experiment design to control for two types of extraneous confounding factors.
4. Evaluate the effectiveness of the new LD based on LA results and in the context of the proposed theoretical framework.

The rest of the introductory section is organized as follows: Section 1.1 provides a more detailed review of existing literature and emerging trends at the cross-section of LA and LD; Section 1.2 introduces the existing LD chosen for the current study, mastery-based online learning modules; Section 1.3 introduces previous LA results on the LD that motivates the current improvements; Section 1.4 describes the improved LD and its relation to the theoretical framework of self-regulated learning; Section 1.5 lists the research questions. The Methods section (Section 2) is organized as follows: Section 2.1 describes the instructional conditions of the current study; Section 2.2 introduces the quasi-experiment design, how it eliminates differences in instructional content, and how it provides some control over variances caused by other extraneous factors (step 3 above); Section 2.3 introduces details of data collection and processing, and section 2.3.3 introduces our method of quantitatively comparing differences in learning tactics adopted by different groups of students by using Hellinger distance as a distance metric. This new method is uniquely suitable for comparing distributions of student learning tactic clusters, which are challenging to compare using traditional statistical methods.

1.1. Previous Research at the Intersection of LA and LD

Learning analytics (LA) focuses on using data about learner actions and learning context to understand and optimize learning and learning environments (SoLAR, 2020). In Macfadyen et al. (2020) suggest that LA can “help learning design research move beyond its current heavy focus on design principles, and begin to evaluate ‘what happens next,’ lending greater rigor and credibility to the field” (p. 7). Earlier efforts in the application of LA to inform learning design often focus on using various analytic methods, such as network analysis, to identify and visualize patterns of student learning processes under a single type of learning design (e.g., Ifenthaler et al., 2018). Some studies also explored teacher reactions to the outcomes and visualizations of their students’ LA and their intention to modify the current learning design (Kaliisa et al., 2020).

There are two notable recent trends in the intersection of learning design and LA. First, to address the challenge of “a paucity of evidence for how learners respond to different learning designs” (Rienties et al., 2017), research has moved from analyzing data collected from a single learning design to comparing student learning behaviour under multiple different learning designs. The different designs can be variations of the same template or are implemented in different components of the same course. For example, Holmes et al. (2019) compared student behaviour, pass rate, and satisfaction across six

ISSN 1929-7750 (online).

different types of learning designs implemented in 55 modules in an online learning setting. Lancaster et al. (2020) compared student learning behaviours in five course sections created from a common course template, yet the instructors of each section implemented their own unique modifications. Fan et al. (2021) compared student learning tactics across seven different learning designs implemented in seven different weekly units of the same MOOC. While those three studies involve different learning designs implemented in different educational settings and examined student behaviour using different methods, all three studies found that differences in learning designs are strongly correlated with differences in student learning behaviour and learning tactics.

Second, an increasing amount of recent LA research has moved beyond analyzing the frequencies of isolated events, such as page access or assessment attempts, to identifying and visualizing patterns in sequences of related learning events that are more indicative of student learning strategies (Huber & Bannert, 2023; Taub et al., 2018, 2022; Wortha et al., 2019). Most of those studies utilize advanced analysis techniques such as sequential pattern mining, clustering, and process mining and interpret the impact of learning design on the learning process through the framework of self-regulated learning (SRL; Saint et al., 2020, 2021; Sonnenberg & Bannert, 2019; Taub et al., 2022).

From an SRL perspective, student learning processes are influenced by multiple internal and external factors (Winne, 2017). Internal factors relate to a student's cognition, metacognition, or motivation. Cognitive factors include processes related to attention or memory capabilities (e.g., working memory capacity). Metacognitive factors relate to one's knowledge and awareness of their own learning and understanding (e.g., activating prior knowledge, making an accurate judgment of how well one understands current content). Motivational factors can include one's goal orientation, task effort or task value, or self-efficacy. External factors are those outside of the learner's control, such as the type of learning environment or available resources students can use while completing a task. The learning design of an online learning environment is one of the major external factors impacting student learning processes.

While presented as separate internal and external factors, the external factors, such as learning design, can impact a student's internal state, such as their motivation, as well as their subsequent steps toward completing a task or learning goal (Winne, 2017). For example, if an online course is designed to allow students to map out how they view course content (e.g., select the sequence of actions, choose which content to read and when), this might impact how they set goals for what they aim to achieve in the course, (i.e., achievement goals; Taub et al., 2022). A prominent recent example of using SRL-based learning analytics to inform learning design is Fan et al. (2021). In this study, different student learning tactics were extracted using trace clustering and network analysis methods, and findings were interpreted according to an SRL framework. The frequency of those tactics was then compared across different learning designs and different performance groups.

Yet despite the improvement in LA methods and theoretical basis, those existing studies are still limited to measuring the correlation between LD and student learning behaviour, primarily due to insufficient control over two types of confounding extraneous factors that could lead to different student learning behaviour. For example, in Fan et al. (2021), the different learning designs were implemented across different weeks of the semester. In Holmes et al. (2019), the learning designs were implemented across 55 modules randomly selected with no control over instructional content or time of implementation. However, in Taub et al. (2022) and Zhang et al. (2021, 2022), students' self-regulated learning strategies were shown to change over the duration of the semester even when the learning design remained the same. Student learning strategies have also been observed to change within a single weekly unit, potentially due to changes in content difficulty or other internal or external factors. In addition, a second type of confounding factor is that in both Holmes et al. (2019) and Lancaster et al. (2020), different LDs were implemented in course sections taken by different student populations and taught by different instructors, and sometimes even taught in different years, all of which could potentially cause detectable differences in student learning processes and outcomes, even under identical LD. Not accounting for the impact of those extraneous factors makes it much harder to attribute the cause of differences in observed learning processes to the differences in the learning design.

1.2. Existing Learning Design

The learning design being investigated in the current study is mastery-based online learning modules. The design is based on the concept of mastery learning first proposed in the 1960s by Keller (Kulik et al. 1974) and Bloom (1968), with the fundamental assertion that most students can achieve mastery on any topic if given enough time and proper support. Early proponents claimed that it would allow 90% of the students to perform at the level previously achieved by only 10% of the students.

Based on the principles of mastery learning, Chen et al. (2018) and Guthrie and Chen (2019b) designed and created online learning modules for a university-level introductory physics course. Each online learning module is focused on explaining one or two basic concepts or developing the skills to solve one kind of problem, designed to be completed in 5 to 30 minutes, depending on the student's prior knowledge and the complexity of the problem. Each individual module consists of an assessment component that tests students' content mastery in 1–2 questions and an instructional component that includes instructional text and practice problems on the topic (see Figure 1A). Upon accessing a module, students are shown the learning objectives of the current module and are required to make an initial attempt on the assessment component before being allowed to access the instructional component. Students can make additional attempts on the assessment component at any time after the first attempt and are not required to access the instructional component. Several modules form a learning module sequence on a general topic, such as conservation of mechanical energy. Within each sequence, students are required to either pass the assessment or use up all attempts before moving on to the next module.

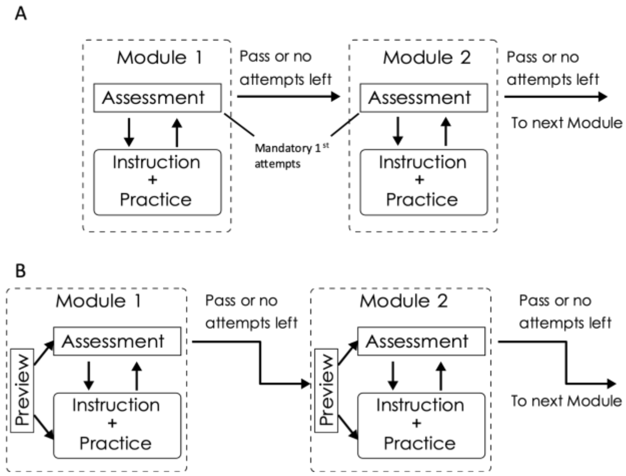

Figure 1. Schematic representation of original and improved learning designs of online learning modules. A: the original learning design with mandatory first attempt on each module. B: The improved new design with a preview stage and an optional first attempt.

According to the original design, the mandatory first attempt serves as a self-assessment tool that allows students who have already mastered the content of the current module to quickly proceed to the next module and motivates students who couldn't pass the first attempt to study the instructional materials. This design feature was inspired by research in "preparation for future learning" (Schwartz et al., 2005) and "productive failure" (Kapur, 2010), which would suggest that failing the initial attempts would allow students to learn better from the instructional materials.

1.3. LA Results from Existing LD

Earlier analysis of student learning processes on mastery-based online learning modules made some key observations regarding the initial attempts. First, students would sometimes make very short assessment attempts (often less than 30 seconds), and a higher frequency of those short attempts correlates with lower course performance. Those short attempts are likely associated with guessing and answer copying (Chen, 2022; Chen et al., 2020) since the assessment would contain problems that would typically take between 5 and 20 minutes for most students to solve. Moreover, abnormally short attempts have been shown to be strongly associated with lower course performance and have been related to answer copying (Alexandron et al., 2017; Warnakulasooriya et al., 2007) or guessing (Chen, 2022) in other studies in many similar online learning environments. Second, frequency of short attempts on the mandatory first attempt increases as the content difficulty increases with time over the semester; the number of students who pass on a short first attempt also increases towards the end of the semester (Zhang et al., 2021). More importantly, an increasing number of students would adopt a strategy of making a short first attempt, accessing the instructional materials, and then making multiple normal-length attempts immediately after accessing the learning materials. The frequency of this strategy increases on harder modules towards the end of a sequence (Taub et al., 2022; Zhang et al., 2022). On the other hand, normal-duration first attempts decrease over the course of the semester, especially on harder modules toward the end of a sequence.

Taken together, those observations suggest that at least some fraction of students are effectively self-assessing their chances of passing the assessment on the first attempt and deliberately submitting guessed answers on the first attempt just to access the learning material. At the same time, it is possible that the mandatory first attempt might have also tempted more students to copy answers from other sources for fear of losing one attempt.

From a self-regulated learning perspective, the original mandatory first-attempt design assumes that students have low levels of self-awareness of their prior knowledge, which was originally inspired by earlier research showing that students are often inaccurate and overconfident in estimating their own exam performance (Foster et al., 2017; Magnus & Peresetsky, 2018; Serra & DeMarree, 2016; Wüst & Beck, 2018). Requiring students to make an initial attempt could facilitate their use of a self-testing strategy and their planning for the subsequent learning task. Yet the LA results did not support this assumption but rather suggested that some students were already self-assessing and planning before their first submission and adjusting their strategy-use according to the difficulty of the assessment problem. For those students, the mandatory first attempt could be an unnecessary hurdle. On the other hand, for students with a performance-dominant goal orientation, the mandatory first attempt might tempt them to find a shortcut by guessing or answer copying to save time and effort.

1.4. Improvement to Mastery Learning Design

One potential improvement for the design of mastery online modules is to make the initial attempt a choice rather than a mandate (see Figure 1B). Students would be shown a problem (very similar to the one in the first assessment attempt) and given the option to either try it first or directly access the learning materials.

From an SRL perspective, this design change could have the following impact on student choice of learning tactics related to their planning or metacognitive monitoring strategies (Winne & Hadwin, 1998, 2008; Zimmerman, 2000):

1. When monitoring their performance, if students are overconfident, they would likely take the first attempt (i.e., feeling of knowing judgment; Greene & Azevedo, 2009) and deem that they do not need to study (even if they do). If they are under-confident, they would likely skip the first attempt and study instead because they feel that they do not know the material (even if they do).
2. When monitoring their performance, if students have good results from prior self-testing (i.e., self-questioning; Greene & Azevedo, 2009), they will likely take the first attempt because they should be able to pass the assessment. If students have bad results from prior self-testing, they will likely skip the first attempt because they need to study more to pass the assessment.
3. When planning, if students outline what they already know about the content (i.e., prior knowledge activation; Greene & Azevedo, 2009), their decision to take or skip the attempt will differ based on their perceived level of their prior knowledge, although the accuracy of making that judgment can vary. If they judge themselves as having high prior knowledge, they are more likely to take the first attempt. If they judge themselves as having low prior knowledge, they are more likely to skip the first attempt and study the material first.
4. When planning, some students may assess how easy it will be to get the correct answer (i.e., ease-of-learning judgment; Dunlosky et al., 2006). If student assess that the assessment will be easy, they will likely take the first attempt. If students assess that it will be difficult, they will likely skip the first attempt and take the time to study.

These assumptions are theoretically based on previous work that empirically tested how students use the processes proposed by theories of SRL and metacognition (Dunlosky et al., 2006; Greene & Azevedo, 2009). It is important to note, however, that these decisions will be made based on student *perceptions* of their prior knowledge or ease of the material, not what they actually know or do not know. For example, a student might judge high prior knowledge of the material, leading to taking the first attempt. However, if their judgment was incorrect, they will likely fail the first attempt.

In addition, our assumptions are based on students who are going to engage in self-regulated learning while completing the course modules. While the decision to skip the first attempt is likely influenced simultaneously by multiple internal factors, an overarching prediction based on SRL frameworks is that if they are effectively self-regulating when engaging with the learning modules, then students with certain internal states (low prior knowledge, low confidence level), are more likely to take the opportunity of skipping the first attempt than students with different internal states (higher prior knowledge, very high levels of confidence). On the other hand, for students who are not actively engaged in self-regulation or whose self-regulation is ineffective, then the decision to skip the first attempt will be made independent of their internal states, or at least with less correlation with their incoming knowledge and confidence. From a learning analytics perspective, we can examine whether students are effectively self-regulating by measuring how the new design impacts existing learning tactics. For example, if low confidence or low prior knowledge is the major determining factor for student decision-making, then under the new design, there will be far fewer students making a failed first attempt since most of them will choose to skip it. However, the number of students making a successful first attempt will not be impacted as much because of their high prior knowledge and/or confidence. On the other hand, if we observe that all existing learning tactics reduce by roughly the same proportion under the new design, then evidence shows that the decision to skip the first attempt is not made based on effective self-regulation.

We also want to note the complex nature of self-regulated learning, such that SRL theory states that there are *several* internal and external factors that might impact self-regulation. These factors can be cognitive, affective, metacognitive, or motivational in nature (Azevedo & Taub, 2020). Since the focus of our paper is self-regulation, we only address cognitive and metacognitive regulatory factors but acknowledge that student decisions to take or skip the first attempt can be affective or motivational in nature as well. We also acknowledge other potential frameworks that can be used to conceptualize our work (e.g., economic model; Barkley & Coffey, 2018; Van Witteloostuijn, 1990). However, we conceptualize this work within SRL theory because of the link between self-regulation and the mastery learning design of the online modules we investigated for this study.

1.5. Research Questions

The current study will answer the following four research questions:

RQ1: *Learning differences caused by extraneous factors*. How much difference in student learning is present between different student populations on the same learning modules with identical learning design?

Answering this question is key to establishing a baseline measure of fluctuation in student learning under identical learning design, providing control over variations caused by extraneous factors, which was absent in previous research.

RQ2: *New tactics by new design*. What new learning tactics are detected because of the new learning design, and how frequently do students adopt those new tactics?

Answering this research question essentially confirms the observation of previous research that different learning behaviour is associated with different LD, but in the context of the current LD of mastery-based online learning. RQ1 and RQ2 also provide the foundation for answering the next two research questions that are key to the current study.

RQ3: *Change in learning tactics*. How do student learning tactics and learning behaviours change in response to the improvement in the LD?

RQ4: *Change in learning outcome and time-on-task*. How do student learning outcomes and time-on-task change in response to the improvement in the LD?

RQ3 and RQ4 are the core research questions of the current study, which are intended to establish a certain level of causal relation between LD and student learning. Following the definition of Lockyer et al. (2013), RQ3 can be seen as mostly detecting the change-to-process measure of learning, while RQ4 gauges the change-in-checkpoint measure.

2. Methods

2.1. Instructional Conditions

2.1.1. Implementation of Online Learning Modules

The online learning modules are created and hosted on the Obojobo learning objects platform, an open-source platform developed by the Center for Distributed Learning at the University of Central Florida. In the current iteration, the assessment component of each module contains one to two multiple-choice problems and permits a maximum of five attempts. The first three attempts are sets of isomorphic problems assessing the same content knowledge with different surface features or numbers. On the fourth and fifth attempts, students are presented with the same problems as in the first and second attempts, respectively, and are awarded 90% of the credit if answered correctly. No problem-specific feedback is provided after each attempt. The instructional component includes text, figures, videos, and practice problems. Each module sequence contains between eight and twelve modules, which students must complete in the order given, with completion defined as either passing the assessment or using up all five attempts.

2.1.2. Demographic and Instructional Backgrounds

Data from the current study are collected from four sections of the same introductory level calculus-based university physics course, with one section taught in Fall 2021 and three sections in Spring 2022. Section names, enrolled students, demographic backgrounds, and mode of instruction are listed in Table 1. In this table, "URM" stands for Under-Represented Minority, classified as students who self-identify as belonging to a race or ethnicity traditionally under-represented in STEM fields, according to the definition of the National Science Foundation; "1st Gen" stands for students whose parent(s)/legal guardian(s) did not complete a bachelor's degree.

Table 1. Demographics and Instructional Conditions of Class Sections Included in the Current Analysis

sections	N	Female	URM	Transfer	1st Gen.	Semester	Course Mode
CtrlSec_1	261	23%	34%	17%	3%	Spring 22	Lecture
CtrlSec_2	115	23%	35%	9%	6%	Spring 22	Lecture
CtrlSec_3	112	35%	33%	23%	16%	Fall 21	Blended
ExpSec	265	25%	41%	15%	8%	Spring 22	Blended

Of the four sections, CtrlSec_3 and ExpSec were taught by the same instructor using a blended classroom format. In those two sections, students watch pre-recorded lecture videos and discuss problem-solving skills during class time. CtrlSec_1 and CtrlSec_2 were taught by two different instructors using a more traditional lecture instructional format. Online learning modules are being used as homework assignments in all four sections (independent of the pre-recorded lecture videos). The homework assignment requirements—including course credit, passing criteria, and due dates—are identical in all four sections since the online homework assignments were copied from one central template course on Canvas. Instructors of CtrlSec_1, CtrlSec_2, and ExpSec had regular weekly meetings during the Spring 2022 semester to ensure synchronous progression of instruction.

2.2. Study Design and Control of Extraneous Factors

Data for the current study were collected from three sequences of online learning modules: Seq1, Seq2, and Seq3, implemented in all four sections. As shown in Table 2, the three sequences were assigned consecutively, covering the course content from week 9 to week 13 of the semester.

Table 2. Number of Learning Modules and the Instructional Content

Sequences	# mod.	weeks	Topic
Seq1	9	9, 10	Linear momentum and collision
Seq2	12	11, 12	Rotational Kinematics
Seq3	8	13	Angular Momentum and Angular Collision

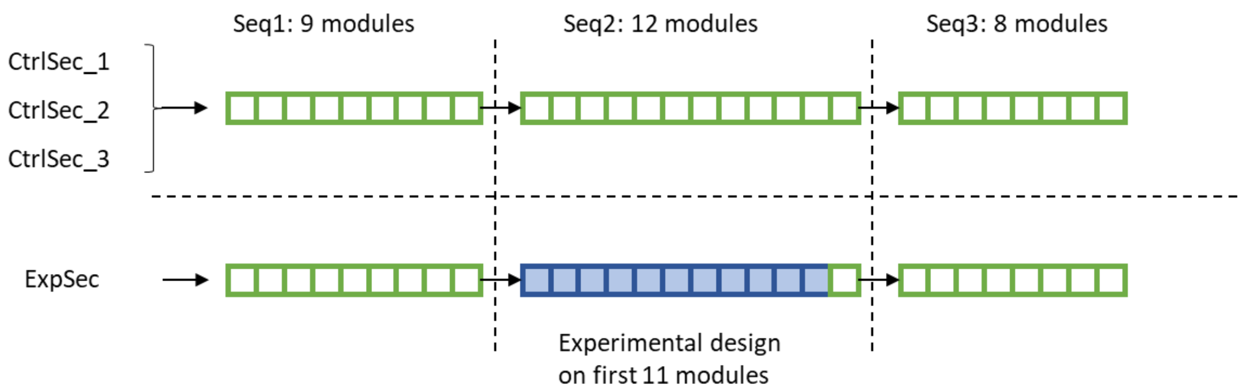

Figure 2. Design of the current study. Each green rectangle represents one original design module, while each blue rectangle represents one experimental design module. The three control sections completed all three module sequences containing original design modules only, while the experiment section completed the two control sequences containing all originally designed modules, and Seq2 with 11 out of the 12 modules under the new design.

The new optional first-attempt design was implemented in the ExpSec, the first eleven online learning modules in Seq2 (see Figure 2). The twelfth module was not replaced since it was designed as an assessment module with no learning material and only an assessment component, so the new design would not make any difference to student behaviour (for readers interested in learning more about module 12, please refer to Whitcomb et al., 2021).

Through this quasi-experimental design, we control for the two types of extraneous variations mentioned in the introduction section that could have significant impact on student learning. First, by using Seq1 and Seq3 as de facto control sequences, we could control for variances that arise from extraneous factors such as instructional modes, demographics, class schedule, or other factors. This is achieved by measuring the normal baseline variance in student learning behaviours and outcomes between all four sections on the two sequences that have identical learning design. The baseline variance measured will be compared to observed differences on Seq2 to ensure that any differences between sections are greater in magnitude than the baseline variations.

There are two main considerations for not including data from weeks 1–8 as part of the baseline measure. First, including data from longer periods of the semester would introduce more extraneous variables that are harder to account for, some of which may be cumulative. For example, holiday schedules are different between the Spring and Fall semesters. When considering only the time periods immediately before and after the administration of the experimental sequence, the impacts of those factors are either not present or much less significant. Second, the computational cost (in particular the memory requirement) for the clustering analysis rapidly increases with the size of the dataset, and three sequences is the upper limit on a current high-end desktop computer. While clustering of larger datasets can be achieved by utilizing advanced cloud computing resources, the resulting method developed will be much harder for average instructors to adopt for frequent evaluation of new learning designs.

Moreover, by using Seq2 as the de facto experiment sequence, and having the new learning design being implemented on only one section, we ensure that the comparison of new and existing learning design is conducted on the same content and in the same relative location in the course (in this case weeks 11 and 12), which eliminated the possible impact of those two extraneous factors.

ISSN 1929-7750 (online).

2.3. Data Collection and Analysis

2.3.1. Data Collection and Data Cleaning

Students' click-stream log data collected from the Obojobo platform is first processed into attempt events and study events. An attempt event starts when the student enters the assessment page of the module and ends when the student clicks the submit button on the assessment page. An attempt event is labelled as a "pass" only if the student correctly answers all questions in the assessment and a "fail" otherwise. A study event starts when the student clicks on any page in the instructional component of the module and ends on either the last click event before a new assessment attempt is initiated or the last record for the student in the module. The duration of the study event is calculated as the sum of all the time spent interacting with each instructional page minus the duration of any inactive periods logged by the platform, including browser close, browser out of focus, and browser inactive after 10 minutes.

2.3.2. Identifying Clusters of Learning Behaviour Using Multi-Level Clustering Methods

The method that we use to process log data and identify clusters of learning behaviour is mostly the same as described in Chen (2022) and Zhang et al. (2022). Here, we give a brief introduction to the data-cleaning process and clustering algorithm and highlight the aspects unique to the current study. Readers interested in a more detailed explanation of the algorithms and justifications for implementation choices are encouraged to refer to earlier papers.

Step 1: Distinguishing between brief and normal attempts. Abnormally short attempts (referred to hereafter as brief attempts) are separated from normal length attempts by fitting the log-duration distribution of all attempts on a single module using a finite mixture of either normal or skewed distribution models using the R package mixsmsn (Prates et al., 2013), following the fitting procedure described in detail in Chen (2022). The brief–normal attempt cutoff is different for each module to account for the fact that some problems take less time to solve than others. We use the term "brief attempts" to be consistent with earlier studies of learning modules (Taub et al., 2022; Zhang et al., 2022).

Step 2: Calculating the distance between two event traces. Each student's interaction with one online learning module is represented by a sequence of events such as {Attempt_1_S, Study, Attempt_2, Attempt_3}. Each sequence ends either when the student passes one of the attempts or when all five allowed attempts are used up. Distance between two sequences is determined by the following seven features that capture student interactions with the learning module and self-regulation process:

1. Total number of assessment attempts
2. Number of attempts before study
3. Fraction of brief attempts among all attempts
4. Whether the first attempt is a brief attempt
5. Whether the last attempt is a brief attempt
6. Did the student abort the module?
7. Whether the first event is a study event

The first six features are identical to what was used in Zhang et al. (2022) and Taub et al. (2022), and the seventh feature is added to differentiate traces from the new optional first-attempt design from the rest of the traces. For feature number six, abort means that the student stopped interacting with the module without passing the assessment or using up all the attempts, which is most often observed on the last module in a sequence.

The distance metric between two event traces is computed using the Gower dissimilarity coefficient (Gower, 1971) since features 1, 2, and 3 are numeric while features 4, 5, 6, and 7 are binary. Following the precedence of our earlier study, the weights for features 5 and 6 are set to 0.5, and all other weights are set to 1.0. This selection of feature weight emphasizes the forethought phase of self-regulation and was previously shown to produce the most well-defined cluster structure among several other sets of possible weight selections that emphasize other phases of the SRL process.

Step 3: K-medoid Trace clustering and process map generation. K-medoid clustering was performed on all traces using the R package cluster (Maechler et al., 2023). The number of clusters is determined by the average silhouette value, a measure of the ratio of intra- and inter-cluster variability described by Rousseeuw (1987). We chose the number of clusters by selecting the maximum average silhouette for a cluster number of less than 10 clusters since higher cluster numbers lead to significantly harder-to-interpret results. The main characteristic of each cluster is visualized by creating process maps (PMs) using the R package processmapR (Janssenswillen et al., 2023), and the top 80% of most frequent traces in each identified cluster was selected to reduce the "spaghetti effects" of process maps, a common visualization procedure (see Janssenswillen et al., 2023; Zhang et al., 2022).

2.3.3. Process and Checkpoint Measures of Student Learning

To evaluate the impact of the new learning design on student learning, we employ both process measures and checkpoint measures, as recommended by (Lockyer et al., 2013). The process measure we choose to examine is the frequency distribution of identified learning tactic clusters on each module, i.e., what fraction of students adopted each identified learning tactic. To quantify the differences between two such frequency distributions measured from any two sections in the study, we use the Hellinger distance, first introduced by Ernst Hellinger in 1909. For any two discrete distributions $P = (p_1, \dots p_k)$ and $Q = (q_1, \dots q_k)$, the Hellinger distance is defined as:

$$H(P,Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{k} \left(\sqrt{p_i} - \sqrt{q_i}\right)^2}$$

The metric is bounded between 0 and 1, and is 1 when distribution P assigns probability 0 to every set to which Q assigns a positive probability, and vice versa.

We will answer RQ3 using the H distance following a three-step process, which is explained in more detail in the results section. First, the baseline H distance variation between different sections under identical LD is measured on the two control sequences on all four sections, as explained in 0. Second, for the experiment section, we use data from the two control sections and construct a hypothetical frequency distribution, assuming that students are engaged in ineffective self-regulation. The process is explained in detail in 0 and 0. Third, we compare this hypothetical distribution to the actual observed distribution on the experimental section to verify whether students are effectively self-regulating and, if so, whether their decision-making is likely based on their prior knowledge and confidence. The analysis process is explained in detail in 3.1.5.

Two types of checkpoint measures are considered in the current study: 1) percentage of students passing each learning module and 2) students' frequency of accessing and time spent on the learning materials of each module. For the first type, since there are multiple ways for a student to pass a module, we chose to focus on the following three most frequently observed types of passing:

Normal Pass After Study (NPAS): Students who passed the module on a normal length attempt after accessing the learning material in the module. In the original learning module design, an NPAS event is associated with successful learning from the learning material.

Normal Pass Before Study (NPBS): Students who passed the module on a normal length attempt before accessing the learning material in the module. An NPBS event is associated with strong incoming content knowledge.

Brief Pass Before Study (BPBS): Students who passed the module on a brief attempt before accessing the learning material in the module. A BPBS event is associated with guessing or answer copying on the first one or two attempts.

The implication of each passing type is inherited from multiple prior studies on the same learning modules. For studying the learning module, we define "study" as accessing the learning materials of each module for more than 30 seconds. The 30-second cutoff was estimated based on a mixture model approach like the one used to determine brief and normal attempts. Since the number of pages and length of each page of the learning material for each module is roughly similar, a uniform cutoff of 30 seconds is used for simpler analysis. For time of accessing the learning material, we measure the "median study time" of each student, which is defined as the median of the time spent on studying learning materials on all the modules studied in one learning module sequence. Only students who engaged with three or more modules in each sequence are included in the analysis.

3. Results

3.1. Learning Tactic Clusters

3.1.1. Identified Learning Tactic Clusters

Eight learning tactics are identified from the trace clustering algorithm performed on the entire dataset. The process maps of learning tactic clusters 1–6 are shown in Figure 3, and clusters 6–8 are shown in Figure 4. Each process map is generated with 80% of the most frequent traces to capture the main feature of the cluster to avoid the so-called spaghetti effect. The order of the clusters was determined randomly based on the order of input data and has no significance. Brief attempts are labelled with an "S" suffix for "short," such as "Attempt_1_S," to distinguish from normal attempts ("Attempt_1"). The number on the arrow between the "Start" node and the first node of each map indicates the number of traces in each process map. The main characteristics of each cluster are summarized in Table 3 in terms of initial attempt, study decision, and attempts after study.

Table 3. Main Characteristics of Each Learning Tactic Cluster

Cluster	Initial Attempt Duration	Study	Passing Attempt Duration
1	Normal (Pass)	NA	Normal
2	Brief (Pass)	NA	Brief
3	Brief	Yes	Normal
4	Normal	No	Brief
5	Normal	Yes	Normal
6*	None	Yes	Normal
7	Brief	Yes	Brief
8	Normal	Yes	Brief

**Note: Cluster 6 is only available under the new design.*

Cluster 1 (initial normal pass): Traces in this cluster pass the module on the first attempt with normal duration before accessing the learning material.

Cluster 2 (initial brief pass): Traces in this cluster pass the module on the first attempt with brief duration.

Cluster 3 (initial brief, study, normal pass): Traces in this cluster make an initial incorrect brief attempt, then access the study material. Most pass the module on their second attempt, while some take 3–4 normal attempts to pass.

Cluster 4 (initial normal, no study, brief pass): Traces in this cluster make an initial normal attempt but then pass the module on one or more following brief attempts without accessing the learning material.

Cluster 5 (initial normal, study, normal pass): Traces in this cluster begin with an initial failed normal attempt, then proceed to study and pass on a normal second attempt.

Cluster 6 (study and normal pass): Traces in this cluster skip the initial attempt and go directly to the study material, then pass the module on the next normal attempt. All the traces in cluster 6 were from the ExpSec in Seq2.

Cluster 7 (initial brief, study, brief pass): Traces in this cluster are like those in cluster 3 in that they start with a brief initial attempt followed by study. However, all of them followed with multiple brief attempts.

Cluster 8 (initial normal, study, brief pass) Traces in this cluster are like those in cluster 5 in that they start with a normal initial attempt followed by study. However, all of them followed with multiple brief attempts.

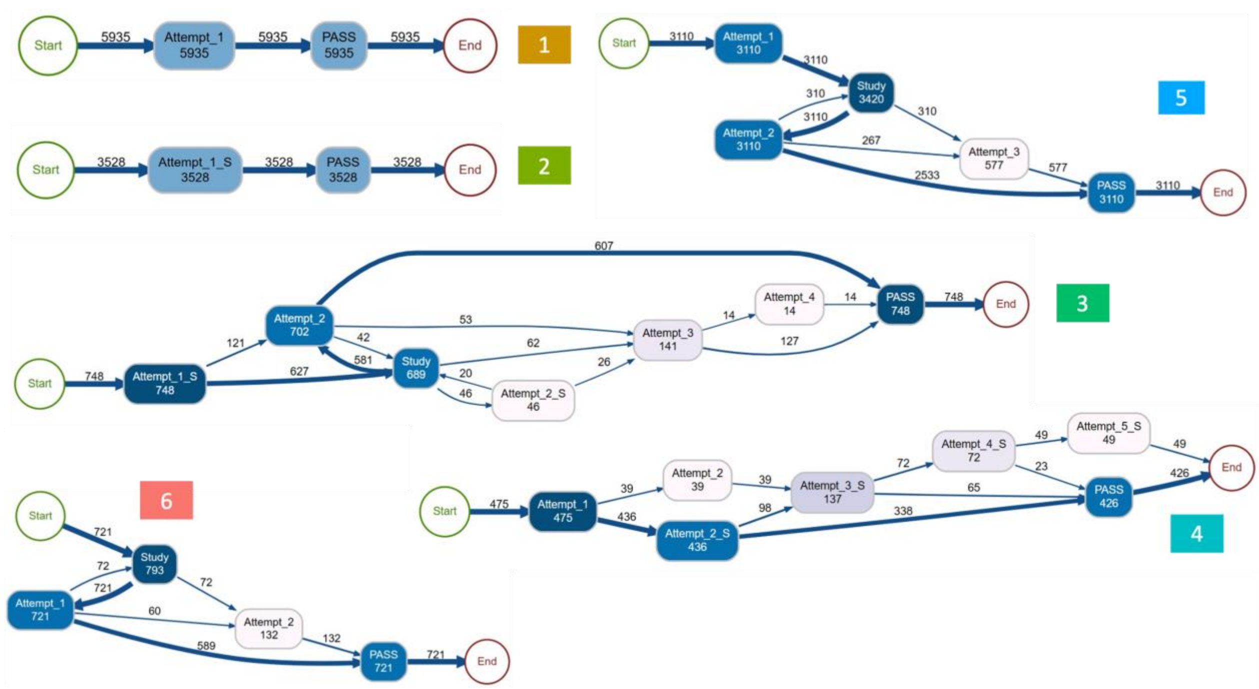

Figure 3. Process maps of identified learning tactic clusters 1–6. Each box represents an attempt, study, or pass event. The colour corresponds to the frequency of the event in the cluster. Each numbered arrow represents a move from one event to another, and both the number and the thickness of the arrow represents the number of traces containing such a move in the cluster.

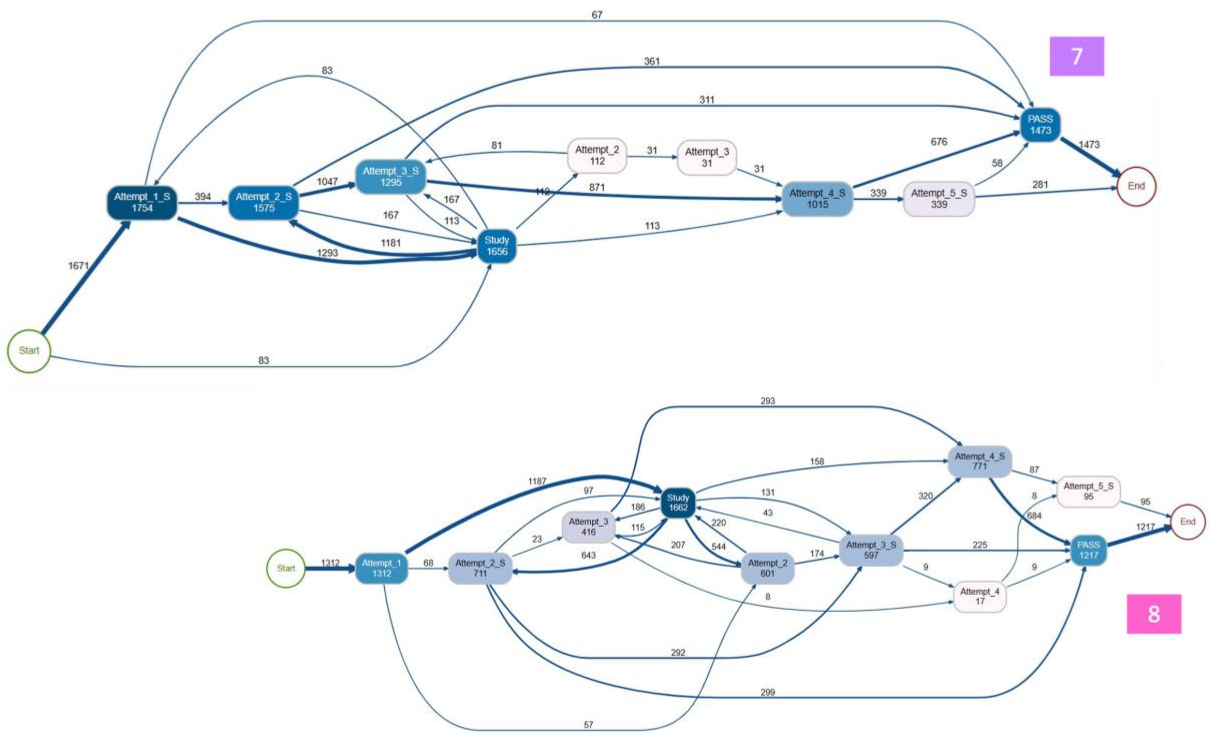

Figure 4. Process maps of identified learning tactic clusters 6–8. Each box represents an attempt, study, or pass event. The colour corresponds to the frequency of the event in the cluster. Each numbered arrow represents a move from one event to another, and both the number and the thickness of the arrow represents the number of traces containing such a move in the cluster.

3.1.2. Similarity of Learning Tactic Cluster Distribution Across Control Sequences

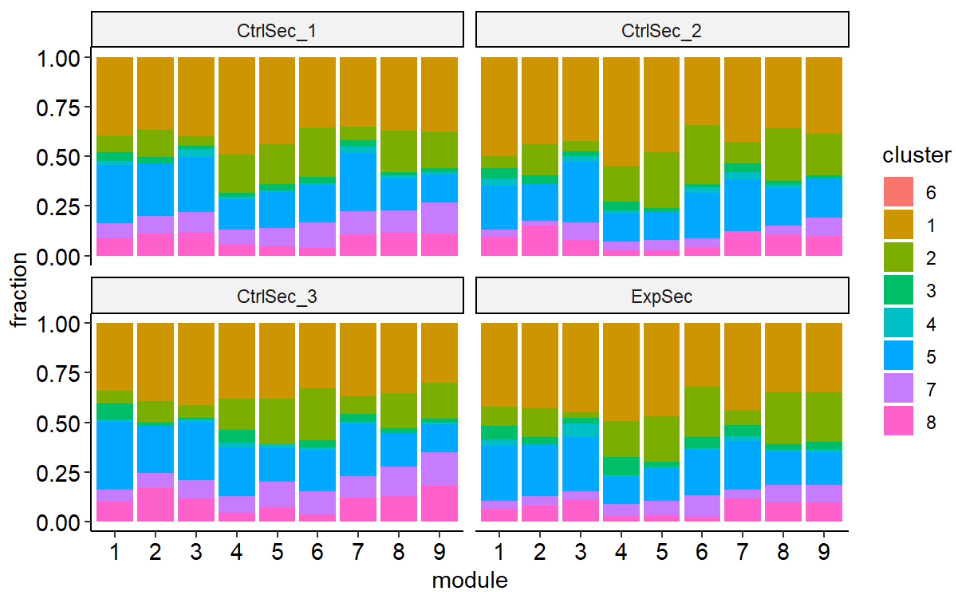

Figure 5. Distribution of learning tactic clusters on each learning module in Seq1 from all four sections.

The stacked bar charts in Figure 5 show the fraction of each strategy cluster on each module in Seq1. The cluster distribution in Seq3 shows a very similar pattern and is not shown here. The cluster distributions between the four sections are quite similar. For any learning tactic cluster that is adopted by, on average, >20% of the student population, the maximum standard deviation of normalized cluster size on any cluster in any module between the four sections is 20% of the mean cluster size or about 7% of the total student population.

To qualitatively examine the uniformity between the four sections on the two control sequences, we calculated the Hellinger distances between every pair of two different sections for each module of the 17 modules contained in the two control sequences. The median Hellinger distances between every two sections are plotted in Figure 6, with the thickness and darkness of each edge corresponding to the median distance.

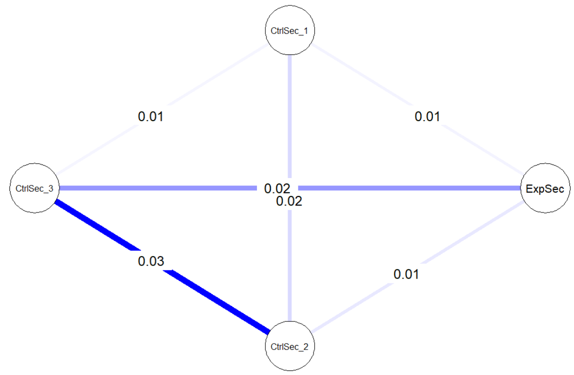

Figure 6. Median Hellinger distances for all 17 modules between every two sections, taken from the two control sequences. The thickness and colour of the lines correspond to the median value.

The Hellinger distances, bound between 0 and 1, are all quite small between each two sections. A Kruskal-Wallis one-way ANOVA on ranks test can be conducted to determine if the set of 17 Hellinger distances between any two sections are consistently larger or smaller than the set of Hellinger distances between another pair of two sections. In other words, the test shows whether the differences in learning tactic cluster distribution between any two sections are consistently larger or smaller than the differences between all other pairs of sections. The test reveals that there is a significant difference in Hellinger distance distributions ($H = 23.2, df = 5, p < 0.01$). Post-hoc analysis using the pairwise Mann-Whitney U test and p-value adjustment confirms that the set of H-distances between CtrlSec_3 and CtrlSec_2 (median H-distance 0.03) is consistently larger ($p < 0.02$) than the set of H-distances between the ExpSec and the other two control sections (median H-distance 0.01). Since CtrlSec_3 is slightly different from ExpSec than CtrlSec_2, we chose to exclude CtrlSec_3 from the analysis in sections 3.1.3 to 3.1.5, which compare the cluster distribution (process measure) on Seq2 with the ExpSec and the control sections. Doing so will make sure that student data from the three remaining sections on the control sequences are as like each other as possible. For those three sections, the differences in Hellinger distance between them are not significantly different according to the same test. Note that data from CtrlSec_3 is still included in the analysis performed on Seq1 and Seq2, as well as on all the checkpoint measures.

3.1.3. Strategy Cluster Distribution on the Experimental Sequence (Seq2)

As shown in Figure 7, the cluster distribution of CtrlSec_1 and CtrlSec_2 is still highly similar on Seq2, with a median Hellinger distance between the two sequences of 0.02, identical to that of Seq1 and Seq3. However, the distribution for the ExpSec is significantly different for the first 11 modules, with 25–40% of the students adopting the “skip first attempt” strategy (cluster 6) and a median Hellinger distance at 0.3.

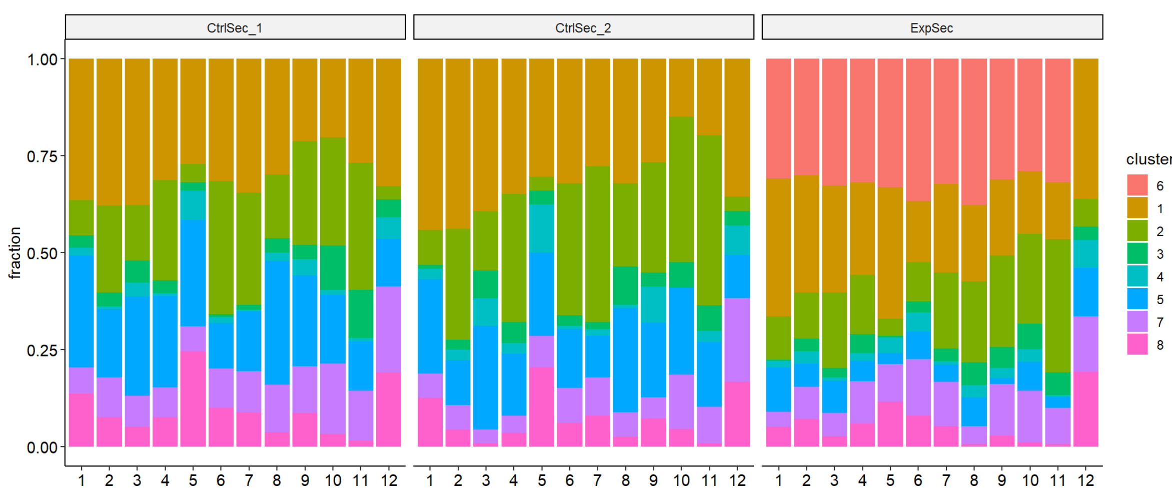

Figure 7. Distribution of learning tactic clusters on the experimental sequence (Seq2) from two control sequences and one experimental sequence.

For any learning tactic cluster (except cluster 6) that is adopted by on average >20% of the student population, the maximum standard deviation for normalized cluster size on any cluster in any module between the three sections is 60% of the cluster size, or about 14% of the student population. The Hellinger distances between ExpSec and the other two sections on the first 11 modules range between 0.2 and 0.4, which are all one order of magnitude greater than what was observed for the control sequences. The maximum Hellinger distance of module 12 is 0.013, which is on par with the observations for the control sequences.

3.1.4. Impact of New Learning Tactic Cluster on Existing Learning Tactic Clusters

To estimate how much the other learning tactic clusters reduce in size as a result of students adopting the new tactic cluster, we first construct a hypothetical tactic distribution, assuming that all students adopt the new tactic cluster at random and independent of the tactic they would have adopted under the original design. Under this hypothesis, the fraction of all the other tactic clusters should be reduced by the same fraction. Therefore, we first take the average of the cluster fraction distributions of CtrlSec_1 and CtrlSec_2, then perform the following transformation for cluster i on module j:

$$\hat{p}_{i,j} = \bar{p}_{i,j} \times (1 - p_{6,j})$$

Here, $\hat{p}_{i,j}$ is the hypothetical cluster fraction for cluster i, with $i \in \{1,2,3,4,5,7,8\}$, on module j, with $1 \leq j \leq 11$. $\bar{p}_{i,j}$ is the average cluster fraction of the two control sequences, and $p_{6,j}$ is the observed fraction of cluster 6 on the ExpSec. The fraction of cluster 6 on the hypothetical cluster distribution is taken as identical to what is observed for the ExpSec.

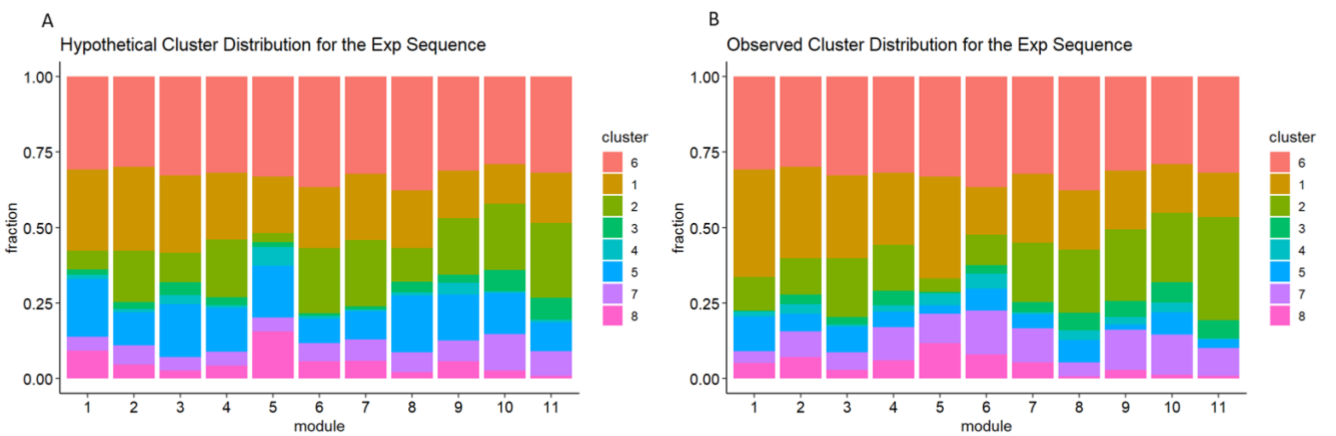

Figure 8. Hypothetical and Observed distribution of learning tactic cluster for the experimental sequence. A: Hypothetical cluster distribution assuming random adoption of new strategy. B: Observed cluster distribution from the experiment section ExpSec.

The hypothetical distribution obtained from this operation is shown in Figure 8A. For better comparison, the observed cluster distribution of the first 11 modules from the ExpSec is shown again in Figure 8B. The justification for constructing such a hypothetical distribution from the two control sequences is that on both Seq1 and Seq3, the cluster distributions between the three sections are essentially indistinguishable from each other. Therefore, we can assert that the average cluster distribution of CtrlSec_1 and CtrlSec_2 would still be indistinguishable from that of the ExpSec, had the new design not been implemented.

For each module, a chi-square goodness of fit test is conducted to test if the observed cluster distribution is significantly different from the hypothetical cluster distribution. The test result is significant for every module after adjusting p-values for multiple comparisons ($p.adj < 0.01, \chi > 22.4$). Since some observed clusters have small numbers of participants on some modules (such as clusters 3, 5, and 8), which may violate the chi-square distribution hypothesis, we conducted the same test by combining all clusters with fewer than 10 counts for the ExpSec. The result is still statistically significant for all modules except module 10.

3.1.5. Quantifying the Heterogeneous Shrinkage of Strategy Clusters

To quantify the shrinkage of clusters as compared to the random adoption hypothetical distribution, we define the "shrink factor" for each strategy cluster i on module j as:

$$\Delta_{i,j} = \frac{\hat{p}_{i,j} - p_{i,j}}{\hat{p}_{i,j}}$$

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

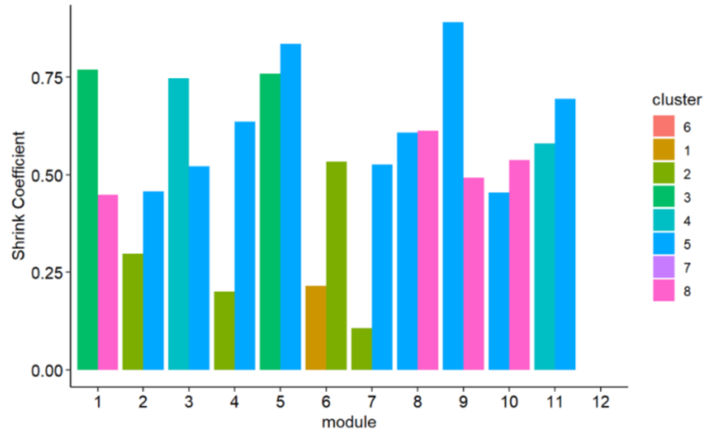

Figure 9. The top two learning tactic clusters that have the largest shrink factor for each of the 11 modules in the learning module sequence. The bar heights correspond to the magnitudes of the shrink factor, and the colour corresponds to the learning tactic cluster.

Here, $\hat{p}_{i,j}$ is the hypothetical fraction for cluster i on module j, while $p_{i,j}$ is the actual observed fraction of the cluster. $\Delta_{i,j} = 1$ when cluster i completely disappeared on module j, while $\Delta_{i,j} = 0$ when cluster i is as big as the hypothetical distribution. $\Delta_{i,j} < 0$ when the observed cluster is bigger than the hypothetical distribution and can theoretically be negative infinity if the hypothetical distribution was predicted to be zero.

In Figure 9 we plot the magnitude of $\Delta_{i,j}$ of the two clusters with the largest positive shrink factor for each of the 11 modules in Seq2. Cluster 5 (normal initial, study, normal pass) appeared nine times on this graph, with $0.45 < \Delta_{5,j} < 0.89$. Cluster 2 (brief initial pass) and cluster 8 (normal initial, study, brief pass) both appeared four times on the graph, with cluster 2 observed more frequently in the first six modules, and cluster 8 among the last five modules. The shrink factor of cluster 2 $(0.11 < \Delta_{2,j} < 0.53)$ is mostly smaller than that of cluster 8 $(0.44 < \Delta_{5,j} < 0.61)$. However, as seen in Figure 8, the size of cluster 8 is much smaller than that of cluster 2 on each module.

In contrast, cluster 1 (pass on normal attempt) is only among the top 2 on module 6 alone. In fact, $\Delta_{1,j} < 0$ for $j \neq \{6, 11\}$. In other words, except for modules 6 and 11, on all other modules the observed cluster 1 is larger than what is predicted by the random adoption model. When compared to the average cluster size of cluster 1 from CtrlSec_1 and CtrlSec_2, cluster 1 from ExpSec is 10–30% smaller on all modules except 6 and 11, where the expected level of shrinkage assuming random adoption is 25–40%.

3.2. Assessment Outcome

As explained in detail earlier, in the current study we measure three different passing types for each learning module: Normal passing after study (NPAS), normal passing before study (NPBS), and brief passing before study (BPBS).

3.2.1. Assessment Outcomes on Seq1 and Seq3

In Table 4, we list the average number of modules passed by each student in each of the three different passing types for the two control sequences, Seq1 and Seq3. The average number of modules passed is very similar between all four sections on the two control sequences.

Table 4. Average Numbers of Modules Passed on Each of the Three Passing Types; Data is Collected from the Two Control Sequences

Sequence	Section	NPAS	NPBS	BPBS
Seq1	CtrlSec_1	1.9	3.4	1.5
Seq1	ExpSec	1.9	3.5	1.7
Seq1	CtrlSec_2	1.8	3.6	2.0
Seq1	CtrlSec_3	1.9	3.1	1.5
Seq3	CtrlSec_1	1.6	2.9	1.6
Seq3	ExpSec	1.6	3.0	1.8
Seq3	CtrlSec_2	1.4	3.2	1.9
Seq3	CtrlSec_3	1.6	2.9	1.6

For each passing type, we counted p_n, the number of students who passed n modules, with n ranging from 0 to the maximum number of modules in that sequence. We then conducted a four-way Friedman's rank sum test for each sequence to examine if the distribution of p_n is significantly different between the four sections. No significant differences were found ($p > 0.5, Q < 2.03$).

Next, we also compared the following three fractions across the four sections for each of the 17 modules on the two control sequences:

1. The fraction of NPAS students among all students who studied the learning material.
2. The fraction of NPBS students among all who finished the module.
3. The fraction of BPBS students among all who finished the module.

The denominator for NPAS passing is different from the other two since NPAS is an indicator of student ability to effectively learn from the learning materials. Therefore, using all students who studied the learning material would ensure that the fraction is not affected by the number of students who initially passed the module without learning from the material.

We define "study" as having spent more than 30 seconds on the learning material of each module, and "finish" as having either passed the module or used up all 5 attempts. For each of the three fractions, we conducted a Fisher's exact test on each module, and the p-values were adjusted using Holm's method (Holm, 1979).

Only one significant difference was found for one passing fraction on one out of the 17 modules: module 9 on Seq1 ($p.adj = 0.03$) for the NPAS fraction. A post-hoc multi-comparison Fisher's exact test with a p-value adjustment revealed that CtrlSec_3 (23%) had a lower NPAS fraction than CtrlSec_2 (52%) and ExpSec (47%; $p.adj = 0.02$).

3.2.2. Assessment Outcome on the Experimental Sequence

In Table 5 we list the average number of modules passed by each student in each of the three different passing categories on the experiment Seq2. The ExpSec has the highest average number of NPAS, and lowest average number of NPBS and BPBS on that sequence. However, none of the differences are statistically significant according to the same Friedman's test ($p > 0.4, Q > 2.7$).

Table 5. Average Numbers of Modules Passed on Each of the Three Passing Types; Data is Collected from the Experiment Sequence

sequence	section	Avg. NPAS	Avg. NPBS	Avg. BPBS
Seq2	CtrlSec_1	2.7	3.8	2.4
Seq2	ExpSec	4.1	3.1	1.9
Seq2	CtrlSec_2	2.5	4.0	3.1
Seq2	CtrlSec_3	3.0	3.6	2.1

On the other hand, when we examine the three passing fractions for each of the 12 modules in this sequence, we found several significant differences between two or more sections, according to Fisher's exact tests and post-hoc pairwise comparison.

The NPAS fraction is significantly different between sections on modules 1, 2, and 11. As shown in Figure 10, post-hoc analysis reveals that on modules 1 and 11, the NPAS fraction of the experiment section is significantly higher than that of CtrlSec_1 ($p < 0.01$). On module 2, the experimental section's NPAS fraction is significantly higher than that of CtrlSec_1 ($p < 0.01$) and CtrlSec_3 ($p < 0.03$). As shown in Figure 11, on modules 2 and 6, the NPBS fraction of ExpSec is significantly lower than all of the control sections by about 15–20% (pairwise Fisher's test with p-value adjustment $p < 0.01$).

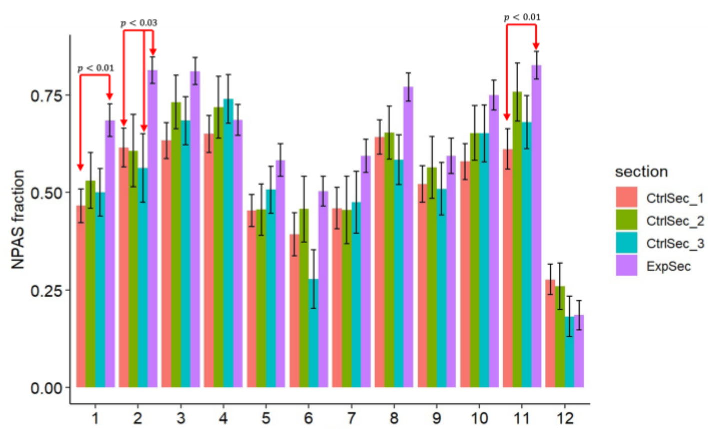

Figure 10. Comparison of NPAS fractions on the experiment sequence between the four sections. Statistically significant differences are highlighted by red arrows.

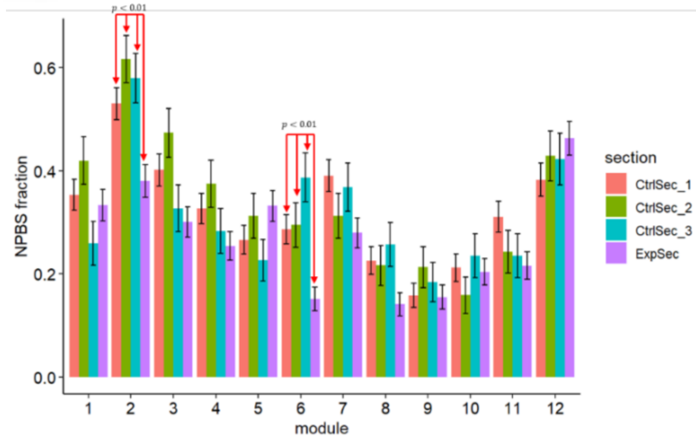

Figure 11. Comparison of NPBS fractions on the experiment sequence between the four sections. Statistically significant differences are highlighted by red arrows.

Finally, on modules 4, 6, and 7, the BPBS fractions are significantly different between the sections. Post-hoc analysis reveals that on module 6, the BPBS fraction of the experimental section is significantly lower than that of CtrlSec_1 and CtrlSec_2 by more than 20%. On modules 4 and 7, the BPBS fraction of ExpSec is 20% lower than that of CtrlSec_2 ($p < 0.01$), as shown in Figure 12. An important technical point that must be mentioned is that while we reported the results from the control and experiment sequences separately, the p-value adjustment of the Fisher's exact tests were conducted with all 29 modules (29 p-values) from all three sequences together for each passing fraction using Holm's method.

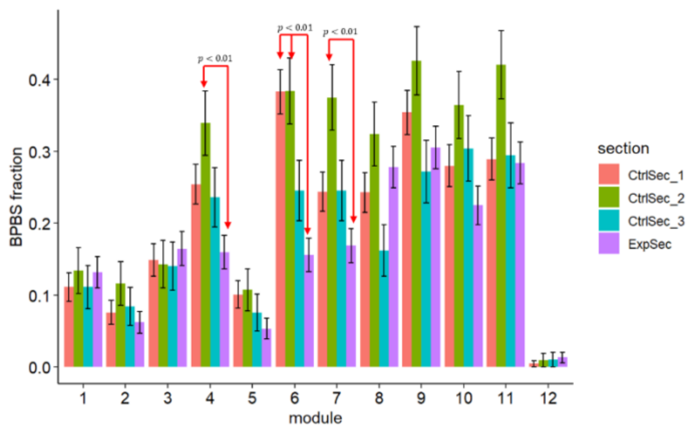

Figure 12. Comparison of BPBS fractions on the experiment sequence between the four sections. Statistically significant differences are highlighted by red arrows.

3.3. Decision to Study the Learning Materials in Each Module

3.3.1. Study Decision and Study Time on Control Sequences

First, we examined whether there is a difference in the number of learning modules in which students chose to study the learning materials. We counted the number of students who studied n number of modules, with n ranging from 0 to the maximum number of modules in each sequence. We conducted a Friedman's rank sum test on the distribution of the number of students that studied n modules across each of the four sections and did not find any statistically significant differences ($p > 0.95, Q < 0.34$) on Seq1 and Seq3.

In addition, we also examined the fraction of students who studied none of the modules, and found that it varied between 26–36% on Seq1, and 34–42% on Seq3. A 2×4 Fisher's exact test is conducted for each of the two control sequences, and none of the sections had a significantly different fraction ($p > 0.14$).

Third, we examined whether there were any differences in students' median study time between the four sections. The average median study time ranged from 670s to 970s for Seq1 and 314s to 442s for Seq3. A Kruskal-Wallis one-way ANOVA on ranks test was used to examine if the distribution of median study time was different between the four sections. For the two control sequences, no significant differences were found ($p > 0.85, H < 0.78$).

3.3.2. Study Decision and Study Time on the Experiment Sequence

On Seq2, the number of learning modules studied was similar for all four sections (Friedman's rank sum test, $p = 0.6, Q = 1.76$). On the other hand, the fraction of students who did not study any module did have significant differences between the four sections (Fisher's exact test, $p = 0.03$). A pairwise post-hoc Fisher's test with p-value adjustment reveals that the fraction of students who studied none of the modules in the ExpSec (18%) is significantly lower than in CtrlSec_2 (32%; $p = 0.03$). Finally, there is also a significant difference on the median study time between the four sections according to the Kruskal-Wallis rank sum test ($p = 0.01, H = 10.6$).

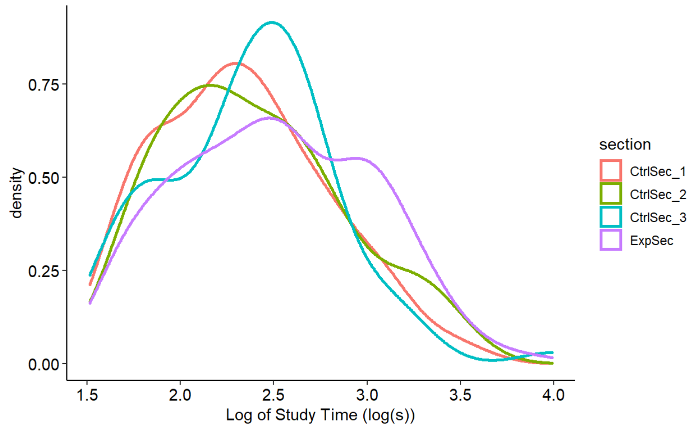

Figure 13. Distribution of median study time for students on the experimental sequence. The median study time is plotted on a log scale.

We plot the density distribution of the median study time for each section in Seq2 in Figure 13. It revealed that students in the ExpSec have overall longer median study times than students in the three control sections. A pairwise Wilcox post-hoc test with p-value adjustment shows that the difference between ExpSec and CtrlSec_1 is statistically significant ($p = 0.01$). The average median study time of the ExpSec is 647s, whereas the average median time for CtrlSec_1 is 397s.

4. Discussion

4.1. Summary of the Findings

We first summarize the main findings of this study in terms of the four research questions.

RQ1: Learning differences caused by extraneous factors. Overall, for the two control module sequences with identical LDs, both the process measures and checkpoint measures showed little difference between the four sections, despite all the extraneous factors that could potentially impact student learning, including student population and instructional mode.

For the process measure, the distribution of identified learning tactic clusters on each module is highly similar; the maximum standard deviation of any cluster on a module is about 7% of the total population. When the differences are quantified using Hellinger distance, the median H distance between any two sections is 0.01. The only exception is the difference between CtrlSec_3 and the other three sections, which is significantly larger. However, the magnitude of the difference is still quite small (median H distance of 0.03).

For checkpoint measures, there were no detectable differences in either the number of modules students decided to study or the mean duration of the study sessions. The three types of passing percentages were also very similar across all four sections, with only one significant difference out of the 51 instances tested (three types of passing percentages on 17 modules).

These results indicate that for mastery-based online learning modules, extraneous factors in the current study may only have a limited impact on students' overall learning processes and learning outcomes under the same design. The only detectable difference is in student learning tactics distribution between one section taught in the Fall 2021 semester and the three sections taught in the Spring 2022 semester. One likely explanation is that students in the Fall 2021 semester were still being impacted by restrictions due to the COVID-19 pandemic, whereas the impact was much less in the Spring 2022 semester. An alternative explanation is that CtrlSec_3 had higher fractions of transfer and 1st generation students, who may be more likely to adopt different learning tactics compared to their peers. Note that one section in the Spring 2022 semester was also taught using the same blended instructional mode as the Fall 2021 semester, so the instructional mode may not be the dominant factor for these differences. More longitudinal data will be needed in future studies to identify the main factors contributing to variations in student learning tactics.

RQ2: New tactics by new design. Only one new learning tactic cluster was detected, which corresponds to the skip first attempt, study, and pass tactic, adopted by about 25–40% of the student population in different modules. What is more noteworthy is the absence of a "skip first attempt, study, and multiple short attempts" tactic cluster. In both the original and the new learning design, making multiple and mostly short attempts after studying is a frequently observed learning tactic used by students (clusters 7 and 8), which indicates that the study event is less effective in helping students master the content. However, for students who chose to skip the first attempt, their learning mostly resulted in successfully passing the assessment.

One possible explanation of this result is that the decision to skip the first attempt is a deliberate one taken predominantly by students who are effectively self-regulating and meaningfully engaging with the learning modules by metacognitively monitoring their emerging understanding as they are learning the content, and therefore are more likely to have successful learning outcomes. On the other hand, students who are less engaged in their learning may not even realize that the design of the modules has changed so they are more likely to continue using their original tactic. Another possibility is that those who skip the first attempt are more likely to have goals oriented toward mastering the content, or the opportunity to skip the first attempt could have caused more students to have mastery-oriented goals. Therefore, their following actions are aligned with this goal, so they are more successful in learning and less likely to guess on the follow-up assessment attempt.

Another possible cause is that many students with higher incoming knowledge chose to skip the first attempt due to having performance-dominant goals and are possibly under-confident. In other words, those who could have passed the assessment without studying chose to study "just to be safe." However, additional analysis shows that while this does happen to some extent, it is unlikely to be the dominant factor, which is explained below.

RQ3: Impact on process measure. The impact of the new learning design on student learning tactics is summarized below.

Adoption of the new tactic is not uniform among students and is likely dependent on the student's internal state, such as incoming knowledge and self-confidence. The observed distribution of identified learning tactics among students was significantly different from what was predicted from random adoption of the new tactic (cluster 6). In other words, the introduction of the new learning design resulted in students abandoning some existing tactics more frequently than other existing tactics.

The "Fail-Study-Pass" tactic (cluster 4) reduced significantly more than expected and had either the largest or the second-largest shrink factor on 9 out of 11 modules. The "Short Attempt and Pass" (cluster 2) tactic reduced significantly more than expected among the first few modules, while the "Normal Fail-Study-brief pass" (cluster 8) reduced significantly more than expected among the last few modules. The reduction of "Normal pass before study" (cluster 1) is much less than expected for all but two modules. In fact, on all but two modules, the observed cluster is actually larger than what was expected under random adoption (with a negative shrink factor).

In short, tactic clusters that involve a normal length, passing first attempt did not shrink or shrank much less than predicted under the new design. On the other hand, tactic clusters that involve either a failed initial attempt, or a brief initial attempt, shrank much more than predicted. The most straightforward explanation of this uneven shrinking in cluster size is that students who would have made a failed initial attempt or simply copied the answer from elsewhere are more likely to take on the opportunity of skipping the first attempt. This would suggest that a considerable number of students are engaged in effective self-regulation during the learning process, frequently assessing the likelihood of passing on the first attempt. Those who are less likely to pass the first attempt are much more likely to not take the first attempt at all, compared to those more likely to pass the first attempt. Notably, on harder learning modules towards the end of the sequence, giving students the opportunity to skip the first attempt may have helped them learn better from the learning material, as the traces that require multiple brief attempts after study to pass the module (cluster 8) shrank significantly under the new tactic. On the other hand, the current analysis cannot exclude other possible and more complicated explanations of the observed distribution. The current simple interpretation will require future follow-up studies to validate it.

RQ4: Impact on checkpoint measure. We did not find any significant differences in the three types of passing percentages (NPAS, NPBS, BPBS) between the four sections on all 11 modules with new LD in the experimental sequence. However, we did detect eight instances of significantly different passing percentages on six individual modules. Post-hoc analysis shows that the significant differences are between the ExpSec and either one or both of the control sections. This is much more frequent compared to what was detected on the two control sequences, which only had one instance of significant passing percentage difference. Overall, students seem to be passing more on normal attempts either before or after study, and less on brief attempts. While the observations are not sufficient to support a conclusive claim that the new learning design improved student learning outcomes, they clearly refuted the original learning design philosophy that a mandatory first attempt would make students learn better.

On the experimental sequence, the ExpSec had fewer students studying none of the modules compared to Ctrl_Sec 2, and the median study time was also longer compared to Ctrl_Sec 1. The observations, insufficient to support a conclusive claim, suggest a trend that students study more under the new learning design.

Finally, it is also worth pointing out that the second control sequence is administered after the experimental sequence in the course. The fact that we did not observe any differences on that sequence suggests that having experienced the new learning design did not change how students engage with the original learning design afterwards.

4.2. General Discussion

4.2.1. Implications for Using Learning Analytics to Inform Learning Design

The current study represents a major step forward from existing studies that use LA methods to examine differences in student learning process under different existing LD (Fan et al., 2021; Holmes et al., 2019; Lancaster et al., 2020). In this study, we created a closed loop between LA and LD. First, a deliberate and focused change to the existing LD was made based on previous LA outcomes and the theoretical framework of self-regulated learning. Second, the new learning design was implemented in an authentic learning situation using a quasi-experimental design, which provides control for two types of extraneous variations. Finally, the loop was completed by using learning analytics to evaluate the impact of the improved design on students' self-regulated learning tactics and learning outcomes. This study contains all three elements key to learning analytics as stated in Motz et al. (2023): 1) it studies an attempt to intervene in an existing learning environment, 2) measures its impact on learning outcomes and learning strategies, and 3) uses data collected from existing education environments. The results of this study did not support the design principles of the existing LD but rather supported the proposed new LD by revealing that many students are capable of effective self-regulation under the new LD. More specifically, we showed that students are capable of effectively deciding whether to skip the first attempt based on their internal state. Such a closed loop procedure will enable instructors and researchers to continuously improve the design of learning environments based on solid evidence from LA.

The quasi-experimental design of this study is a significant improvement over previous studies on LA and LD in terms of explicitly accounting for the impact of extraneous factors such as instructional content, time of instruction, and differences in student population and instructional mode. The results suggest that the fluctuations in both process and checkpoint measures caused by extraneous factors are much smaller compared to the differences observed under the experimental condition. While the level of control on extraneous confounding factors in the current study design is weaker than a randomized controlled trial or similar natural experimental designs (Craig et al., 2017; Felker & Chen, 2023), it has a major advantage of being very easy to implement, and introduces minimum disruption to the teaching and learning process, making it widely applicable in many different settings. Similar study design can be widely adopted in future studies to examine the extent to which the current observations are replicable under different instructional contexts and for different LDs.

Finally, the current study extends the application of process mining and trace clustering algorithms to comparing the differences in student learning tactics under different learning designs, by using Hellinger distance as a quantitative metric between two distributions. This is an improvement over existing studies using similar analysis approaches either to describe the learning tactics under one type of learning design or to only make qualitative comparisons of differences in learning tactics.

4.2.2. Implications for Instructors

For instructors, the current study suggests that the optional first attempt design is a viable, and possibly superior, design for mastery-based online learning modules compared to the original LD. While more rigorous study is required to establish a firm causal relationship between change in learning design and change in student learning, the current results are sufficient to show that the new design is at least not harmful for student learning. More specifically, the naïve concern of some instructors that almost all students will skip the first attempt since they are all score oriented is not supported by the observation. Rather, students demonstrated relatively accurate levels of self-assessment. In addition, the original hypothesis that mandating an initial attempt would improve learning from the subsequent learning materials is also not supported by the current observations. Rather, it seems that an optional first attempt might result in slightly higher passing rates and longer study time. Finally, it is possible that giving students the choice of whether to take the initial attempt gave them an opportunity to self-regulate and could have prompted more students to self-assess. Therefore, the new design might be beneficial as a method to foster self-regulation among students.

The current study also signifies the importance of taking data collection and data analysis into account when implementing new learning designs. Rather than implementing the new design in all learning modules, the authors intentionally kept the two sequences identical between two semesters and used data from those sequences to control for the impact of extraneous factors. Future instructors who intend to make data-driven decisions for improving learning design could use similar strategies to improve the reliability of data collection and data analysis.

4.3. Directions for Future Research

In the current study, a hypothetical learning tactic distribution was generated based on data from the two control module sequences compared to the observed distribution. It is in principle similar to but less rigorous than a "differences in differences" method commonly used for analyzing natural experiments, but computationally much easier since only data from three sequences needs to be processed using trace clustering. A complete "differences in differences" analysis can be performed in follow-up studies either with significantly higher computation power, or more efficient clustering algorithms that would allow for clustering of student data on multiple module sequences.

Alternatively, one could also implement the new learning design on the last few modules in a sequence and use data from the first few modules as a control to conduct a "differences in differences" analysis. In addition, it would be possible to create a predictive model, such as a second order hidden Markov model (Guthrie & Chen, 2019a), that could predict individual student learning tactics under existing learning design, and compare that to the same student's observed learning tactics under the new learning design. All the above-mentioned methods would improve the estimate of the impact of the new learning design on students' self-regulation and learning outcomes.

Another major caveat is that the current study relied exclusively on the analysis of trace data. This limited the authors' ability to discern between different internal states and SRL procedures that could lead to the same observed outcome. In future studies, a mixed-method design involving surveys of students' self-regulation can be implemented to gain further insight into student decision-making processes and how they are influenced by various internal states. If we conduct student interviews, for example, this will confirm if students are engaging in self-regulated learning strategies to inform their decisions to engage with the course modules and make or skip attempts.

In addition, one caveat in the interpretation of the current study is that the "initial normal pass" cluster was assumed to contain mostly students who had higher incoming knowledge of the content and high self-confidence. This was supported by previous studies on the same or similar modules, indicating that passing on a normal length first attempt is highly correlated with overall strong performance in class (Chen et al., 2020; Taub et al., 2022). However, one cannot exclude the possibility that rather than solving the problem, some students could have spent the time searching for answers online. With the rapid development of AI, it is even less clear in future studies that the submitted answer actually came from the student. Future studies will need to find new and innovative ways to reliably assess student knowledge mastery in online learning environments.

Finally, the current study attempts to answer the question of "which learning design is better for student learning?" Yet the best answer to that question could well be "it depends": students with different internal states may need different learning designs that provide different levels of external support and incentive to achieve better learning. Today's intelligent tutoring systems already customize problems and feedback according to student performance data. Future learning systems could provide different students with different learning designs that contain different levels or types of scaffolding for self-regulation, based on student learning tactics on prior content identified through learning analytics.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The publication of this article received financial support from National Science Foundation Grant No. DUE-1845436. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Acknowledgments

The authors would like to thank Dr. Rui Xie of the University of Central Florida for extensive discussion and advice on the use of Hellinger Distance and related statistical methods. The authors would also like to thank the Learning Systems and Technology Team led by Dr. Francisca Yonekura for creation and maintenance of the Obojobo Learning Objects system.

References

Alexandron, G., Ruipérez-Valiente, J. A., Chen, Z., Muñoz-Merino, P. J., & Pritchard, D. E. (2017). Copying@scale: Using harvesting accounts for collecting correct answers in a MOOC. *Computers & Education*, *108*, 96–114. https://doi.org/10.1016/j.compedu.2017.01.015

Azevedo, R., & Taub, M. (2020). The challenge of measuring processes and outcomes while learning from multiple representations with advanced learning technologies. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives*. Routledge. https://doi.org/10.4324/9780429443961-34

Barkley, A., & Coffey, B. K. (2018). An economic model of student learning. *Journal of Agricultural and Applied Economics*, *50*(4), 503–525. https://doi.org/10.1017/AAE.2018.13

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, *1*(2).
Chen, Z. (2022). Measuring the level of homework answer copying during COVID-19 induced remote instruction. *Physical Review Physics Education Research*, *18*(1), 010126. https://doi.org/10.1103/PhysRevPhysEducRes.18.010126
Chen, Z., Garrido, G., Berry, Z., Turgeon, I., & Yonekura, F. (2018). Designing online learning modules to conduct pre- and post-testing at high frequency. *2017 Physics Education Research Conference Proceedings*, 26–27 July 2017, Cincinnati, OH, USA (pp. 84–87). https://doi.org/10.1119/perc.2017.pr.016
Chen, Z., Xu, M., Garrido, G., & Guthrie, M. W. (2020). Relationship between students' online learning behavior and course performance: What contextual information matters? *Physical Review Physics Education Research*, *16*(1), 010138. https://doi.org/10.1103/PhysRevPhysEducRes.16.010138
Craig, P., Katikireddi, S. V., Leyland, A., & Popham, F. (2017). Natural experiments: An overview of methods, approaches, and contributions to public health intervention research. *Annual Review of Public Health*, *38*, 39–56. https://doi.org/10.1146/annurev-publhealth-031816-044327
Dunlosky, J., Baker, J. M. C., Rawson, K. A., & Hertzog, C. (2006). Does aging influence people's metacomprehension? Effects of processing ease on judgments of text learning. *Psychology and Aging*, *21*(2), 390–400. https://doi.org/10.1037/0882-7974.21.2.390
Fan, Y., Matcha, W., Uzir, N. A., Wang, Q., & Gašević, D. (2021). Learning analytics to reveal links between learning design and self-regulated learning. *International Journal of Artificial Intelligence in Education*, *31*(4), 980–1021. https://doi.org/10.1007/s40593-021-00249-z
Felker, Z., & Chen, Z. (2023). Reducing procrastination on introductory physics online homework for college students using a planning prompt intervention. *Physical Review Physics Education Research*, *19*(1), 010123. https://doi.org/10.1103/PhysRevPhysEducRes.19.010123
Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions. *Metacognition and Learning*, *12*(1), 1–19. https://doi.org/10.1007/S11409-016-9158-6
Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857–871. https://doi.org/10.2307/2528823
Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology*, *34*(1), 18–29. https://doi.org/10.1016/j.cedpsych.2008.05.006
Guthrie, M. W., & Chen, Z. (2019a). Adding duration-based quality labels to learning events for improved description of students' online learning behavior. In C. F. Lynch, A. Merceron, M. Desmarais, & R. Nkambou (Eds.), *Proceedings of the 12th International Conference on Educational Data Mining* (EDM2019), 2–5 July 2019, Montréal, Quebec, Canada (pp. 560–563). International Educational Data Mining Society. https://ssrn.com/abstract=3522718
Guthrie, M. W., & Chen, Z. (2019b). Comparing student behavior in mastery and conventional style online physics homework. *2019 Physics Education Research Conference Proceedings*, 24–25 July 2019, Provo, UT, USA (pp. 190–195). https://www.per-central.org/items/detail.cfm?ID=15275
Hellinger, E. (1909). Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen [New justification of the theory of quadratic forms of infinite variables]. *Journal für die Reine und Angewandte Mathematik*, *136*, 210–271. http://eudml.org/doc/149313
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. http://www.jstor.org/stable/4615733
Holmes, W., Nguyen, Q., Zhang, J., Mavrikis, M., & Rienties, B. (2019). Learning analytics for learning design in online distance learning. *Distance Education*, *40*(3), 309–329. https://doi.org/10.1080/01587919.2019.1637716
Huber, K., & Bannert, M. (2023). Investigating learning processes through analysis of navigation behavior using log files. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528-023-09372-3
Ifenthaler, D., Gibson, D., & Dobozy, E. (2018). Informing learning design through analytics: Applying network graph analysis. *Australasian Journal of Educational Technology*, *34*(2), 117–132. https://doi.org/10.14742/AJET.3767
Janssenswillen, G., van Hulzen, G., Depaire, B., Mannhardt, F., Beuving, T., & Urvikalia Hasselt University. (2023). *processmapR: Construct process maps using event data* (Version 0.5.3) [Data set]. https://CRAN.R-project.org/package=processmapR
Kaliisa, R., Kluge, A., & Mørch, A. I. (2020). Combining checkpoint and process learning analytics to support learning design decisions in blended learning environments. *Journal of Learning Analytics*, *7*(3), 33–47. https://doi.org/10.18608/jla.2020.73.4
Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science*, *38*(6), 523–550. https://doi.org/10.1007/s11251-009-9093-x

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(6161), 935–937. https://doi.org/10.1126/science.1238056

Kulik, J. A., Carmichael, K., & Kulik, C.-L. (1974). The Keller plan in science teaching: An individually paced, student-tutored, and mastery-oriented instructional method is evaluated. *Science*, *183*(4123), 379–383. https://doi.org/10.1126/science.183.4123.379

Lancaster, A., Moses, S., Clark, M., & Masters, M. C. (2020). The positive impact of deliberate writing course design on student learning experience and performance. *Journal of Learning Analytics*, *7*(3), 48–63. https://doi.org/10.18608/jla.2020.73.5

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

Macfadyen, L. P., Lockyer, L., & Rienties, B. (2020). Learning design and learning analytics: Snapshot 2020. *Journal of Learning Analytics*, *7*(3), 6–12. https://doi.org/10.18608/jla.2020.73.2

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2023). *cluster: Cluster analysis basics and extensions* (R package version 2.1.6) [Data set]. https://CRAN.R-project.org/package=cluster

Magnus, J. R., & Peresetsky, A. A. (2018). Grade expectations: Rationality and overconfidence. *Frontiers in Psychology*, *8*, 2346. https://doi.org/10.3389/fpsyg.2017.02346

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*(4), 516–534. https://doi.org/10.1109/TLT.2018.2868673

Motz, B. A., Bergner, Y., Brooks, C. A., Gladden, A., Gray, G., Lang, C., Li, W., Marmolejo-Ramos, F., & Quick, J. D. (2023). A LAK of direction: Misalignment between the goals of learning analytics and its research scholarship. *Journal of Learning Analytics*, *10*(2), 1–13. https://doi.org/10.18608/jla.2023.7913

Persico, D., & Pozzi, F. (2015). Informing learning design with learning analytics to improve teacher inquiry. *British Journal of Educational Technology*, *46*(2), 230–248. https://doi.org/10.1111/BJET.12207

Prates, M. O., Lachos, V. H., & Cabral, C. R. B. (2013). mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software*, *54*(12), 1–20. https://doi.org/10.18637/jss.v054.i12

Rienties, B., Nguyen, Q., Holmes, W., & Reedy, K. (2017). A review of ten years of implementation and research in aligning learning design with learning analytics at the Open University UK. *Interaction Design and Architecture Journal*, *33*, 134–154. https://doi.org/10.55612/s-5002-033-007

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Saint, J., Fan, Y., Singh, S., Gašević, D., & Pardo, A. (2021). Using process mining to analyse self-regulated learning: A systematic analysis of four algorithms. *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK ’21), 12–16 April 2021, Irvine, CA, USA (pp. 333–343). ACM Press. https://doi.org/10.1145/3448139.3448171

Saint, J., Whitelock-Wainwright, A., Gašević, D., & Pardo, A. (2020). Trace-SRL: A framework for analysis of microlevel processes of self-regulated learning from trace data. *IEEE Transactions on Learning Technologies*, *13*(4), 861–877. https://doi.org/10.1109/TLT.2020.3027496

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). Information Age Publishing.

Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students’ desired grades predict their biased grade predictions. *Memory & Cognition*, *44*(7), 1127–1137. https://doi.org/10.3758/S13421-016-0624-9

SoLAR. (2020). *What is Learning Analytics?* Society for Learning Analytics Research. https://www.solaresearch.org/about/what-is-learning-analytics/

Sonnenberg, C., & Bannert, M. (2019). Using process mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior? *Computers in Human Behavior*, *96*, 259–272. https://doi.org/10.1016/j.chb.2018.06.003

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

Taub, M., Banzon, A. M., Zhang, T., & Chen, Z. (2022). Tracking changes in students’ online self-regulated learning behaviors and achievement goals using trace clustering and process mining. *Frontiers in Psychology*, *13*, 813514. https://doi.org/10.3389/fpsyg.2022.813514

Van Witteloostuijn, A. (1990). Learning in economic theory: A taxonomy with an application to expectations formation. *Journal of Economic Psychology*, *11*(2), 183–207. https://doi.org/10.1016/0167-4870(90)90003-R

Warnakulasooriya, R., Palazzo, D. J., & Pritchard, D. E. (2007). Time to completion of web-based physics problems with tutoring. *Journal of the Experimental Analysis of Behavior*, *88*(1), 103–113. https://doi.org/10.1901/jeab.2007.70-06

Whitcomb, K. M., Guthrie, M. W., Singh, C., & Chen, Z. (2021). Improving accuracy in measuring the impact of online instruction on students' ability to transfer physics problem-solving skills. *Physical Review Physics Education Research*, *17*(1), 010112. https://doi.org/10.1103/PhysRevPhysEducRes.17.010112

Winne, P. H. (2017). Cognition and metacognition within self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 36–48). Routledge. https://doi.org/10.4324/9781315697048

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227–304). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410602350-12

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Lawrence Erlbaum Associates.

Wortha, F., Azevedo, R., Taub, M., & Narciss, S. (2019). Multiple negative emotions during learning with digital learning environments: Evidence on their detrimental effect on learning from two methodological approaches. *Frontiers in Psychology*, *10*, 2678. https://doi.org/10.3389/fpsyg.2019.02678

Wüst, K., & Beck, H. (2018). "I thought I did much better": Overconfidence in university exams. *Decision Sciences Journal of Innovative Education*, *16*(4), 310–333. https://doi.org/10.1111/DSJI.12165

Zhang, T., Taub, M., & Chen, Z. (2021). Measuring the impact of COVID-19 induced campus closure on student self-regulated learning in physics online learning modules. *Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (LAK '21), 12–16 April 2021, Irvine, CA, USA (pp. 110–120). ACM Press. https://doi.org/10.1145/3448139.3448150

Zhang, T., Taub, M., & Chen, Z. (2022). A multi-level trace clustering analysis scheme for measuring students' self-regulated learning behavior in a mastery-based online learning environment. *Proceedings of the 12th International Conference on Learning Analytics and Knowledge* (LAK '22), 21–25 March 2022, Online (pp. 197–207). ACM Press. https://doi.org/10.1145/3506860.3506887

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7