Curricular Concept Maps as Structured Learning Diaries: Collecting Data on Self-Regulated Learning and Conceptual Thinking for Learning Analytics Applications

Ville Kivimäki[1], Joonas Pesonen[2], Jani Romanoff[3], Heikki Remes[4], Petri Ihantola[5]

Abstract

The collection and selection of the data used in learning analytics applications deserve more attention. Optimally, selection of data should be guided by pedagogical purposes instead of data availability. Using design science research methodology, we designed an artifact to collect time-series data on students' self-regulated learning and conceptual thinking. Our artifact combines curriculum data, concept mapping, and structured learning diaries. We evaluated the artifact in a case study, verifying that it provides relevant data, requires a limited amount of effort from students, and works in different educational contexts. Combined with learning analytics applications and interventions, our artifact provides possibilities to add value for students, teachers, and academic leaders.

Notes for Practice

- Data about constructs related to self-regulated learning, such as motivation, emotion, and metacognitive experiences, is highly relevant in learning analytics applications.

- While this data is difficult to capture automatically, learning diaries can gather process data about students' internal states.

- The main contribution of this article is to present a methodology in which concept maps are used as learning diaries to gather meaningful data for learning analytics applications.

- In the future, this method of data gathering can be combined with different kinds of learning analytics interventions, including personalized feedback at scale.

[1]Corresponding author Email: ville.kivimaki@helsinki.fi Address: School of Engineering, Aalto University, Otakaari 4, 02150 Espoo / P.O. Box 14100, FI-00076 Aalto, Finland; Faculty of Educational Sciences, University of Helsinki, Siltavuorenpenger 1–5, 10 / P.O. Box 9, 00014 University of Helsinki, Helsinki, Finland, https://orcid.org/0000-0003-4939-7400

[2]Corresponding author Email: joonas.pesonen@helsinki.fi Address: Faculty of Educational Sciences, University of Helsinki, Siltavuorenpenger 1–5, 10 / P.O. Box 9, 00014 University of Helsinki, Finland, https://orcid.org/0000-0003-4166-4762

[3]Address: Department of Mechanical Engineering, Aalto University, Puumiehenkuja 5 / P.O. Box 14300, 00076 Aalto, Finland, https://orcid.org/0000-0002-4642-0225

[4]Address: Department of Mechanical Engineering, Aalto University, Puumiehenkuja 5 / P.O. Box 14300, 00076 Aalto, Finland, https://orcid.org/0000-0002-5030-3494

[5]Address: Faculty of Educational Sciences, University of Helsinki, Siltavuorenpenger 1–5, 10 / P.O. Box 9, 00014 University of Helsinki, Finland, https://orcid.org/0000-0003-1197-7266

1. Introduction

Research literature related to learning analytics applications reveals that surprisingly little consideration has been given to which data is meaningful to collect; this is underlined by Winne's (2017) notion that out of several widely cited descriptions of learning analytics, none answer the question of which data should be gathered for input to methods that generate learning analytics. In a review of student-facing dashboards, Bodily & Verbert (2017) found that the most often used types of data in dashboards were (1) the number of times a resource was accessed (in 75% of articles), (2) data on student mastery as measured by assessment (in 37% of articles), and (3) data on time spent accessing resources (in 30% of articles). It is clear that the

availability of data may affect which data is used in learning analytics. For example, data on resource use and time spent is typically automatically logged by learning management systems.

At the same time, a growing body of evidence suggests that learning analytics applications need a better grounding in the learning sciences (Jivet, Scheffel, Drachsler, & Sprecht, 2017; Marzouk et al., 2016; Sedrakyan, Malmberg, Verbert, Järvelä, & Kirschner, 2018). As concluded by Jivet et al. (2017) in their review, self-regulated learning (SRL) seems to be the core theory informing the design of learning analytics dashboards. Winne (2017) describes on one hand how SRL theory can guide which kinds of data to gather for learning analytics and on the other hand how learning analytics can support SRL. Another crucial aspect of learning is how learners assimilate new concepts and propositions into their existing conceptual frameworks. While this question has been addressed in concept mapping research (Novak & Musonda, 1991; Daley & Torre, 2010; Nesbit & Adesope, 2006), the potential to use data gathered with concept mapping in learning analytics applications seems to be underutilized. In this paper, we explore this underutilized potential by applying SRL theories to develop a concept mapping tool to provide meaningful data for learning analytics applications.

The aim of the current research is to design a methodology by which data on students' SRL and conceptual thinking can be collected. Our research questions are as follows:

> RQ1: How can we collect data about constructs related to SRL, such as motivation, emotion, and metacognitive experiences?

> RQ2: In a real course setting, which data collection methods would be viable to implement, considering the workload of the exercise and students' willingness to share their data at scale?

Design science research (DSR; Hevner, 2007) was chosen as a research approach because it is seen as a suitable method for "education research where an information and communications technology artifact is designed and created as a central part of the research process" (Chard, 2017). In section 2, we will situate our research with related literature. In section 3, Methodology, we will elaborate on the research approach and analysis methods we used. We describe the design artifact in section 4 and evaluate and discuss it in section 5.

2. Related Literature

SRL refers to how learners systematically activate and sustain their cognitions, motivations, behaviours, and affects toward the attainment of their learning goals (Greene & Schunk, 2017). As Winne (2010) states, self-regulating learners "actively research what they do to learn and how well their goals are achieved by variations in their approaches to learning." SRL is a core conceptual framework for understanding the cognitive, motivational, and emotional aspects of learning (Panadero, 2017).

Meta-analytical evidence suggests that properly designed SRL interventions can improve learning (Dignath & Büttner, 2008; Panadero, 2017; Sitzmann & Ely, 2011). Interventions to enhance students' SRL include, for example, self-assessment (Panadero & Alonso-Tapia, 2013; Panadero, Jonsson, & Botella, 2017), concept mapping (Gurlitt & Renkl, 2010; Nesbit & Adesope, 2006; Novak & Musonda, 1991), learning diaries (Klug, Ogrin, Keller, Ihringer, & Schmitz, 2011; Schmitz & Wiese, 2006), and learning analytics dashboards (Bodily & Verbert, 2017; Jivet et al., 2017; Sedrakyan et al., 2018). In this study, we are at the intersection of this knowledge base. Next, we will focus on the most relevant literature related to our design process.

2.1. SRL and Learning Analytics

Learning analytics are typically built on log data automatically captured by digital tools. However, it is difficult or impossible to automatically capture data about SRL constructs that describe students' internal states, such as motivation, emotion, and metacognition. Traditionally, self-report questionnaires have been used to measure students' internal states, but they fail to catch the dynamic nature of students' SRL (Panadero, Klug, & Järvelä, 2016). One solution is to use self-reported process data, namely structured learning diaries (Figure 1). Diaries permit real-time recording of learning processes and have high ecological validity (Schmitz, Klug, & Schmidt, 2011).

While the field of learning analytics attempts to improve students' learning by collecting, analyzing, and reporting data about learners and their contexts (Siemens, 2013), data from these digital tools is also crucial to the SRL theories themselves. Panadero et al. (2016) present three waves in the history of SRL measurement. While the first and second waves focused on self-reports and process data, respectively, the third wave combines measurement and intervention within the same tools (Panadero et al., 2016). The data produced by these third-wave measurement and intervention tools is usually time-series data of students' learning processes, including variables related to motivation, emotions, learning strategies, goal attainment, and so on (Panadero et al., 2016). Examples of third-wave tools are structured learning diaries (Schmitz et al., 2011) and software tools MetaTutor (Azevedo et al., 2013), OurPlanner and OurEvaluator (Järvelä et al., 2015), and nStudy (Winne &

Hadwin, 2013). In addition to collecting records on SRL, they also aim to affect students' self-regulation (Panadero et al., 2016).
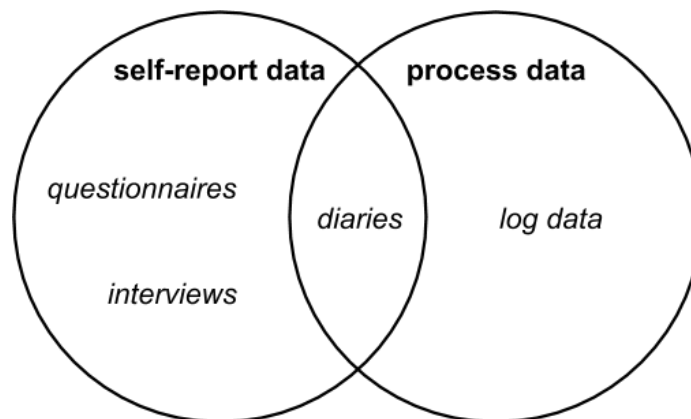
Figure 1. Diaries as self-reported process data

There is great potential in combining SRL theory and measurement tools with approaches of learning analytics and educational data mining. First, educational data mining can help to operationalize constructs, investigate multivariate relationships among constructs, and model and understand student trajectories related to SRL (Winne & Baker, 2013). Second, SRL theory can guide which kinds of measures are meaningful to capture, how and when to deliver the analytics, and what kind of information should be provided to students to support SRL (Winne, 2017). Moreover, Schumacher & Ifenthaler (2018) found that students expect learning analytics to support their planning and organization of learning processes, provide self-assessments, deliver adaptive recommendations, and produce personalized analyses of their learning activities, all features that are strongly related to SRL theories.

Most SRL models can be conceptualized into three phases: preparatory, performance, and appraisal (Panadero, 2017). In a review of several SRL models by Puustinen & Pulkkinen (2001), this trichotomy was broken down into common subprocesses as follows:

- Preparatory: task analysis, planning, activation of goals, setting goals
- Performance: monitoring and controlling
- Appraisal: reflecting, regulating, adapting

Self-monitoring is a subprocess of the performance phase and can be found in several SRL models (Hadwin, Järvelä, & Miller, 2011; Pintrich, 2000; Zimmerman & Paulsen, 1995). Self-monitoring can further be divided into informal self-monitoring (casual observation or spontaneous reaction) and formal self-monitoring (systematic judgments and observations that are recorded) (Zimmerman & Paulsen, 1995). An example of formal self-monitoring is structured learning diaries (Schmitz et al., 2011), which at the same time are both measurement instruments of SRL and intervention tools for enhancing students' SRL (Panadero et al., 2016).

2.2.   Structured Learning Diaries

While learning diaries in general pursue the goal of reflecting the learning content or one's learning behaviour, structured learning diaries achieve this by using standardized question items that are answered repeatedly. Items can be, for example, Likert items or open questions, and they may be newly developed or items from existing questionnaire instruments (Klug et al., 2011). Existing items may need to be adapted for the diary format (e.g., from "When learning …" to "Today, while learning …") (Schmitz & Wiese, 2006). Nonverbal items are also possible, such as measuring emotions with pictures (Bradley & Lang, 1994).

Structured learning diaries have been shown to increase students' SRL. Schmitz & Wiese (2006) found that learning diaries combined with training sessions on SRL skills improved engineering students' self-efficacy, effort, and distraction handling. Schmitz & Perels (2011) found a positive linear trend of self-regulation when grade 8 students in a German grammar school used structured diaries during math homework. However, Fabriz, Dignath-van Ewijk, Poarch, & Büttner (2014) found that keeping a structured learning diary without any further intervention did not improve self-regulation, whereas students who kept the diary and at the same time received further information on self-regulation showed increases in strategy use and self-efficacy. Further, Dignath-van Ewijk, Fabriz, & Büttner (2015) found that keeping a weekly learning

diary without further intervention improved metacognitive skills and metacognitive attitude, although the effects were not as strong as expected.

In addition to being an SRL intervention, structured learning diaries are a successful measurement instrument for studying constructs that change in time (Panadero et al., 2016), that is, studying psychological states instead of traits (Geiser, Götz, Preckel, & Freund, 2017). For example, much of the research on academic emotions (Pekrun, Goetz, Titz, & Perry, 2002) has been conducted with large samples at one point in time, thus assessing emotions as traits. Only recently have academic emotions been studied with real-time assessments (Ketonen, Dietrich, Moeller, Salmela-Aro, & Lonka, 2018).

A major challenge with structured learning diaries is that they require significant effort from the subject (Klug et al., 2011). Therefore, the number of items to include in the diary must be carefully considered (Schmitz & Wiese, 2006). Furthermore, it is important to make the potential effect of the instrument on learning explicit, so that students' willingness to work with the tool is increased (Panadero et al., 2016). Based on prior research, general information about self-regulation may even be imperative for the intervention to be successful (Fabriz et al., 2014).

Literature on combining learning analytics and learning diaries is sparse. There are examples of applying analytical methods to the different kinds of reflective texts that are often found in learning diaries. Munezero, Montero, Mozgovoy, & Sutinen (2013) successfully used sentiment analysis to track emotions in students' learning diaries. Kovanović et al. (2018) used text analytics methods to assess students' reflective writings.

2.3.   Concept Mapping and Curriculum Data

The method of concept mapping, developed in 1972 at Cornell University (Novak & Musonda, 1991), is a widely used graphical tool for structuring and presenting knowledge. In this approach, concepts are illustrated with nodes (e.g., circles), while connections between concepts are illustrated with (labelled) edges. Concept mapping is based on theories of learners' cognitive structures and meaningful learning (Ausubel, 1963). These foundations are still visible in a later survey of concept maps in medical education by Daley & Torre (2010), who find promoting meaningful learning to be one of the four main ways to use concept maps in education. According to the survey, concept maps also function by providing an additional resource for learning, by enabling instructors to provide feedback to students, and by supporting assessment.

All of the aforementioned ways of using concept maps in education can be relevant to SRL. In a meta-analysis of learning with concept maps, Nesbit & Adesope (2006) conclude that many reported benefits of concept mapping are, indeed, related to improvements in the various aspects of SRL. According to the authors, the hypothesized ways that concept maps support learning are manifold. For example, because concept maps are unordered, using them may lead to the use of deep learning strategies when compared to repetitive reading of textual information, where surface learning is more likely. This is supported by Chmielewski & Dansereau (1998), who find that the use of knowledge maps may lead to increased ability to construct structured information in tasks that do not involve concept maps.

Moreover, deeper engagement with concept maps is often connected to better learning results. Gurlitt & Renkl (2010), for example, compare students who labelled connections in a predefined concept map to those who (in addition to labelling) had to create the connections themselves. The students who created connections were cognitively more engaged, had higher self-efficacy, and performed better on the learning test. Regarding the use of concept mapping for assessment, possible criteria can be quantitative (e.g., number of concepts, number of levels of hierarchy, number of cross-links) or qualitative (e.g., quality of concept labels, proposition quality, quality of hierarchy) (Strautmane, 2012).

In higher education, students often use concept maps in the context of a single lecture or a full course. This is natural, as teachers (who facilitate the use of concept maps in their courses) have significant freedom in delivering courses within the boundaries of the curriculum. Moreover, course activities are rarely designed in collaboration with other courses. Borders between the courses are defined by curricula, which in the broader context can be seen as all the learning that is planned and guided by the school (Kerr, 1968).

Curriculum theories differentiate between intended curriculum (the vision presented in official documents), implemented curriculum (the actual process of teaching and learning), and attained curriculum (the learning experiences and learning outcomes of the learners) (van den Akker, 2003). The intended curriculum is traditionally presented as text in official documents, where relations between different parts of the curriculum are defined implicitly and are thus difficult to use within analysis algorithms. However, via educational mapping, the curriculum may also be presented as a network model. Willcox & Huang (2017) define educational mapping as the "process of analyzing an educational system to identify entities, relationships and attributes." Entities (programs, modules, courses, concepts, learning outcomes, etc.) and relationships between them (a course has a module as parent, a course has a prerequisite of a learning outcome, etc.) form a graph structure, which can then be used to represent, visualize, and analyze educational data at scale (Willcox & Huang, 2017).

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)

109

3. Methodology

In this paper, we design an IT artifact as a solution for the problem framed in section 1. For this we use a DSR approach, introduced in 2004 by Hevner, March, Park, & Ram (2004). Here, the environment and knowledge base set a conceptual framework for building an artifact that is evaluated in terms of both applicability and research (Figure 2). In section 1, we defined environmental problems for which we need a solution (relevance cycle). In section 2, we conducted a selective literature review to raise possible solutions for the problem (rigour cycle). In section 4, we present the design artifact based on our findings. Finally, in section 5, we present the results of the artifact evaluation study and propose issues for further iteration rounds for the artifact design (design cycle).
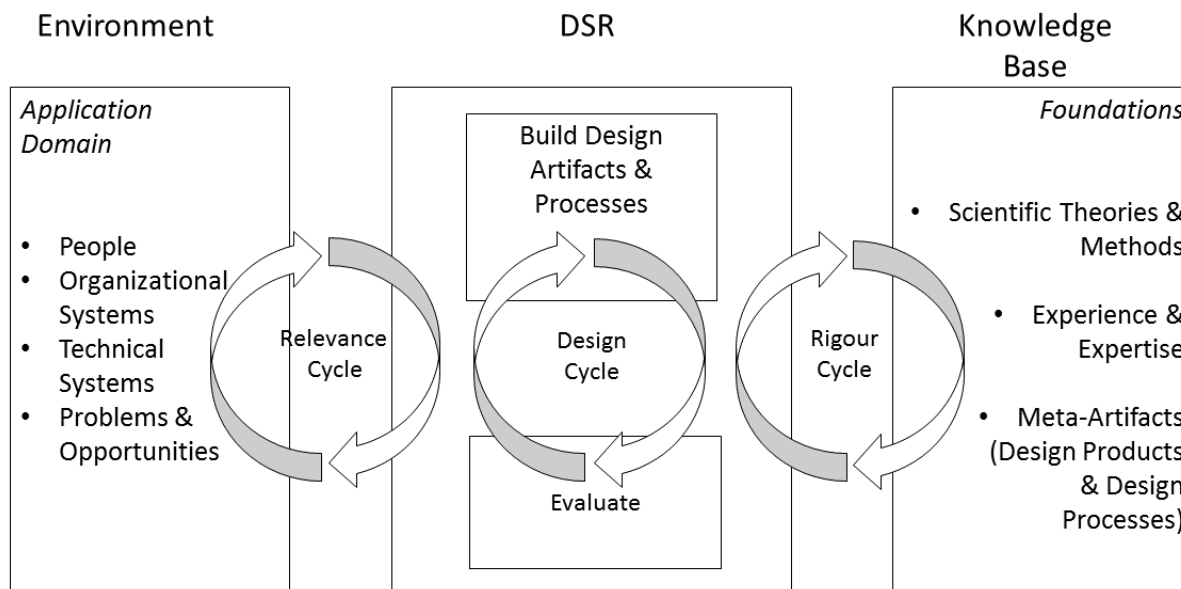

Figure 2. Cyclical view of DSR (adapted from Hevner, 2007)

Hevner et al. (2004) define seven features as guidelines for implementing DSR methodology. A viable artifact must be produced (1). It should be relevant to the defined problem (2), and the utility, quality, and efficacy must be demonstrated and evaluated (3). The design must provide contributions to research (4) and utilize rigorous methods in the construction and evaluation of the artifact (5). In addition, the design process should use what is available to reach the goal while acknowledging the environmental constraints (6). Finally, the research must be presented to relevant audiences (7).

DSR offers a framework in which scientific theories and methods can be conducted (Weber, 2012). Ideally, DSR contributes to both real-life outcomes and research findings. This aligns with the goals of our study, as we aim to develop a research-based methodology for collecting meaningful data to be used in learning analytics applications.

Good DSR often begins by identifying and representing opportunities and problems in an actual application environment (Hevner, 2007). These opportunities and problems are presented in the introduction section of this paper. For the rigour cycle, we conducted a selective literature review of well-known SRL interventions: self-assessment, concept mapping, learning diaries, and learning analytics dashboards. During the process, we narrowed our focus to concept mapping and structured learning diaries.

Our research aims to produce a data collection instrument that produces data related to students' SRL and is viable in a real course setting at scale. In our design, this rules out the possibility of physiological measurements. Building on our experience with the application domain (environment) as well as previous research (knowledge base), we designed an instrument, which we present in detail in section 4.

To help us design and evaluate the artifact, we came up with the following evaluation criteria:

A. *Data relevance:* The data gathered with the design artifact is related to students' self-regulation and conceptual thinking.
B. *Student effort:* The effort required from students is reasonable, so that students are willing to use the artifact.
C. *Scalability:* The artifact can be used in different kinds of educational contexts.

To evaluate the artifact, we carried out a case study involving one engineering course and one business course at a research-oriented European university (see Table 1). Both courses consisted of lectures, assignments, and a summative assessment. The courses were selected based on temporal opportunity and teachers' willingness to participate. Participating students were awarded extra points. To use the design artifact, students were required to download and install a concept mapping application and submit files created with the application every week through the learning management system used in the courses.

All participants were informed of the research and were allowed to opt out of the research at any time. Activities related to this experimental study were optional assignments, introduced in the learning management system under the title "Dynamic Feedback System" and associated with extra points. Students were instructed to participate actively, but the content as such would not affect the number of extra points awarded. These activities were demonstrated in the introductory lecture of the course.

Table 1. Case Study Context and Student Demography

	Course 1: attending (participating)	Course 2: attending (participating)
Field	Engineering	Business
Time	Spring 2018	Spring 2018
Duration	14 weeks	7 weeks
Workload	5 credits, ~135 hours	5 credits, ~135 hours
Number of students	15 (12)	44 (44)
Age	M = 26 (27)	M = 26 (26)
	Mdn = 24 (25)	Mdn = 25 (25)
	SD = 5.9 (6.5)	SD = 3.8 (3.8)
Gender	1 female = 7% (0)	19 female = 47% (19/47%)
	14 male = 93% (12/100%)	25 male = 57% (25/47%)
Grade point average (scale 0–5, where 5 is the highest grade)	M = 3.9 (3.9)	M = 4.4 (4.4)
	Mdn = 3.9 (3.9)	Mdn = 4.5 (4.5)
	SD = 0.52 (0.57)	SD = 0.4 (0.4)
Weight of extra points in the final grading	3%	2%

Regarding evaluation criterion A, we used qualitative content analysis to analyze the data gathered by the design artifact during the case study. Two human coders read students' open responses (coding units), and they set the response categories and descriptions. Next, each response was assigned to one category by each of two coders separately. An interrater reliability analysis using the kappa statistic was performed to determine consistency among coders. Finally, the relation between response categories and theory (self-regulation and conceptual thinking) was elaborated.

Regarding evaluation criterion B, we collected and analyzed data on the usage of the instrument during the case study. The number of submissions and topic relations was collected each week to see if use of the artifact decreased over time.

Regarding evaluation criterion C, we compared two different contexts and the effect of the context on the usage of the tool.

4. Artifact Description

In this section, we introduce the designed artifact, Dynamic Feedback System (DFS), a methodology and a toolset that combine curriculum mapping with structured learning diaries to gather data to be used in learning analytics applications. First, we present a general description of the design artifact, and then we describe how the artifact was used in the case study. We present the results of the case study in section 5.

Through educational mapping, different curriculum entities (programs, modules, courses, concepts, learning outcomes, etc.) and their relations can be extracted. Relations between programs, modules, and courses are usually defined in student information systems, but structures below the course level usually exist only as text organized under a few subheadings (prerequisites, learning outcomes, course content, literature, etc.). Text mining–based approaches may provide some help for extracting different entities and relations. Even if it would be technically feasible to extract course topics from an official curriculum, this might not be desirable because official curricula are often written on a rather general level to allow teachers

freedom in instructional design. Therefore, we decided to extract the topic material from course descriptions, after which teachers could comment or change the topics related to their courses. We used this teacher-reviewed content to generate the curricular concept maps (Figure 3).
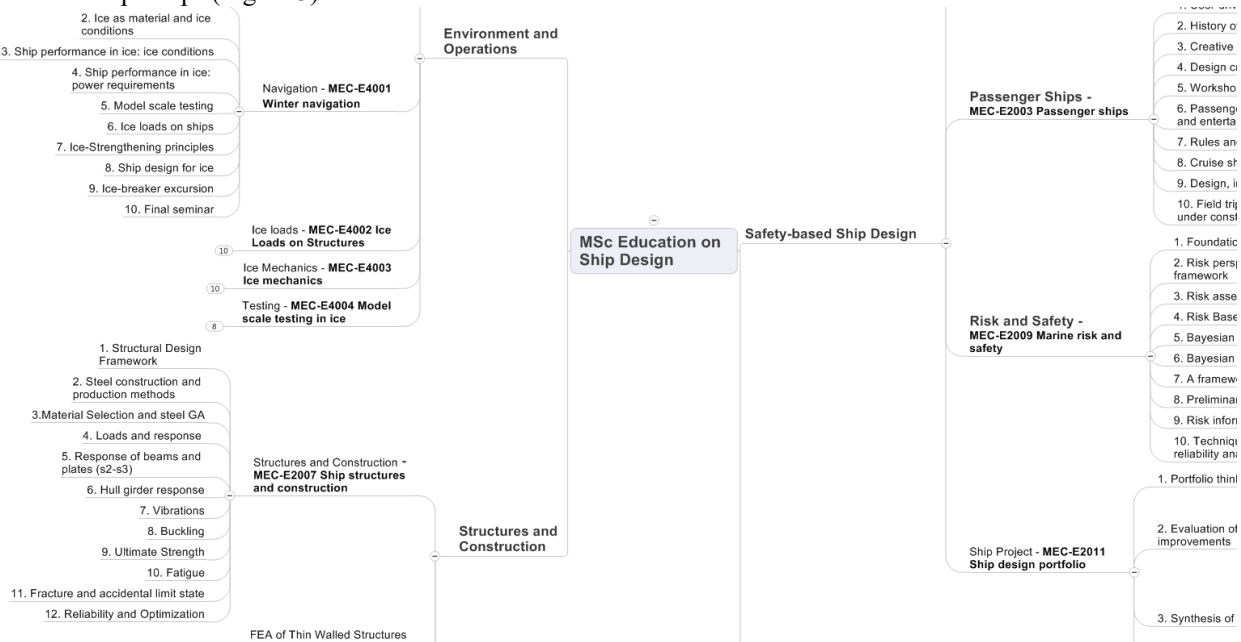

Figure 3. Extract of a curricular concept map template for Course 1

The curricular concept map template (Figure 4) is a tree structure, with the top-level node being the title of the program in question. The second level represents modules, the third level courses, and the fourth level course topics. A typical curricular concept map template for a degree program includes dozens of courses and several hundred topics. A typical template file has fewer than 10 modules, 100 or fewer courses, and many hundreds of topics.
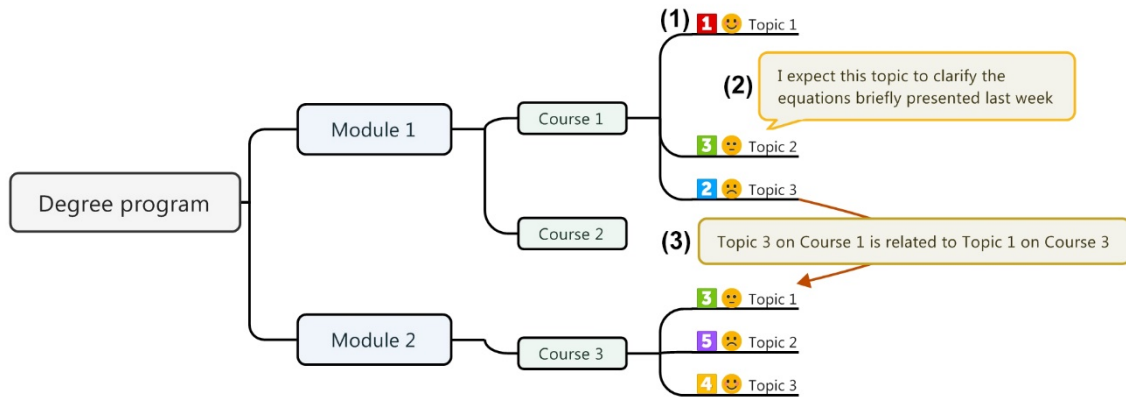

Figure 4. Curricular concept map as a structured learning diary, including responses to standardized items (1), responses to open items (2), and topic relationships (3)

Templates are provided for each student of the program in question. In addition to providing a visual view of the curriculum contents, students can interact with the visuals by enriching the data in an analogous way with structured learning diaries. There are three main categories of interacting with the map: (1) answering standardized items related to a certain topic, (2) answering open questions related to a certain topic, and (3) creating and explaining relations between different topics. In Table 2, these student actions are presented in more detail and mapped with SRL phases and actions (see Puustinen & Pulkkinen, 2001).

Table 2. DFS Application Design against SRL Theory

| SRL phase	Action	DFS features from the student's point of view
Preparatory	Task analysis, planning, activation of goals, setting goals	See an overview of all content in my degree program. Prioritize topics. Define my own competence in topics. Comment on topics.
Performance	Monitoring, controlling	See the progress of my own learning process on course and program level. Comment on topics.
Appraisal	Reflecting, regulating, adapting	Update my own competence in topics. Draw connections between topics. Comment on topics.

To analyze and visualize the data of hundreds of curricular concept maps in a meaningful way, data must first be extracted and then aggregated. These data sets can then be analyzed further by researchers and utilized in learning analytics applications, for example, as learning analytics dashboards or to provide personalized feedback at scale.

In the case study, we included three standardized items inside each topic: priority, judgment of learning, and emotion (see Figure 5). First, priority was assessed with number icons: 1 for "very high priority," 2 for "high priority," 3 for "low priority," 4 for "very low priority," and 5 for "priority not selected." Second, fraction icons were used to assess students' judgments of learning, representing 0%, 25%, 50%, 75%, and 100%. Third, four emoticons (neutral, happy, sad, and angry) were used to assess emotions. "Priority not selected," "0%," and "neutral" were used as initial default icons. Students were told that the icons could be changed at any time during the course.

Figure 5. Students' answers to the pictorial question items in Course 2

In addition to these standardized items, there was also one open question item, for which we did not directly specify a question but let students use the comment function in any way they found relevant to them (see Figure 6).
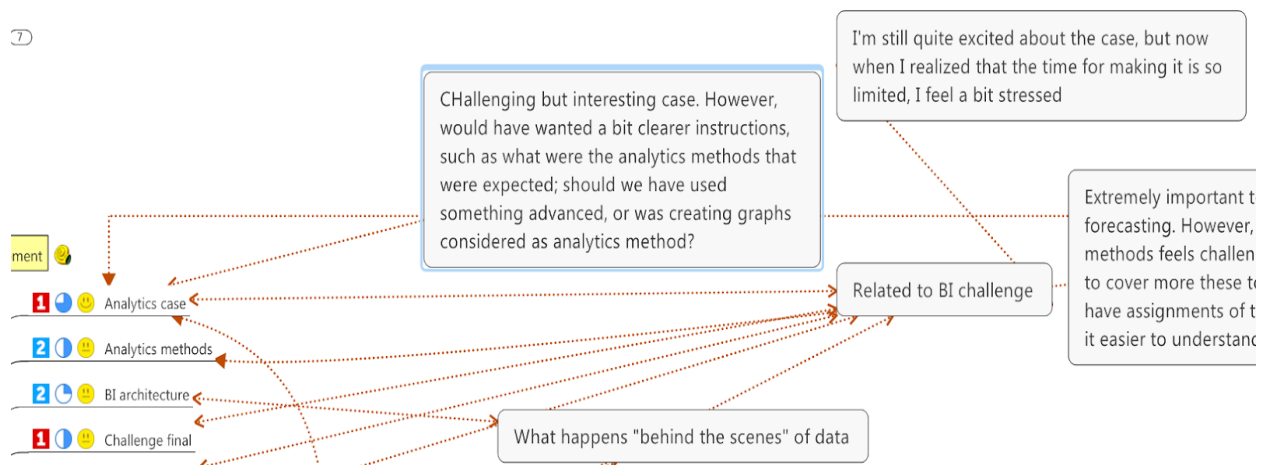
Figure 6. Students' answers to open questions in Course 2

The related topics item was introduced using the "Insert relationship" functionality of the software, which allows students to create connections between topics and include an explanation of the relationship (see Figure 7).
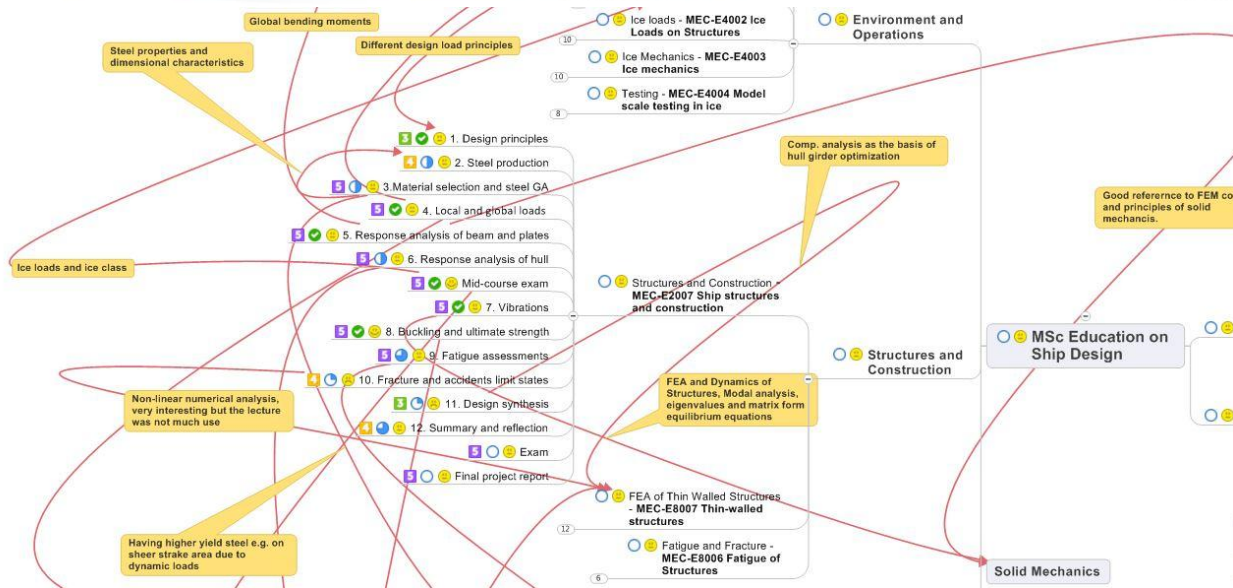
Figure 7. An extract of students' curricular concept map from Course 1

Two data sets were programmatically extracted from concept maps. First, we extracted a data set that includes each student's assessment of learning and emotions for each topic each week as well as each student's open responses (Table 3). Second, we extracted a data set that includes topic relationships as marked by each student each week (Table 4).

Table 3. An Extract from the Topics Data Set

Student	Week	Topic	Emotion	Priority	Judgment of learning	Comment
1	1	BI architecture	Happy	High	25%	""
1	2	BI architecture	Neutral	High	75%	"Data science now is more restricted by the existing IT architecture […]"
1	3	BI architecture	Neutral	High	75%	"Data science now is more restricted by the existing IT architecture […]"

Table 4. An Extract from the Topic Relations Data Set

Student	Week	From	To	Explanation
2	3	BI architecture	Wider analytics vision	"Architecture to be developed based on the vision"
2	3	Competitive market intelligence	BI architecture	"Increasing data volume is one of the challenges for data management and analysis."
3	4	Competitive market intelligence	BI architecture	"Managing data requires a well-defined structure of processes."

The extracted data was used to build dashboards for teachers and students. The first dashboard uses the topics data set and lists different topics, students' judgments of learning, and emotions related to them (see Figure 8). The second dashboard uses the topic relations data set to show a network of related courses (Figure 9) based on the topic relations reported by students. However, evaluation of these dashboards is outside the scope of the current paper.
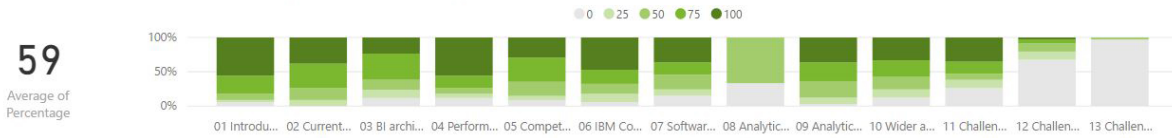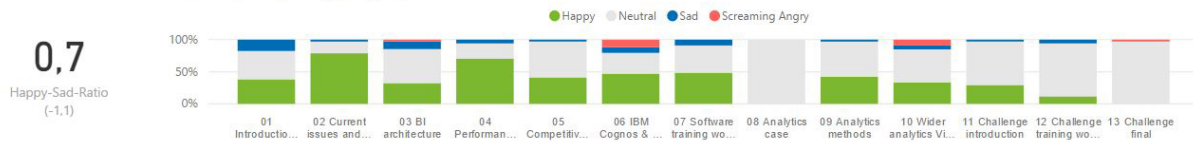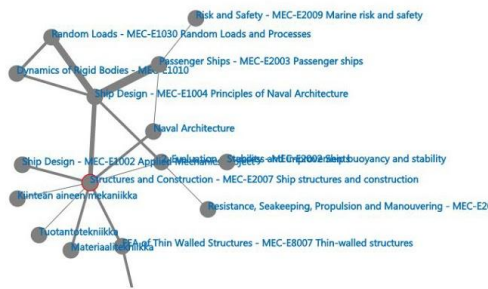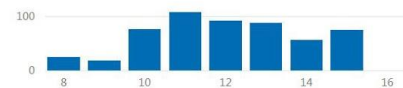
Figure 8. A learning analytics dashboard displaying students' judgments of learning, priorities, and emotions regarding different topics
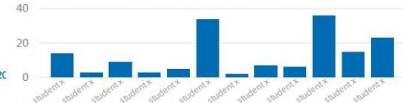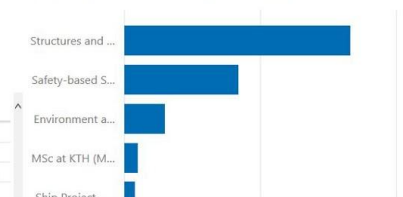

Figure 9. Course relationships network

This artifact was designed based on the literature review. Next, the artifact is evaluated against the evaluation criteria by using the case study methodology described therein.

5. Evaluation and Discussion

In this section, we evaluate our design artifact against the evaluation criteria presented in section 3. Afterwards, we discuss the results of the evaluation.

5.1. Evaluation of the Design Artifact

According to evaluation criterion A, the data gathered with the design artifact should be related to students' self-regulation and conceptual thinking. Using qualitative content analysis, we recognized five categories of open responses: keywords,

replicating content, reflecting on content, reflecting on process, and providing feedback (Table 5). Each response was then separately categorized by two human coders (Table 6). The most common categories for responses were keywords (43.6% and 41.5%) and replicating content (30.7% and 29.6%). After that came reflecting on process (13.9% and 14.5%) and reflecting on content (7.9% and 10.8%). Least common was providing feedback (3.9% and 3.6%). The interrater reliability for the coders was found to be substantial, with kappa = 0.731 ($p < 0.001$).

Table 5. Categorization of Open Responses Related to Topics

Category	Description	Examples
Keywords	Student uses keywords.	"GDPR," "Entity Matching"
Replicating content	Student replicates or summarizes lecture content.	"Exploratory models focus on historical relationships and minimum bias."
Reflecting on content	Student reflects on content learned.	"Started much much earlier than expected. In the 18th century!"
Reflecting on process	Student reflects on their learning process.	"I wasn't in lecture and haven't still watched video → Idea to watch lecture in Monday."
Providing feedback	Student provides feedback on instruction.	"Slides were good and gave a good overview of the topic. Could have more technical content though."

Table 6. Number of Responses in Each Category

Category	Keywords	Replicating content	Reflecting on content	Reflecting on process	Providing feedback	Total
Human coder 1	360 (43.6%)	253 (30.7%)	65 (7.9%)	115 (13.9%)	32 (3.9%)	825 (100%)
Human coder 2	342 (41.5%)	244 (29.6%)	89 (10.8%)	120 (14.5%)	30 (3.6%)	825 (100%)

Relationship responses were coded and categorized in a similar way (Table 7 and Table 8). The most common categories were "Overlapping" (51.4% and 36.0%) and "Transfer" (27.7% and 41.1%). The interrater reliability for the raters was found to be moderate, with kappa = 0.415 ($p < 0.001$).

Regarding the relevance of this to students' self-regulation and conceptual thinking, there is an apparent connection. We argue that the category "Reflecting on process" is linked to the appraisal phase in SRL models. The category "Reflecting on content" and the relationship categories "Overlapping," "Common theme," "Transfer," and "Hierarchy" are related to students' conceptual thinking.

According to evaluation criterion B, the effort required from students should be reasonable, so that students are willing to use the artifact. For Course 1, the response rate showed a decreasing trend from 80% (12 out of 15) in the first week to 53% (8 out of 15) in the final week (Table 9). In Course 2, the trend was similar, with 100% (44 out of 44) response rate in the first week and 61% (27 out of 44) in the final week (Table 10).

Regarding design artifact scalability (evaluation criterion C), no fundamental differences were detected between the two contexts included in the case study in relation to use of the design artifact.

Table 7. Categorization of Open Responses Related to Topic Relationships

Category	Description	Examples
Overlapping	Connected nodes contain common elements, stated in the relationship label.	Competitive market intelligence ↔ Business analytics: "Golden record management"
Common theme	Connected nodes have a common theme, stated in the relationship label.	Business analytics ↔ Performance marketing: "Both topics concern the importance of business analytics to ensure your business is profitable and as efficient as possible."
Transfer	Knowledge of the node is beneficial while learning about the connected node.	BI architecture ↔ Ecommerce plan and architecture: "BI architecture also important here."
Hierarchy	The node can be considered to be a part of the connected node.	Data manipulation ↔ Information production process "Data manipulation is a part of data supply."

Table 8. Content Analysis on Topic Entries

Category	Overlapping	Common theme	Transfer	Hierarchy	Total
Human coder 1	150 (51.4%)	45 (15.4%)	81 (27.7%)	16 (5.5%)	292 (100%)
Human coder 2	105 (36.0%)	46 (15.8%)	120 (41.1%)	21 (7.2%)	292 (100%)

Table 9. Students' DFS Usage in Course 1

Week	1	2	3	4	5	6	7	8	9	10	11	12	13
Number of diary submissions	12	11	11	10	12	8	7	9	5	8	9	8	8
Number of topic relations	112	65	101	85	118	83	79	99	73	94	98	97	99
– per student	9.3	5.9	9.2	8.5	9.8	10.4	11.3	11.0	14.6	11.8	10.9	12.1	12.4

Table 10. Students' DFS Usage on Course 2

Week	1	2	3	4	5	6	7
Number of diary submissions	44	43	36	35	32	31	27
Number of topic relations	131	197	203	249	251	263	272
– per student	3.0	4.6	5.6	7.1	7.8	8.5	10.1

5.2. Discussion

A major issue with learning analytics applications is that they often fail to reach the impact stage, that is, change behaviour. We argue that this is partly because the data used in the applications is often selected based not on learning science but on other factors (e.g., availability). To create impactful learning analytics applications, we should place more emphasis on considering what kind of data is gathered in the first place and how the student is involved in creating or enriching the data in the process. The theory of SRL provides an excellent framework for these considerations.

Our evaluation results show that the design artifact produced relevant data related to students' self-regulation and conceptual thinking. Over 10% of the open responses expressed traces of reflection on the learning process. In addition, the topic relationship responses provided valuable information on students' conceptual thinking. However, the case study did not include clear instruction to students on how to use the open response items, resulting in a wide variety of answers. In the future, carefully considered and theory-informed definitions of the question items are needed. For example, to gather data on the preparatory phase of SRL, a question item would need to be formulated accordingly.

A few additional characteristics define the data collected. First, temporal aspects are central: structured learning diaries are filled in weekly, allowing us to analyze and visualize learning as a process. Temporal analytics is recognized as an area of synergistic potential and productive dialogue between learning sciences and learning analytics (Knight, Wise, & Chen, 2017), and time-series data allows us to study students' psychological states instead of traits (Geiser et al., 2017).

Second, all gathered data is domain-specific, since all the data points connect to curriculum topics defined by the teacher in charge of the course. This is important, because students' relations with learning content may vary drastically: the same student may be anxious about one topic and at the same time be very excited about another. In addition, having a common vocabulary for the topics is beneficial when using the data in learning analytics applications and interventions.

Third, the data on topic relationships provides many novel opportunities. Constructing meaning by assimilating new concepts and propositions into existing conceptual frameworks is at the heart of human learning. Online concept mapping provides a way to gather data on this process. As presented by Strautmane (2012), there are multiple ways of using concept mapping data to assess students' knowledge — thus, this data could also be seen as a measure of learning outcomes.

As found in the literature review, one key issue with structured learning diaries is that they require a lot of effort from the student. Our results show a decreasing trend in the response rate as the course progresses. We suspect that this is at least partly due to the process, which required students to download a desktop application and send attachments instead of directly interacting with the map inside the learning management system. Moreover, whereas in most of the studies cited in section 2.2 the standardized question items are domain-general, in DFS they are domain-specific; each standardized item is repeated within each course topic. This adds to the effort required of students. It is also one of the strengths of the current approach, so an optimal balance must be found.

One interesting area concerns data privacy, and here we should focus on studies related to the willingness of students to share their data. In their learning diaries, students were willing to share experiences that were quite personal. Based on our experience, it seems that students might be willing to tell institutions how they are doing, as opposed to having institutions make such conclusions based on students' digital fingerprints.

5.3.   Future Directions

The goal of this paper was to design a methodology by which data on SRL and conceptual thinking could be acquired during a course. This methodology contributes to recent development in the field of educational data mining and learning analytics, where a shift toward a more holistic view of a student has been recognized in data collection (Hellas et al., 2018). When connected with learning analytics applications and interventions such as dashboards or personalized feedback at scale (Pardo et al., 2017), DFS could potentially include and affect institutional, instructional, and learner-level actors. The applicability of the experience-type data collected with DFS should be studied as a part of the teaching process and in curriculum design. We suggest the following aspects to be studied further:

SRL feedback loop: While studying, the student carries out formal self-monitoring, which is documented in a structured learning diary. Data from all students is gathered and used as a basis for learning analytics interventions aimed at supporting the development of SRL skills. An example is detecting each student's level of reflection in the diary and then providing personalized suggestions.

Adaptive teaching feedback loop: As part of instructional design, the teacher needs to decide on how the intended curriculum will actually be implemented and choose topics to include in the curricular concept map. The learning diaries built by students on the curricular concept maps in turn act as feedback, enabling the teacher to adapt their teaching.

Curriculum development feedback loop: Curricular concept maps are prepared based on the official curricula, directing students' self-monitoring. Data from these self-monitoring activities is used to provide information on areas that are well functioning, contradictory, overlapping, and/or isolated within the curriculum, providing an evidence base for further curriculum development.

6. Conclusions

In this study, we have presented the design of DFS, a toolset that combines curriculum mapping, concept mapping, and structured learning diaries to produce data related to students' SRL and conceptual thinking. This data, combined with learning

analytics applications and interventions, provides possibilities to add value for students, teachers, and academic leaders.

The design of DFS utilized DSR methodology, where theory and practice both play key roles in the design process of an artifact. Interaction between theory and practice is the essence of learning analytics. In this paper, we combined these into a viable artifact, which was evaluated in a case study. Even though the artifact's user interface was not given much attention in the design phase, the tool was rather easily adopted by students. We expected students to be familiar with concept mapping and learning diaries. Based on our experience, we suggest that the design of learning analytics tools should utilize processes already familiar to the target audience and integrate these applications closely with processes that already exist in the institution.

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

References

Ausubel, D. P. (1963). *The Psychology of Meaningful Verbal Learning*. Oxford, U.K.: Grune & Stratton.

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 427–449). New York: Springer. https://dx.doi.org/10.1007/978-1-4419-5546-3_28

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, *10*(4), 405–418. https://dx.doi.org/10.1109/TLT.2017.2740172

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://dx.doi.org/10.1016/0005-7916(94)90063-9

Chard, S. M. (2017). Using design science in educational technology research projects. *University of Mindanao International Multidisciplinary Research Journal, 2*(1).

Chmielewski, T. C., & Dansereau, D. F. (1998). Enhancing the recall of text: Knowledge mapping training promotes implicit transfer. *Journal of Educational Psychology*, *90*(3), 407–413. https://dx.doi.org/10.1037/0022-0663.90.3.407

Daley, B. J., & Torre, D. M. (2010). Concept maps in medical education: An analytical literature review. *Medical Education*, *44*(5), 440–448. https://dx.doi.org/10.1111/j.1365-2923.2010.03628.x

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*(3), 231–264. https://dx.doi.org/10.1007/s11409-008-9029-x

Dignath-van Ewijk, C., Fabriz, S., & Büttner, G. (2015). Fostering self-regulated learning among students by means of an electronic learning diary: A training experiment. *Journal of Cognitive Education and Psychology*, *14*(1), 77–97. https://dx.doi.org/10.1891/1945-8959.14.1.77

Fabriz, S., Dignath-van Ewijk, C., Poarch, G., & Büttner, G. (2014). Fostering self-monitoring of university students by means of a standardized learning journal—A longitudinal study with process analyses. *European Journal of Psychology of Education*, *29*(2), 239–255. https://dx.doi.org/10.1007/s10212-013-0196-z

Geiser, C., Götz, T., Preckel, F., & Freund, P. A. (2017). States and traits. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment*, *33*(4), 219–223. https://dx.doi.org/10.1027/1015-5759/a000413

Greene, J. A., & Schunk, D. H. (2017). Historical, contemporary, and future perspectives on self-regulated learning and performance. In *Handbook of Self-Regulation of Learning and Performance* (pp. 17–32). New York: Routledge.

Gurlitt, J., & Renkl, A. (2010). Prior knowledge activation: How different concept mapping tasks lead to substantial differences in cognitive processes, learning outcomes, and perceived self-efficacy. *Instructional Science*, *38*(4), 417–433. https://dx.doi.org/10.1007/s11251-008-9090-5

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. *Handbook of Self-Regulation of Learning and Performance*, *30*, 65–84. https://psycnet.apa.org/record/2011-12365-005

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)

119

Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V. V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., & Liao, S. N. (2018). Predicting academic performance: A systematic literature review. In *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (ITiCSE 2018 Companion), 2–4 July 2018, Larnaca, Cyprus (pp. 175–199). New York: ACM. https://dx.doi.org/10.1145/3293881.3295783

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems, 19*(2), Article 4. https://aisel.aisnet.org/sjis/vol19/iss2/4/

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. https://doi.org/10.2307/25148625

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., Koivuniemi, M., & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development: ETR & D*, *63*(1), 125–142. https://dx.doi.org/10.1007/s11423-014-9358-1

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In *Data Driven Approaches in Digital Education, Proceedings of the 12th European Conference on Technology Enhanced Learning* (EC-TEL 2017), 12–15 September 2017, Tallinn, Estonia (pp. 82–96). Lecture Notes in Computer Science, Springer. https://dx.doi.org/10.1007/978-3-319-66610-5_7

Kerr, J. F. (1968). *Changing the Curriculum*. (J. F. Kerr, Ed.). London, U.K.: University of London Press.

Ketonen, E. E., Dietrich, J., Moeller, J., Salmela-Aro, K., & Lonka, K. (2018). The role of daily autonomous and controlled educational goals in students' academic emotion states: An experience sampling method approach. *Learning and Instruction*, *53*, 10–20. http://dx.doi.org/10.1016/j.learninstruc.2017.07.003

Knight, S., Wise, A. F., & Chen, B. (2017). Time for change: Why learning analytics needs temporal analysis. *Journal of Learning Analytics*, *4*(3), 7–17. https://dx.doi.org/10.18608/jla.2017.43.2

Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., & Dawson, S. (2018). Understand students' self-reflections through learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (LAK '18), 5–9 March 2018, Sydney, NSW, Australia (pp. 389–398). New York: ACM. https://dx.doi.org/10.1145/3170358.3170374

Klug, J., Ogrin, S., Keller, S., Ihringer, A., & Schmitz, B. (2011). A plea for self-regulated learning as a process: Modelling, measuring and intervening. *Psychological Test and Assessment Modeling*, *53*(1), 51–72. https://psycnet.apa.org/record/2011-13343-004

Marzouk, Z., Rakovic, M., Liaqat, A., Vytasek, J., Samadi, D., Stewart-Alonso, J., Ram, I., Woloshen, S., Winne, P. H., & Nesbit, J. C. (2016). What if learning analytics were based on learning science? *Australasian Journal of Educational Technology*, *32*(6). https://dx.doi.org/10.14742/ajet.3058

Munezero, M., Montero, C. S., Mozgovoy, M., & Sutinen, E. (2013). Exploiting sentiment analysis to track emotions in students' learning diaries. In *Proceedings of the 13th Koli Calling International Conference on Computing Education Research*, 14–17 November 2013, Koli, Finland (pp. 145–152). New York: ACM. https://dx.doi.org/10.1145/2526968.2526984

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*(3), 413–448. https://dx.doi.org/10.3102%2F00346543076003413

Novak, J., & Musonda, D. (1991). A twelve-year longitudinal study of science concept learning. *American Educational Research Journal*, *28*(1), 117–153. https://dx.doi.org/10.3102/00028312028001117

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*, 422. https://dx.doi.org/10.3389/fpsyg.2017.00422

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students (in Spanish). *Electronic Journal of Research in Educational Psychology, 11*(2), 551–576. https://dx.doi.org/10.14204/ejrep.30.12200

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. https://dx.doi.org/10.1016/j.edurev.2017.08.004

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*, *60*(6), 723–735. https://dx.doi.org/10.1080/00313831.2015.1066436

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2017). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology: Journal of the Council for Educational Technology, 50*(1), pp. 128–138. https://dx.doi.org/10.1111/bjet.12592

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)

120

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91–105. https://dx.doi.org/10.1207/S15326985EP3702_4

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 451–502). San Diego, CA, USA: Academic Press. https://dx.doi.org/10.1016/B978-012109890-2/50043-3

Puustinen, M., and Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research, 45,* 269–286. https://dx.doi.org/10.1080/00313830120074206

Schmitz, B., Klug, J., & Schmidt, M. (2011). Assessing self-regulated learning using diary measures with university students. In D. H. Schunk & B. Zimmerman (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 251–266). New York: Routledge. https://dx.doi.org/10.4324/9780203839010

Schmitz, B., & Perels, F. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition and Learning*, *6*(3), 255–273. https://doi.org/10.1007/s11409-011-9076-6

Schmitz, B., & Wiese, B. S. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology*, *31*(1), 64–96. https://dx.doi.org/10.1007/s11409-011-9076-6

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior, 78,* 397–407. https://dx.doi.org/10.1016/j.chb.2017.06.030

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2018). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, in press. https://dx.doi.org/10.1016/j.chb.2018.05.004

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *The American Behavioral Scientist*, *57*(10), 1380–1400. https://dx.doi.org/10.1177/0002764213498851

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin*, *137*(3), 421–442. https://dx.doi.org/10.1037/a0022777

Strautmane, M. (2012). Concept map-based knowledge assessment tasks and their scoring criteria: An overview. In A. J. Cañas, J. D. Novak, & J. Vanhear (Eds.), *Concept Maps: Theory, Methodology, Technology — Proceedings of the Fifth International Conference on Concept Mapping* (Vol. 2, pp. 80–88), 17–20 September 2012, Valletta, Malta. http://cmc.ihmc.us/cmc2012papers/cmc2012-p113.pdf

van den Akker, J. (2003). Curriculum perspectives: An introduction. In J. van den Akker, W. Kuiper, & U. Hameyer (Eds.), *Curriculum Landscapes and Trends* (pp. 1–10). Dordrecht, Netherlands: Springer.

Weber, S. (2012), Comparing key characteristics of design science research as an approach and paradigm. *Proceedings of the 2012 Pacific Asia Conference on Information Systems* (PACIS 2012), 11–15 July 2012, Ho Chi Minh City, Vietnam (paper 180). http://aisel.aisnet.org/pacis2012/180

Willcox, K. E., & Huang, L. (2017). Network models for mapping educational data. *Design Science*, *3*, E18. https://dx.doi.org/10.1017/dsj.2017.18

Winne, P. H. (2010). Bootstrapping learner's self-regulated learning. *Psychological Test and Assessment Modeling*, *52*(4), 472–490. https://www.psychologie-aktuell.com/fileadmin/download/ptam/4-2010_20101218/08_Winne.pdf

Winne, P. H. (2017). Learning analytics for self-regulated learning. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 241–249). Society for Learning Analytics Research. https://dx.doi.org/10.18608/hla17.021

Winne, P. H., & Baker, R. S. J. d. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *JEDM|Journal of Educational Data Mining*, *5*(1), 1–8. https://jedm.educationaldatamining.org/index.php/JEDM/article/view/28

Winne, P. H., & Hadwin, A. F. (2013). nStudy: Tracing and supporting self-regulated learning in the internet. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 293–308). New York: Springer. https://dx.doi.org/10.1007/978-1-4419-5546-3

Zimmerman, B. J., & Paulsen, A. S. (1995). Self-monitoring during collegiate studying: An invaluable tool for academic self-regulation. *New Directions for Teaching and Learning*, *1995*(63), 13–27. https://dx.doi.org/10.1002/tl.37219956305