



Concept maps as assessment for learning in university mathematics

Tanya Evans¹ · Inae Jeong¹

Accepted: 10 January 2023 / Published online: 1 March 2023
© The Author(s) 2023

Abstract

A concept map is a visualisation of a group of related concepts with their relationships identified by directed arrows, which can be viewed as an externalisation of a schema encoded in a learner's long-term memory. Concept maps have become popular learning tools in various educational settings. However, the existing research has not fully explored the effectiveness of concept mapping as a learning tool nor analysed its utility as an assessment tool in mathematics. This study reports on the implementation and evaluation of a novel assessment, concept mapping, in a university mathematics course ($N=219$). We investigated relationships between concept mapping performance and two major outcome variables: academic achievement and assessment self-efficacy (a two-factor measure comprising *Comprehension and Execution* and *Emotional Regulation*). Hierarchical multiple regression showed that concept mapping performance explains a statistically significant amount of variance in both the final exam scores and the *Emotional Regulation* factor of assessment self-efficacy after accounting for other conventional coursework assessments. The explained variance in the final exam suggests that concept mapping used as an assessment tool can detect and evaluate learners' distinct capabilities that are not discerned by conventional assessments. Hence, concept mapping's potential to assess conceptual understanding warrants further investigations. Moreover, the association with the emotional regulation efficacy measure suggests that concept mapping as a learning activity involves more positivity about the ability to succeed in facing challenges than a typical assignment. This potentially indicates more perseverance and effortful learning while actively engaged in meaning-making and the type of relational reasoning elicited by concept mapping.

Keywords Concept map · Assessment · Mathematics education · Conceptual understanding · Self-efficacy · Undergraduate mathematics

✉ Tanya Evans
t.evans@auckland.ac.nz

Inae Jeong
ijeo175@aucklanduni.ac.nz

¹ Department of Mathematics, University of Auckland, Auckland, New Zealand

1 Introduction

Concept maps are becoming a popular educational tool used by teachers and learners across many disciplinary areas. A concept map can be broadly defined as a node-link diagram in which each node represents a concept, and each link represents the relationship between the two concepts it connects, with labels on the lines to specify the relationships (Gurlitt & Renkl, 2010; Schroeder et al., 2018). Diagrams resembling concept maps have been used by philosophers and logicians for many centuries (Nesbit & Adesope, 2013), but the term “concept map” and the idea to use it as an educational tool originated with Joseph Novak and his colleagues in the 1970s (Novak, 1990). The main assumption made by the advocates was that constructing concept maps is an effective way to promote meaningful learning. The key feature of a high-quality concept map is a clear hierarchical representation of concepts arranged with more general concepts placed higher on the map and linked to more specific concepts placed lower on the map (Novak & Cañas, 2008). Novak and colleagues have further recommended that effective concept maps should include horizontal cross-links to elaborate on the relationships other than generality/specificity. Figure 1 illustrates three different categories of concept maps for “Flowers”, according to the classification of structural quality formulated by Kinchin et al. (2000). The first type is “spoke” (A in Fig. 1), where the main concept is placed at the centre, and the only connections found in the map are between each derived concept and the central concept, resembling the shape of a spoke. The second type is “chain” (B in Fig. 1), which has a linear structure, developing only a single strand of a specific aspect of the main concept. Both “spoke” and “chain” are not considered good examples of effective concept mapping. The third type, “net” (C in Fig. 1), is regarded as the best of the three because, in addition to a hierarchy depicting more specific concepts of each strand from the central concept downwards, it also reveals cross-link connections between the concepts developed in different strands.

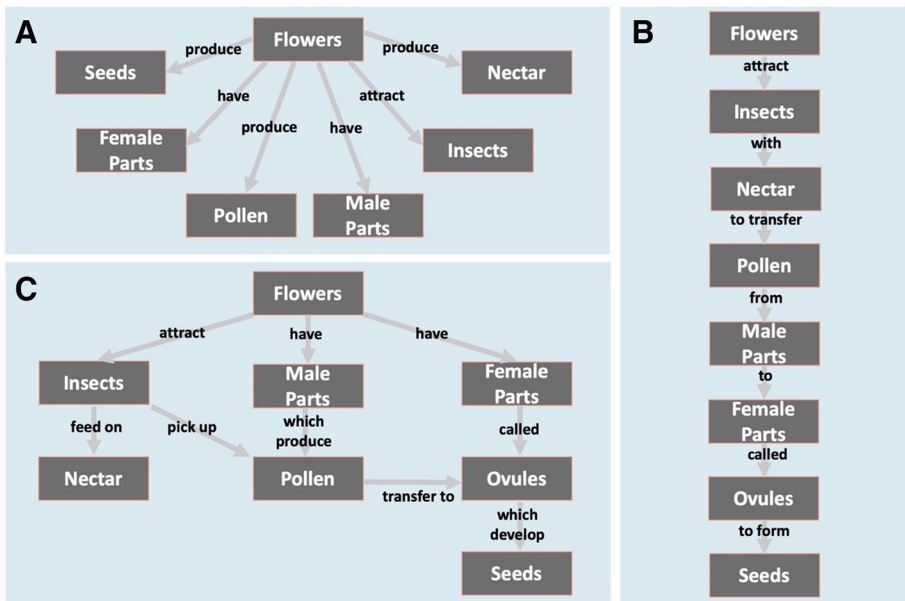


Fig. 1 Structural types of concept maps, adapted from Kinchin et al. (2000)

Educators can use concept mapping in two major ways. First, an expert can devise a concept map as a learning tool to be provided as a guide for learners to study. Alternatively, students can be tasked to create a map by themselves. Supporting the original premise of early proponents, the extant research shows that concept mapping activity has a much greater impact in the latter setting, particularly at a postsecondary level (Schroeder et al., 2018). This is likely to be due to the meaningful engagement required from the mapper when creating their concept map. This action involves the so-called higher-order learning activities such as organising and synthesising. As a result of engaging in such activities, the concept mapping process is believed to enable high-quality learning (Nesbit & Adesope, 2006; O'Day & Karpicke, 2020; Schroeder et al., 2018). However, no integrative cognitive theory has yet been employed to explain the processes by which students learn with concept maps (Gurlitt & Renkl, 2010; Nesbit & Adesope, 2006, 2013).

Despite the lack of theory, many studies have been published demonstrating promising results about the concept mapping utility in educational settings. A recent meta-analytic study showed that learning with concept maps produced a moderate, statistically significant effect ($g = 0.58$, $p < .001$) compared to other conventional forms of studying in a synthesis of 142 studies (Schroeder et al., 2018). Of the 142 studies included in the meta-analysis (Schroeder et al., 2018), 118 were classified as STEM. Most investigated the use of concept mapping in learning natural sciences: biology, physics, and chemistry. At the tertiary level, three studies investigated concept mapping as a learning tool in statistics courses (Chiou, 2009; Lambiotte et al., 1993; Sas, 2008). However, none of the studies included in the meta-analysis considered concept mapping in the tertiary mathematics context.

A small corpus of qualitative case studies exists in the mathematics education literature, which illustrates various implementations and suggests the potential viability of concept mapping in mathematics classrooms (Afamasaga-Fuata, 2009; Baroody & Bartels, 2000; Gallenstein, 2011; Ryve, 2004; Wilcox & Sahloff, 1998; Williams, 1998). There is evidence that examples of concept mapping are used in teacher training in the UK, USA and Australia (Afamasaga-Fuata, 2009; Ollerton & Watson, 2001; Prestage & Perks, 2013; Schmittau, 2004), although such use may not be fully reported in research journals. Some work has been done to describe the training needed to enable students to construct concept maps in a Grade-8 Chinese classroom (Jin & Wong, 2010) and unpack the nature of the conceptual understanding held by this group through the use of concept maps (Jin & Wong, 2015, 2023).

In summary, it appears that, specifically for mathematics, the existing research has not fully explored the effectiveness of concept mapping as a learning tool nor analysed its utility as an assessment tool. The present study attempts to make progress by reporting on incorporating concept mapping into a university mathematics course as a weekly task. We employ an exploratory study methodology with two aims: (1) to analyse the relationship between student concept mapping performance and student performance on the final exam and (2) to investigate the relationship between concept mapping performance and a fundamental psychometric construct linking cognitive and affective domains — learner self-efficacy. Before introducing the study, we first outline theoretical justifications in support of concept mapping as a learning and assessment tool based on the educational psychology and mathematics education perspectives and briefly overview the existing literature on concept maps and self-efficacy.

1.1 Theoretical and empirical foundations

1.1.1 Concept mapping as an effective learning strategy

Our understanding of the mechanisms involved in learning from a cognitive perspective has advanced substantially over the last five decades by employing mathematical modelling and experiential testing of those models (Atkinson & Shiffrin, 1968). Simplistically, a modal model of Human Cognitive Architecture could be used to illustrate the current research state relevant to our context (Sweller, 2008; Sweller et al., 2019). Figure 2, modified from Inglis and Mejía-Ramos (2021), illustrates the three main components of human cognitive architecture and their interactions (Atkinson & Shiffrin, 1968; Clark & Paivio, 1991; Fiorella & Mayer, 2015; Mayer & Moreno, 2003). The first component, sensory memory, allows for the incoming sensory information (such as what we see, hear, touch, smell, etc.) to be stored long enough for the selected information to be transferred to working memory (second component). Even after the stimuli have ceased, impressions of sensory information could be retained in sensory memory for short periods of time, provided they are selected for further processing in the working memory by the mechanism of attention (Dehaene, 2020).

Working memory represents the domain where all conscious thinking happens — it represents a state of awareness while processing information. Crucially, working memory has two limitations: first, the duration for which the information can be stored there is very

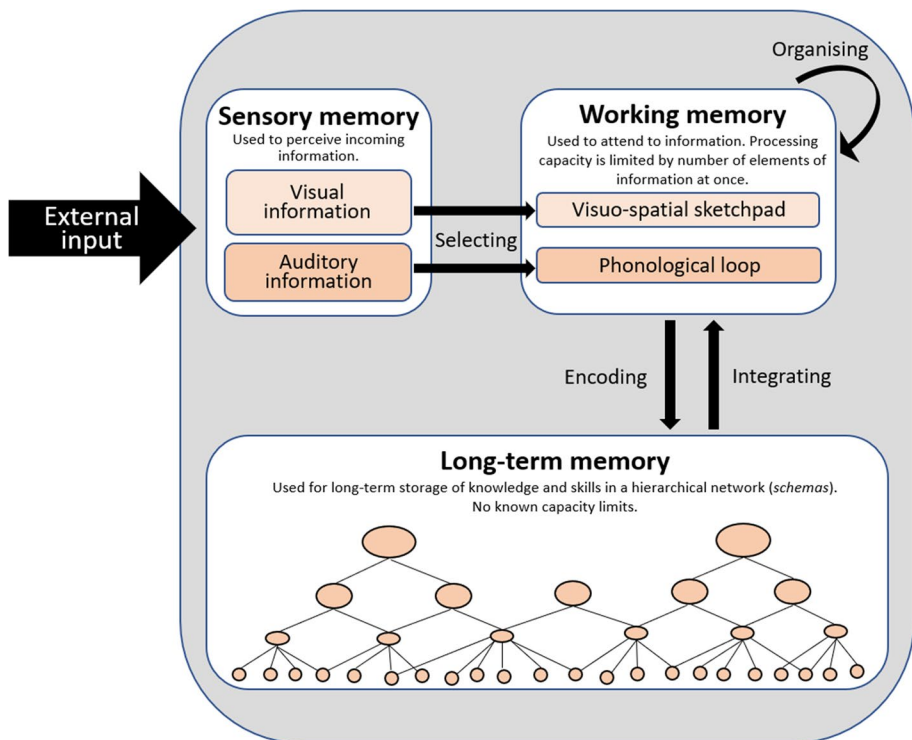


Fig. 2 A modal model of human cognitive architecture (modified from Inglis and Mejía-Ramos (2021))

short; almost all of the information is lost after 20 s without rehearsal (Peterson & Peterson, 1959; Sweller, 2021). Second, working memory is extremely limited in capacity. It has been known since the mid-1950s that working memory can only hold about seven items of information (plus or minus two) (Miller, 1956). However, only about three to four items of information can be processed simultaneously while mentally combining, comparing, or manipulating the items somehow (Cowan, 2001). As an example, on average, we can remember about seven random digits, but if asked to reorder them from, say, highest to lowest, the successful completion of the task would be challenging unless the number of digits is reduced.

The process of learning happens through working memory. There are two ways information can enter working memory. By focusing attention on certain incoming sensory information, one might consciously process a stimulus from sensory memory. The information can also come from the third memory system, long-term memory. Long-term memory is a central component in human cognitive architecture; it represents a repository of an enormous network of complex units of closely linked snippets of information, called schemas. We now know that our sense of self comes from that vast amount of information stored in long-term memory (Sweller, 2021). It is well-evidenced that the main reason experts outperform novices is because of the superior quality of their schemas that organise a sizable body of knowledge and index it by a large number of patterns that, on recognition, guide the expert in a fraction of a second to the relevant parts (Chi et al., 1981; Ericsson & Lehmann, 1996; Larkin et al., 1980; Sweller et al., 1998). These schemas constitute long-term memory and, when required, are retrieved and integrated into working memory (Chase & Simon, 1973). Long-term memory appears to have no practical capacity limits and, thus, could be used to overcome the temporal limitations of working memory. If you have stored a schema in long-term memory, you can repeatedly reintegrate it into working memory, transcending the 20-second limit. Furthermore, long-term memory also functions as a bypass with the capacity limits of working memory. Suppose snippets of information are organised into a coherent schema and stored in long-term memory. In that case, the whole schema can be brought into working memory as one item of information to be manipulated and integrated with other information units.

Since establishing the central role of long-term memory, educational psychologists and neuroscientists have devoted a great deal of attention to understanding how we learn most effectively. It is well understood that in educational contexts, learning is not just an additive process of storing new snippets of information in memory, as in a computer (Fiorella & Mayer, 2015). Rather, learning depends on two major factors: (1) what is presented and how and (2) the cognitive processing that the learner is actively engaged in during learning. Thus, learning is viewed as a generative activity. This conception of learning is a natural theoretical advance built on the coalescence of the cognitive revolution and constructivist ideas about the importance of meaning-making while learning. This well-evidenced theory envisions effective learning to comprise three stages: (1) learners actively select the relevant aspects of incoming information by paying attention, (2) which is followed by organising this information into a coherent cognitive structure in working memory, and (3) integrating cognitive structures with relevant prior knowledge activated from long-term memory (Fiorella & Mayer, 2015).

Mathematics education research further developed these conceptions to distinguish between different types of mathematical understanding and knowledge. These ideas can be traced back to Skemp's (1976) distinction between "relational understanding" and "instrumental understanding" of mathematics. The latter describes a limiting yet commonly occurring understanding based on knowing a set of rules without understanding

the reasons, whereas “learning relational mathematics consists of building up a conceptual structure (schema)” (Skemp, 1976, p. 14). Later research on mathematical thinking divided mathematics knowledge into two types: procedural knowledge and conceptual knowledge (Hiebert & Lefevre, 1986), with the latter defined as “knowledge that is rich in relationships. It can be thought of as a connected web of knowledge, a network in which the linking relationships are as prominent as the discrete pieces of information” (pp. 3–4). This dichotomy was subsequently questioned and reconceptualised by Star (2005) to include the depth dimension (accounting for quality).

The latest mathematics cognition perspective assumes that conceptual understanding is developed when a sufficiently well-organised schema has been encoded into long-term memory (Inglis & Mejía-Ramos, 2021). From this perspective, concept mapping is assumed to make learning efficient because it promotes meaningful learning and requires learners to engage deeply with the material by focusing on the organisational structure of a set of related concepts and producing elaborative connections among them (Gurlitt & Renkl, 2010; Nesbit & Adesope, 2006). The term “knowledge elaboration” is often used in reference to meaningful learning, emphasising the importance of using prior knowledge to expand and refine new insights utilising processes such as organising, restructuring, interconnecting, integrating new elements of information, and identifying relations between them (Kalyuga, 2009). Research has shown that knowledge elaboration is the key mechanism behind the success of well-known learning strategies such as self-explanations (Chi et al., 1994) and elaborative interrogation (Dunlosky et al., 2013). Viewed from this perspective, the similarities between knowledge elaboration strategies and the processes involved in high-quality concept mapping are recognisable. According to Karpicke and Blunt (2011, p. 772), “concept mapping bears the defining characteristics of an elaborative study method: It requires students to enrich the material they are studying and encode meaningful relationships among concepts within an organized knowledge structure.” In other words, the construction of a high-quality concept map involves high-order thinking necessary to explicate not only generality/specificity relations (e.g., concept maps A and B in Fig. 1) but also cross-links — the connections between concepts derived from different strands developed from the main concept (e.g., C in Fig. 1). Such constructivist activity is likely to foster meaningful mathematics learning and promote “relational understanding” in the sense of Skemp (1976) by activating efficient processes in working memory in selecting relevant information (from learning resources), integrating it with the existing schemas in long-term memory, and reorganising this knowledge into a new, bigger and better-organised schema, which could subsequently be encoded into long-term memory.

1.1.2 Concept mapping as an assessment tool

Research on assessment distinguishes between summative assessments, designed to determine academic progress after a set unit of material (i.e., assessment of learning) and formative assessments, designed to monitor student progress during the learning process and provide feedback (i.e., assessment for learning) (Chappuis & Stiggins, 2002). However, this distinction has not been a primary focus of recent tertiary mathematics studies. Contrary to the assumption that formative and summative assessment approaches are incompatible, a recent article demonstrated how these assessment forms could be combined in university mathematics teacher education (Buchholtz et al., 2018). Thus, the central consideration in mathematics education at a tertiary level is about what type of reasoning is elicited and assessed by various tasks. Specifically, in line with the distinction between procedural and

conceptual knowledge, an important question is how to design valid and reliable methods to assess different types of understanding (e.g., instrumental vs. relational in the sense of Skemp (1976)). Research is conclusive that the way in which procedural knowledge is assessed has become relatively standardised. In contrast, the supposedly designed tasks to assess conceptual knowledge do not always align with theoretical claims about mathematical understanding (Crooks & Alibali, 2014). Moreover, an extensive overview of undergraduate mathematics assessment revealed that most questions on exams and coursework could be categorised as “imitative” — questions that can be solved by performing prescribed algorithms and recalling analogous (if not identical) solutions (Bergqvist, 2007; Iannone et al., 2020; Iannone & Simpson, 2011). Researchers pointed out that this issue is amplified in service mathematics courses designed for non-mathematics majors (Mac an Bhaird et al., 2017). A comparable service course serves as a setting for the present study, in which we investigate the feasibility of concept mapping as an assessment tool to diversify measuring capacity and as a learning strategy to foster relational reasoning.

1.2 Self-efficacy

Self-efficacy is a commonly used psychometric construct introduced by Bandura (1977) as part of his social cognitive theory. Self-efficacy is defined as a set of one’s beliefs about their capability to perform and complete a particular task (Bandura, 1997). The construct has long been considered to play an essential role in achievement motivation, and much past research has shown that self-efficacy is a significant predictor of academic performance (Pajares & Graham, 1999; Zimmerman, 2000). Since the 1970s, a large body of research based on Bandura’s conception of self-efficacy has been developed, making this construct one of the most studied and utilised in educational psychology research.

Mathematics-specific self-efficacy has been given special consideration in mathematics education research and is generally viewed as a belief about one’s capacity for doing mathematics (Pajares & Miller, 1995). Methods for measuring it have been developed and used to demonstrate that students with higher self-efficacy tend to show greater interest, effort, persistence, help-seeking behaviour, and, ultimately, greater mathematics achievement than those who feel less efficacious (Pajares & Graham, 1999; Pajares & Kranzler, 1995; Pajares & Miller, 1995; Schukajlow et al., 2019; Skaalvik et al., 2015; Williams & Williams, 2010).

1.2.1 Types of self-efficacy

Some researchers make a distinction between academic self-efficacy, which is defined as “general perceptions of academic capability” (Richardson et al., 2012, p. 356), and performance self-efficacy, which is defined as “perceptions of academic performance capability” (Richardson et al., 2012, p. 356). When students face a situation, the type of self-efficacy which influences the student depends on whether the situation is familiar. If the situation is familiar, the student will use previous experience to predict their performance based on their performance self-efficacy. However, if the situation is unfamiliar, students tend to rely on the more generalised competency beliefs to predict their performance, which is known as academic self-efficacy (Zimmerman et al., 1992). Specifically for mathematical problem-solving in a high school classroom, Pajares and Kranzler (1995) demonstrated that performance self-efficacy has a strong direct effect on mathematics anxiety and performance even when controlling for the general mental ability.

More recent research sought to develop our understanding of assessment-related self-efficacy with a focus on both students' beliefs about content-specific tasks and their beliefs around assessment-taking, such as their beliefs about whether or not they are good at taking tests. For example, it is not uncommon to hear people lament that they are "not good at taking a test" or "I know I can do it, just not on an exam". To incorporate this more nuanced perspective into the construct of self-efficacy, researchers developed a new instrument, the Measure of Assessment Self-Efficacy (MASE), designed to assess beliefs related to assessment preparation and performance, which is most relevant to research undertaken in a natural educational setting such as a university course (Riegel et al., 2022).

1.2.2 Sources of self-efficacy

There are ways to improve one's self-efficacy. Bandura (1997) suggested four sources of self-efficacy: (1) mastery experience, (2) vicarious experience, (3) social persuasion, and (4) emotional and psychological states. Mastery experience, which can be thought of as former attainment, was reported to have the most substantial influence (Usher & Pajares, 2008). This source is especially powerful when the experience involves overcoming a challenge.

1.2.3 Self-efficacy and concept mapping

Only a few studies have investigated the effect of concept mapping on self-efficacy. For example, a study by Chularut and Debacker (2004) used a randomised pretest-posttest control group design to examine the impact of concept mapping on students learning English as a second language in the USA. The researchers found that the students in the concept mapping group showed a slightly greater self-efficacy increase than those in the alternative learning strategy group. Similar results were reported by Khajavi (2012), who undertook a study involving students enrolled in an English reading comprehension course in Iran, demonstrating a substantial effect size of the concept mapping intervention. Analysing the effect on self-efficacy from a different angle, Gurlitt and Renkl (2010) showed that different concept mapping activities (label-provided-lines vs. create-and-label-lines) resulted in significantly different self-efficacy measures in a study of German psychology students.

In summary, although it is naturally anticipated that concept mapping could positively impact self-efficacy, this topic has rarely been studied, and no such studies were undertaken in mathematics education contexts. From the perspective of mathematics learning theories, the type of reasoning elicited by concept mapping is conducive to forming a relational/conceptual understanding of mathematics, which could enhance self-efficacy as a mastery experience source. Furthermore, self-efficacy could be impacted through the mechanism of metacognition: regular concept mapping tasks prompt students to be critically aware of their thinking and learning and reflect on the depth of their understanding. This is likely to impact their perceptions of themselves as mathematics thinkers and learners, thereby affecting their self-efficacy (Coutinho, 2008).

1.3 Research questions

The present study had two broad goals. The first is to report on the design, development, and deployment of a new type of assessment in a large mathematics course at the university

level. Students were tasked to construct a concept map weekly either as part of a problem-set during their tutorials (practical problem-solving sessions) or as a marked assignment. The second goal was to evaluate this initiative using student coursework data and their responses to a questionnaire (administered at the end of the semester) by conducting statistical analyses to determine and compare the effects of various assessment forms on student performance and self-efficacy. Specifically, we sought to answer the following research questions:

RQ1. Does student concept mapping performance explain a statistically significant amount of variance in the final exam scores after accounting for other coursework assessments?

RQ2. Does student concept mapping performance explain a statistically significant amount of variance in their self-efficacy (assessment-related) after accounting for other coursework assessments?

2 Methods

2.1 Research site

The research site of this study was a large research-intensive university (University of Auckland, New Zealand). The study was implemented in a stage-II undergraduate mathematics course that serves the needs of students majoring in various disciplines such as finance, economics, chemistry, physics, computer science, statistics, etc. The contents covered in this course are Calculus II, Linear Algebra II, and Introduction to Ordinary Differential Equations. As for most courses in this university, this course is delivered over 12 teaching weeks with the repeated structure of three 1-hour lectures and one 1-hour tutorial (25 to 30 students per room working on problems) each week. The lectures include active learning activities such as interactive whole-class questioning/discussions and think-pair-share tasks.

2.2 Participants and study description

In the trial semester, out of the 355 students enrolled in the course, 35 students studied overseas, completing the course online; the remaining 320 studied locally. Tutorials are an essential component of this course and have been implemented in an active learning format for over a decade. During tutorials, students are encouraged to interact and solve various problems in small groups. All students are expected to attend a tutorial each week, starting from the 2nd week of the semester. In addition to this work, students must submit their solutions to a weekly “marked problem”, which are traditional written assignments except for the two that were concept mapping tasks. Hence, study participants were expected to construct a concept map weekly, either as part of the tutorial question set or when assigned as a marked problem.

Specifically, two marked problems were assigned, asking students to construct concept maps on the following topics: in Week 3 on *Series* and Week 7 on *Vector Spaces*. A template, called a Knowledge Organiser, was provided to the students as a guide. Prior to creating a concept map, students were prompted to state the definition of a given concept and provide at least two examples and at least one non-example. Students were allowed a week

to complete the task. Out of ten concept maps that students completed during the course, only the two assigned as “marked problems” were collected for marking.

In the first lecture of the course, the two authors of this study provided a short introductory session on concept mapping. In addition, an example of a concept map on Linear Systems (a prerequisite topic for the course) was shared with the students online so that students could refer to it anytime. However, after collecting student submissions for marking in Week 3, the quality of the concept maps indicated that some students did not fully understand the instructions for this task or found it very difficult to construct a concept map. Hence, the first author, who is a lecturer of the course, gave another explanation during a lecture to demonstrate how to construct a high-quality concept map.

A model solution was provided to the students, noting that there can be many variations of a high-quality concept map (Fig. 3). In order to assess students’ work on concept maps, a rubric was developed and validated by the researchers, which was reported in a separate study (Jeong & Evans, 2021). The validation methodology was based on comparing four rubrics, developed according to previous research, by examining their reliability and validity. Rubric 1 was based on a qualitative assessment of intrinsic qualities of a concept map (such as the type of concept map: spoke, chain, net (Kinchin et al., 2000)), combined with perceived effort on a discrete 4-point scale. The other three rubrics were analytical, counting the number of concepts (Rubric 2), relationships (Rubric 3) and the (inverse) ratio between them (Rubric 4). Regression modelling was used to compare the extent to which the four rubrics explained the variance in the scores on a concept map question on the final exam. (The final exam included a question on concept mapping — see Figs. 4 and 5). Two rubrics were identified as most valid: Rubrics 1 and 4. However, taking into account reliability considerations, Rubric 4 was selected as the most suitable, ensuring objectivity in assigning scores: The ratio method outputs a numerical value equal to the inverse ratio of the number of concepts used in the concept map to the number of relationships identified. For example, the score for the concept map in Fig. 3 is obtained by calculating the inverse ratio of the number of concepts (=25) and the number of relationships between

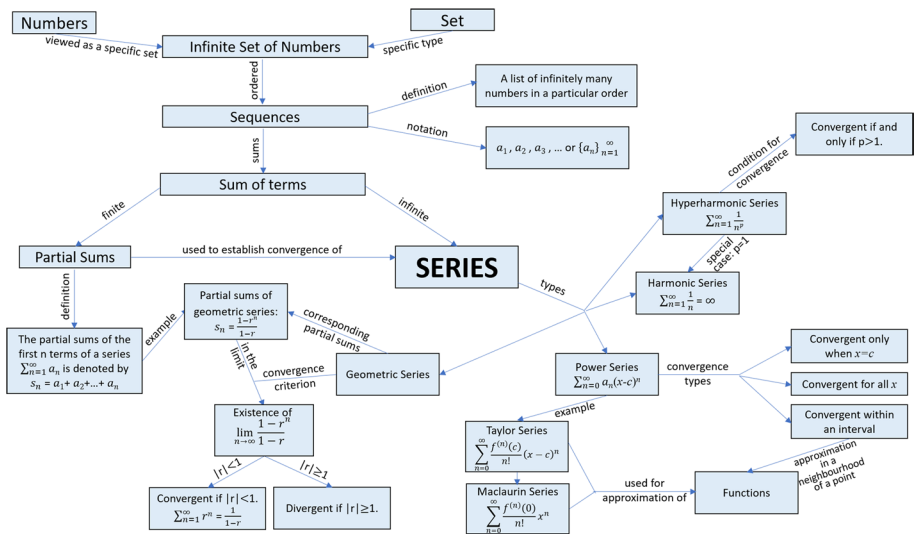


Fig. 3 Concept map on Series

This question has three parts - make sure you answer all of them.

Consider this incomplete concept map. Fill out the three blanks: The main concept **A** in the grey ellipse is . The concept missing in grey box **B** is . The missing connecting relation labelled by **C** is .

Fig. 4 Concept map question in the final exam

them (=30). Hence the score is $30/25 = 1.2$. According to this method, for a fixed number of concepts, the higher scores indicate high-level elaboration, with more related concepts connected vertically and horizontally in a net-type formation. On the other hand, fewer

Orthonormal matrix	The Wronskian	excludes is a linear combination of is a span of generalises to generalises to
Markov chain	Set of solutions of a linear differential equation	
Least squares solutions	Series	
Vector space	General solution of $Ax=b$	
Partial derivative	Least squares solutions	
Discrete dynamical system	Sequence	
Symmetric matrix	Normal equations	
Square matrix	Set of solutions of a linear differential equation	
Vector space	Set of solutions of a linear differential equation	

Fig. 5 In-line choice options for A, B, and C for identifying missing elements of the concept map in Fig. 4

concepts used limit the upper bound of the inverse ratio since the maximum number of connections for a graph with n nodes is $\frac{n(n-1)}{2}$ (in the case of a complete graph). Hence, the maximum inverse ratio for a graph with n nodes is $\frac{n-1}{2}$. An extreme example is a concept map with only two concepts connected by one relation with a score of $\frac{1}{2} = 0.5$. (More information can be found in Jeong & Evans, 2021).

2.3 Exam question on concept mapping

The final exam comprised 30 multi-choice questions, with one question on concept mapping organised as three in-line choice options worth three marks (10% of the final exam). The overarching idea of this question was to assess students' consolidated macro view of the concept of *Vector Space* as a mathematical structure used to generalise physical spaces such as \mathbb{R} , \mathbb{R}^2 , \mathbb{R}^3 , and then \mathbb{R}^n ; to define fundamental subspaces related to matrices (such as Nullspace, Column space, Eigenspace) and to be utilised in solving differential equations. The concept map question was presented as a completed concept map with a few missing concepts and relationships for students to identify (Fig. 4).

Students were presented with three in-line choice questions with multi-choice options to replace objects A, B, and C in the concept map. The multi-choice options given for each of A, B, and C are shown in Fig. 5, with the correct answers at the bottom of the lists. In addition, considering that it may be the first time students would have encountered such a question in an examination, a mock exam containing a similar question was provided at the end of the semester.

3 Data collection and availability statement

The data was collected from the Learning Management System (Canvas) and a questionnaire administered at the end of the semester through Qualtrics. All students ($N=355$) were invited to participate in the study but not all consented. The final dataset contained 219 participants. The raw dataset is available at <https://doi.org/10.17608/k6.auckland.19618014>.

3.1 Coursework marks

Coursework marks considered for this study included the following:

- 1 final exam (50%).
- 1 mid-semester test (20%).
- 30 quizzes (15%).
- 10 marked problems (10%).
- 10 tutorial participation marks (5%).

The 30 quizzes were short multi-choice assessments conducted online, assessing basic comprehension of material introduced in a previous lecture, starting from Week 2 (Evans et al., 2021). The lowest four marks were disregarded, with the sum of the top 26 quiz marks counted towards the final course mark.

The ten marked problems were short written assignments due at the end of each week, submitted for marking and feedback. As explained above, two of the marked problems

were on concept maps, one on *Series* and the other on *Vector Spaces*. The ratio method (described above) was used to assign scores to students' concept maps.

The exam comprised 30 multi-choice questions, with more than a third (12 points out of 30) attempting to assess conceptual understanding (including the question on filling in the blanks in a concept map described in 2.3). Exam questions are available at <https://doi.org/10.17608/k6.auckland.19618014>.

3.2 Self-efficacy measures

In order to measure students' self-efficacy in this study, a tailor-made instrument measuring tutorial-related self-efficacy was used: the Measure of Assessment Self-Efficacy for Mathematics Tutorials (MASE-T), which was based on the general assessment self-efficacy instrument, MASE (Riegel et al., 2022). The instrument explicitly measures assessment-related self-efficacy and comprises two factors: (1) *Comprehension and Execution* and (2) *Emotional Regulation*. The MASE items assess participants' beliefs in their ability to understand, perform, and emotionally regulate while studying for and during various types of course assessment. Instruments to measure exam-related self-efficacy (MASE-E) and quiz-related self-efficacy (MASE-Q) were developed and validated (Riegel et al., 2021, 2022), providing general guidelines for adapting the instrument to any other type of assessment. In this study, the guidelines were used for tutorials since they involved mathematical problem-solving without the stress associated with high-stakes assessments such as tests and exams. Students were prompted with the following tutorial scenario:

Please envision yourself in the scenario: you are enrolled in a mathematics course that has a weekly tutorial worth 0.5% of your final grade. The tutorial contains four questions on a topic you studied in your course. In considering taking part in this tutorial, please rate the extent to which you agree or disagree with the following statements. (The items comprising the scales are listed in Table 1).

Responses to statements were measured using a slider scale from 1 to 100 (where 1=Cannot do at all, 50=Moderately sure can do, and 100=Highly certain can do). Based on confirmatory factor analysis of students' responses collected in week twelve

Table 1 Tutorial-related self-efficacy measure (MASE-T)

Factor/Item

Comprehension and Execution

1. I can understand the content and skills needed for the assessment.
2. When preparing for the assessment, I can organize my time well.
3. I can understand the questions in the assessment.
4. I can correctly answer the questions on the assessment.
5. During the assessment, I can answer the questions within the time constraints.

(Scale $\alpha=0.917$)

Emotional Regulation

6. During my preparation, I am able to cope with my negative emotions toward the assessment.
7. Even when I struggle while studying, I am able to stay positive about my ability to succeed.
8. During the assessment, I am able to cope with my negative emotions toward the assessment.

(Scale $\alpha=0.910$)

of the semester, a two-factor model was confirmed, which was consistent with the general structure of MASE; the model offered an acceptable fit ($\chi^2/df=2.203$, CFI=0.986, SRMR=0.0238). The internal consistency of the scales was confirmed using Cronbach's alphas: $\alpha=0.917$ for the *Comprehension and Execution* factor and $\alpha=0.910$ for the *Emotional Regulation* factor, with values higher than 0.7 reflecting the reliability of the scales.

4 Results

In addressing the study's research questions, hierarchical multiple regression analyses were undertaken. Table 2 summarises the descriptive statistics and correlations of the variables included in the analyses.

4.1 Concept map score as a predictor of the final exam score (RQ1)

A hierarchical multiple regression analysis was undertaken to investigate the ability of concept map scores to explain the variance in the final exam scores after controlling for other more traditional assessment components. Preliminary analyses were conducted to ensure no violation of the assumptions of normality, linearity, and homoscedasticity. Additionally, the correlations among the predictor variables included in the study were examined, presented in Table 2. Quiz scores were not included in the analysis due to high correlations with both tutorial marks and marked problems. Most correlations were weak to moderate, ranging between $r=.17$, $p<.01$ and $r=.43$, $p<.001$ with one exception: borderline high correlation between tutorial scores and marked problems ($r=.70$, $p<.001$). Moreover, the addition of the tutorial scores was found to reduce the adjusted R^2 , thus justifying its removal as a predictor in the final model (see Table 3).

In the first step of hierarchical multiple regression, the test score was entered as the most likely predictor of the exam score. This model was statistically significant $F(1, 217)=37.49$; $p<.001$ and explained 14.7% of the variance in the final exam scores (adjusted $R^2=0.143$). The addition of marked problem scores in Step 2 resulted in a statistically significant increase in adjusted R^2 of 0.050, $F(1, 216)=14.43$, $p<.001$. In the final model (Step 3), after controlling for the test and marked problem scores, the addition of the concept map scores resulted in a statistically significant increase in adjusted R^2 of 0.025, $F(1, 215)=4.89$, $p=.028$. All three assessment types made a significant unique

Table 2 Descriptive statistics and correlations of factors in the hierarchical multiple regressions

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7
1. Final exam	60.27	17.19							
2. CE factor of MASE-T	64.54	18.39							
3. ER factor of MASE-T	65.59	19.51							
4. Total MASE-T	64.93	18.24							
5. Test	66.80	19.31	0.38***	0.40***	0.38***	0.41***			
6. MP	55.80	19.94	0.38***	0.34***	0.24***	0.31***	0.43***		
7. Tutorial	87.14	22.89	0.22**	0.25***	0.18**	0.23***	0.30***	0.70***	
8. Concept map	0.66	0.52	0.26***	0.17**	0.22**	0.19**	0.19**	0.31***	0.31***

Note. $N=219$. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. *CE* Comprehension and Execution, *ER* Emotional Regulation, *MP* score marked problem score

Table 3 Summary of Model 1

Variable	Final exam score					
	Step 1		Step 2		Step 3	
	<i>B</i>	β	<i>B</i>	β	<i>B</i>	β
Constant	37.45***		31.73***		31.10***	
Test score	0.34***	0.38	0.24***	0.27	0.23***	0.26
MP score			0.22***	0.26	0.19**	0.22
Concept map score					4.70*	0.14
R^2_{adj}	0.143***		0.193***		0.208***	
ΔR^2_{adj}	0.143***		0.050***		0.025*	

Note. $N = 219$. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. MP score marked problem score

contribution to the final model (Table 3), with the concept map score recording a standardised Beta value of 0.14, $p < .05$, indicating that after accounting for the effects of all other variables, with every increase by one standard deviation in the concept mapping score, the final exam score increases by 0.14 standard deviation. In summary, the concept map scores significantly increased the proportion of explained variance in the final exam score above and beyond what was accounted for by the test and marked problem scores alone.

4.2 Concept map score as a predictor of the tutorial-related self-efficacy (RQ2)

Hierarchical multiple regression analyses (Model 2a and Model 2b) were conducted to ascertain whether or not concept mapping tasks make an independent contribution to explaining the variance in the *Tutorial-Related Self-Efficacy* construct, which comprises two factors: *Comprehension and Execution* (CE) and *Emotional Regulation* (ER). Initially, we used all coursework assessment components as predictors, but these initial models were not feasible due to a multicollinearity problem between quizzes and marked problems, and also between quizzes and tutorial marks, indicated by high correlations (which were greater than 0.7). Therefore, a decision was made to remove the quiz scores from the model. After removing quizzes, none of the predictors had pairwise correlations higher than 0.7, and all other model assumptions were satisfied.

4.2.1 Model 2a (comprehension and execution factor of MASE-T)

The summary of Model 2a is presented in Table 4, in which the full model (Step 4) with the test, marked problem, tutorial and concept map scores predicting the *Comprehension and Execution* factor of tutorial self-efficacy (CE of MASE-T) was statistically significant, $R^2 = 0.198$, $F(4, 214) = 13.21$, $p < .001$; adjusted $R^2 = 0.183$. The first model (Step 1), where only the test score predicts CE of MASE-T, was statistically significant with adjusted $R^2 = 0.158$, $F(1, 217) = 41.98$, $p < .001$. The addition of marked problem scores resulted in a significant increase of adjusted R^2 of 0.030, $F(1, 216) = 9.04$, $p < .01$. However, the addition of tutorial marks and concept maps scores resulted in a reduction of adjusted R^2 . The only significant predictor in the final model was the test score.

Table 4 Summary of Model 2a

Variable	Comprehension and Execution in tutorial self-efficacy							
	Step 1		Step 2		Step 3		Step 4	
	<i>B</i>	β	<i>B</i>	β	<i>B</i>	β	<i>B</i>	β
Constant	38.94***		34.08***		33.24***		33.30***	
Test score	0.38***	0.40	0.30***	0.31	0.30***	0.31	0.30***	0.31
MP score			0.19**	0.20	0.17*	0.19	0.17	0.18
Tutorial mark					0.02	0.02	0.01	0.02
Concept map score							1.66	0.05
R^2_{adj}	0.158***		0.188***		0.185***		0.183***	
ΔR^2_{adj}	0.158***		0.030**		-0.003		-0.002	

Note. $N = 219$. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. MP score marked problem score

4.2.2 Model 2b (emotional regulation factor of MASE-T)

The summary of Model 2b, explaining the variance in the *Emotional Regulation* factor of tutorial self-efficacy (ER of MASE-T), is given in Table 5. The full model (Step 4) with the test, marked problem, tutorial and concept map scores predicting ER of MASE-T was statistically significant, $R^2 = 0.169$, $F(4, 214) = 10.89$, $p < .001$; adjusted $R^2 = 0.154$. The first model (Step 1), where the only predictor is the test score, is statistically significant with adjusted $R^2 = 0.141$, $F(1, 217) = 36.69$, $p < .001$. In Step 2, the addition of marked problem scores resulted in a non-significant increase of adjusted R^2 of 0.003, $F(1, 216) = 1.76$, $p = .19$. The addition of tutorial participation marks in Step 3 resulted in a small reduction in the adjusted R^2 . However, the addition of concept map scores in the final Step 4 resulted in a statistically significant increase in adjusted R^2 of 0.014, $F(1, 214) = 4.40$, $p = .037$. This indicates that the concept map score explains a statistically significant amount of variance in ER of MASE-T over and above what was explained by all other assessment components. Importantly, only the test and concept map scores have a significant effect on ER of MASE-T in the final model, with the standardised Beta values of $\beta = 0.33$, $p < .001$ and

Table 5 Summary of Model 2b

Variable	Emotional Regulation in Tutorial Self-Efficacy							
	Step 1		Step 2		Step 3		Step 4	
	<i>B</i>	β	<i>B</i>	β	<i>B</i>	β	<i>B</i>	β
Constant	39.93***		37.59***		36.44***		36.65***	
Test score	0.38***	0.38	0.34***	0.34	0.34***	0.34	0.33***	0.33
MP score			0.09	0.09	0.07	0.07	0.05	0.05
Tutorial mark					0.03	0.03	0.01	0.01
Concept map score							5.26*	0.14
R^2_{adj}	0.141***		0.144***		0.140***		0.154***	
ΔR^2_{adj}	0.141***		0.003		-0.004		0.014*	

Note. $N = 219$. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. MP score marked problem score

$\beta = 0.14$, $p = .037$, respectively. This means that, out of all low-stakes assessments, only the concept mapping regression coefficient is statistically significantly different to 0, inferring that there is a positive linear relationship with the ER of MASE-T after accounting for the effect of all other variables.

In summary, student concept mapping performance explains a statistically significant amount of variance in the *Emotional Regulation* factor of self-efficacy after accounting for other coursework assessments, which is an unexpected and somewhat surprising result.

5 Discussion

One of the major aims of this study was to test whether student concept mapping performance explains a statistically significant amount of variance in the final exam scores after accounting for other coursework assessments (RQ1). Hierarchical multiple regression was used to ascertain this, which is nested modelling with a way to control for an individual contribution of a variable to explain variance in the dependent variable above and beyond what was accounted for by other variables and to test the significance of the improvement. First, all three assessment types considered (Test ($\beta = 0.26$), Marked problems ($\beta = 0.22$), and Concept maps ($\beta = 0.14$)) made significant unique contributions to the final model (Table 3). Second, the addition of concept map scores in the final step of the nested modelling has significantly improved the model's ability to explain variance in the final exam scores above and beyond what was accounted for by the test and marked problem scores alone. This finding supports the hypothesis that concept mapping is a principally different type of assessment, detecting learners' capabilities that are not discerned by regular assessments. Taken together with the latest mathematics cognition perspective, which assumes that conceptual understanding is developed when a sufficiently well-organised schema has been encoded into long-term memory, it is plausible to suggest that concept mapping (as an externalisation of a schema) may have the potential to assess conceptual understanding. This would be useful since the limitations of conventional assessments in evaluating conceptual understanding are widely accepted. Many attempts to design special instruments to assess conceptual understanding have been made. For example, a Calculus Concept Inventory (CCI) was developed by Epstein (2007), which has been widely used in research and practice for over a decade (Epstein, 2013). However, serious criticism has recently been levelled at the validity of the CCI measure (Gleason et al., 2019). Moreover, in a study involving foundation mathematics students in the UK, the low internal consistency for both the subset and full set of CCI items suggested that the instrument does not measure a single construct (Bisson et al., 2016), thus demonstrating that accurate evaluation of conceptual understanding via an instrument remains a largely elusive goal.

From the theoretical perspective, this is also a plausible hypothesis. The construction of a high-quality concept map involves explicating not only generality/specificity relations but also cross-links — the connections between concepts derived from different strands developed from the central concept (Fig. 1 C). An ability to construct such a map is likely to correspond to a “relational understanding” in the sense of Skemp (1976), reflecting a superior conceptual understanding.

However, a major limitation of this study is the lack of robust evidence to make generalisable claims about the utility of concept mapping as an assessment tool for conceptual understanding. Future research could focus on a detailed analysis of qualitative features of concept mapping that differentiate it from conventional assessments and by comparing it

with the existing valid methods for assessing conceptual understanding, such as a comparative judgement (CJ) approach, which was successfully used for assessing conceptual understanding of some mathematical topics: derivatives, p-values and use of letters in algebra (Bisson et al., 2016, 2020; Jones et al., 2019).

The other aim of this study was to investigate whether student concept mapping performance explains a statistically significant amount of variance in their self-efficacy after accounting for other coursework assessments (RQ2). Hierarchical multiple regression was used to compare two models: Model 2a (predicting the *Comprehension and Execution* factor of tutorial self-efficacy) and Model 2b (predicting the *Emotional Regulation* factor of tutorial self-efficacy). Both models were significant, albeit with only a small percentage of self-efficacy variance explained as indicated by the values of R^2_{adj} in the range of 0.154–0.183. However, one noteworthy difference between the models is that the change in R^2_{adj} in the final step (adding concept mapping) is only significant in Model 2b. This indicates that concept mapping performance explains a statistically significant amount of variance in the *Emotional Regulation* factor of self-efficacy after accounting for other coursework assessments. This finding was somewhat unexpected. It is known that the main source of self-efficacy is mastery experience, which is especially powerful when the experience involves succeeding on challenging tasks (Bandura, 1997; Usher & Pajares, 2008). To that end, it was reasonable to expect that concept mapping performance would contribute to explaining the *Comprehension and Execution* factor of self-efficacy because of the nature of the concept mapping activity. The type of reasoning elicited by concept mapping is conducive to forming a relational/conceptual understanding of mathematics, which could lead to enhanced appraisals of comprehension. Moreover, the comprehension marker of self-efficacy could have been impacted through metacognition, as explained in Section 1.2.3: concept mapping tasks prompt students to be critically aware of their thinking and learning and to reflect on the depth of their comprehension. However, our result did not confirm this hypothesis. Instead, the regression modelling pointed to the importance of investigating emotional regulation as part of the cognitive processes involved in learning with concept mapping.

The role of emotions and beliefs in task completion has been recently identified as an important area of research on mathematics-related affect (Hannula, 2002; Hannula et al., 2016; Schindler & Bakker, 2020; Zan et al., 2006). From a broader perspective, an extensive body of research has demonstrated that students' emotions profoundly affect their academic engagement and performance, identifying a particular impact of positive and negative moods on task completion (Pekrun & Linnenbrink-Garcia, 2012). Given these considerations and our findings, we can hypothesise that successful concept mapping activity could enhance learners' emotional regulation appraisals due to the distinct characteristics of the activity. It could be that concept mapping is superior to other conventional types of assessment through its influence on the manifestation of positive emotions experienced by a learner when a comprehensive concept map is completed. The self-reflection on learning that happens at that moment could be more profound than, say, after completing a multi-choice test or submitting an assignment with problems mimicking worked examples provided. Through the repeated experience of regularly constructing high-quality concept maps during the course of the semester, the positive emotions induced by the learning activity could accumulate to substantial improvement in the learner's affect.

However, our results are limited in that they do not provide any evidence about the causality of the relationship. It could be that the direction of the relationship is the other way around. Learners with better emotional regulation produce higher quality concept maps. Given our results, we can only conclude that there is a positive linear relationship between

the *Emotional Regulation* factor of self-efficacy and concept mapping performance after accounting for the effect of other assessments. Still, we can not ascertain that the learning activity has caused an improvement in the *Emotional Regulation* factor measure. The measure consists of items pertaining to the suppression of negative emotions and recording the extent to which students agree with: “Even when I struggle while studying, I am able to stay positive about my ability to succeed”. A likely explanation is that learners who believe that when they struggle during a learning activity and can remain positive and succeed are more likely to persevere in the face of a challenge and not give up.

An important finding from our analysis is that among the low-stakes course assessments, only concept maps are positively associated with the *Emotional Regulation* factor of self-efficacy, controlling for other variables. In other words, no linear relationship with the conventional coursework assignments (weekly marked problems) was identified. There is no difference in the *Emotional Regulation* measure between those students who do well and those who do not on a typical homework assignment, controlling for other assessments. This contrasts with the concept mapping indicators, which, perhaps, points from yet another perspective that concept mapping is a principally different learning activity compared to a typical mathematics task. The theoretical perspective, outlined in Section 1.1.1, offers justification for this conclusion because success in a concept mapping task requires perseverance and sustained effort in deliberate meaning-making through deep engagement with the material by focusing on the organisational structure of a set of related concepts to produce elaborative connections among them as part of externalisation and reformation of a schema. Such effortful meaning-making could be a missing facet of many traditional mathematical tasks often assigned to learners.

6 Final remarks

Much has been said about the need to seek research-grounded solutions to improve practice. A particular focus has been placed on classroom-based interventions¹, which are rarely undertaken and evaluated in mathematics education (Stylianides & Stylianides, 2013). To that end, one of the goals of our study was to report on the design, development, implementation, and evaluation of a novel type of assessment in a university mathematics course. The design principle of our intervention is generalisable and transferable to other educational domains as a blueprint for an assessment structure and related instruction that could be utilised in mathematics education broadly. Stylianides and Stylianides (2013) propose three dimensions of evaluation of classroom interventions: (1) how amenable it is for scaling up, (2) how practicable it is for curricular integration, and (3) how capable it is of producing long-lasting effects. Evaluated this way, our intervention can arguably be deemed effective for the first two criteria: the number of students utilising concept mapping is unlimited; concept maps can be easily integrated into the existing curricula structures in a practicable way. Determining long-lasting effects is a more complex challenge, and ongoing research is needed. Future research could build on evidence presented in this study to investigate effectiveness of concept mapping as an assessment and a learning tool in different and/or properly controlled settings.

¹ The term “intervention” is used similarly to its standard use in medicine where an intervention denotes action taken to improve a situation (Stylianides & Stylianides, 2013) and does not necessitate an experimental design.

Acknowledgements We thank Kaitlin Riegel (University of Auckland) for expert advice on MASE-T validation.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This work was supported by the Faculty of Science Research and Development Grant (FRDF grant number: 3720159).

Data availability The raw dataset analysed during the study and exam questions are available at <https://doi.org/10.17608/k6.auckland.19618014>.

Declarations

Ethics approval Ethics approval was granted by the University of Auckland Human Participants Ethics Committee on 25/02/2021 for 3 years (reference Number UAHPEC21976).

Informed consent Data used in the study was obtained through informed consent. Assurance was provided that participation in the project was completely voluntary and confidential.

Competing interests The authors declare no competing interests.

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

References

- Afamasaga-Fuata, K. (2009). *Concept mapping in mathematics*. Springer. <https://doi.org/10.1007/978-0-387-89194-1>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2), 191. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. W H Freeman/Times Books/ Henry Holt & Co.
- Baroody, A. J., & Bartels, B. H. (2000). Using concept maps to link mathematical ideas. *Mathematics Teaching in the Middle School*, 5(9), 604–609. <https://doi.org/10.5951/mtms.5.9.0604>
- Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *The Journal of Mathematical Behavior*, 26(4), 348–370. <https://doi.org/10.1016/j.jmathb.2007.11.001>
- Bisson, M. J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education*, 2(2), 141–164. <https://doi.org/10.1007/s40753-016-0024-3>
- Bisson, M. J., Gilmore, C., Inglis, M., & Jones, I. (2020). Teaching using contextualised and decontextualised representations: examining the case of differential calculus through a comparative judgement technique. *Research in Mathematics Education*, 22(3), 284–303. <https://doi.org/10.1080/14794802.2019.1692060>
- Buchholtz, N. F., Krosanke, N., Orschulik, A. B., & Vorhölter, K. (2018). Combining and integrating formative and summative assessment in mathematics teacher education. *ZDM Mathematics Education*, 50(4), 715–728. <https://doi.org/10.1007/s11858-018-0948-y>
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational leadership*, 60(1), 40–44.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, 4(1), 55–81. [https://doi.org/10.1016/0010-0285\(73\)90004-2](https://doi.org/10.1016/0010-0285(73)90004-2)

- Chi, M. T., Glaser, R., & Rees, E. (1981). *Expertise in problem solving*. In: Pittsburgh Univ PA Learning Research and Development Center.
- Chi, M. T. H., De Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477. [https://doi.org/10.1016/0364-0213\(94\)90016-7](https://doi.org/10.1016/0364-0213(94)90016-7)
- Chiou, C. C. (2009). Effects of concept mapping strategy on learning performance in business and economics statistics. *Teaching in Higher Education*, 14(1), 55–69. <https://doi.org/10.1080/13562510802602582>
- Chularut, P., & Debacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology*, 29(3), 248–263. <https://doi.org/10.1016/j.cedpsych.2003.09.001>
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210. <https://doi.org/10.1007/BF01320076>
- Coutinho, S. (2008). Self-efficacy, metacognition, and performance. *North American Journal of Psychology*, 10(1), 165–172.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.
- Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, 34(4), 344–377. <https://doi.org/10.1016/j.dr.2014.10.001>
- Dehaene, S. (2020). *How we learn: the new science of education and the brain*. Penguin Books Limited. <https://books.google.co.nz/books?id=R9C2DwAAQBAJ>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Epstein, J. (2007). Development and validation of the calculus concept inventory. In *Proceedings of the Ninth International Conference on Mathematics Education in a Global Community* (pp. 165–170).
- Epstein, J. (2013). The Calculus Concept Inventory - measurement of the effect of teaching methodology in mathematics. *Notices of the American Mathematical Society*, 60, 1018–1027. <https://doi.org/10.1090/noti1033>
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47(1), 273–305.
- Evans, T., Kensington-Miller, B., & Novak, J. (2021). Effectiveness, efficiency, engagement: Mapping the impact of pre-lecture quizzes on educational exchange. *Australasian Journal of Educational Technology*, 37(1), 163–177. <https://doi.org/10.14742/ajet.6258>
- Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107707085>
- Gallenstein, N. L. (2011). Mathematics concept maps: assessing connections. *Teaching Children Mathematics*, 17(7), 436–440. <https://doi.org/10.5951/teachmath.17.7.0436>
- Gleason, J., Bagley, S., Thomas, M., Rice, L., & White, D. (2019). The calculus concept inventory: a psychometric analysis and implications for use. *International Journal of Mathematical Education in Science and Technology*, 50(6), 825–838. <https://doi.org/10.1080/0020739x.2018.1538466>
- 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.
- Hannula, M. S. (2002). Attitude towards mathematics: emotions, expectations and values. *Educational Studies in Mathematics*, 49(1), 25–46. <https://doi.org/10.1023/A:1016048823497>
- Hannula, M. S., Di Martino, P., Pantziara, M., Zhang, Q., Morselli, F., Heyd-Metzuyanım, E., Lutovac, S., Kaasila, R., Middleton, J. A., & Jansen, A. (2016). Attitudes, beliefs, motivation and identity in mathematics education: an overview of the field and future directions. *Springer Nature*. <https://doi.org/10.1007/978-3-319-32811-9>
- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. *Conceptual and procedural knowledge: The Case of Mathematics*, 2, 1–27.
- Iannone, P., & Simpson, A. (2011). The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 30(4), 186–196. <https://doi.org/10.1093/teamat/hrr017>
- Iannone, P., Czichowsky, C., & Ruf, J. (2020). The impact of high stakes oral performance assessment on students' approaches to learning: a case study. *Educational Studies in Mathematics*, 103(3), 313–337. <https://doi.org/10.1007/s10649-020-09937-4>
- Inglis, M., & Mejía-Ramos, J. P. (2021). Functional explanation in mathematics. *Synthese*, 198(S26), 6369–6392. <https://doi.org/10.1007/s11229-019-02234-5>

- Jeong, I., & Evans, T. (2021). Embedding concept mapping into university mathematics: comparison and validation of marking rubrics. In *Proceedings of the 13th Southern Hemisphere Conference on the Teaching and Learning of Undergraduate Mathematics and Statistics* (pp. 2–16). <https://doi.org/10.17608/k6.auckland.20330460.v1>
- Jin, H., & Wong, K. (2010). Training on concept mapping skills in geometry. *Journal of Mathematics Education*, 3(1), 104–119.
- Jin, H., & Wong, K. Y. (2015). Mapping conceptual understanding of algebraic concepts: an exploratory investigation involving grade 8 chinese students. *International Journal of Science and Mathematics Education*, 13(3), 683–703. <https://doi.org/10.1007/s10763-013-9500-2>
- Jin, H., & Wong, K. Y. (2023). Complementary measures of conceptual understanding: a case about triangle concepts. *Mathematics Education Research Journal*, 35(1), 153–174. <https://doi.org/10.1007/s13394-021-00381-y>
- Jones, I., Bisson, M., Gilmore, C., & Inglis, M. (2019). Measuring conceptual understanding in randomised controlled trials: can comparative judgement help? *British Educational Research Journal*, 45(3), 662–680. <https://doi.org/10.1002/berj.3519>
- Kalyuga, S. (2009). Knowledge elaboration: a cognitive load perspective. *Learning and Instruction*, 19(5), 402–410. <https://doi.org/10.1016/j.learninstruc.2009.02.003>
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775. <https://doi.org/10.1126/science.1199327>
- Khajavi, Y. (2012). Influencing EFL learners' reading comprehension and self-efficacy beliefs: the effect of concept mapping strategy. *Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras*. <https://doi.org/10.30827/digibug.31953>
- Kinchin, I. M., Hay, D. B., & Adams, A. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, 42(1), 43–57. <https://doi.org/10.1080/001318800363908>
- Lambiotte, J. G., Skaggs, L. P., & Dansereau, D. F. (1993). Learning from lectures: Effects of knowledge maps and cooperative review strategies. *Applied Cognitive Psychology*, 7(6), 483–497. <https://doi.org/10.1002/acp.2350070604>
- Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208(4450), 1335–1342. <https://doi.org/10.1126/science.208.4450.1335>
- Mac an Bhaird, C., Nolan, B. C., O'Shea, A., & Pfeiffer, K. (2017). A study of creative reasoning opportunities in assessments in undergraduate calculus courses. *Research in Mathematics Education*, 19(2), 147–162. <https://doi.org/10.1080/14794802.2017.1318084>
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2), 81. <https://doi.org/10.1037/h0043158>
- Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: a meta-analysis. *Review of Educational Research*, 76(3), 413–448.
- Nesbit, J. C., & Adesope, O. O. (2013). Learning through visual displays. In G. Schraw, M. T. McCrudden, & D. Robinson (Eds.), *Concept maps for learning: theory, research, and design* (pp. 303–328). IAP Information Age Publishing.
- Novak, J. D. (1990). Concept mapping: a useful tool for science education. *Journal of Research in Science Teaching*, 27(10), 937–949. <https://doi.org/10.1002/tea.3660271003>
- Novak, J. D., & Cañas, A. J. (2008). *The theory underlying concept maps and how to construct them*. In Florida Institute for Human and Machine Cognition.
- O'Day, G. M., & Karpicke, J. D. (2021). Comparing and combining retrieval practice and concept mapping. *Journal of Educational Psychology*, 113(5), 986–997. <https://doi.org/10.1037/edu0000486>
- Ollerton, M., & Watson, A. (2001). *Inclusive mathematics 11–18*. Continuum.
- Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, 24(2), 124–139. <https://doi.org/10.1006/ceps.1998.0991>
- Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*, 20(4), 426–443. <https://doi.org/10.1006/ceps.1995.1029>
- Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: the need for specificity of Assessment. *Journal of Counseling Psychology*, 42(2), 190–198. <https://doi.org/10.1037/0022-0167.42.2.190>

- Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259–282). Springer. https://doi.org/10.1007/978-1-4614-2018-7_12
- Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58(3), 193. <https://doi.org/10.1037/h0049234>
- Prestage, S., & Perks, P. (2013). *Adapting and extending secondary mathematics activities*. <https://doi.org/10.4324/9780203462386>
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353.
- Riegel, K., Evans, T., & Stephens, J. (2021). Predicting mathematics exam-related self-efficacy as a function of prior achievement, gender, stress mindset, and achievement emotions. In S. S. Karunakaran & A. Higgins (Eds.), *2021 Research in Undergraduate Mathematics Education Reports* (pp. 255–263). <http://sigmaa.maa.org/rume/Site/Proceedings.html>
- Riegel, K., Evans, T., & Stephens, J. M. (2022). Development of the measure of assessment self-efficacy (MASE) for quizzes and exams. *Assessment in Education: Principles Policy & Practice*, 29(6), 729–745. <https://doi.org/10.1080/0969594X.2022.2162481>
- Ryve, A. (2004). Can collaborative concept mapping create mathematically productive discourses? *Educational Studies in Mathematics*, 56(3), 157–177. <https://doi.org/10.1023/b:educ.0000040395.17555.c2>
- Sas, M. (2008). *The effects of students' asynchronous online discussions of conceptual errors on intentionally flawed teacher-constructed concept maps*. UNLV Retrospective Theses & Dissertations. 2800. <https://doi.org/10.25669/ewio-y711>
- Schindler, M., & Bakker, A. (2020). Affective field during collaborative problem posing and problem solving: a case study. *Educational Studies in Mathematics*, 105(3), 303–324. <https://doi.org/10.1007/s10649-020-09973-0>
- Schmittau, J. (2004). Uses of concept mapping in teacher education in mathematics. AJ Canãs, JD Novak & Gonázales (Eds.), *Concept Maps: Theory, Methodology, Technology. Proceedings of the First International Conference on Concept Mapping*.
- Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing Concept Maps: a Meta-analysis. *Educational Psychology Review*, 30(2), 431–455. <https://doi.org/10.1007/s10648-017-9403-9>
- Schukajlow, S., Achmetli, K., & Rakoczy, K. (2019). Does constructing multiple solutions for real-world problems affect self-efficacy? *Educational Studies in Mathematics*, 100(1), 43–60. <https://doi.org/10.1007/s10649-018-9847-y>
- Skaalvik, E. M., Federici, R. A., & Klassen, R. M. (2015). Mathematics achievement and self-efficacy: relations with motivation for mathematics. *International Journal of Educational Research*, 72, 129–136. <https://doi.org/10.1016/j.ijer.2015.06.008>
- Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77(1), 20–26.
- Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 36(5), 404–411.
- Stylianides, A. J., & Stylianides, G. J. (2013). Seeking research-grounded solutions to problems of practice: classroom-based interventions in mathematics education. *ZDM Mathematics Education*, 45(3), 333–341. <https://doi.org/10.1007/s11858-013-0501-y>
- Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- Sweller, J. (2008). Human cognitive architecture. *Handbook of Research on Educational Communications and Technology*, 35, 369–381.
- Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Sweller, J. (2021). *Why inquiry-based approaches harm students' learning*. The Centre for Independent Studies.
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in School: critical review of the literature and future directions. *Review of Educational Research*, 78(4), 751–796. <https://doi.org/10.3102/0034654308321456>
- Wilcox, S. K., & Sahloff, M. (1998). Assessment: another perspective on concept maps: empowering students. *Mathematics Teaching in the Middle School*, 3(7), 464–469. <https://doi.org/10.5951/mtms.3.7.0464>

- Williams, C. G. (1998). Using concept maps to assess conceptual knowledge of function. *Journal for Research in Mathematics Education JRME*, 29(4), 414–421. <https://doi.org/10.5951/jresematheduc.29.4.0414>
- Williams, T., & Williams, K. (2010). Self-efficacy and performance in mathematics: reciprocal determinism in 33 nations. *Journal of Educational Psychology*, 102(2), 453–466. <https://doi.org/10.1037/a0017271>
- Zan, R., Brown, L., Evans, J., & Hannula, M. S. (2006). Affect in mathematics education: an introduction. *Educational Studies in Mathematics*, 113–121. <https://doi.org/10.1007/s10649-006-9028-2>
- Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: the role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29(3), 663–676. <https://doi.org/10.3102/00028312029003663>
- Zimmerman, B. J. (2000). Self-efficacy: an essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82–91. <https://doi.org/10.1006/ceps.1999.1016>

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