STEM Education
https://www.aimsciences.org/journal/A0000-0006

STEM Education, 1 (4): 256–278
DOI: 10.3934/steme.2021017
Received: March 2021
Accepted: November 2021

*Article*

# Non-routine mathematical problem-solving: creativity, engagement, and intuition of STEM tertiary students

**Tanya Evans[1]*, Sergiy Klymchuk[2], Priscilla E. L. Murphy[2], Julia Novak[1], Jason Stephens[3] and Mike Thomas[1]**

[1] Department of Mathematics, University of Auckland, Auckland, New Zealand; j.novak@auckland.ac.nz (J.N.); moj.thomas@auckland.ac.nz (M.T.)
[2] Department of Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand; sergiy.klymchuk@aut.ac.nz (S.K.); priscilla.murphy@aut.ac.nz (P.E.L.M.)
[3] School of Learning Development and Professional Practice, University of Auckland, Auckland, New Zealand; jm.stephens@auckland.ac.nz (J.S.)

* **Correspondence:** t.evans@auckland.ac.nz; Tel: +64-9-923-8783

Academic Editor: Zlatko Jovanoski

**Abstract:** This study set out to evaluate an intervention that introduced a period of non-routine problem-solving into tertiary STEM lectures at four tertiary institutions in New Zealand for 683 students. The aim was twofold: to attempt to increase student engagement and to introduce them to the kind of domain-free abstract reasoning that involves critical, creative, and innovative thinking. This study was conducted using a mixed-methods approach, utilizing different types of instruments to gather data: comprehensive student pre- and post-test questionnaires, a content validation survey for the questionnaires, focus group interviews (student participants), open-ended questionnaire (lecturer participants), and naturalistic class observations. The main findings are as follows. Students' behavioural engagement was significantly greater during the intervention. Perceptions of the utility value of the activity improved at the end of the semester for all students. There were no significant changes in students' convergent thinking (problem-solving), intuition, or creativity (originality, fluency, and elaboration traits of the divergent thinking) during the course, probably due to the relatively short timescale of the intervention. However, overall, the results of the investigation suggest that with a relatively small effort, teachers can improve STEM student engagement by devoting a few minutes per lecture on non-routine problem-solving. This is something that can be easily implemented, even by those who primarily teach in a traditional lecturing style.

**Keywords:** engagement, puzzle-based learning, STEM education, tertiary education, mixed-methods

## 1. Introduction

In 2012, the New Zealand Government identified, as a priority, the need to address the undersupply of students studying STEM subjects for delivering its Business Growth Agenda1. Low engagement and retention rates in STEM subjects contribute to the shortage of STEM graduates, producing a negative impact on the New Zealand economy.

A substantial number of STEM tertiary students terminate their studies during the first year, not because the courses are too difficult but, anecdotally, because they are too dry and boring. There is a myriad of possible explanations why students may drop out, one of which is as unfortunate as it is avoidable: they fail to find them engaging enough [1, 2]. There are specific terms to describe this phenomenon, such as academic disengagement and disinterest [1].

In addition to specific subject-based knowledge, many high-tech companies require strong problem-solving and thinking skills from their employees. To select the best of the best, companies often use puzzles as part of a job interview process. It is assumed that the ability to solve puzzles relates to the creative thinking needed for solving real-life problems in innovative ways. A classic example is Microsoft, where the goal of a job interview is to assess general problem-solving ability rather than specific content knowledge. Microsoft sees parallels between the reasoning used to solve puzzles and the thought processes involved in solving real, innovative problems [3])

Some efforts have been made to develop pedagogical strategies employing non-routine problem-solving as part of a tertiary learning experience. Positive feedback from students, lecturers, and researchers has been reported from several studies (see, for example, [4], [5], [6], [7]). The authors of the above studies noted that interesting non-routine problems, including puzzles, potentially can improve students' creativity, curiosity and enhance their conceptual understanding, problem-solving strategies, critical thinking skills and lateral "outside the box" thinking.

There are clear links between skills developed by solving non-routine problems and the professional skills required in many workplaces. Parhami [8] argues that puzzling problems are plentiful in all research arenas regardless of discipline. Moreover, he points out that many engineering problems are puzzle-like; hence engineering students should be exposed to these types of problems during their education.

The problem of transfer of non-routine problem-solving to novel domains has been investigated by Evans, Thomas and Klymchuk [6], as a pilot study that served as a foundation for the present study. They recognised and conceptualised the main phenomenon of the study guided by Pugh and Bergin's [9] synthesis of motivational influences on the transfer of learning. Their analytic tool aligned with self-efficacy, one of the four motivational constructs identified by Pugh and Bergin as critical in transfer. The investigation of the mechanisms involved in the transfer of learning in the context of non-routine problem solving was undertaken through the lens of self-efficacy, which is defined by Bandura (1997, p. 3), as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments". In the context of transfer, "self-efficacy usually refers to confidence in the ability to do or learn a skill that can transfer to another domain" ([9], p. 153). An extensive body of knowledge has been amassed from numerous studies pertaining to the mechanisms by which self-efficacy influences the transfer of learning. In summary, current research reports that self-efficacy is positively associated with the transfer of learning through mechanisms that include its influence on

[1] See https://img.scoop.co.nz/media/pdfs/1208/BGAprogressreportbuildinginnovationaugust2012.pdf

cognitive engagement (metacognition, use of effective strategies) and persistence [9, 10].

The researcher's interest in this topic is rooted in a well-known paradox: on the one hand, it is widely assumed that studying mathematics equips learners with problem-solving skills in a broad sense. On the other hand, there is plenty of anecdotal evidence from employers complaining about the lack of creativity in problem-solving of never-before-seen problems by mathematics and computer science graduates. These students tend to get fixated on rigorous and precise solutions disregarding contextual requirements that make this approach unsuitable. Arguably, this inability to be flexible and creative is a manifestation of the problem of transfer that has not been well understood in the context of mathematics at a high level. This prompted the researchers to zoom in on the issue through a novel construct in an attempt to conceptualise the specificity of students' confidence in regard to applying creativity to solving non-routine problems. In their study, Evans, Thomas and Klymchuk [6], defined the novel construct of *lateral thinking self-efficacy,* explored its discriminant validity, and used it as an investigation tool. The study uncovered a significant difference between genders when lateral thinking self-efficacy was observed. The proportion of males reporting confidence in solving nonroutine problems was significantly greater than the proportion of females, despite no significant association between gender and self-reported non-routine problem-solving performance. This major misalignment raises critical questions about equity with regard to employability attributes and prospects for females in particular. The regrettable part of these inferences is how heavily they factor into important life decisions, such as the types of employment females seek to apply for. Could an unfortunate implication be that female STEM graduates are less inclined to apply for jobs at companies like Microsoft, Google, and other technology companies, which use puzzles as part of their screening interviews? No research has been conducted yet to investigate this further.

Moreover, little progress has been made to date in research on how to enhance *lateral thinking self-efficacy.* It is known that mastery experience is one of the main factors contributing to the development of self-efficacy [11]. Hence, a research-informed way to enhance *lateral thinking self-efficacy* is to incorporate non-routine problem-solving into tertiary education. However, it is not clear in what format and dosage (type/number of problems, spacing interval, and duration of treatment) this should be done. To that end, there is a substantial research gap as none of the studies mentioned above have thoroughly investigated the impact of regular use of non-routine problem solving on learners.

In addressing this gap in the literature, this study was set up to evaluate an intervention that introduced a period of non-routine mathematics problem-solving into tertiary STEM lectures. The study had two aims; first, attempting to increase student engagement and second, introducing learners to abstract reasoning that involves critical, creative, and innovative thinking. It involved a partnership between several institutions across different STEM disciplines, providing diverse expertise and experience. Specially, the study involved lecturer-participants, mathematicians, physicists, computer scientists, educational psychologists (scholars and practitioners) from Auckland University of Technology, the University of Auckland, Manukau Institute of Technology, and Whitireia New Zealand.

## 2. Research questions

Our project's primary aim was to evaluate a strategic and innovative pedagogical intervention based on the regular use of non-routine problem-solving in undergraduate STEM courses.

In fulfilling this aim, the project sought to answer four research questions:

***RQ1***: Was use of the lecture-based intervention activity of non-routine problem-solving associated with an increase in engagement?

***RQ2***: Does the integration of non-routine problem-solving in lectures affect participants' ability to inhibit erroneous intuitive thinking and exhibit creative thinking?

***RQ3***: Are any observed effects moderated by individual differences such as demographic characteristics or prior achievement?

***RQ4***: What were students' perceptions of the use of non-routine problem-solving as part of their lectures?

## 3. Theoretical frameworks

The research considered the potential effects of non-routine problems on tertiary STEM students' engagement in lectures and students' ability to exhibit creative thinking while inhibiting erroneous intuitive thinking. In this study, a non-routine problem is defined as one for which the students do not have a ready-made method of solution they can apply, but the content is still within the scope of their knowledge base [12]. These problems include puzzles that cannot be solved by rote, are presented entertainingly, look deceptively simple but have unexpected answers and counterintuitive solutions, and hence are invaluable to promote reasoning that involves both logical and creative thinking [5]. These could also include paradoxes, which have a surprising, unexpected, counterintuitive statement that presents as invalid but in fact, is true, and sophisms, involving intentionally invalid reasoning that presents as formally correct despite containing a subtle mistake or flaw. Some illustrative examples of such non-routine problems are:

**The Race.** In a 100-metre race, athletes A, B, and C all run at a uniform pace. If A beats B by 10 metres and B beats C by 10 metres, by how much does A beat C? (Many students answer 20 metres, but the correct answer is 19 metres. For a solution, see the Appendix)

**Tennis Tournament.** Fifty six players start a (singles) tennis tournament. How many matches will be played if a player who loses a match leaves the tournament?

**Fraction of a Square.** What fraction of the large square is shaded in Figure 1? (After a few minutes, students are provided with a hint: rotate the shaded square by 45 degrees). See Appendix for a solution.

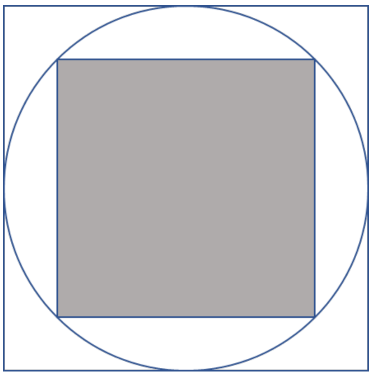

**Figure 1.** Example of a non-routine problem: finding a shaded fraction of the large square

The pedagogical approach informing the project was the Puzzle-Based Learning approach developed by Michalewicz and Michalewicz [13], which has been successfully adopted in numerous educational settings worldwide. Below we present the theoretical considerations related to the three major aspects of the project: engagement, intuition, and creativity.

### *3.1. Engagement*

The first area of focus in this research was the potential effects of non-routine problem-solving on student engagement. The relationship between engagement and student participation, learning, and performance has been a growing area of investigation in mathematics and other STEM disciplines [14-16]. Although engagement has been seen as related to motivation, engagement and motivation are generally viewed as distinct constructs [15]. Definitions of engagement usually include active involvement with the content of a learning activity [17], and often categorise it into behavioural, cognitive, and emotional dimensions, which align with doing, thinking, and feeling.

Behavioural engagement refers to participation inside and outside the lecture theatre, including student conduct, persistence, participation, group work, concentration, and commitment to the course. Its theoretical foundation stems from expectancy-value theory [18], which postulates that personal beliefs about one's competence in learning can determine specific behaviour. For instance, learners may display cooperative behaviour in class if they perceive themselves to be competent enough to perform challenging activities.

Cognitive engagement relates to the thinking that students do while engaged in a learning task activity [17]. A large body of mathematical problem-solving literature has been developed over the years, which emerged with the work of Alan Schoenfeld [19] in the 80s. On the basis of his investigations of the complexities involved in utilising Pólyas's heuristic problem-solving [20], he developed a theory of human decision making, capturing, in particular, non-routine decision making [21] as a cognitive process. Being cognitively engaged involves students' personal investment in learning, their self-regulatory strategies [14], flexibility in problem-solving, preference for hard work, and positive persistence in the face of failure [22], as well as a willingness to exert the effort necessary to comprehend complex ideas [23]. In general, cognitive engagement is considered to have an effect on learning outcomes. The positive relationship of engagement to learning raises the importance of creating engaging learning experiences for students.

The tasks that teachers employ can impact engagement if the tasks create opportunities for learners to engage in activity that is interesting, meaningful, and challenging [24], leading to improved engagement [17]. Tasks that are both interesting and emphasise higher-order skills and real-world applications have been shown to lead to better engagement [14]. Other research [25] has proposed that non-routine questions, or tasks where the approach is not immediately obvious, such as those employed in our study, will encourage deeper student engagement.

### *3.2. Intuition*

The second focus area in this research was to investigate the potential effects of non-routine problem-solving on students' ability to inhibit erroneous intuition. Most mathematicians consider intuition as an essential aspect of their work, with Burton [26] reporting, "Intuition, insight or instinct was seen by most of the seventy mathematicians whom I interviewed as a necessary component for developing knowing" (p. 31). There is a complex relationship between intuition (which may be holistic or integrative) and logic/rigour in mathematics (which may be detailed or analytic). Both are important and complement each other, with Fischbein, who introduced intuition as a research domain in mathematics education, pointing out "the intuitive structures are essential components of every form of active understanding and of productive thinking" [27]. However, some researchers have warned that intuition can mislead and create misconceptions, and therefore sometimes should be inhibited [28-30]. Thereby, they showed that it is essential to be able to inhibit erroneous intuition when solving problems. In this case, inhibition refers to "the executive function in charge of stopping or overriding a mental process" [31] or the act of suppressing distracting information and unwanted responses [32].

In his highly influential book for non-experts, "Thinking, fast and slow", Daniel Kahneman [33] presents an excursion of the mind, providing an overview of the two systems that determine the way we think and make choices. System 1 is fast, intuitive, and emotional, whereas System 2 is slower, more deliberative, and logical. Kahneman exposes the profound faults and biases of fast thinking (System 1), and explains the significant impact of intuitive impressions on our thoughts and behaviour.

Some intervention studies have been conducted (e.g., [28],[34]) in an attempt to improve students' ability to overcome intuitive interference, such as introducing a warning to overcome automatic processing of certain variables [34]. Attridge and Inglis [28] found that prefacing tasks where intuition might mislead students with a task that is slow and effortful, thus requiring the use of working memory, was beneficial to inhibiting erroneous intuition. Similarly, Babai et al. [34] found that an explicit warning during task activity activated inhibitory control mechanisms and helped students overcome the interference. In this research, we studied the effect of an intervention employing non-routine problem-solving on inhibiting intuition. In order to examine effects on inhibition of intuition, we used an extended and slightly modified Cognitive Reflection Test (CRT) [35] and a convergence task based on selected counterintuitive puzzles from Klymchuk [36] (see https://doi.org/10.17608/k6.auckland.13370942.v1 for the instruments used).

### *3.3. Creativity*

The project's third primary focus was on whether exposure to non-routine problem-solving could influence students' creativity. Although it has proven relatively difficult to define creativity precisely as there are more than 100 definitions of creativity in the literature, a number of intellectual qualities that may contribute to it have been considered. Among these is the divergent-production ability that is related to idea generation and elaboration abilities. Another creative thinking source is a transformation ability, which pertains to one's ability to connect old and new knowledge and hence create new ideas. One crucial characteristic of this ability is a readiness to be flexible, where flexibility leads to reorganisation and reinterpretation [37].

Students who have developed flexibility or versatility [38] in their thinking tend to be able to go beyond routine competencies, be more innovative and creative, understand why procedures work, and modify or invent new procedures [39]. These creative aspects of versatile thinking may arise from an ability to translate within and between representations, to interact conceptually with representations, to switch between perceptions of a mathematical entity as a process or an object, and to exploit the power of visual schemas by linking them to relevant analytical ones [40].

Creativity is a process that is often required in the solving of non-routine problems since there is no available standard approach. It has been suggested [41] that a creative approach to problem-solving involves 1) preparation (where the problem is "investigated in all directions"); 2) incubation (unconscious processing of information, cognitive processing); 3) illumination (a sudden insight or discovery of a solution); and 4) verification (deliberate testing and evaluating of the solution or idea). It is worth noting that each of these aspects of the creative problem-solving process requires a high level of student engagement. Students may go through various stages of this creative thinking process recursively since the most appropriate solutions will require further insightful thinking and verification. Creativity in our project was understood to be what Leikin [42] (see also [43]) calls "relative creativity" and Sriraman and Haavold [44] refer to as "everyday creativity" (compared with "extraordinary creativity" or exceptional knowledge that changes perceptions of the world). To measure creativity, we used the highly influential theory of creativity developed by Guilford [45]), who considered creative thinking as divergent thinking based on fluency, flexibility, originality, and elaboration. When solving problems, fluency is identified as the ability to generate a great number of solutions; flexibility the ability to suggest a variety of approaches; originality the ability to propose unusual approaches; and elaboration the ability to organise the details of an idea and implement it.

In summary, in this project we utilised non-routine problem-solving in lectures and analysed its influence on the focus areas of student engagement, intuition, and creativity. The relationship between engagement and student learning is well established and creativity, which has links to engagement, is crucial to problem-solving of all kinds. Further, while intuition is important in solving problems, it can sometimes misguide, whereby it needs to be inhibited. These focus areas led to the following design for the study.

## 4. Methods

In this project, a pedagogical innovation was used affecting 683 undergraduate Year 1 and Year 2 students who were enrolled in twelve STEM courses studying astronomy, computing, engineering, and mathematics at four tertiary institutions in New Zealand: Auckland University of Technology (152 students), the University of Auckland (440 students), Manukau Institute of Technology (25 students), and Whitirea New Zealand (66 students). Lecturer-participants posed one or two non-routine problems for students to solve in the middle of every lecture. These tasks were a bolt-on and did not depend on the lecturers' teaching methods. Apart from these tasks, predominantly, the lecturers involved in the trail used lecturer-centred traditional approaches.

When presented with a task, the students could work in a small group or individually. After a few minutes, lectures asked students if they came up with a solution and let them explain their ways of thinking. If students did not volunteer any solutions, then the lecturer provided an explanation. This activity lasted for 3-5 minutes per lecture. Thus, the scale of our intervention was small, amounting to less than 10% of lecture time. We estimate that, in total, the intervention involved 30 non-routine problems over a 12-week semester. Other aspects of the course delivery (tutorials, assignments, tests, and examinations) were unaltered.

### *4.1. Study structure*

The distinct nature of the research questions necessitated adoption of various investigation methods. In total, five sub-studies were undertaken with the following data collection and analyses (Table 1).

### *4.2. Participants and procedures*

During the first week of the second semester of 2018, all students were invited to participate in the study via a verbal announcement in a lecture and an online invitation. Out of the 683 students, 100 students (approximately 15% of those invited) volunteered, but only 64 students completed both pre-test and post-test questionnaires (a 9% response rate). Thus, the sample in Sub-study 2 included 36 males (56%), 27 females (42%), and one gender diverse (2%); 58 were younger than 25 years old (91%), and the remaining six participants were over 25 years old (9%). In addition to this sample of participants who completed both questionnaires, there were two smaller samples of participants: 1) Sub-study 5: 12 student participants who volunteered for a one-hour task-based interview (these are numbered S1-S12 in the analysis), and 2) Sub-study 4: nine lecturers involved in the intervention who completed the lecturer questionnaire by email. Participating students were rewarded with gift vouchers for their time. The project was approved by the University of Auckland Human Participants Ethics Committee on 05/06/2018 (Ref 021387).

**Table 1.** Data collection and analyses (5 sub-studies)

| Sub-study | Design | Data collection | Sample | Data analysis |
|---|---|---|---|---|
| 1 | Naturalistic observation | Observers (detached, non-participant) used two observation protocols in class at the start and the end of the trial semester. Protocol 1 was designed to validate the fidelity of implementation; Protocol 2 to record occurrences of student off-task behaviour | Over 600 students studying mathematics, astronomy and computer science in New Zealand tertiary institutions | Quantitative, Friedman test |
| 2 | Quasi-experimental, comparative | Two questionnaires at the start (T1) and the end (T2) of the trial semester comprising various instruments | 683 students from four tertiary institutions were invited to participate, but only 64 students completed both questionnaires | Quantitative, Descriptive, ANOVA tests, Regression analysis |
| 3 | Descriptive, comparative | As part of the questionnaire at T2, an instrument assessing student overall engagement in the course during the trial semester based on [23] | Identical sample as in phase 2 | Quantitative, Descriptive, ANOVA test |
| 4 | Descriptive | Survey of lecturer-participants (open-ended questions) | Nine lecturers participating in the trial | Qualitative |
| 5 | Descriptive | Focus group interviews of student-participants (tasks and open-ended questions) | Twelve volunteering students (in five small groups) | Qualitative |

### *4.3. Instruments*

Several different types of instruments were used to gather data to address the research questions. All of them—the student pre-test (in week 1) and post-test (in week 12) questionnaires, content validation survey questions for the pre- and post-test questionnaires, focus group interview questions, the lecturer questionnaire, and class observation protocol—can be found at https://doi.org/10.17608/k6.auckland.13370942.v1.

The student pre- and post-questionnaires were designed to measure several psychological constructs and capacities (e.g., engagement, response inhibition, creativity) using questions from the following sources:

**engagement**—three instruments: (1) observation protocols for observations of five lectures (in total); (2) validated questionnaire of Ahlfeldt et al. [23], which was developed based on the USA's National Survey of Student Engagement (NSSE) [46] as a 'global' measure of course engagement;

**inhibition of intuition**—an extended and slightly modified Cognitive Reflection Test (CRT) that included the original CRT [35] as well as two questions inspired by Babai et al. [34];

**convergent thinking**—the Convergence Task based on selected counterintuitive puzzles from Klymchuk [36]; and

**divergent thinking/creativity**—an Alternative Uses Test from the well-established Torrance Test of creativity [47] based on Guilford's model of creativity [45]. Specifically, students were asked to list as many uses as possible for either a brick (Form A) or a shoe (Form B) in a short time.

Given the importance of student engagement in this project, the foregoing self-reporting was triangulated by lecturer feedback and lecture observations. Detached, non-participant observers were present in five of the lectures in three different tertiary institutions and recorded the time spent in class on non-routine problems, the number of problems introduced in each lecture, the quality of delivery, and the active engagement of the students. This active engagement was measured using two protocols. The first protocol was designed to evaluate the fidelity of implementation in order to establish whether the intervention was being implemented as intended. The second observation protocol was used to assess learners' behavioural engagement in class at various time points: immediately before, during, and right after the intervention (non-routine problem-solving activity). Using this protocol, observers visually scanned a sample of 30 students (approximately) per lecture, registering all instances of students' off-task behaviour and making records in a chart. These observations were repeated frequently (about every five minutes) with a minimum of two scans before the non-routine problem-solving activity and two after the end of it. The two protocols were used during a 30 to 45 minute time period by one or two observers. The two observers were members of the research team selected to undertake the observations. The first three lecture observations were conducted by both observers simultaneously. This was done to establish inter-observer reliability before they could undertake independent observations (Sub-study 1).

In addition, the nine lecturers involved in the intervention provided feedback on the engagement of their students by email response to a questionnaire (Sub-study 4).

The SPSS 25 program was used for all statistical analyses, including Chi-square tests, *t*-tests, Friedman test, parametric test, and Repeated Measures ANOVAs.

## 5. Results

### *5.1. Student engagement*

As detailed in the preceding section, student engagement was evaluated in several ways. Specifically, in this section we present results from 1) in-lecture (real-time) observations of students' behavioural engagement, 2) students' self-reported responses on pre- and post-questionnaires concerning their academic engagement, and 3) lecturer reports of student engagement.

#### 5.1.1. Behavioural engagement: Results from lecture observations (Sub-study 1)

In answering the first research question (RQ1), we report that the difference in behavioural engagement was significantly different during the lecture, with greater engagement observed during the non-routine problem-solving activity (the intervention). First, the data from the protocol for assessing implementation fidelity was analysed. It demonstrated acceptable levels of dosage (type/number of non-routine problems used), adherence (establishing that the activity was used as expected during a lecture), and quality of the delivery. While there were slight variations across lectures, all assessed indicators were above the threshold for every lecture that was observed. Given this assurance of the implementation fidelity, the data from the second observation protocol was analysed to evaluate the changes in student off-task behaviour. The instances of student off-task behaviour were recorded if students were not paying attention, being distracted by mobile devices, or snoozing.

The data came from observations of five lectures from different courses at three different tertiary institutions: Auckland University of Technology, the University of Auckland and Manukau Institute of Technology. The data analysis from five independent lecture observations revealed that behavioural engagement was significantly greater during the intervention time period compared to the time period before and after the intervention. As illustrated in Figure 2, the off-task behaviour followed a distinctive V-shaped pattern (decreasing from Before to During, then increasing from During to After) and was observed across all five lectures.

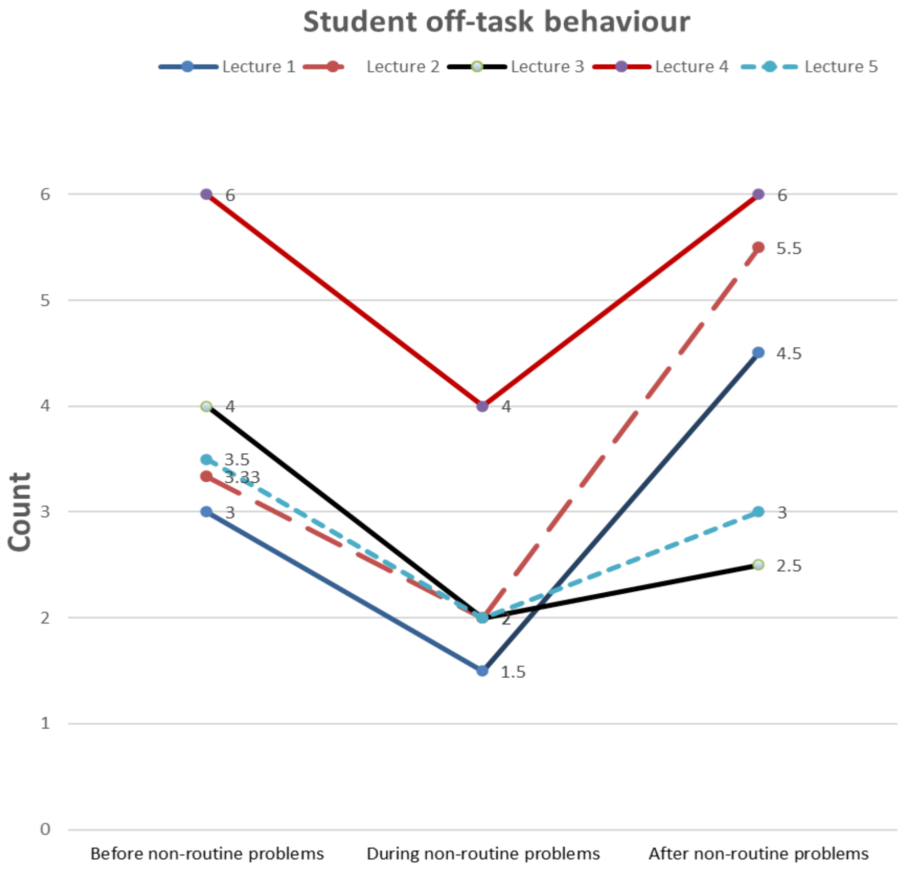


**Figure 2.** Off-task behaviour patterns, reported as the average number of students per observation in a group of 30 students (approximately)

*Notes: the data points correspond to the rate of off-task behaviour among a group of 30 students for each of the five lectures observed (recorded prior, during, and straight after the intervention).*

A statistically significant difference in the off-task behaviour recorded during three time periods in class was confirmed by the Friedman test, $\chi^2(2) = 7.90$, p = .019. Specifically, the observation data revealed more instances of off-task behaviour before the intervention (Median = 3.5) than during (Median = 2); a median difference decrease of 1.5 units. This followed by a median difference increase of 2 units of off-task behaviour after the intervention (Median = 4.5) (p = .053 for both pairwise comparisons for Before-During and During-After, with Bonferroni adjustment correcting for multiple tests).

The observed consistency in data across all observed lectures in five different courses show a noteworthy pattern of the rise and fall of student engagement. In summary, these data provide strong evidence that the intervention activity of non-routine problem-solving caused a boost in student engagement during a lecture. Plausible explanations for the observed pattern include the following. First, student engagement could have been improved due to a novelty effect; second, a problem that is only likely to take three minutes to solve is not expected to be perceived as a barrier to engagement. Hence, the perceived ease/time required is also likely to impact willingness to engage significantly.

5.1.2. Academic engagement: Results from student questionnaires (Sub-study 3)

To further examine student engagement, we utilised a commonly used instrument to measure student overall engagement during the trial semester developed by Ahlfeldt et al. [23]. The instrument was developed to establish a somewhat 'global' measure of learners' engagement at a class level, designed to explore three broad areas: (1) levels of cooperative learning, (2) levels of cognitive challenge, and (3) the development of personal skills. The instrument comprised 14 items with the sum of these three groupings of questions producing the Engagement Score (ES) for each student.

The results of our analysis demonstrated a broad range of ESs (24 to 55), with a mean of 34.85. According to the aggregated scores reported in Ahlfeldt et al. [23], this corresponds to a moderate level of engagement. However, we note that the current study's mean ES (34.85) is higher than the overall mean reported in the comparable category of the 'Science and Mathematics' courses, with the overall mean of 33.57, reported from studies undertaken in seven mathematics/science university courses (Ahlfeldt et al. [23], p. 16). Despite this benchmark measure registering a slightly higher (non-significant) value, this result cannot be taken as evidence in meaningful comparison due to the lack of information about a population norm for an equivalent population of students.

Moreover, we included an additional item to the end-of semester questionnaire in order to ascertain student engagement with the intervention as a self-report measure. The item was a 4-point Likert scale with a question: 'With respect to this course, about how often have you worked on solving puzzles and/or creative problems during class?' In answering research question 3 (RQ3), we used a one-way ANOVA to investigate group differences with respect to prior achievement. No statistically significant effect of prerequisite grades was observed ($F(2, 52) = 1.61$, $p = .610$); the effect size (partial $\eta^2 = .058$). However, as shown in Figure 3, the highest level of engagement was reported by participants with the lowest grades in the prerequisite course ($M = 3.08$, $SD = .73$), whereas students in the middle grade band reported the lowest level ($M = 2.45$, $SD = 1.07$). This is a somewhat surprising and unexpected finding that should be investigated further. However, we acknowledge that this finding may have nothing to do with the intervention.

Some of these results were previously reported by the researchers [48], together with analysis utilising the lens of the engagement framework by Kahu and Nelson [2], suggesting that an intervention to incorporate non-routine problem-solving in university STEM lectures has changed the educational interface in a positive way. The researchers argued that amid wider participation in higher education, enabling student success necessitates targeted considerations of the group of students who are likely to underperform. Given the results of the investigation, it is plausible to suggest that a method of lecture delivery that utilises the inherent engagement effect of non-routine problem-solving could positively influence student engagement—an effect that appears to extend to a group of low-achieving students.

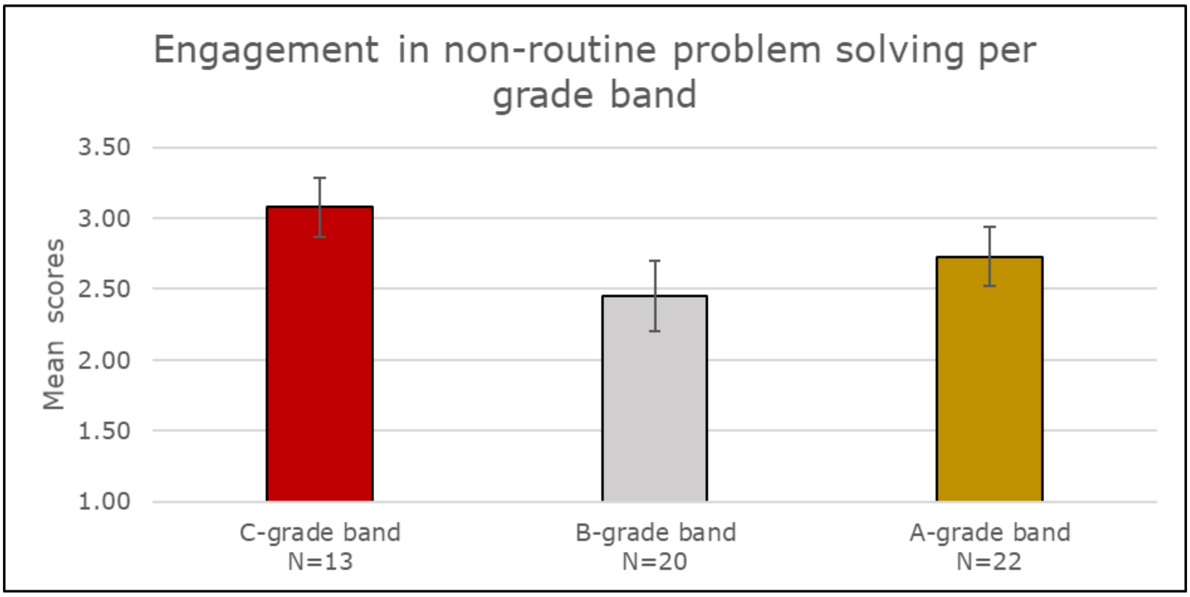


**Figure 3.** Level of engagement in non-routine problem-solving activities by prior achievement ($N$=55)

*Notes. Item: With respect to this course, about how often have you worked on solving puzzles and/or creative problems during class? (4=Very Often, 3=Often, 2=Occasionally, 1=Never)*

*Error bars: +/- 2 SE*

### *5.2. Student intuitive and convergent thinking*

As detailed in Table 2, results from paired-sample $t$-tests indicated that participants' intuitive thinking did not change significantly over time. Specifically, the mean CRT score for all participants at Time 2 was .80, a small increase of .04. There was some variation in mean change based on the sequencing of test forms participants received. For example, while participants who completed Form A of the CRT (the original version) at both Time 1 and 2 earned the highest overall mean scores, they experienced a decrease in mean score over time (.87 at T1 and .84 at T2). Conversely, participants who completed Form B at T1 and Form A at T2 started with the lowest mean score at T1 (.68) but experienced the greatest increase over time ($\Delta M$ = .13). Although not statistically significant, the latter increase constitutes a small to medium size effect (Cohen's $d$ = 0.48). In short, it appears that Form A and B were not equivalent forms (the latter proving more difficult than the former), and that the notable increase demonstrated by those who completed Form B at T1 and Form A at T2 may likely be the result of between-form differences in item difficulty.

Similarly, results from paired-sample $t$-tests indicated that participants' convergent thinking did not change significantly over time (Table 3). Specifically, the mean score for all participants at T1 was .55 and at T2 was .53, a small decrease of .02. As seen in Table 2, there was some variation in mean change based on the sequencing of test forms participants received. This time it was participants who completed Form B both times who obtained the highest overall mean scores (.64 at T1 and .61 at T2) and experienced the greatest decrease in mean scores (0.04). Participants who completed Form A at T1 and Form B at T2 experienced the greatest increase over time ($\Delta M = .05$). However, none of the differences was statistically significant ($p > .05$), and their magnitudes were very small (Cohen's $d < 0.30$). In short, there were no notable changes in participants' convergent thinking (problem-solving) ability over time, regardless of form sequencing.

**Table 2.** Statistics of CRT by time and form

| | | Time 1 | | Time 2 | | Difference (Time 2 - Time 1) | | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | *n* | *M* | SD | *M* | SD | *M* | SD | *t*-value | Cohen's *d* |
| All Participants | 41 | 0.76 | 0.28 | 0.80 | 0.25 | 0.04 | 0.25 | 0.89 | 0.15 |
| Form A - Form A | 14 | 0.87 | 0.19 | 0.84 | 0.22 | (0.03) | 0.19 | (0.62) | (0.14) |
| Form B - Form B | 7 | 0.71 | 0.32 | 0.71 | 0.28 | 0.00 | 0.38 | 0.00 | 0.00 |
| Form A - Form B | 5 | 0.76 | 0.26 | 0.72 | 0.33 | (0.04) | 0.09 | (1.00) | (0.13) |
| Form B - Form A | 15 | 0.68 | 0.32 | 0.81 | 0.23 | 0.13 | 0.24 | 2.09 | 0.48 |

*Notes.* Parenthetical values (in red) are negative.Cohen's $d = (M_{T2} - M_{T1}) / SQRT( (SD_{T1}^2 + SD_{T2}^2) / 2 )$

In summary, there were no significant main effects for time on either participants' intuitive or convergent thinking. However, there was one interesting finding—a potential moderation effect for gender on the CRT. Specifically, whereas females did not demonstrate any notable change over time on the CRT (from .80 at T1 to .79 at T2), males demonstrated a small gain over time (from .74 at T1 to .81 at T2). While the observed difference was not statistically significant, the partial eta-squared value (0.027) signifies a small effect for the interaction (see Table 4) and suggests that the intervention may have been more effective for males than females.

**Table 3.** Statistics of convergent thinking by time and form

| | | Time 1 | | Time 2 | | Difference (Time 2 - Time 1) | | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | *n* | *M* | SD | *M* | SD | *M* | SD | *t*-value | Cohen's *d* |
| All Participants | 42 | 0.55 | 0.25 | 0.53 | 0.28 | (0.02) | 0.29 | (0.39) | (0.07) |
| Form A - Form A | 15 | 0.53 | 0.23 | 0.52 | 0.29 | (0.02) | 0.33 | (0.19) | (0.06) |
| Form B - Form B | 7 | 0.64 | 0.20 | 0.61 | 0.13 | (0.04) | 0.09 | (1.00) | (0.21) |
| Form A - Form B | 5 | 0.55 | 0.41 | 0.60 | 0.29 | 0.05 | 0.33 | 0.34 | 0.14 |
| Form B - Form A | 15 | 0.52 | 0.24 | 0.48 | 0.32 | (0.03) | 0.33 | (0.40) | (0.12) |

*Notes.* Parenthetical values (in red) are negative.Cohen's $d = (M_{T2} - M_{T1}) / SQRT( (SD_{T1}^2 + SD_{T2}^2) / 2 )$

**Table 4.** CRT Effect for Gender (transformed variable: average)

| | | Time 1 | | Time 2 | | | | | Observed |
|---|---|---|---|---|---|---|---|---|---|
| Variable | *n* | *M* | SD | *M* | SD | Test | *F* | Partial $\eta^2$ | Power |
| Male | 23 | 0.74 | 0.30 | 0.81 | 0.23 | Time | 0.53 | 0.014 | 0.110 |
| Female | 17 | 0.80 | 0.24 | 0.79 | 0.28 | Gender | 0.07 | 0.002 | 0.058 |
| Total | 40 | 0.77 | 0.28 | 0.80 | 0.25 | Time * Gender | 1.06 | 0.027 | 0.171 |

5.2.1. Lecturer questionnaire

One question from the lecturer questionnaire on intuition was "Did you observe any change in the intuitive nature of the students' initial attempts to answer the puzzles over the time?" Six out of nine lecturers reported that they observed a change. They commented:

L1: When students noticed the problem solving often depended on the interpretation of the question, they started thinking around what was being asked. It opened the problem up for them in terms of the way they looked at the question.

L2: On the initial meetings where a puzzle was used, the students tried to solve the puzzle just by looking at it and thinking. The later meetings, they tried to use pictures made on paper. There was also more activity, when students discussed among them the possible solution at the latest meetings.

L3: They tended to discuss more at the start and bounce ideas off each other which they weren't doing at the start.

L4: Students became more cautious with using their intuition as they saw it often mislead in puzzles.

L5: Students were more willing to share their intuitions when solving the puzzles.

One lecturer did not observe any changes in intuition and two lecturers replied: "Not sure".

#### 5.2.2. Focus group interviews: Intuitive thinking

To answer RQ4, students' perceptions were analysed qualitatively as part of focus group interviews. The 12 students in the focus groups worked on solving given tasks as well as constructing their own. When asked to describe the kinds of thinking required to solve them, some referred to the need to control one's thinking in some way, saying:

S7: When I'm doing a puzzle I will think about it at least twice … it may contain some traps.

S11: I often get caught out using intuition straightaway … [I] need to try using a more pragmatic approach, which would decrease the effect of my over-eagerness to find the solution.

S12: You … must not take your first intuitive-designed answer.

This idea of inhibiting intuition was central and the students agreed that it is essential. They also expressed the need to get an overview of the problem as part of the creative process.

S3: Solving puzzles though one can have a strategy, don't have formulaic approaches.

S10: Solving puzzles allows one to approach problems as a whole … instead of short term step by step.

S12: Creating something new: a perspective.

This desire for a strategic overview was linked by S11 to the need to inhibit intuitive thinking.

S11: Innovation helps to hold back intuition by developing a new strategy.

### *5.3. Student creativity*

The impact of the intervention on students' creativity was measured via a divergence task, some questions in the self-assessment section (Sub-study 2), feedback from the lecturer questionnaire (Sub-study 4), problem-solving and problem-posing tasks during the focus group interviews (Sub-study 5).

#### 5.3.1. The divergence task

The divergence task was based on the Alternative Use Test. Students were asked to generate in a relatively short time as many alternative uses of a brick (form A) or a shoe (form B) as possible. Based on Guilford's [45] model of creativity, we measured the following traits of divergent thinking: fluency, elaboration, and originality. The originality score used was calculated based on the originality of the words or ideas generated in participant responses. As shown in Table 5, results from paired-sample t-tests indicated no significant differences in participants' fluency, elaboration, or originality scores (all p's > .05). That said, there were very small increases across time for both fluency and originality (Cohen's *d*'s = 0.18 and 0.19, respectively).

| | | Time 1 | | Time 2 | | Difference (Time 2 - Time 1) | | | |
|---|---|---|---|---|---|---|---|---|---|
| Variable | *n* | *M* | SD | *M* | SD | *M* | SD | *t*-value | Cohen's *d* |
| Fluency | 43 | 7.35 | 4.31 | 8.09 | 4.14 | 0.74 | 3.27 | 1.49 | 0.18 |
| Elaboration | 43 | 2.56 | 3.13 | 2.49 | 2.33 | (0.07) | 3.20 | (0.14) | (0.03) |
| Originality | 43 | 10.52 | 7.38 | 12.00 | 8.36 | 1.48 | 6.89 | 1.57 | 0.19 |

*Notes.* Parenthetical values (in red) are negative.Cohen's *d* = $(M_{T2} - M_{T1}) / \mathrm{SQRT}( (SD_{T1}^2 + SD_{T2}^2) / 2 )$

5.3.2. Self-assessment section

In this section of the pre- and post-surveys, participants were asked to use a five-point Likert-type scale to indicate how strongly they agreed or disagreed with two statements related to creativity:

- Solving puzzles can enhance your creativity.
- Solving puzzles makes you a more innovative thinker.

The vast majority of students agreed with both statements at both times. Specifically, 87.8% of participants agreed or strongly agreed with the first statement at T1 and slightly more at T2 (90.2%). Similarly, while 82.9% of participants agreed or strongly agreed with the second statement at T1, slightly more did so at T2 (90.2%). The survey section also included the following open-ended question on creativity: "Do you think solving puzzles can enhance your creativity?" The vast majority of participants (86% at T1 and 84% at T2) agreed that solving puzzles could enhance their creativity. Common survey responses to both questions 1 and 2 defined creativity in terms of the need to "think outside the box": 12 times (pre-survey), and 15 times (post-survey). About half the participants (47% at T1, 57% at T2) thought they were less likely to focus on a single heuristic to get the correct answer: "It forces you to look outside the obvious first glance answer and think about other possibilities outside of the obvious."

Other responses to this question (42% at T1, 45% at T2) focused on notions of thinking differently. They stated that instead of applying a formula, they needed to access their prior knowledge and employ creative thinking in the problem-solving: "It teaches us to rely on our knowledge but not necessarily follow the formula of what we've been taught, but instead a good balance of prior knowledge and thinking `outside of the box'." They agreed there was a need to think beyond commonly accepted ideas ("It helps you see that there are a lot of different ways to doing the same task or one standard way of doing a range of things") and to interpret the questions in different ways ("It requires us to look at the question in different perspective and visualise the situation in our head").

5.3.3. Lecturer questionnaire

The following question from the lecturer questionnaire addressed creativity: "Did you observe any change in creativity of the answers over the time you used puzzles?" Three lecturers answered "yes", five "no", and one "not sure". Typical comments from the lecturers who answered "no" were "not enough time" for the intervention and/or the course was "too short". Typical comments from the lecturers who answered "yes" were as follows: "Students were trying more diverse and sometimes crazy ideas and approaches"; "Students were more playful and less reserved when working with the puzzles."

5.3.4. Focus group interviews: Creativity

With respect to RQ4, of the 12 students in the five focus groups, 83% (10) agreed that innovative thinking was related to solving non-routine problems. One focus group, in particular, seemed to do better than the others in producing a task involving some creativity. This group comprised S7, S8, and S9 and was given the following task:

*Make up an interesting problem based on the information below. You may add extra information if required.*
*A car, A, is travelling due East at 30km/h. A second car, B, is travelling South West at 50km/h.*

The response from this group was as follows: *The radius of a sphere is 400km. Cars A and B start at the same place and time and move in their designated directions and speeds. What is the sum of the distances that both have travelled before they meet again?*

Setting aside the practical considerations of their problem, it is instructive to consider the group's thought processes. They stated that they wanted to produce an "interesting" or non-standard problem and that doing this was related to being creative.

S7: Brainstorm to have some interesting idea. Should be something unusual.
S8: Creativity. The task is to create an interesting problem therefore creativity is important.
S9: To come up with something interesting and out of the box.

Moreover, all of them described this kind of creative thinking as "out of the box". In addition to the quote from S9 above, they also said:

S7: Skills like thinking more carefully and thinking out of the box should be useful.
S8: My issue is I try to solve things like routine questions instead of thinking outside the box.
S9: The thinking is a lot more out of the box. You really have to stretch your brain more creatively.

It seems clear that the students recognised that solving non-routine problems is different from solving routine ones. It requires them to use different kinds of thinking, such as approaching the problems with a holistic strategy rather than linearly, thinking more creatively, and attempting to inhibit their intuitions.

### *5.4. Student perceptions: Employability*

To answer RQ4, student perceptions of the impact of the intervention on their employability were evaluated via the self-assessment section (Sub-study 2) and focus group interviews (Sub-study 5).

5.4.1. Self-assessment section

In this section of the pre- and post-questionnaires, participants were asked to use a five-point Likert-type item to indicate their agreement or disagreement with the statement: “The knowledge and skills related to solving puzzles will be useful to me in the future.” The vast majority of students agreed with this at both times: 80.2% at T1 and significantly more at T2 (90.3%).

Triangulating this data was the open-ended question in the section on future benefits from solving puzzles: “Do you think solving puzzles can benefit you in the future?” The vast majority of participants (94% at T1, 92% at T2) agreed that solving puzzles could benefit them in the future. Some participants (32% pre-test, 28% post-test) specified particular ways they thought puzzles would be beneficial in their future careers, such as ”I hope to pursue a career in data science and science technologies”; “Absolutely, I am aiming to work as an Astronomer—possibly in the field of astrophysics, and math/physics are concerned with a multitude of real-world puzzles.” They also believed that employers expected employees to utilise problem-solving skills and that they would test these skills in job interviews: “I think employers look for people with good problem-solving skills”; “Employers look for people who have problem-solving skills. You are tested on this in interviews.”

In terms of personal growth and development, many participants (47% pre-test, 59% post-test) commented that they enjoyed solving puzzles, felt a sense of achievement, and thought it developed their creativity as well as their problem-solving skills: “I really enjoy them and get joy from them and a feeling of satisfaction and self-worth and pride when completing them”; “It gives me a feeling of achievement”; and “It enhances creative and problem-solving skills.” The utility value rates were higher than enjoyment rates.

Further, some participants (21% pre-test, 11% post-test) stated that puzzles could enhance problem-solving skills in their daily lives, at work (“It builds skills that are useful in the workplace/everyday life”), and when solving real-world problems (“You will learn problem-solving skills that you can apply in the real world”; “The world is evolving and there will be many puzzles to solve in this life, both as a society as a whole and personally”). Moreover, they expected their problem-solving skills to help with solving complex life problems (“Faced non-traditional problems in the real world”).

5.4.2. Focus group interviews: Employability

All 12 students agreed that the novel kinds of thinking gained from solving non-routine problems would be useful in other areas of life, citing as examples logistical issues, transport (such as flights), food packaging, climate change, house building, and other business environments.

### *5.5. Summary of results*

We summarise here the main findings of the research that answer the research questions.

- Engagement
  - Students’ behavioural engagement was significantly greater during the intervention. The evidence showed that they found the problems more engaging than the lecture itself, with fewer instances of off-task behaviour observed.
  - The group of students with C grades in prerequisite courses demonstrated high engagement with non-routine problem-solving.

- Intuition
    - Even though students saw the importance of inhibiting intuitive thinking when necessary to solve non-routine problems, the inhibition control did not change significantly over time.
- Convergent thinking
    - There were no significant changes in students' convergent thinking (problem-solving) ability over time.
- Creativity
    - There were no significant changes in students' creativity (originality, fluency, and elaboration traits of divergent thinking) over the course of the intervention.
- Group differences
    - Grades in prerequisite courses did not significantly influence over time student self-efficacy in, emotional disposition toward, or perceived value of non-routine problem-solving.
    - The results suggest that the intervention may have been more effective for males than females, and this hypothesis could be investigated in future research.
- Student perceptions about learning
    - Students agreed that solving non-routine problems was useful for their learning and enhanced their creative and innovative thinking abilities. They talked about the need to "think outside the box" and think holistically rather than be focused on a linear approach.
    - Perceptions of the utility value of non-routine problem-solving improved at the end of the semester for all students.
    - The students strongly agreed that solving non-routine problems in their courses would benefit their future learning and careers, and other areas of life.

We anticipated a high level of student engagement during the non-routine problem-solving intervention, and this was realised. The feedback by the vast majority of the students also confirmed that they saw valuable benefits from the activities for their future employment. We also expected that the intervention would enhance students' creativity and ability to inhibit their erroneous intuition. However, the increase in students' creativity and ability to control their intuition before and after the intervention was not statistically significant. One possible reason for this could be the small timescale of the intervention with simply not enough time to exhibit such a change. Although the intervention was regular, 2-3 times a week for three months, the total time spend on non-routine problem-solving was only about 1-2 hours. Having spent 10+ years in formal education where there was little or no attention paid to creativity in STEM subjects, it is plausible to expect that students need more time to unlock their creativity. Many researchers point out that (primary) intuition is strongly present and very resistant to change (e.g., [27],[29]). Hence, the crucial factors enabling creativity and the ability to inhibit intuition, when necessary, require researchers' further attention.

## 6. Implications and recommendations for practice

Learners' success has always been and always will be a central focus for researchers in higher education. Hence, interventions aimed at improving student engagement in tertiary institutions are valuable as behavioural and cognitive engagement is often used as an indicator of improved learning ([2],[16]). A primary outcome of this research is that an intervention employing non-routine problem-solving has significantly improved student engagement during its implementation in lectures and induced positive attitudes to the solution process. Hence, we would recommend that tertiary institutions consider initiating a similar pedagogical practice as part of their STEM subjects. While a few universities may already offer optional seminars or even compulsory courses based on Puzzle-Based learning for their first-year STEM students, this is far from the norm. It may be the case that many lecturers teaching large classes presume that adopting an approach to teaching that promotes student engagement is too hard, given their limited resources, and the substantial investment in designing and developing new resources and methods. However, the outcomes of this trial suggest that lecturers can foster learners' engagement with a small-scale developmental effort by simply spending a few minutes on non-routine problem-solving activity during lectures. The implementation effort is relatively minor, even for those lecturers primarily teaching in a teacher-centred "traditional" style.

In conclusion, we reiterate that the students in our study were convinced that the experience of solving non-routine mathematics problems in lectures is of value to them, both for their current learning and for their future education and careers. Furthermore, they enjoyed the experience and had a positive emotional response to it, making them feel more capable of solving non-routine problems. The observation that the students with lower prior achievement had high level of engagement should add to the appeal of the approach for many institutions because it can increase the student retention rate. It appears that those students were emotionally engaged and interested in non-routine problem-solving rather than disengaged, perhaps reducing the odds of them dropping out of courses [1]. All of these positive findings contribute to the strength of the recommendation that other educators give the introduction of this kind of intervention serious consideration.

## 7. Limitations

One of the key limitations of this research is that it was a reasonably short timescale intervention. We could not spend more time on non-routine problem-solving since it was outside the standard course content and was not integrated with the assessment structure. Future studies should investigate a non-routine problem-solving intervention that is meaningfully integrated into the curriculum and instruction. There was also a relatively low response rate (leading to overall small sample size) despite a very positive attitude of the vast majority of students towards the intervention and their high engagement level. Possible reasons for this include the requirement to attend most lectures to be eligible for the study (typical attendance was around 50%), the comprehensive nature of pre- and post-test questionnaires, and the challenging tasks that some students were not confident to do. Due to a small sample size, we sometimes were not able to make enough observations to detect significant differences, for example in the parametric test. Another limitation was the use of a non-probability sampling—a convenience sampling method with students self-selected for the study was used instead of random sampling.

**Funding**

This work was funded by the Teaching and Learning Research Initiative (New Zealand government fund). Raw data and the data collection instruments can be found at https://doi.org/10.17608/k6.auckland.13370942.v1

## References


1. Blondal, K.S. and S. Adalbjarnardottir, Student Disengagement in Relation to Expected and Unexpected Educational Pathways. *Scandinavian Journal of Educational Research*, 2012. 56(1): 85-100. https://doi.org/10.1080/00313831.2011.568607.
2. Kahu, E.R. and K. Nelson, Student engagement in the educational interface: understanding the mechanisms of student success. *Higher Education Research & Development*, 2018. 37(1): 58-71. https://doi.org/10.1080/07294360.2017.1344197.
3. Poundstone, W., *How would you move Mount Fuji? Microsoft's cult of the puzzle - How the world's smartest companies select the most creative thinkers*. 2000: Little Brown and Company.
4. Falkner, N., R. Sooriamurthi, and Z. Michalewicz, Teaching puzzle-based learning: Development of basic concepts. *Teaching Mathematics and Computer Science*, 2012. 10(1): 183-204.
5. Thomas, C., et al., Puzzle-based Learning of Mathematics in Engineering. *Engineering Education*, 2013. 8(1): 122-134. https://doi.org/10.11120/ened.2013.00005.
6. Evans, T., M.O.J. Thomas, and S. Klymchuk, Non-routine problem solving through the lens of self-efficacy. *Higher Education Research & Development*, 2020: 1-18. https://doi.org/10.1080/07294360.2020.1818061.
7. Klymchuk, S., Puzzle-based learning in engineering mathematics: students' attitudes. *International Journal of Mathematical Education in Science and Technology*, 2017. 48(7): 1106-1119. https://doi.org/10.1080/0020739X.2017.1327088.
8. Parhami, B., A puzzle-based seminar for computer engineering freshmen. *Computer Science Education*, 2008. 18(4): 261-277. https://doi.org/10.1080/08993400802594089.
9. Pugh, K.J. and D.A. Bergin, Motivational Influences on Transfer. *Educational Psychologist*, 2006. 41(3): 147-160. https://doi.org/10.1207/s15326985ep4103_2.
10. Ford, J.K., et al., Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology*, 1998. 83(2): 218-233. https://doi.org/10.1037/0021-9010.83.2.218.
11. Usher, E.L. and F. Pajares, Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research*, 2008. 78(4): 751-796. https://doi.org/10.3102/0034654308321456.
12. Selden, A., et al., Why can't calculus students access their knowledge to solve non-routine problems?, in *Research in collegiate mathematics education: IV. Issues in mathematical education*, E. Dubinsky, A. Schoenfeld, and J. Kaput, Editors. 2000, pp. 128-153: American Mathematical Society.
13. Michalewicz, Z. and M. Michalewicz, *Puzzle-based learning: An introduction to critical thinking, mathematics, and problem solving*. 2008: Hybrid Publishers.
14. Fredricks, J.A., Engagement in School and Out-of-School Contexts: A Multidimensional View of Engagement. *Theory Into Practice*, 2011. 50(4): 327-335. https://doi.org/10.1080/00405841.2011.607401.
15. Fredricks, J.A., M. Filsecker, and M.A. Lawson, Student engagement, context, and adjustment: Addressing definitional, measurement, and methodological issues. *Learning and Instruction*, 2016. 43: 1-4. https://doi.org/https://doi.org/10.1016/j.learninstruc.2016.02.002.
16. Watt, H.M.G. and M. Goos, Theoretical foundations of engagement in mathematics.

*Mathematics Education Research Journal*, 2017. 29(2): 133-142. https://doi.org/10.1007/s13394-017-0206-6.

17. Helme, S. and D. Clarke, Identifying cognitive engagement in the mathematics classroom. *Mathematics Education Research Journal*, 2001. 13(2): 133-153. https://doi.org/10.1007/BF03217103.
18. Eccles, J., Expectancies, values, and academic behaviors, in *Achievement and Achievement Motives: Psychological and Sociological Approaches.*, J.T. Spence, Ed. 1983, pp. 75-146: W.H. Freeman.
19. Schoenfeld, A.H., *Mathematical Problem Solving*. 1985: Academic Press.
20. Pólya, G., *How to Solve it?* 1945, Princeton, NJ: Princeton University Press.
21. Shoenfeld, A.H., Research Commentary: Reflections of an Accidental Theorist. *Journal for Research in Mathematics Education JRME*, 2010. 41(2): 104-116. https://doi.org/10.5951/jresematheduc.41.2.0104.
22. Connell, J.P. and J.G. Wellborn, Competence, autonomy, and relatedness: A motivational analysis of self-system processes, in *Self processes and development.* 1991, pp. 43-77. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
23. Ahlfeldt, S., S. Mehta, and T. Sellnow, Measurement and analysis of student engagement in university classes where varying levels of PBL methods of instruction are in use. *Higher Education Research & Development*, 2005. 24(1): 5-20. https://doi.org/10.1080/0729436052000318541.
24. Ryan, A.M., Peer Groups as a Context for the Socialization of Adolescents' Motivation, Engagement, and Achievement in School. *Educational Psychologist*, 2000. 35(2): 101-111. https://doi.org/10.1207/S15326985EP3502_4.
25. Jaworski, B., et al., An Activity Theory Analysis of Teaching Goals versus Student Epistemological Positions. *The International Journal for Technology in Mathematics Education*, 2012. 19: 147-152.
26. Burton, L., Why Is Intuition so Important to Mathematicians but Missing from Mathematics Education? *For the Learning of Mathematics*, 1999. 19(3): 27-32.
27. Fischbein, E., Intuition and Proof. *For the Learning of Mathematics*, 1982. 3(2): 9-24.
28. Attridge, N. and M. Inglis, Increasing cognitive inhibition with a difficult prior task: implications for mathematical thinking. *ZDM*, 2015. 47(5): 723-734. https://doi.org/10.1007/s11858-014-0656-1.
29. Thomas, M.O.J., Inhibiting intuitive thinking in mathematics education. *ZDM*, 2015. 47(5): 865-876. https://doi.org/10.1007/s11858-015-0721-4.
30. Trémolière, B. and W. De Neys, When intuitions are helpful: Prior beliefs can support reasoning in the bat-and-ball problem. *Journal of Cognitive Psychology*, 2014. 26(4): 486-490. https://doi.org/10.1080/20445911.2014.899238.
31. MacLeod, C.M., The concept of inhibition in cognition, in *Inhibition in cognition.* 2007, pp. 3-23. Washington, DC, US: American Psychological Association.
32. Gilmore, C., et al., The role of cognitive inhibition in different components of arithmetic. *ZDM*, 2015. 47(5): 771-782. https://doi.org/10.1007/s11858-014-0659-y.
33. Kahneman, D., *Thinking, fast and slow*. Thinking, fast and slow. 2011, New York, NY, US: Farrar, Straus and Giroux.

34. Babai, R., E. Shalev, and R. Stavy, A warning intervention improves students' ability to overcome intuitive interference. *ZDM*, 2015. 47(5): 735-745. https://doi.org/10.1007/s11858-015-0670-y.
35. Frederick, S., Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 2005. 19(4): 25-42. https://doi.org/10.1257/089533005775196732.
36. Klymchuk, S., *Money puzzles: On critical thinking and financial literacy*, Maths Press, 2001.
37. Guilford, J.P., Creativity: Yesterday, Today and Tomorrow. *The Journal of Creative Behavior*, 1967. 1(1): 3-14.
38. Thomas, M.O.J., Developing versatility in mathematical thinking. *Mediterranean Journal for Research in Mathematics Education*, 2008. 7(2): 67-87.
39. Hatano, G. and Y. Oura, Commentary: Reconceptualizing School Learning Using Insight From Expertise Research. *Educational Researcher*, 2003. 32(8): 26-29. https://doi.org/10.3102/0013189X032008026.
40. Thomas, M.O.J. Conceptual representations and versatile mathematical thinking. in *ICME-10: The 10th International Congress on Mathematical Education*. 2004. Copenhagen, Denmark.
41. Wallas, G., *The art of thought*. 1926: Harcourt, Brace & Company.
42. Leikin, R., in *Exploring Mathematical Creativity Using Multiple Solution Tasks*. 2009, pp. 129-145: Brill | Sense.
43. Leikin, R. and D. Pitta-Pantazi, Creativity and mathematics education: the state of the art. *ZDM*, 2013. 45(2): 159-166. https://doi.org/10.1007/s11858-012-0459-1.
44. Sriraman, B. and P. Haavold, Creativity and giftedness in mathematics education: A pragmatic view, in *First compendium for research in mathematics education*, J. Cai, Ed. 2016: National Council of Teachers of Mathematics.
45. Guilford, J.P., Traits of creativity., in *Creativity and its cultivation*, H.H. Anderson, Ed. 1959, pp. 142-161: Harper & Brothers.
46. *National Survey of Student Engagement*, Indiana University Center for Postsecondary Research and Planning, 2000.
47. Torrance, E., *Education and the creative potential*. 1963: University of Minnesota Press.
48. Evans, T., et al., Engagement of undergraduate STEM students: the influence of non-routine problems. *Higher Education Research & Development*, 2020: 1-17. https://doi.org/10.1080/07294360.2020.1835838.

## Appendix

Solutions to the non-routine problems:

**The Race.** If A beats B by 10 metres, that means that when A finishes B is at the 90 metre mark, meaning that B's pace is 90% of A's. Similarly, C's pace is 90% of B's. Thus, C's pace is $0.9\times0.9\times100=81\%$ of A's, so that when A is at the 100 metre mark, C is only at the 81 metre mark and A beats C by 19 metres.

**Tennis tournament.** Fifty-five (55 players to be eliminated need 55 matches. Illustrating 'start at the end' problem-solving strategy.)

**Fraction of a Square.** First, rotate the shaded area 45 degrees. Then draw the diagonals of the shaded square to conclude that it is one half of the large square.