”Just-in-time teaching” in undergraduate mathematics

Kristina Juter and Jan-Fredrik Olsen

We compared five groups of students to investigate the effects of ”Just-in-time teaching” (JiTT), a method designed to both help students keep up with the often fast pace of undergraduate calculus and to deepen their learning. In total, 137 students participated in the study. The outcome is discussed in terms of conceptual and procedural knowledge in relation to examination and other assessment tasks. We observed an improvement on the assessed items and a shift in study habits.

Teaching university mathematics at the undergraduate level often means providing a learning environment suitable for a large number of students. It is a challenge for the lecturer to get a sense of how the students are keeping up the pace and whether they grasp the topics taught. *Just-in-time teaching* (Novak, Patterson, Gavrin & Christian, 1999) is a method to help students actively follow the course with a focus on conceptual understanding. It is also designed to give the lecturer a diagnostic tool so that he or she has up to date information on the struggles of his or her students.

Simkins and Maier (2010) edited a collection of reports where the use of JiTT in physics, biology, geosciences, economics, history and the humanities are discussed. For each field, the question ”Does JiTT help?” is posed, and answered positively, with a varying degree of stringency.

In this article we posed the following research questions:

> What effects, if any, does Just-in-time teaching have for the students in an undergraduate calculus course?
>
> What effects does Just-in-time teaching have on learning prospects for the students and implementation requirements for the teacher?

Kristina Juter, *Kristianstad University*
Jan-Fredrik Olsen, *Lund University*

Juter. K. & Olsen, J.-E. (2014). ”Just-in-time teaching” in undergraduate mathematics. *Nordic Studies in Mathematics Education*, 19 (2), 77–96.

Related to the questions are the circumstances under which the teaching occurs, such as lectures, problem sessions and the teacher's commitment to the course. To answer the questions, we studied, both quantitatively and qualitatively, a calculus class of 137 students in a Swedish university. The class was split into an experimental group and a control group. Four of the students were interviewed and the teacher's work and intentions were documented. We looked for effects on the students' abilities to solve problems, to explain concepts to both fellow students and lay-persons, as well as any effect on their study habits. We also looked for effects, in a more qualitative sense, on the learning situation for the students and the impact for teaching.

We remark that in the context of secondary school education, Bloom (1984) observes what is coined the 2-sigma effect. Namely, under one-on-one tutoring in science classes, 98 % of pupils perform as well, or better, than the average pupil under traditional teaching methods. In a world of increasing use and development of technology, not only is an efficient undergraduate science education important, but it has the potential to be improved. A natural question is therefore, how close to the 2-sigma effect can we now get?

The structure of the paper is as follows. First, we present the theoretical framework of the paper. Next, we describe the group participating in the study before moving on to descriptions of planning and implementation. After discussing the empirical results, we describe 4 short interviews with students participating in the study and end the paper with a brief discussion of our findings.

Theoretical framework

A significant part of learning mathematics includes solving tasks. This is true to the point that understanding a concept, and being able to solve tasks involving the concept, is regarded as synonymous by most students. Below we review theories for considering the problem of deep versus surface learning.

Procedures and concepts

Hiebert and Lefevre (1986) distinguish between the conceptual and technical aspects of a concept. Indeed, they define *conceptual knowledge* (p. 3) to be a web of information well-linked together with numerous and meaningful connections, and *procedural knowledge* (p. 6) to be knowledge requiring an input that the learner recognizes and is able to perform a single procedure or a string of procedures on to achieve a desired outcome with only a minimum of conceptual understanding.

Strong and valid connections between concepts, i.e., conceptual knowledge, help learners understand more as new information is embedded in, and supported by, their existing knowledge (Hiebert & Carpenter, 1992).

Tall (2004, 2008) introduces a model describing mathematical development of three different types, called worlds. *The conceptual-embodied world* has an emphasis on exploring activities. *The proceptual-symbolic world* focuses on concepts' dual features as objects and processes expressed in symbols where the objects, processes and symbols together are denoted *procepts* (Gray & Tall, 1994). *The formal world* is where mathematical properties are deduced from the formal language of mathematics in definitions and theorems.

It is clear that students' mathematical knowledge develops in various situations, e.g. at lectures, in study groups or in situations outside the university context. In the terminology of Tall, a person's *concept image* (as defined by Tall & Vinner, 1981) develops through experiences in the three worlds. Students' solutions to tasks, reasoning and other mathematical actions, correct or erroneous, can all be seen as traces of their concept images.

A student needs to have a solid conceptual knowledge to be able to move from the first of Tall's worlds to the next. Although the three worlds in one sense describe a hierarchy of increased mathematical maturity, the development may be too rapid in some cases, rendering the student with a proceptual-symbolic thinking strategy without a firm enough embodiment, leading to a shallow learning strategy.

Students' strategies for learning mathematics

Lithner (2003, 2004) investigated textbook exercises used in basic calculus courses in Sweden. The exercises covered limits, derivatives and integrals in various applications. He uses the term *identification of similarities* to denote reasoning based on recognition of surface properties followed by an attempted solution based on imitating previously seen procedures. He states that identification of similarities is easier than more conceptually demanding reasoning. *Plausible reasoning* (Lithner, 2004) is when the reasoning is based on intrinsic properties of the mathematical objects with an aim towards a plausible truth (the reasoning need not be complete or accurate). Lithner claims that most exercises done by students only require reasoning based on identification of similarities which may leave the students believing that mathematical activity is only mimicking other people's efforts.

Another influence on students' ways of learning mathematics is assessment. Bergqvist (2007) shows that exams (the tasks in the study involved differentiability, extreme values and proof) at universities in Sweden generally are constructed so that the students are able to pass them using only *imitative reasoning*. This means that the students do not need to have more than an instrumental understanding of the topics assessed at the exams. This result combined with Lithner's on textbooks (2004) implies that there is no inspiration for students to learn mathematics conceptually, in the sense of Hiebert and Lefevre (1986). Penglase (2004) investigated students learning calculus at a university where an alternative assessment method was tested to change their focus on imitative reasoning to a more conceptual attitude. The course lecturer let the students practice self-assessment and reflection and strived for the students to have a deep understanding of calculus. They were also encouraged to see mathematics as a humanly constructed means to make sense of the world, rather than mere rules used to solve a set of predetermined tasks. Penglase (2004) concludes that there is a need for a consistent focus through all elements of a mathematics course to facilitate for the students to obtain a deeper learning strategy.

Just-in-time teaching

Proposed by Novak and Patterson (1999), *Just-in-time teaching* (JiTT) is a technique that is being applied in several disciplines, and is designed to encourage deep learning (Simkins & Maier, 2010).

The idea is to make students prepare more efficiently for class at home, and to offer a diagnostic tool for the teacher to assess the understanding of the students prior to class. Ideally, this leads to a classroom situation where students and instructors have a common understanding of the difficulties in the material to be covered.

The typical implementation of the technique is to assign reading assignments with one or two conceptual questions to be handed in the day before class by means of some online learning platform. This should take the students about 30–60 minutes. The instructor then looks through responses before class, adjusting the plan for the lecture where appropriate. Moreover, the instructor may include a few (anonymized) answers as examples, linking the lecture to the students own responses.

Novak and Patterson (2010) explain that in order to encourage students to complete the JiTT assignments, it is advisable for these assignments to count towards about 5–10 % of the final grade in the course. However, this may be counter-productive as summative assessment tends to favor procedural over conceptual understanding. Another difficulty is that in a large class, it is problematic from a time management perspective for the instructor to give detailed individual feedback on the reading assignments. For these reasons, it is suggested that they be assessed on a simple scale (e.g. 0–5) based solely on effort.

Description of the class

The current study was performed on a first semester class of calculus at a Swedish university with about 140 students. The class corresponds to half of the workload of the students over the course of one semester (the other being taken by an introductory course in algebra). Below, we describe both the structure of the course, the students and the teachers.

The structure of the course

The usual mode of teaching in the course is as follows. Twice a week, new material is covered in plenary lectures of 2 hours each, followed by problem sessions in groups of at most 30 students (also 2 hours each) that contain a mix of hands-on activities and blackboard demonstrations. Corresponding lectures and problem sessions are not on the same day. Also, every second week a homework assignment consisting of 6–7 problems similar to the ones worked in the problem sessions are to be done in groups of 3–5 students. The homework is graded and handed back to the students.

The curriculum follows chapters 1 to 9 in the Calculus textbook of Adams and Essex (2010). This means that limits, continuity, differentiation, Taylor expansions, integration, differential equations and series are covered. Emphasis is placed on mathematical stringency, i.e. the epsilon-delta definition of the limit is discussed at length, as are proofs of the intermediate value theorem, Taylor’s theorem and the fundamental theorem of calculus.

The final examination consists of two parts. First, there is the written exam consisting of what could be called routine exercises which require various degrees of procedural skill. Getting 50 % of the marks on this exam yields the grade ”Godkänd” (pass). Second, a student receiving more than 60 % on this exam is given the option to take an oral exam to get ”Väl godkänd” (pass with distinction). A typical oral exam lasts one hour and requires the student to give an epsilon-delta proof, to prove a theorem from differential or integral calculus (e.g. the mean value theorem, Taylor’s formula or the fundamental theorem of calculus), and a theorem on the convergence of infinite series.

The students

The course is offered twice a year. The study was done during the spring in which the group of students is rather heterogeneous. Out of a total number of 137 students included in the study, 54 were not registered in any study program, 14 were mathematics majors, while 69 students were registered on other programs within the faculty of science. The latter group consisted almost exclusively of students with a physics-related major (64 out of the 69). We remark that these students had no dedicated mathematics course in their first semester.

Since the pace and contents of the course are rather ambitious, the course is considered one of the more difficult on campus. Indeed, in the semester of the study, 155 students enrolled, 110 showed up for the exam, 54 passed the exam and an additional 21 passed the do-over exam.

We remark that out of the 155 enrolled students, we only included the 137 students who submitted at least one homework assignment. We did this in order to exclude students not actively following the course during the semester. As a consequence, out of the 110 students who took the final exam, 20 (5 passed and 15 failed) were excluded. These were all students who did the exam as a re-take. Still, 5 students who did the exam as a re-take were included in the study (3 passed and 2 failed). From the above, it follows that out of the 90 students included in the study, and who took the exam, 49 passed.

The teachers

The teaching team consisted of one lecturer and four teaching assistants (PhD students in mathematics and junior faculty). The teaching assistants were each responsible for a problem session group of about 30 students each. As is described below, these were the control groups in the study. The experimental group, as well as the plenary lectures, was led by the lecturer responsible for the course. It is important to clarify that this was the second author of the paper (Olsen).

As responsible for the course, the second author planned the contents of all plenary lectures and all problem sessions. He wrote the suggested solution to all homework, and was responsible for maintaining the Moodle-based course website. There, also Geogebra-applets used in the course were made available. This means that the material covered across the various problem session groups was mostly identical. Each teaching assistant was responsible for correcting the homework of the students in his or her group.

The first author (Juter) was not involved in any teaching or planning of the course. For this reason, to counter any sense among the students that the performance on the study-related tasks would be used as part of a summative assessment, it was the first author who informed the students of the various aspects of the experiment, handed out the relevant tests and performed the subsequent interviews.

We point out that to be able to control for the potential positive effect of having the second author of this paper as a teacher, the experimental group was disbanded 4 weeks into the course (the course lasted 19 weeks) and the students distributed among the other groups. The second author of the paper, then assumed responsibility of one of the remaining groups. This allowed a comparison of results between a quiz in the fifth week of the course, and on the final exam in week 19 (see figure 5 below).

The study

The overall plan

To implement the Just-in-time teaching method on a subset of the students, we first randomly divided the 137 enrolled students into 5 problem session groups of around 30 students each. As is usual, each of these groups is taught by a member of the faculty, or a PhD student. We label these as groups 1 to 5. Groups 1 through 4 were the control groups, while group 5 was the experimental group.

The Just-in-time teaching treatment was applied to group 5 only during the second and third weeks of the course. This was exactly when limits and continuity were being covered (4 problem sessions). Ahead of each of the problem sessions the students were to answer two questions of mostly conceptual nature, requiring a minimum of computation. These "mini-tests", as we call them, were to be handed in on the Moodle-based e-learning platform used by the class. We describe these tests, which were used to plan the subsequent problem session, in the following section.

As mentioned in the previous section, after these two weeks, group 5 was disbanded and the students distributed among the remaining 4 groups. Here, we note that due to the traditional structure of plenary lectures complemented by smaller problem sessions at the university in question, it was natural to use the problem sessions to apply JiTT to a subset of the students. As a consequence, however, our implementation of the JiTT method differs from how it is usually prescribed, namely, as a tool used in connection with plenary lectures.

To measure the effect (if any) of the pedagogical treatment, a quiz was administered to all students during a plenary lecture in the fifth week of the course. They were given the final 30 minutes of the lecture to work on the quiz (they were allowed extra time if requested). The quiz had two parts, one with 5 rather openly formulated mathematics tasks about limits and continuity. The open formulation was to allow the students to use whatever representations they wanted. The tasks were labelled as conceptual or procedural, as described in the theoretical section, to reveal the students' conceptions. The conceptual tasks were:

1 Explain what a limit is to someone who has not encountered the term before.

2 What does it mean for a function to be continuous at a point?

3 Give an example of a function which has a limit at a point, but which is not continuous at that point.

The procedural tasks were to compute the following two limits:

4 $\lim_{x\to\infty}\frac{x^2+\sin x}{\sqrt{x^4+1}}$ 5 $\lim_{x\to 0}\frac{|2-x|-|2+x|}{x}$

In addition, the score from the following question was split evenly between the conceptual and procedural parts:

6 Formulate and explain the epsilon-delta definition of a limit.

The final part of the quiz addressed how the students worked with the course. The aim of this part was to see if there were any differences between the groups. The students got 8 statements to agree with or not on a scale of 1–4:

a I prepare well before problem sessions.

b I prepare well before lectures.

c After problem sessions, I revise what we did and work further with what I did not understand or had time to do.

d After lectures, I revise what we did and work further with what I did not understand or had time to do.

e I usually study with other students.

f I prefer to study on my own.

g I feel that I spend a sufficient amount of work on the calculus course.

h I feel that time is not enough for my work on the calculus course.

Out of the 137 students included in the study, 90 did the quiz. Out of these, 76 gave their names. A total of 61 students both gave their names and took the final exam.

The first author performed in-depth interviews with 4 students from the experimental group to obtain further information on how the students perceived and worked with the JiTT method. The students were selected to represent various marks on the course and have an even gender composition (two female and two male). One interview was conducted with two students in a pair, and the other two individually for practical reasons on the students' behalf. Each interview lasted 30–40 minutes and comprised questions from the quiz, the mini-tests as well as the main part about their study habits and views on JiTT. In the section on results, we also take into account the results on the final exams as formerly described.

The experimental group

We now describe in greater detail the treatment applied to the experimental group. In each of the four problem sessions dealing with limits a mini-test containing two questions was published on the Moodle-based e-learning platform used in the course.

The first mini-test was as follows (translated from Swedish):

1 Describe in your own words what it means that $lim\, f(x) = L$ when x approaches some number a. Does the value of the function at the point $x = a$ matter?

2 Assume that both $lim\, f(x) = 0$ and $lim\, g(x) = 0$ as x approaches a. What can be said about $f(x) / g(x)$? You are encouraged to draw on examples in your explanation.

In the Moodle-platform, the students could answer by means of a simple text box and a submit button. This made the use of mathematical symbols difficult, and encouraged explanations using mostly words and a minimum of formalism. The deadline was set to 18:00 the evening before the problem session, although this was not strictly enforced.

There were 22, 21, 24 and 22 responses, respectively, on each of the four mini-tests. Fifteen students responded to all four tests, 7 to three of the tests, 4 to two of the tests and 1 student to only 1 test.

In the subsequent problem session, the teacher prepared a sheet with "instructive" answers to the above questions. The sheet contained both false and correct statements. This was not only intended to give examples of both faulty and correct arguments, but was also supposed to give the students a sense of the class actually being planned according to their current level of proficiency. As an example, the following answers were chosen as representative of the first question on the first mini-test:

> *lim* $f(x) = L$ as x goes towards a means that the value of the function goes toward L as x gets closer to a, but that it will never reach the exact value since the point that is exactly in a is undefined. It is so despite the fact that we can make a pretty good guess of what it is supposed to be, but it is not possible to know exactly.
>
> The expression describes the value, L, that the function, $f(x)$, gets closer to as x approaches the number a. The value of the functions (sic!) when x is a does not have to be L or even be defined. We seek the value of the function when x gets closer to a, but not when x is a.

About 15–30 minutes were set aside to discuss these answers, before discussing more or less the same exercises as in the other problem session groups.

Empirical findings

Recall that the students were divided into 5 groups out of which group 5 was the experimental group. After the experiment, group 5 was disbanded and the students allocated among the 4 remaining groups. We label the resulting 4 groups as groups 11 to 14.

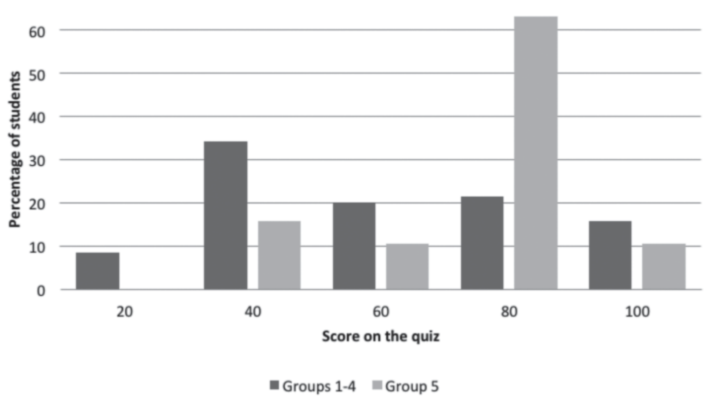

Figure 1. *The percentage of students obtaining a given score on the quiz (out of a maximum 18 points)*

A first analysis of the quiz

In figure 1, we compare the performance of the students in the control groups to the students in the experimental group on the quiz. It appears that the performance of the students in the two groups differed. First, the mean value of group 5 is higher than that of groups 1–4. Perhaps as important, the variance in group 5 appears much smaller. A t-test shows that the mean of group 5 is higher with a one sided p-value of 0.004, with the variance being smaller with an f-test p-value of 0.03.

Next, we consider the conceptual questions and the procedural questions separately (see previous section).

In figure 2, we see that most of the difference in the performance between group 5 and groups 1–4 can be linked to the conceptual score. As for the total score, the improved means are statistically significant, while the reduced variance is significant only for the conceptual score.

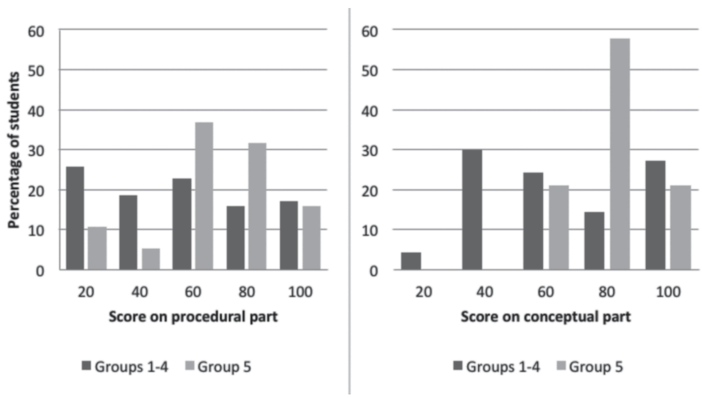

Figure 2. *Histograms of the score of the students with respect to the procedural part of the quiz (left-hand side) and the conceptual part (right-hand side)*

We now turn to the section on the quiz related to study habits. Perhaps surprising, in figure 3, there appears to be little difference in the reported study habits, at least in terms of a one sided t-test (equal variances) with a level of 5 % on the averages. Closest to being significant was the statement ”I usually work with others” which yielded a one-sided p-value of 0.07. Other statistics being close to significant were ”I prepare well before plenary lectures” and ”I revise after the plenary lectures” with p-values of 0.14 and 0.11, respectively. The statement ”I prefer to work alone” had a p-value of 0.19.

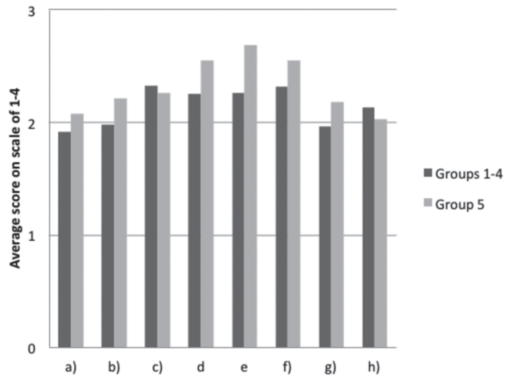

Figure 3. *The averages of the students' answers to the various work habit related questions on the quiz*

Table 1. *The correlation of total score on the quiz and working habits.*

Groups	a	b	c	d	e	f	g	h
1–4	0.26	0.21	0.14	0.15	-0.12	0.31	0.50	-0.19
5	0.39	0.04	0.56	0.10	0.37	-0.13	0.001	-0.05

Note. For the Pearson-r statistic to be significant in a two-sided hypothesis test at a level of 5 % the values for groups 1–4 have to be above 0.23 and for group 5 above 0.46.

Table 1 shows the correlation of the total score on the quiz and working habits. We note that the highest correlations are to be found in questions c) and g), and that the correlations are reversed with respect to the groups. This is illustrated in figure 4. In other words, in the experimental group, revising after a problem session had a much higher effect than in the other groups. Revising after plenary lectures seems to be of lower significance in both groups. Moreover, it is curious that the perception that the student has spent "a sufficient amount of time revising" is highly correlated to success in groups 1–4, while this seems to be irrelevant to the students in group 5. Pointing in the same direction is the fact that for groups 1–4, then question a) ("I prepare well before problem session") is positively, and significantly, correlated to the quiz score, while this is not the case for the experimental group (although the correlation is positive). Finally, question f) ("I prefer to work alone") is significantly and positively correlated for groups 1–4, and non-significant for the experimental group.

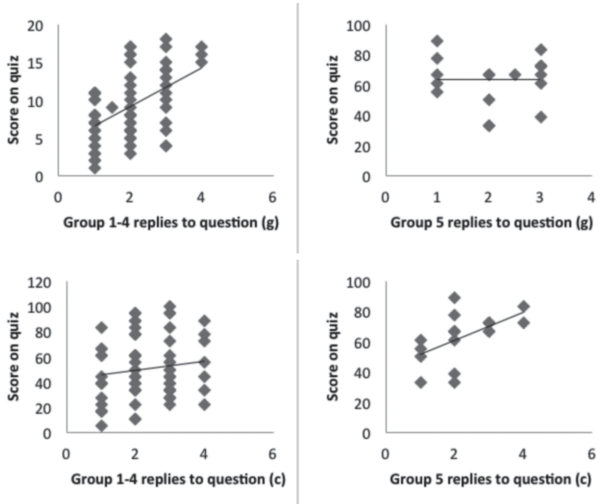

Figure 4. *Scatter plots of the replies to work habit question g) "It feels as if I spend a sufficient amount of time on the course" and c) "After problem sessions, I revise what we did and work further with what I did not understand or had time to do".*

Including the final exam in the analysis

We now continue the analysis taking the results of the exam into account. This not only allows us to see whether the experimental treatment had any effect on the exam result, but it also allows us to check the performance of the teacher of the experimental group when not using the JiTT method.

In the left-hand graph of figure 5, we show the distribution of exam scores of the students groups 1–4 and group 5. Although the students from group 5 performed better than the rest on the exam, a closer

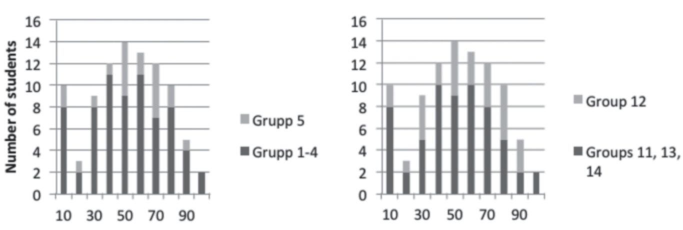

Figure 5. *The distribution of the exam scores of group 5 and group 12, respectively, among the course exam results.*

analysis shows that this is not statistically significant. A student t-test gives a p-value above 0.3.

After group 5 was disbanded, the teacher of group 5 took over group 12 in the new enumeration mentioned above. In the right-hand graph of figure 5, we compare the exam results of group 12 to the rest of the class. It is clear that no dramatic shift, as seen in figures 1–3, has taken place. Indeed, as in the previous paragraph, a student t-test shows no significant difference between the distributions, although it has a rather low one sided p-value of about 0.13.

Finally, we include the following analysis that addresses the way we decided to mark the quizzes. Our philosophy put more value on a complex, but possibly incoherent, concept image than is usual, i.e., a concept image that comprises concepts and relations between them elaborated enough to enable the students to reason mathematically, although with some short-comings, both logical and formal. However, at an early point in a course, this may be appropriate as the material is still very new to the student. In figure 6, we show the correlation between the quiz score and the score on the final exam.

We note that if you got more than 50 % of the points on the quiz, then you had a 82 % chance of getting more than 50 % of the score on the exam, while if you got less than 50 % on the quiz, you had a 23 % chance of getting 50 % or more on the exam.

Four students' thoughts of the experiment

We now describe the four student-interviews. We call the two female students Anna and Britt, and the two male students Calle and Didrik. They were all physics majors.

Anna: She got the highest mark on the exam. She liked the Just-in-time teaching and wanted to have JiTT in the whole course. It forced her to read the course book to manage the tests with difficult tasks. She spent about one hour for each mini-test and used the course book as help. She submitted her answers the night before deadline. The exercises were not directly useful for the exam since she thought they gave more theoretical understanding. After the period with JiTT, she said she became lazy and went back to not following the course as strictly. She had some difficulty when she did the mini-tests. She was uncertain of what to explain and what her answers actually meant. The minimalistic feedback to the JiTT assignments made it hard for her to know if her own work was correct. However, since all correct answers to all questions were given so she thought it was OK. She did not use the feedback or answers again in the course. Anna solved the suggested course tasks, sometimes after consulting the course book, but more often she used her lecture notes. She found

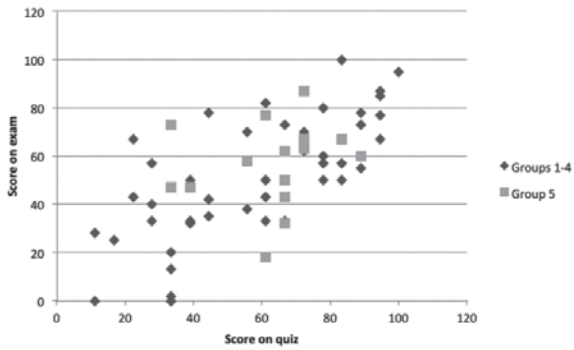

Figure 6. *Scatter plot of quiz score versus exam score*

Note. If pooled together, the scatter plot yields a Pearson correlation coefficient of 0.65 which is statistically significant with a one-sided p-value of less than 0.005.

the course book too structured making it hard to find the important parts among the other parts, but she did like the vast number of examples. She worked alone, since it was more efficient and she had not found anyone at her level to work with. The problem sessions were her best learning support and she attended the problem sessions at the end of the course when she found that she needed them. At first, Anna did not like mathematics, but after a while she started to understand and became fond of the subject, and found the various tasks easier to complete.

Britt: She passed the exam, but with the lowest mark and was not pleased. She was positive to the JiTT and thought it was less demanding than she expected at first. She usually submitted her answers at the last minute. When she had done the JiTT-tasks she continued with the rest of the course work, which she thought was a positive effect of the JiTT. Britt used the book as support for answering the mini-tests. She answered the JiTT-assignments alone, but worked in groups on the rest of the homework (2–10 students, all in physics), and sometimes alone to get peace to work. Britt studied less after the mini-tests, but thought this could also be due to an algebra exam parallel to the calculus course. She would have preferred JiTT over normal teaching throughout the entire course since she lacked self-discipline and became forced to keep the pace. She used the feedback she got from the JiTT when studying for the exam. She read ahead to prepare for difficult lectures, but did not like the book, as it was similar everywhere making it hard to discern the important parts. She solved all recommended problems in the book and some more. Problem sessions were best for learning, but she liked the lectures as well. Britt thought that the mini-tests took a long time from the course, which went fast as she had expected it to do. She thought that it was hard to know when a proof is done although this became easier at the end of the course.

Calle and *Didrik*: Calle did not pass the first exam, but passed the second without margins. Didrik passed the first exam with a good margin, but not the highest mark. They both liked the JiTT since it forced them to read before each problem session. They both needed to prepare before the lectures to be able to follow them and the mini-tests worked well as preparation tools. Both students submitted their answers a couple of hours before deadline, both alone. They usually worked together with one other student. Didrik used the course book and Wikipedia, Calle mostly used the course book. They read their feedback from the JiTT and then they looked at their work again. Calle needed the feedback to get the solutions completely right. They both agreed that the mini-tests did not require a lot of time and they preferred JiTT the entire course to be forced to follow the pace. The tests were more to understand the theory than to pass the exam. The problem sessions were most useful for learning and they attended all of them. They thought the book was good with many details and accessible not only to mathematicians. Calle and Didrik were both confused about how to give proofs. In connection with the mini-tests, they were also uncertain about when a proof is complete. They thought it became clearer at the end of the course. Calle stressed that he wanted to understand and not just repeat something. The course was mainly as they had expected it to be.

These four cases are included to give a sense of how the students perceived the JiTT in the course. The students showed many similarities even though they performed differently at the exam, e.g. the uncertainty about giving proofs and understanding when a proof is done. All students except Anna used their feedback from the JiTT again when studying other parts of the course. Anna, Calle and Didrik thought that the mini-tests helped them theoretically with the topics of the course, and all four students claimed to have a better understanding of giving proofs at the end of the course, due to the mini-tests or otherwise. The mini-tests worked as a catalyst for the students' work with the course as they kept the pace.

Discussion

In the introduction, we posed questions of what effects JiTT has on an undergraduate course in mathematics. To answer the questions, we split the students into an experimental and a control group. The main difference between the groups was observed on a quiz administered shortly after the application of the experimental treatment (see figures 1 and 2). We saw a clear improvement in the performance of the experimental group, both in terms of a shifted mean and a reduced variance. On the final exam, however, we observed no significant effect (figure 5). An obvious explanation for this is that the exam was given 3 months after the 2-week experimental treatment ended.

Closely related to the question of whether JiTT has any effect, is the question whether the observed effect is only due to the specific teacher assigned to the experimental group. This becomes especially important, since this teacher both gave the plenary lectures and is the second author of the current paper. To control for this, we disbanded the experimental group and let the second author take over one of the non-experimental groups (group 12). As can also be seen from figure 5, the performance of this group was identical with that of the class as a whole.

In relation to the potential positive effect of the assigned teacher, we mention an effect that would indicate a difficulty involved in the implementation of the JiTT method which was both encountered in this study, and which has also been reported in previous implementations of the JiTT method (e.g., Hill 2004; Alexandros Sopasakis, private communication, September 4 2013), namely, time-management. For various reasons, most students handed in their answers to the mini-tests between 20:00 and midnight, instead of before 18:00. At first, it did not seem reasonable to deny them this. What this meant was that the teacher read through the various answers of the students at 5 am, and worked quite hard to compile a list of ”instructive” answers and changing notations for them to be readable. The resulting lack of sleep lead to the teacher perceiving a reduction of his own performance.

Taking the interviews into consideration, we get a suggestion as to why the JiTT appears to have a positive learning effect. Similar to the students in Penglase’s (2004) study, the students in the experiment group in this study had some kind of ongoing assessment forcing them to be more active. Indeed, all the students interviewed mentioned this in some way. As a result, their conceptual skills were clearly improved, and, to a certain extent, were their procedural skills. Most likely, the reason for this disparity is simply that the mini-tests focused on conceptual skills. However, this effect did not last until the final exam. An explanation for this could be that the students were well aware of the structure of the exam which traditionally has a focus on procedural knowledge. This means that all students need to master are routine strategies to pass the course. However, it should be mentioned, that more ambitious students need to pass a more conceptually focused oral exam to get the highest grade.

We did not observe any statistically significant change in the reported study habits of the students in the experimental group (figure 3). What we did observe, however, was a change in correlations between the result on the quiz and working habits (table 1). For instance, correlation with revising after problem sessions was significantly stronger in the experimental group. In other words, it seems that the students were able to make more sense of them. Also, in the experimental group there was no correlation between having spent enough time on the course and the performance on the quiz. This correlation was quite strong in the control groups. A reason for this may be that more efficient learning strategies in the JiTT problem sessions led to the students not having to spend a large amount of time outside of class to achieve the same, and in fact, superior results. It is curious to note, however, that the students in the control group do not seem to be aware of this themselves. Also pointing in this direction, was that for the experimental group, the answer to the question of having prepared well before problem sessions was not significantly correlated with a good result on the quiz, while the contrary was true for the control group. The final significant correlation was the apparent positive effect for the control group between preferring to work alone and good quiz scores. While we have no evidence to support this, it would be tempting interpreting this as saying that the students in the experimental group had adopted a study technique more efficient when working in groups, or at least, were less inefficient than the control group.

As remarked in the description of the class, the student-body was quite heterogeneous. Indeed, 47 % of the 137 students participating in the study had already completed a first semester of physics, albeit without taking any dedicated course in mathematics. It is well-known that these students tend to outperform the others on the exam. In particular, this term they made up 63 % of the students passing the course. For the mathematics majors, both numbers were 10 %. The remaining students had no registered major. For this reason, it is important to remark that 52 % of the students in group 5 were physicists, while in the control groups the average number of physicists was 45 %.

Finally, we mention the seemingly paradoxical fact that group 5 did not do significantly better than the control groups on the final exam despite outperforming them on the quiz, which appears to be a good predictor for success. One possible explanation, which should be kept in mind throughout this paper, is that the number of students from group 5 (16) that took both the quiz and the exam is very small. If we were to speculate, then another possible explanation could be that JiTT empowered students who do not thrive under normal teaching methods to perform well on the quiz. When the JiTT treatment was stopped, these students may then have fallen back to their normal routine, as implied in the interviewed students' answers. In addition, the strong correlation between the quiz score and the score on the final exam could be taken as an indication that if the study was implemented over the entire semester, then a positive effect on the exam scores would have been seen. However, from this study, it is not possible to know whether this would actually have happened.

It is interesting to note that the two students who gave the sample answers to the first mini-test, shown earlier, performed nearly identically on the final exam; the student giving the ”good” answer got 50 %, while the other student got 47 %. Unfortunately, these students did not wish to be interviewed for the study. However, out of the students we did interview, all but Calle gave a very good answer to the sample question.

References

Adams, R. A. & Essex, C. (2010). *Calculus: a complete course*. Toronto: Pearson Canada.

Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *Journal of Mathematical Behavior*, 26 (4), 348–370.

Bloom, B. S. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13 (4), 4–16.

Gray, E. & Tall, D. (1994). Duality, ambiguity and flexibility: a procedural view of simple arithmetic. *The Journal for Research in Mathematics Education*, 26 (2), 115–141.

Hiebert, J. & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.

Hiebert, J. & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: an introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: the case of mathematics*. London: Lawrence Erlbaum.

Hill, S. (2004). *Just-in-time teaching*. PowerPoint lecture presented at 2004 Cottrell scholars’ conference, Tucson. Retrieved from http://www.phys.ufl.edu/~hill/talks/Cottrell%2004.pdf

Lithner, J. (2003). Students’ mathematical reasoning in university textbook exercises. *Educational Studies in Mathematics*, 52, 29–55.

Lithner, J. (2004). Mathematical reasoning in calculus textbook exercises. *Journal of Mathematical Behavior*, 23, 405–427.

Novak, G. M., Patterson, E. T., Gavrin A. D. & Christian W. (1999). *Just-in-time teaching: blending active learning with web-technology*. Upper Saddle River: Prentice Hall.

Novak, G. M. & Patterson, E. T. (2010). An introduction to Just-in-Time Teaching (JiTT). In S. Simkins & M. Maier (Eds.), *Just-in-Time Teaching: across the disciplines, across the academy* (pp. 3–23). Sterling: Stylus Publishing.

Penglase, M. (2004). Learning approaches in university calculus: the effects of an innovative assessment program. In I. Putt, R. Faragher & M. McLean (Eds.), *Mathematics education for the third millennium: towards 2010* (Proceedings of the 27th annual conference of the Mathematics Education Research Group of Australasia) (pp. 446–453). Sydney: MERGA.

Simkins, S. & Maier, M. (Eds.) (2010). *Just-in-Time Teaching: across the disciplines, across the academy*. Sterling: Stylus Publishing.

Tall, D. (2004). Thinking through three worlds of mathematics. In M. Johnsen Høines & A. B. Fugelstad (Eds.), *Proceedings of the 28th conference of the International Group for the Psychology of Mathematics Education* (pp. 281–288). Bergen University College.

Tall, D. (2008). The transition to formal thinking in mathematics. *Mathematics Education Research Journal*, 20 (2), 5–24.

Tall, D. & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12, 151–169.

Kristina Juter

Kristina Juter works as an associate professor in mathematics education at Kristianstad University in Sweden. She got her doctoral degree in mathematics and learning at Luleå University of Technology in 2006. Her research interests are, among others, students' conceptual development at undergraduate calculus courses and mathematics teacher identity.

kristina.juter@hkr.se

Jan-Fredrik Olsen

Jan-Fredrik Olsen has a PhD in pure mathematics from NTNU in Norway (2009), and is currently working as an associate professor in mathematics at Lund University in Sweden. While his research interests lie primarily in the meeting point between complex analysis and Fourier analysis, he has a strong interest in undergraduate mathematics education and the potential of technology to improve student motivation and learning.

jan-fredrik.olsen@math.lu.se