CEJOR (2016) 24:1009–1030
DOI 10.1007/s10100-015-0416-5

ORIGINAL PAPER

Kaizen teaching and the learning habits of engineering students in a freshman mathematics course

Jonas Gallenkämper[1] · **Bernhard Heim**[2] · **Judith Kreuzer**[1] · **Florian Rupp**[2] · **Paul von Stockhausen**[2] · **Nils Viet**[2]

Published online: 2 September 2015

Abstract Which are the teaching methods that actually contribute to the learning of mathematics? The answer to this certainly is the holy grail of didactic and pedagogy, and should be supported by large scale statistical evidence. Our article aims at providing an initial step into this direction by first illustrating a teaching paradigm that is suited for the generation of large scale data sets: based on industry best practice quality assurance standards we introduce the Kaizen teaching paradigm which enforces Kolb's reflective learning cycle on the students' side. Second, we present and analyze the data we obtained through our pilot implementation at a engineering freshman mathematics course in the Sultanate of Oman. These emphasize the effectiveness of Kaizen teaching and once again show the necessity of continuous learning. A practice that seems to be forgotten in traditional university engineering courses due to the mere

✉ Jonas Gallenkämper
jonas.gallenkaemper@rwth-aachen.de

Bernhard Heim
bernhard.heim@gutech.edu.om

Judith Kreuzer
judith.kreuzer@matha.rwth-aachen.de

Florian Rupp
florian.rupp@gutech.edu.om

Paul von Stockhausen
paul.stockhausen@gutech.edu.om

Nils Viet
nils.viet@gutech.edu.om

1 Lehrstuhl A für Mathematik, RWTH Aachen, 52056 Aachen, Germany

2 Department of Mathematics and Science, German University of Technology in Oman, Halban Campus, Muscat, Sultanate of Oman

size of the audience. In particular it seems that a Markovian estimator for students' performance may have to be considered.

Keywords Student centred learning · Blended learning · Kaizen learning · Kolb cycle · Teaching in the Middle East

1 Introduction

Teaching, contrary to learning, is an intentional process that conveys an objective or a goal. As Rodriguez and Fahara (2010) clearly highlight, teaching at university environment adopts various models that have to do with personal characteristics of the teacher, the institutional mission, the work environment, relationships with administrators and alternate factors that occasionally are not considered strictly related to teaching such as those mentioned by Dunn and Dunn (1998), like time or the number of students per course.[1]

Though, up to now a complete survey of the teaching impacts of a university level course over a significantly long period of time was rarely conducted. In the winter term 2013/14 we had the chance to conduct such a longer term study while introducing a tailored new teaching paradigm: The philosophy of Kaizen learning that we adopted from our experience at quality control in different global industrial companies. We transformed Kaizen, the philosophy of continuous improvement through strict quality assurance and sample testing, to the university classroom (cf. Heim et al. 2014). This article describes the structural set-up so that this was possible and analyses the success factors provided by direct student's feedback and objective data analysis. It is interesting to note that the adoption of industry principles in education already lead to a predecessor of our Kaizen strategy: In 1984 David Kolb invented a reflective learning cycle based on quality assurance methodologies (cf. Kolb 1984), and reinterpreted previous work on adult learning and group dynamics of Kurt Lewin from the 1940s (cf. Lewin 1951). In that respect Kaizen teaching is the continuous guidance and reinforcement by the lecturer of the students Kolb learning cycle. We refer as well to references Howard et al. (1996), Svinicki and Dixon (1987), Dede (2009) for hands-on adaptations and applications of Kolb's learning cycle.

The novelty and importance of this article is comprised by introducing—in an example based way—the paradigm of Kaizen teaching and giving as well as analyzing the statistical results of a comprehensive case study on the effectiveness and appreciation of Kaizen teaching. Our results strongly suggest a high correlation between continuous learning and testing with the final results the students achieved. In view of increasing pedagogical requirements this teaching methods seems to help maintaining proper standards by allowing a majority of todays students pass the course at the same time.

The article is structured as follows: Proceeding first to a short overview of the course "Mathematics I for Engineering" at the German University of Technology in Oman

[1] We refer to the literature, like Bostrm (2011), Coffield et al. (2004), Honey and Mumford (1986) for more details on the over 70 teaching styles that are currently subject to vivid discussion.

which was the foundation of our case study, we discuss the educational background of our students as well as the types of data we collected and how the students scored in Sect. 2. It is here that we give a first analysis of the data with respect to their distributions and causes for skewness. Next, in Sect. 3 we apply regression techniques to distill correlations between the Kaizen approach and the students' performance. It is interesting to note that the pre-knowledge of the students as tested in one of the very first lectures is not very significant for the students final scoring. On the other hand continuous tests and in particular those tests that were prior had a significant impact on the results of the test conducted next. Like with athletes this may allow a Markovian estimator for students: todays performance is a good predictor for tomorrows success and if one is falling behind extra effort has to be taken immediately. Section 4 then shows the impacts of Kaizen teaching on the students stressing again the importance of continuous learning and repetition during teaching. Finally, Sect. 5 concludes with a short summary and outlook.

2 The student's performance data and its interpretation

Let us start with some information about the course we gave at the winter term 2013/14 and its students (cf. Heim et al. 2014): The syllabus of "Mathematics I for Engineering" (MATH I) at the German University of Technology in Oman (GUtech) consists of (a) mathematical notations, numbers and elementary logic, (b) function basics and trigonometric functions, (c) solution methods for linear systems of equations, (d) vector spaces, vectors, linear mappings and matrices and e) determinants and diagonalization of matrices. The 102 students who took this course came from 12 different countries (84 from Oman, 6 from India to mention just the two largest nationality groups), 76 of them already attended a pre-university program, and the ratio between female and male students was 63–37 %. Moreover an introductory mathematics test at the start of MATH I showed a wide variety of basic mathematical understanding and skills of the students.

Moreover, when reviewing the educational background of our students, there seem to be certain factors critical to their success that may be grouped into four categories:

Family structure Omani families are rather large (although the birth-rate declined from over 7 to now at a bit over 2) such that especially the girls have to help at home. Moreover there is constant disturbance by mobile devices.

Common practices in schools Schools tend to prize memorization above reflection and creative thinking, teachers are an unquestioned authority and pupils actually prefer to be told exactly what they have to do instead of thinking by themselves (cf. Sidani and Thornberry 2009, Ali and Camp 1995). Moreover, teachers are expected to help students pass the course and not to make their lives too hard. With respect to international knowledge surveys, like the "Third International Mathematics and Science Study 2011", Omani pupils score rather poorly compared to other Arabic nations and the world average (cf. IEA 2011). As coming from Germany, our first impressions and attempts to deal with this situation of teaching in a completely different ethnic background certainly required some calibration (cf. Opera 2010). On the other hand, our subject "Mathematics I for Engineering" (MATH I) demands a progression from calculus to mathematics, it requires analytic skills and geometrical sense.[2]

Existence of save options The plethora of fossil resources (oil & gas) and policies that are focused on the exploitation of these seem not the lead to a diversification of perspectives (cf. Thomas 2012). Moreover, Omanization policies more or less guarantee every good graduate a well-rewarded position in the public sector (e.g. with full pension after 25 years of service).

Further issues Schools emphasize on arithmetic and algebra, in particular geometric thinking and spatial sense and drawing capabilities are extremely poorly developed amongst students. Especially for freshmen there may be difficulties with English as the language of education and writing from left to right.

Returning to the specifics of MATH I it is worth to state that this course is a service course for engineering students, hence a burden for the most, the curricula of whom already allocate only little time for self-studies and the preparation of homework assignments.

One of the major challenges for the 2013/14 course was the huge number of students (3-times as much as in the previous courses) and thus the required massive teaching and didactic efforts to enable learning of abstract mathematical structures in such a large class environment. This was achieved by a set of structural measures for the organization of the course, a consistent cybernetic test feedback loop strategy best termed as Kaizen learning, as well as incorporation of student activation in a plethora of ways.

2.1 The student's performance data

A comprehensive set of data was collected to measure the students performance and the effects of different elements of the MATH I course:

Several *State of Knowledge Tests* were conducted to check the student's awareness of basic mathematical concepts that they should know from high school, like fractions, absolute value, differentiation, story problems, etc. The miserable results at the State of Knowledge Test 1 (first lecture) warned us that students will be facing severe problems with the new topics in MATH I as fundamental prerequisites were missing. So the speed of the lecture was adjusted and exercises dealing with those prerequisites were incorporated into the course. Further State of Knowledge Tests in the 6th lecture week and in the 11th lecture week were used to collect information on the students progress on these topics.

[2] As usual the term calculus refers to the set of recipes and is taught to students more or less in a form of copying a presented solution idea to another set of problems which may not even require an in depth understanding of the problem itself. Calculus consist of the basics and the correct way of applying formulas. Although calculus is an essential first step towards mathematics it is to be clearly separated from mathematics. Mathematics has the characteristics of a language for the sciences and demands the autonomous realization of solutions for problems of a previously unknown type. Thus the transfer of techniques has to be taught as well as the modeling of scientific and engineering scenarios.

Online pre-learning exercises and homework assignments can be considered as a small distant learning universe in itself supported by the platform MUMIE.[3] During their pre-learning activities the students get in a playful and hands-on way into touch with new topics that are explained in depth in the next lecture. Hence, they are already aware of the new concepts and may be motivated to understand them better as they already know about their usefulness. The pre-learning exercises are composed of very easy tasks, like the transposition of a 3×2-matrix.

For every homework assignment there was an optional demonstration part including further explanations that serve as a training at which the students can practise and get instant feedback on their answers (if the type of the exercise permits). Then, in theory, the students started the mandatory homework assignments. If the task permits, like checking the results of an Gaussian elimination, the exercise is corrected and graded automatically, if not (e.g. induction), the students hand in their solutions to the tutors who correct them. This provides the opportunity of a more in-depth cause-effect search in finding errors the students were falling for and of an individual consultation about these.

Lecture and tutorial tests These two test types were designed for instant feedback about the progress of the course. The short tests at the beginning or end of the lecture check the understanding of new topics and allow us to react in the tutorial groups with dedicated examples before the students go for their homework assignments. These tests consumed about 10 min and contained very easy exercises which the students were allowed to solve with their lecture notes.

In the tutorial groups some more tricky tests were conducted about the topics discussed in the prior week to the tutorial group. The aim of these tests is to see if the students are able to solve problems alone after successfully working on the homework assignments.

Mock and Midterm Exam Grading of the MATH I course was based on 40 % of the midterm and on 60 % of the final exam. The midterm was conducted in the 8th lecture week containing the topics of sets, functions, absolute values, complex numbers, inequalities, induction, the Gaussian algorithm and boolean algebra. Its level of difficulty was similar to the tutorial tests. In order to acclimatize our freshmen to exams at university, we provided a mock exam one week before the midterm. The mock exam was barely timed and more difficult than the actual exam, so the students could recognize those topics they have to practise more intensely. At the actual midterm exam, the students were granted more time, as we focussed on what the students know and not how fast they reach the solution.

The continuous testing has the positive effects of imposing a strict routine on the students to continuously perform during the term and allows for a cybernetic teaching approach that measures students achievements and enables us to correct deviating developments immediately, cf. Fig. 1.

[3] MUMIE is an acronym for "Multimedial Mathematics in Engineering" and was designed jointly by the RWTH Aachen, TU Berlin and TU Munich, see www.mumie.net for further information. The version used at GUtech enables users to explore and learn mathematics by themselves and gave students the chance to practice together with a full material collection of the lecture's topics and definitions.

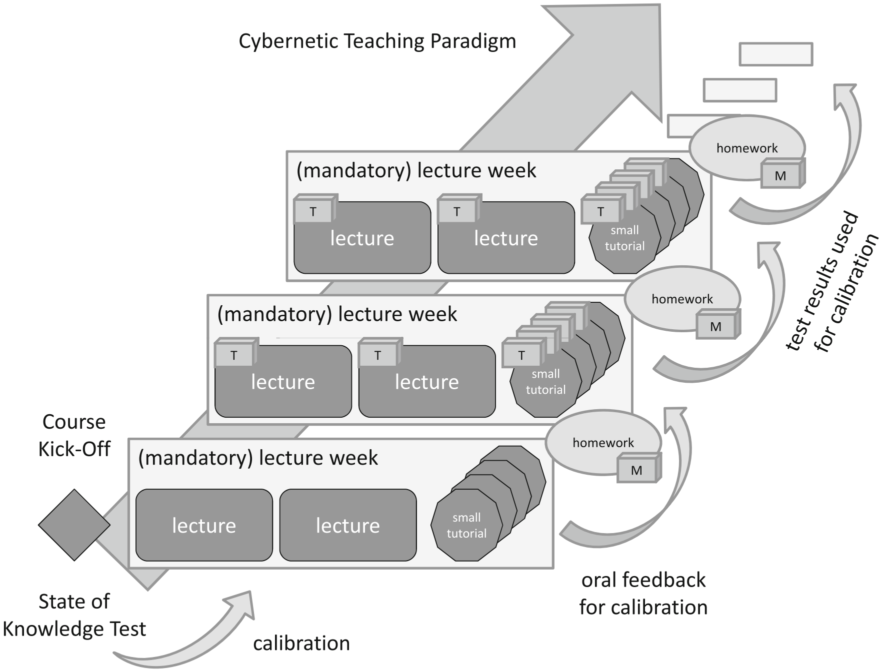

Fig. 1 The cybernetic teaching approach that is in particular realized by the different lecture and tutorial tests and the immediate feedback gained through them

In industry such small cycles of “plan–do–study–act” and then progress are commonly accepted and associated to W. Edwards Deming’s philosophy of Kaizen (Kai = change, zen = for the better), a productivity improving idea, as a continuous change for the best. It is a philosophy of continuous improvement to reduce waste and thereby achieve better efficiency. Important is that Kaizen should be implemented top down. Higher management levels as well as simple workers need to have the same mindset of continuous improvement to achieve significant results. The described amplitude of short tests allowed us to carry this philosophy in a meaningful way to academia, and led us to (more or less) daily evaluations of the course and its continuous improvement. As already stated in the introduction these small cycles are typically referred to Kolb’s learning cycles (cf. Kolb 1984) in the didactics literature, where they are interpreted in terms of “observe–reflect–conceptualize–act” (cf. Svinicki and Dixon 1987).

2.2 Interpretation of the performance data

In order to analyse the significance of these tests in their own right, we give two graphical indicators for each of them while no data clearance is performed upfront (which may be necessary later on to clarify certain relationships among the data). The first is a cumulative histogram on the overall achievements and the second a Quantile-Quantile-plot (Q-Q-plot) of the thus received data. Q-Q-plots are used here to identify the normal distribution of a given random variable by comparing the quantiles of our data sets (ordinal axis) with those of the normal distribution (abscissa axis). If a data set is normally distributed all points generated in a Q-Q-plot lie on the identity line as both, the data set and the normal distribution have the same quantiles. Variations from this behaviour show the deviation of the data set away from normal distribution, see e.g. Chambers et al. (1983).

Figure 2 displays these graphics for the State of Knowledge Test 1. They allow us to conclude a rather poor performance of the students in this test as well as normal distribution of the data. So our student body has a normally distributed knowledge of the relevant prerequisites in mathematics that is centred about a poor mean.

The State of Knowledge Test 1 was conducted during the first lecture to check the mathematical prerequisites of the students (fractions, binomial formula, powers, etc.). A surprising outcome was that it seemed for the majority of the students to be no problem to simplify fractions where binomial formulas are involved. Though, more than 28 % were not able to simplify fractions with numbers without the use of a calculator (which was prohibited during the test). This dependence on a calculator and the consequent reliance on it seems to be a giant drawback from the mathematical school education (not only in Oman).

Figure 3 shows the results of the State of Knowledge Test 2 that consisted of questions similar to that in the first State of Knowledge Test in order to compare the students advancement in the necessary preliminaries of mathematics. The data exhibit a clear skew towards higher grades, as can be witnessed on the high occupation of the largest quantile in the Q-Q-plot. Here, it is rather difficult to speak of normally distributed data. Compared with the situation in the first State of Knowledge Test, the skew feature of the Q-Q-plot in Fig. 3 can be interpreted as the results of learning and especially the impact of teaching. It may be fair to consider this raise in performance in basic mathematics skills as a by-product of MATH I that should not be underestimated.

Figure 4 shows the results of the online pre-learning tests. Here, we see a large occupation of the tails whereas the region around the mean follows a normal distribution

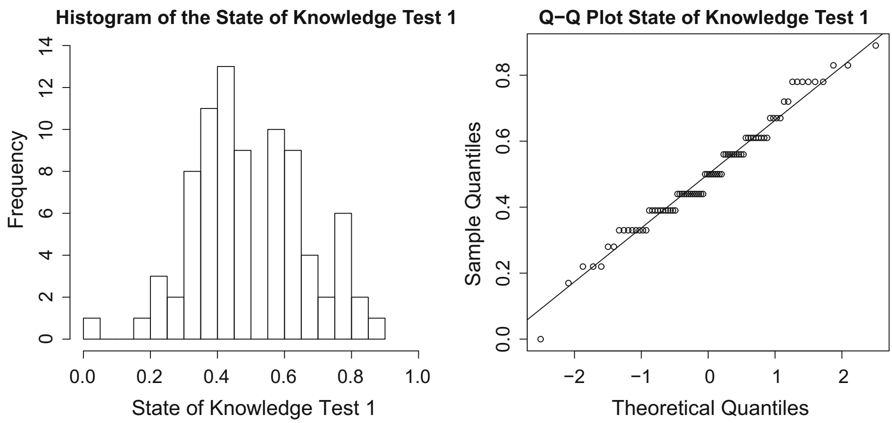

Fig. 2 Histogram and Q-Q-plot for the points achieved in the State of Knowledge Test 1

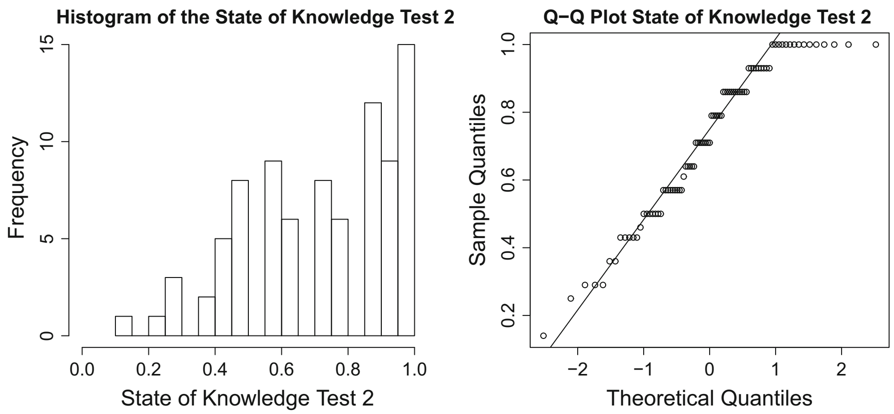

Fig. 3 Histogram and Q-Q-plot for the points achieved in the State of Knowledge Test 2

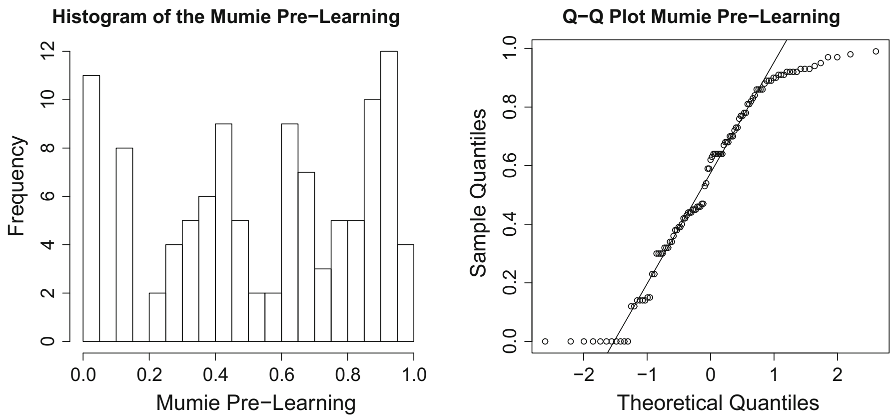

Fig. 4 Histogram and Q-Q-plot for the points achieved in the online pre-learning tests

quite well. Altogether this, of course, does not allow to classify these data as normally distributed. The heavy tails can be explained due to the set-up of the pre-learning questions in the e-learning environment. First of all, as being the case also for the homework assignments, a non-negligible number of students had problems logging into the e-learning platform or simply decided not to consult the platform regularly. Moreover, the pre-learning tests are designed as a kind of teaser for a repetition of the lectures and as a motivation for the homework assignments. So high grades in the pre-learning tests are intended.

Figure 5 illustrates the results of the e-learning homework assignments provided in the platform MUMIE. Here, a skew of the distribution towards good achievements is clearly visible which is explained by the easy nature of the questions imposed and the positive effects of the pre-learning exercises which set the mind of the students for the actual homework assignments. As for the pre-learning tests, technical difficulties may explain the considerable number of the students performing badly.

The results of the lecture tests are shown in Fig. 6. They seem to fit a normal distribution quite well for middle and higher quantiles, though a heavy tail due to poor achievements is clearly noticeable. This may be explained by those students who did not attend the lecture regularly.

Figure 7 displays the results of the tutorial tests that seem to have a rather similar distribution to the results of the homework assignments with their peak at good performance and the two heavy tails. Again, the weak performance tail may be explained by missing students, whereas the good performance tail may be seen as an indicator of the learning process by solving the homework assignments (as it was intended for this kind of tests).

It is interesting to mention, that the pre-learning, the homework assignments as well as the two test types described here show a similar deviation from the normal distribution in terms of their quantiles as well as histograms. This emphasises, once again, Mandelbrot's insight that real data are (far) from normal distribution and for

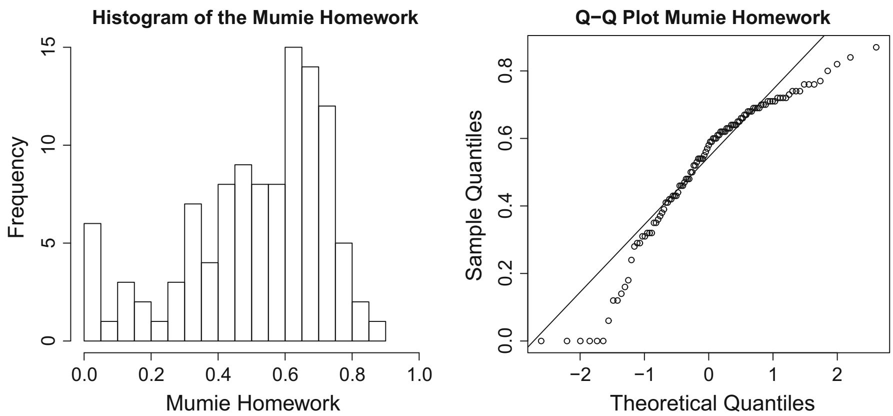

Fig. 5 Histogram and Q-Q-plot for the points achieved in the e-learning homework assignments

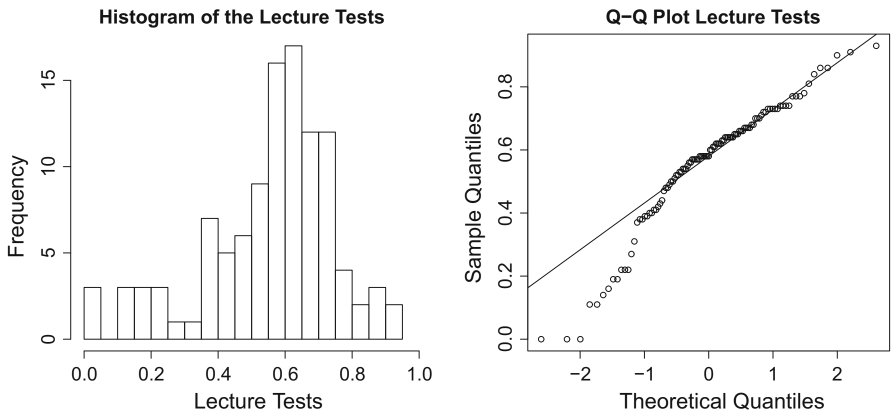

Fig. 6 Histogram and Q-Q-plot for the points achieved in the lecture tests

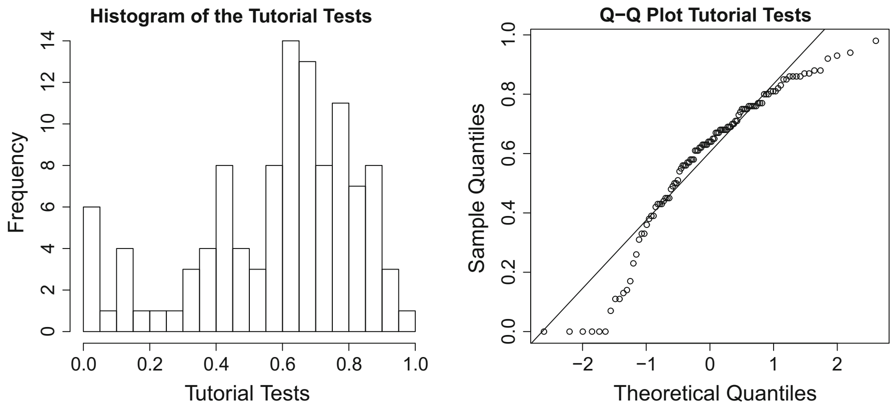

Fig. 7 Histogram and Q-Q-plot for the points achieved in the tutorial tests

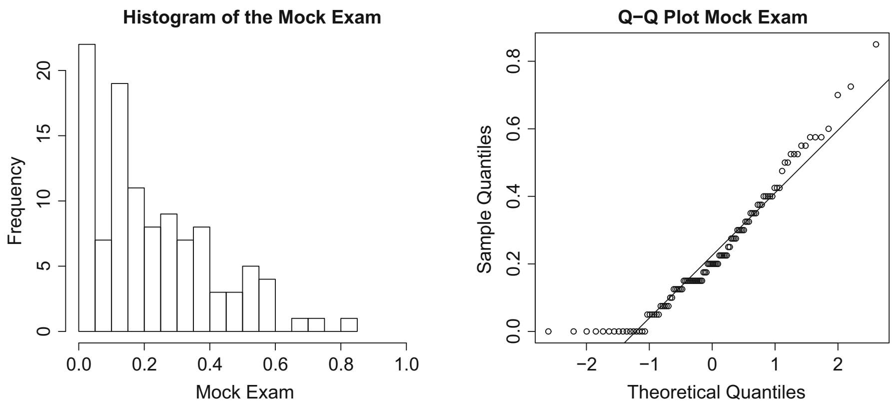

Fig. 8 Histogram and Q-Q-plot for the points achieved in the mock exam

certain have heavy tails (cf. Mandelbrot 1983). A first hypothetical explanation of this similarity of the data may lie within the student's enthusiasm to follow the course: There seems to be a large number who actively participates and thus takes the easy to get credits, whereas another large group seems to be rather inactive and is not really interested in following MATH I.

Figure 8 gives the results of the mock exam. Due to the nature of the mock exam as a tough test with strong time restrictions the data are as expected: the majority of the students did not perform well, and the histogram looks like the mirror image of the data of the State of Knowledge Test 2.

Figure 9 shows the results of the midterm exam. Both, the histogram as well as the Q-Q plot of the points achieved in the midterm exam suggest that those data are normally distributed with a mean centered about sufficient performance. The considerable number of easy questions in the midterm may help to explain the higher grades as well as the kick the students got through the mock exam.

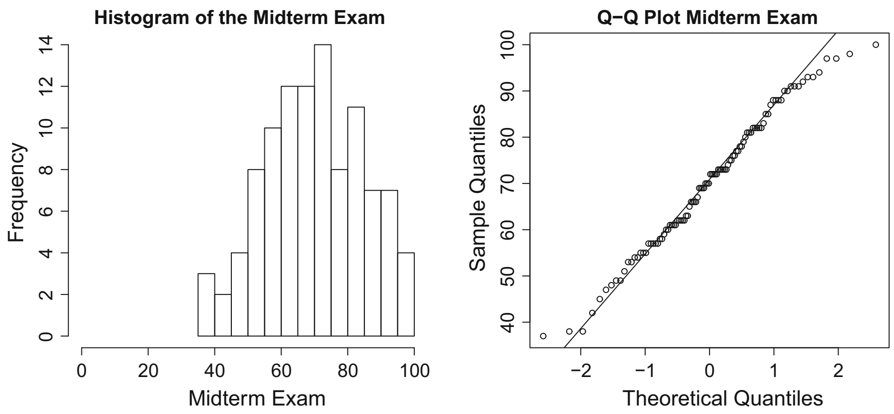

Fig. 9 Histogram and Q-Q-plot for the points achieved in the midterm exam

3 Correlation between the course methods and the students' performance

Since the successful application of Gauss in 1801 linear regression, i.e. the least square method, is a popular tool to fit empirical data to linear laws. Given empirical variables $X_1, X_2, \ldots$ one can construct, for instance, a linear model $Y = a + b_1 X_1 + \varepsilon$, a quadratic model $Y = a + b_1 X_1^2 + \varepsilon$, or a multi-linear model $Y = a + b_1 X_1 + b_2 X_2 + \cdots + \varepsilon$ with constant coefficients $a, b_1, b_2, \ldots$ that governs the empirical realization. The condition under which this method can be applied to gain such correlation insights is that the (residual) error ε, which measures the deviation of the empirical data from the linear model, is distributed normally (cf. Rupp and Brandmeier 2010). For an introducing text on correlation analysis see, e.g., Kleinbaum and Kupper (1978), pp. 71, or Sachs (2002), pp. 493.

Thus, the striking prerequisite for a regression analysis is that the (residual) error is normally distributed, otherwise an error measure that aims to identify the deviation from a mean value based on a least square approximation does not make much sense (unless weightings of the least square deviation are introduced). In the following, we give linear and non-linear regression approaches for certain combinations of our data and test the fulfilment of the normal distribution of the residual errors using Q-Q-plots.

Table 1 compares our data sets from the previous section with each other such that the test which was conducted before the other is taken as the hypothetical influence factor. The table also gives two statistical key indicators, the correlation coefficient of the respective setting as well as the R^2-value. The correlation coefficient between two variables X and Y, or more precisely the Pearson product-moment correlation coefficient between two (quadratically integrable random) variables, is a measure of the linear correlation (dependence) between the two variables. It is defined as the covariance of these two variables divided by the product of their standard deviations. The R^2-value is the square of the correlation coefficient, and often called the coefficient of determination. It estimates the fraction of the variance in Y that is explained by X in a simple linear regression.

Table 1 Key indicators of regression analyses performed with the test data that are suitable for such a statistical analysis

Comparison	Correlation	R^2	Figures
State of Knowledge 1 versus mock exam	0.466	0.208	10a
State of Knowledge 1 versus midterm exam	0.377	0.142	10b
Pre-learning versus Mumie homework	0.707	0.5	11
Pre-learning versus tutorial tests	0.398	0.158	–
Pre-learning versus mock exam	0.238	0.057	–
Pre-learning versus midterm exam	0.104	0.011	–
Mumie homework versus tutorial tests	0.598	0.358	12
Mumie homework versus mock exam	0.456	0.208	–
Mumie homework versus midterm exam	0.474	0.225	–
Lecture tests versus mock exam	0.597	0.356	13a
Lecture tests versus midterm exam	0.437	0.191	13b
Tutorial tests versus mock exam	0.593	0.352	14a
Tutorial tests versus midterm exam	0.521	0.272	14b
Mock exam versus midterm exam	0.726	0.527	15

It is good practice to discard too low correlation coefficients as "not significant" and categorise correlation coefficients between 0.4 and 0.5 as "moderately significant", coefficients between 0.5 and 0.7 as "significant" and those above 0.7 as "highly significant"—see, e.g., Sachs (2002), pp. 536, on confidence regions for correlation coefficients. Finally, for convenience, in Table 1 the labels of the figures are displayed that show the graphical outputs of these regression analysis.

Figure 10 shows the results of a correlation analysis for the linear impact of the factor State of Knowledge Test 1 (which resembles the pre-knowledge of our students) on the mock midterm exam (a) and on the midterm exam (b). With a correlation coefficient of 0.466 ($R^2 = 0.208$) subject to the mock and one of 0.377 ($R^2 = 0.142$) subject to the midterm one may conclude that pre-knowledge is useful but does not really help for the first test at university. As it should be, new knowledge has to be acquired during the course of the lecture.

Concerning the impact of pre-learning, we see that this measure really worked at the intended place and has the most significant correlation with homework assignments where a correlation factor of about 0.707 and a R^2-value of approximately 0.5 is displayed, see Fig. 11. This is as expected, as the pre-learning sets the mind for the students and mentally prepares them for the homework. Moreover, those who did not go for the pre-learning did not bother about the homework assignments either (or simply lacked internet access to work on these electronic exercises). Further studies have to reveal what keeps students busy and active during the course.

The remaining correlations of pre-learning with the tutorial tests, the mock exam and the midterm exam seem to be rather insignificant so that we do not discuss them here at the main part of the article, but provide their short regression analysis at the appendix.

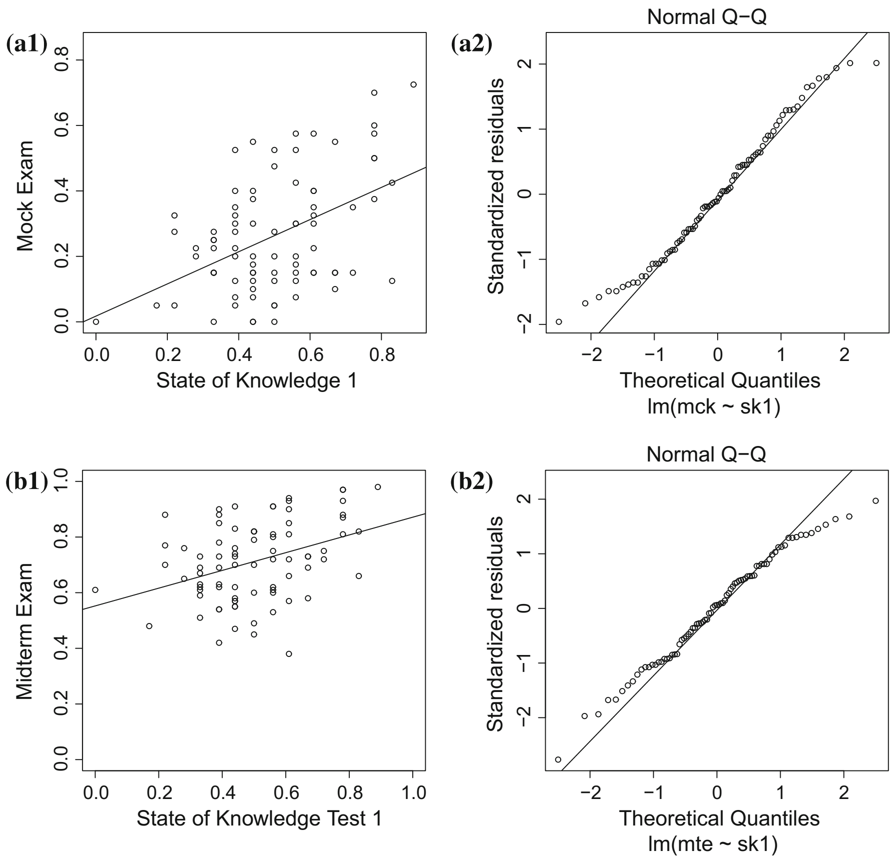

Fig. 10 Regression and Q-Q plots for **a** State of Knowledge Test 1 versus Mock Exam, and **b** State of Knowledge Test 1 versus Midterm Exam

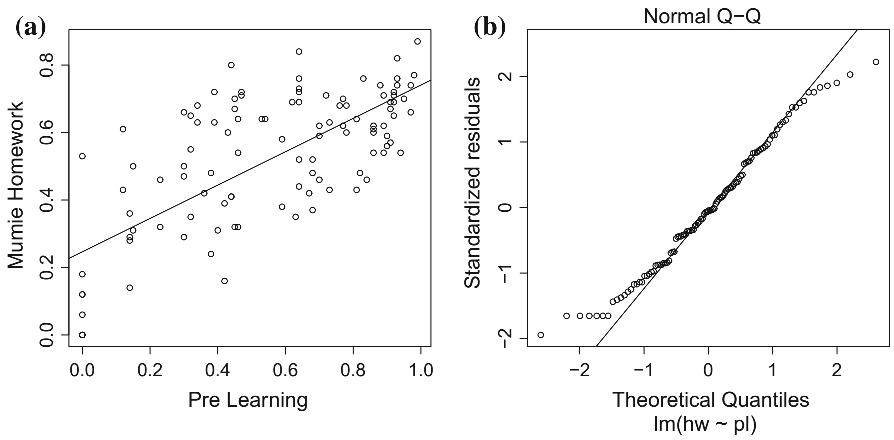

Fig. 11 Regression (**a**) and Q-Q plots (**b**) for “pre-learning versus Mumie Homework”

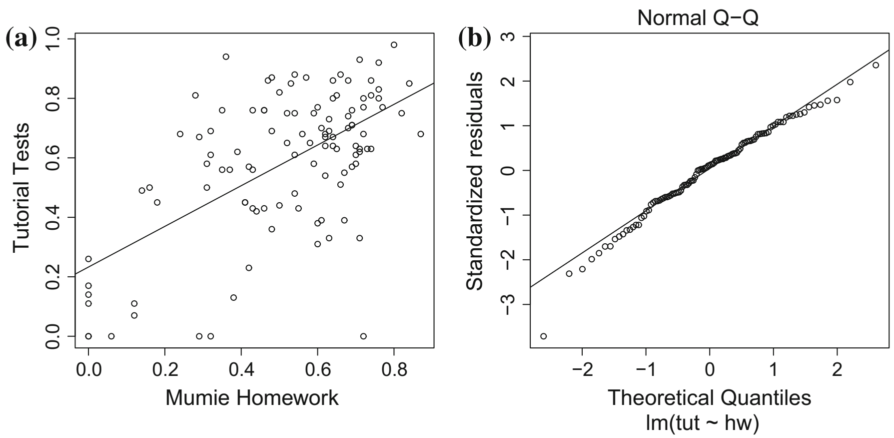

Fig. 12 Regression (**a**) and Q-Q plots (**b**) for Mumie Homework versus tutorial tests

Again and as intended the most significant influence of the homework assignments is with the tutorial tests (again we give the remaining regression analyses at the appendix). For homework assignments as a linear explanation factor of the results achieved at the tutorial tests we obtain a significant correlation coefficient of 0.598 ($R^2 = 0.358$), see Fig. 12. The failure of a higher correlation may be explained by the insights the students gained when they performed the homework and than re-thought it (from different angles and with solutions) during the tutorials. Thus, the factor of learning should, in some sense, be included in further studies to obtain better explanations for the performance of short tests based on homework assignments.

Next, Fig. 13 displays the results of such a regression analysis for the linear impact of the factor tutorial tests on the mock midterm exam (a) and on the midterm exam (b). As explained, the tutorial tests can be viewed as a direct measurement of the students' understanding of the course topics as they examine the learning process through the homework assignment and their solutions. We see a significant linear correlation expressed by the correlation coefficients of 0.593 ($R^2 = 0.352$) with respect to the mock and of 0.521 ($R^2 = 0.272$) subject to the midterm exam. These results seem to support the hypothesis that continuous learning and staying tuned to the topics of the lecture pays off to a large degree.

When inspecting the data clouds one may wonder if a non-linear fit may lead to better results. Indeed, our analysis suggests a slightly better fit of the mock results Y_{mock} with an exponential curve over the results from the tutorial tests T, see Fig. 11c1:

$$Y_{mock} \sim a \cdot \exp(b \cdot T).$$

where $a \approx 0.03602$ and $b \approx 2.85119$. For this exponential fit the residual sum of squares is 2.226 compared to 2.472389 for the linear model. Analogously, a quadratic model seems to fit the results from the tutorial tests better with the midterm results Y_{mid}, see Fig. 13c2:

$$Y_{mid} \sim a \cdot T^2 + b \cdot T + c .$$

where $a \approx 79.84$, $b \approx -45.63$ and $c \approx 64.39$. Here, the residual sum of squares reads as 1.350014 for the quadratic model compared to 1.612961 for the linear model. Thus, it seems that we can speak of an even non-linear correlation induced from the tutorial tests on the results of the mock and the mid term exam.

Interestingly enough, the lecture tests display a behaviour very similar to the tutorial tests, cf. Fig. 14. Taking them as a linear factor for the success at the mock exam leads to a correlation coefficient of 0.597 ($R^2 = 0.356$) and as a factor for the success at the midterm, respectively, to a correlation coefficient of 0.437 ($R^2 = 0.191$). One may argue that the lecture test resemble the capability of the students to grasp new ideas in the lecture and therefore assume that those who are fast in understanding are good at exams also. The somewhat low correlation for this very appealing hypothesis questions it, though. There seems to be more to the subject of learning than just understanding something quickly and applying it to easy examples while having it in one's short-time memory.

Again, the data clouds suggest to try to leave the classical paths of linear regression. Figure 14c shows the non-linear regression plots depending on the lecture tests, which are again better fitting than the linear regression. Again we used a exponential regression for the mock exam

$$Y_{mock} \sim a \cdot \exp\left(b \cdot L\right) + c .$$

where $a \approx 0.05751$, $b \approx 2.57971$ and $c \approx -0.03541$. For this exponential fit the residual sum of squares is 2.294 compared to 2.453855 for the linear model. Analogously, a quadratic model was better fitting for the midterm exam Y_{mid}, see Fig. 13c2:

$$Y_{mid} \sim a \cdot L^2 + b \cdot L + c .$$

where $a \approx 0.8779$, $b \approx -0.4944$ and $c \approx 0.6655$. Using the quadratic regression, we achieved a residual sum of squares of 1.582254 compared to 1.791039 for the linear model.

The main finding of our analysis is the strong correlation between the mock and the midterm exam with a correlation coefficient of 0.726 ($R^2 = 0.527$), see Fig. 15. This means that there is almost a linear dependency between the mock results and those of the midterm, and about half of the variation we recognize in the midterm results seems to be due to this linear relation to the mock results. This supports significantly the hypothesis that effective learning during the short period between the mock and the midterm seems to be rather impossible although it helps in special cases. It seems that the continuity of learning and the early preparation for an exam are indispensable factors for the success of students.

Additional information to understand some features of this data cloud are that all students achieved over 27 points (out of 100) at the midterm and that all eight students that scored with zero points in the mock were invited to an oral exam. This seems to have had a great influence on their learning behaviour and gave them a great performance boost. It is worth to note that after we distributed the midterm grades

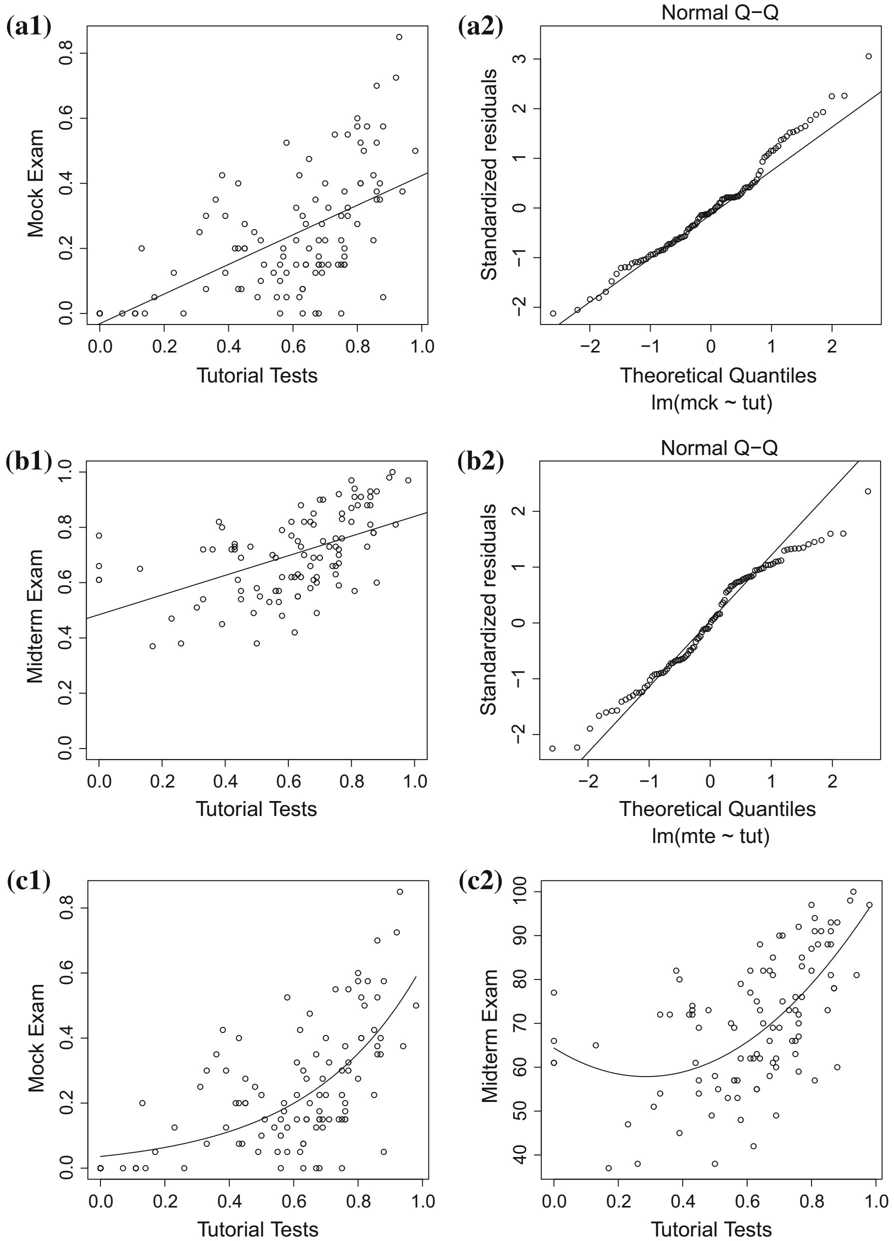

Fig. 13 Regression and Q-Q plots for **a** tutorial tests versus mock exam, and **b** tutorial tests versus midterm exam. The *third row* of graphs shows an exponential fit for the data “tutorial tests versus mock exam” in *c1* and a exponential fit for the data “tutorial tests versus mock exam” in *c2* as discussed in the text

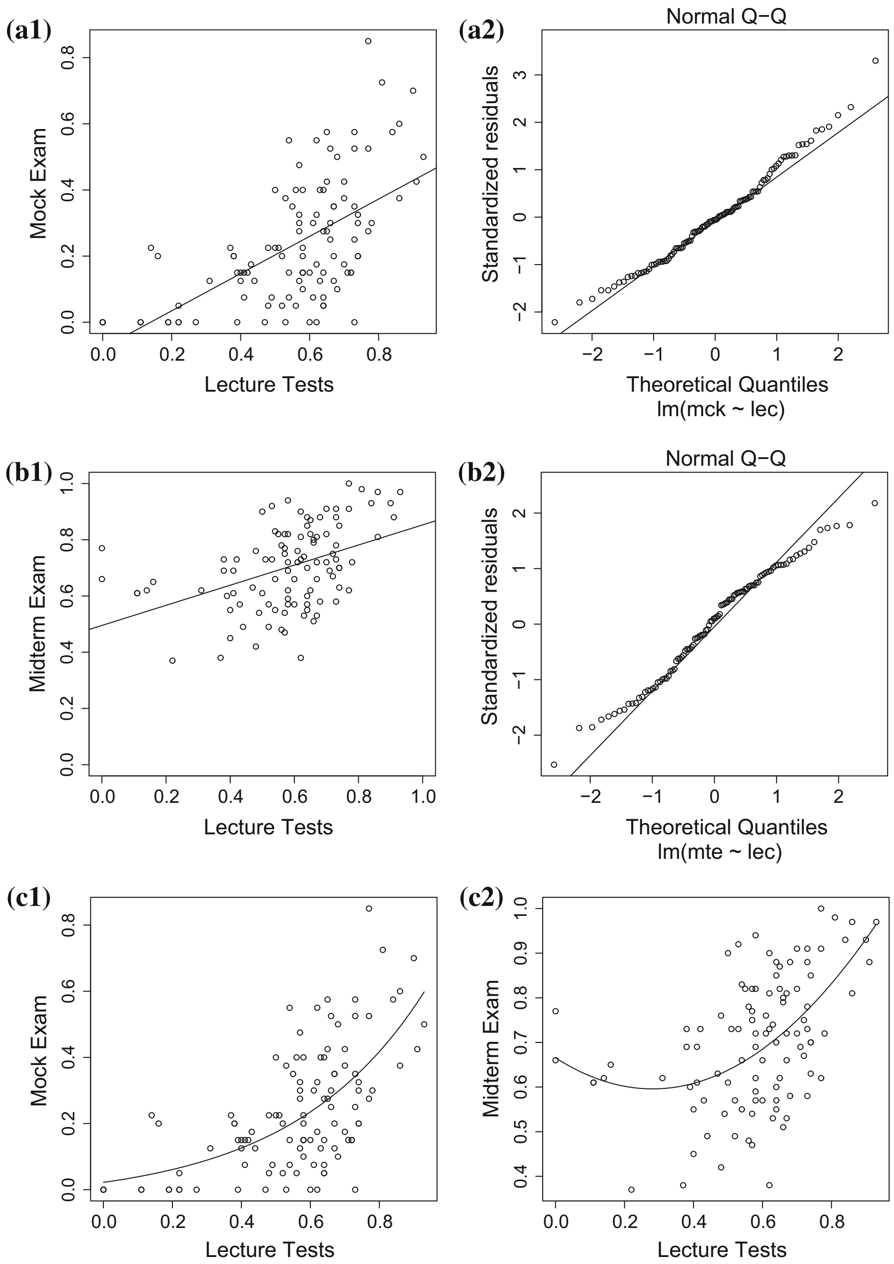

Fig. 14 Regression and Q-Q plots for **a** Lecture Tests versus mock exam, and **b** lecture tests VS midterm exam. The *third row* of graphs shows an exponential fit for the data “lecture tests versus mock exam” in *c1* and a exponential fit for the data “lecture tests versus mock exam” in *c2* as discussed in the text

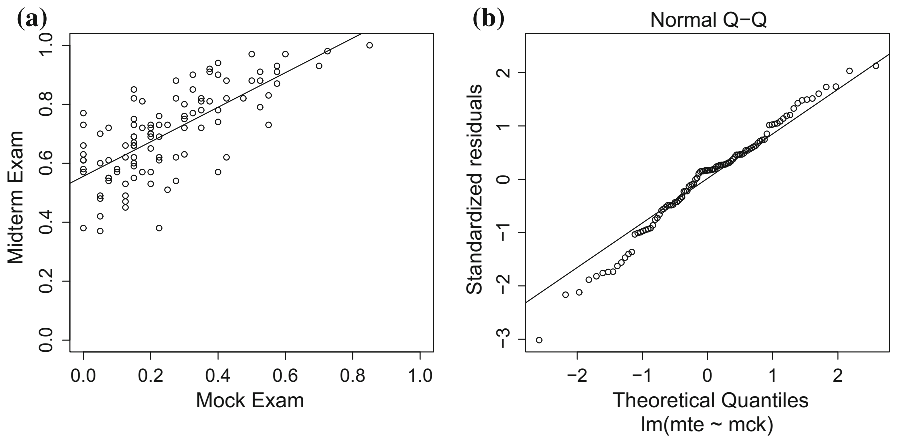

Fig. 15 Regression (**a**) and Q-Q plots (**b**) for “Mock Exam versus Midterm Exam”

some other students envied those who had the oral and asked if they could have an oral exam also next time.

4 How did Kaizen teaching affect the students?

Figure 16 illustrates the test scores of three model students from the first test after the State of Knowledge Test 1 to the midterm exam (as the last data point in the graph). It is noticeable that the student who achieved nearly 90 points in the midterm exam (dashed line) which is equivalent to an A has performed way above average over the whole semester. The average student (solid line) missed a few tests, but was still able to achieve sufficient scores in the other ones. In the midterm exam he/she received a $B-$. The 3rd student (dotted line) started the semester pretty strong, but seemed to have lost track later in the lecture. This combined with several missed tests led to an F in the exam.

Moreover, we witness a performance drop of all students around the 10th test and it seems to be there when the separation of the three students actually took place. The best of them was able to get back to high percentages quickly and further defeats were adjusted. The 2nd student began to struggle with the material and oscillated in performance. The 3rd student more or less lost connection to the lecture and could not find the way back to the performance shown before this decline. Figure 16 thus shows clearly that once you are lost in mathematics and you do not recoup, then you are lost for the rest of the semester. Moreover, as shown for the 2nd student, facing the struggles and challenges of the lecture continuously pays-off at the end.

In that respect, it is fair to conclude that without the continuous feedback and test loops installed (as it is the case at a traditional course that just provides a final exam) the 2nd student may have also be lost very soon. Moreover, when inspecting the initial phase of the students’ performance in Fig. 16, we see that both the 2nd and 3rd student

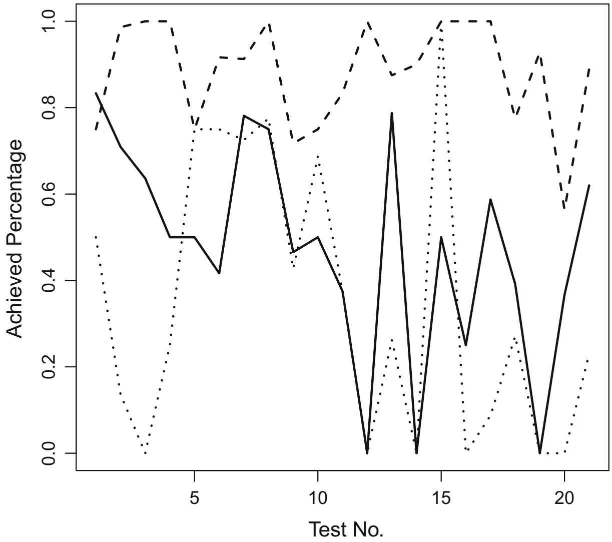

Fig. 16 Test scores of three model students from the beginning of the semester until the midterm exam

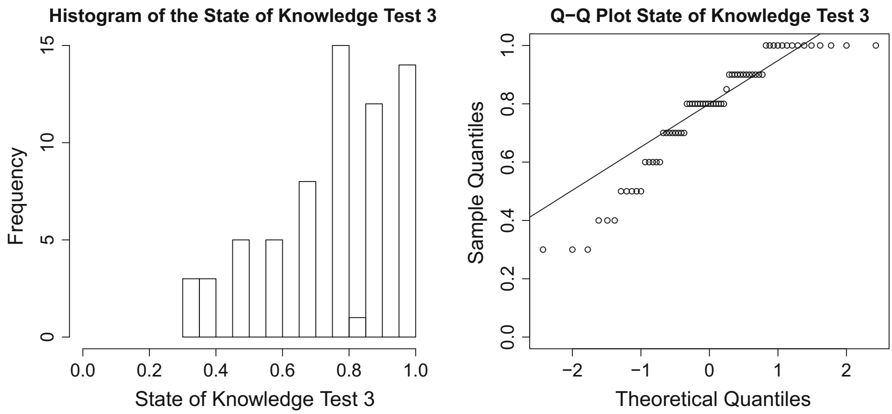

Fig. 17 Histogram and Q-Q-plot for the points achieved in the State of Knowledge Test 3

were starting with negative slope, but at some point seemed to realize this peculiar trend and then pushed themselves up again.

The State of Knowledge Test 1 is our baseline for measuring the effects of teaching. In this section, we inquire which factors influence the performance in the State of Knowledge Tests 2 and 3 and if they are correlated with our efforts.

Figure 17 displays the results of the State of Knowledge test 3 conducted in the 11th week of the lecture period. Here, it is striking that like in the State of Knowledge Test 2 the results are shifted significantly to better grades. In addition, none of the students scored $<30\,\%$ this time.

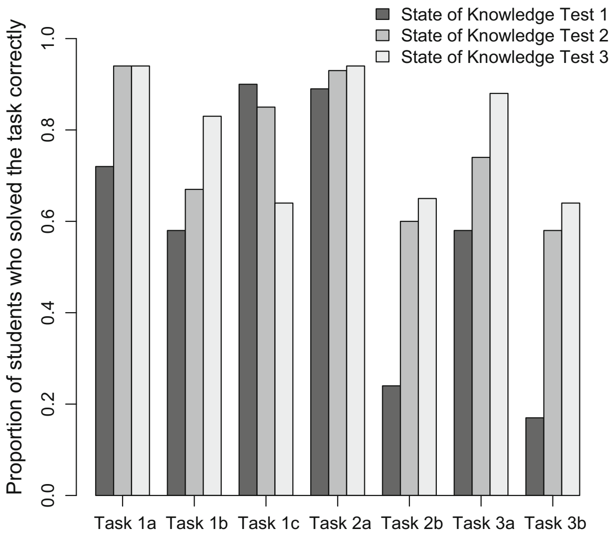

Fig. 18 The achieved scores in the three state of knowledge tests in comparison

The first State of Knowledge Test consisted of 18 subtasks to check the students' skills in basic mathematical operations. Seven of theses were repeated two times with a slightly increasing difficulty throughout the semester. Task 1 was about cancelling fractions: Each time the first subtask (task 1a) was a basic fraction of real numbers below 100. The second subtask (task 1b) was a fraction of exponentiations of numbers below 5. In task 1c, the numerator was a quadratic expression and the denominator was an irreducible factor such that the students had to use a binomial formula to cancel the fraction.

The second task dealt with absolute values. In part "a", the absolute value of a difference of real numbers had to be calculated and in the second subtask the students had to solve an equation of a small real number and an absolute value of a difference of x and a real number. The third task checked whether the students know basic values of cosine and sine at points like 0 and multiples of $\pi/2$. As can be seen in Fig. 18, the performance of the class improved from each test to the next in every model task, except in task 1c a change for the worse is visible. This may be explained by the kind of tasks in the 2nd and 3rd State of Knowledge Test. The binomial fractions from the first test were complicated with a factor such that the pattern learned before university could not be applied as easy as in the initial test anymore.

5 Résumé

Based on Kolb's insight of a specific reflective learning cycle students have to go through in order to actually learn and understand (cf. Kolb 1984), we outlined and applied the teaching method of Kaizen teaching/ learning where the lecturer takes the position of continuously guiding and reinforcing this cycle throughout the complete course (cf. Heim et al. 2014). Despite the complete outline of the course which may serve for further callibrations of the Kaizen paradigm, a comprehensive case study with the about 100 participants of the course was conducted to distill the strength and weakness of the Kaizen framework in a real course context.

During our discussion it became clear that interesting correlations between the Kaizen approach and the students' performance seem to exist. In particular, the pre-knowledge of the students as tested in one of the very first lectures is not very significant for the students final scoring. This, indicates that the majority of students were able to learn significantly (otherwise only the best would have scored reasonably and thus a correlation between pre-knowledge and the final results should have been clearly visible). Moreover, continuous tests and specifically those tests that were prior had a significant impact on the results of the test conducted next. As we stated in the introduction, this situation is comparable to that of an athlete and may allow a Markovian estimator for students: todays performance is a good predictor for tomorrows success and if one is falling behind extra effort has to be taken immediately.

Certainly this study on, say, university operations research covers just the starting point and it would be interesting to view the performance of students over even longer periods. Indeed surveys taken during the summer term 2014 with the same students showed the validity of the concept and in particular the relevance of continuous learning. However, the correlations between the mock and the real exam surprisingly did not turn out to be as strong as expected. As we saw in personal conversations, the students started to prepare themselves very carefully for the coming mock exams and not only for the real exams, and an additional (externally imposed) policy that the best of the mock and the exam would count for the final grade also lead to a weakening of the correlation.

References

Ali AJ, Camp RC (1995) Teaching management in the Arab world: confronting illusions. Int J Educ Manag 9(2):10–17

Bostrm L (2011) Students' learning styles compared with their teachers' learning styles in secondary schools. Inst Learn Styles J 1:17–38

Chambers J, Cleveland W, Kleiner B, Tukey P (1983) Graphical methods for data analysis, Wadsworth

Coffield F, Ecclestone K, Hall E, Moseley D (2004) Learning styles and pedagogy. A systematic and critical review. Learn Skills Res Cent. www.lsrc.ac.uk

Dede S (2009) The teacher's educational leadership roles according to Kolb's theory of learning. Humanity Soc Sci J 4(2):153–163

Dunn R, Dunn K (1998) The complete guide to the learning styles inservice system, Needham heights. Allyn and Bacon, MA

Heim B, Rupp F, Viet N, Stockhausen Pv, Gallenkämper J, Kreuzer J (2014) Driving student-centred calculus: results of a comprehensive case study for Kaizen-learning in the Sultanate of Oman. Int J Math Educ Sci Technol 46(3):354–369

Honey P, Mumford A (1986) A manual of learning styles. Peter Honey, Maidenhead

Howard RA, Carver CA, Lane WD (1996) Felder's learning styles, Bloom's taxonomy, and the Kolb learning cycle: tying it all together in the cs2 course, SIGCSE '95 2/96, Philadelphia, PA, pp 227–231

International Association for the Evaluation of Educational Achievement (IEA) (2011) TIMSS international results in mathematics, chap 2, TIMSS & PIRLS Lynch school of Education, Boston College

Kleinbaum DG, Kupper LL (1978) Applied regression analysis and other multivariable methods. Duxbury Press, Belmont

Kolb D (1984) Experiential Learning. Prentice Hall, Englewood Cliffs, NJ

Lewin K (1951) Field theory and learning. In: Cartwright D (ed) Field theory in social science: selected theoretical papers. Social Science Paperbacks, London

Mandelbrot BB (1983) The fractal geometry of nature. W. H. Freedman, San Francisco

Opera J (2010) Multiethnic teaching methods in higher education: perception, practice, and benefits. J Multicult Educ 6(3):34–62

Rodriguez AL, Fahara MF (2010) Online teaching styles: a study in distance education. Int J Univ Teach Fac Dev 1(2):1–14

Rupp F, Brandmeier R (2010) Benchmarking procurement functions: causes for superior performance. Benchmarking Int J 17(1):5–26

Sachs L (2002) Angewandte statsitik—anwendung statistischer methoden, 11th edn. Springer, Berlin, New York, Heidelberg

Sidani YM, Thornberry J (2009) The current Arab work ethic: antecedents, implications and potential remedies. J Bus Ethics 91(1):35–49

Svinicki MD, Dixon NM (1987) The Kolb model modified for classroom activities. Coll Teach 35(4):141–146

Thomas R (2012) Maths scores in Oman among weakest in the world: OECD, Muscat Daily, 1 April, Muscat