Identifying Consistent Variables in a Heterogeneous Data Set: Evaluation of a Web-Based Pre-Course in Mathematics

Katja Derr
Baden-Wuerttemberg Cooperative State University Mannheim, Germany
katja.derr@dhbw-mannheim.de

Abstract: E-learning has made course evaluation easier in many ways, as a multitude of learner data can be collected and related to student performance. At the same time, open learning environments can be a difficult field for evaluation, with a large variance in participants' knowledge level, learner behaviour, and commitment. In this study the effectiveness of a mathematics pre-course administered to four cohorts of prospective students at a technical faculty in Germany was evaluated. Deficits in basic mathematics knowledge are considered one risk factor regarding graduation in STEM-related subjects, thus the overall goal was to investigate if the pre-course enabled "at risk" students to improve their starting position. A data analysis was performed, relating students' preconditions when entering university, their attitude towards mathematics, and their use of learning strategies with further study success. The strongest determinant of first year performance were results in a diagnostic pretest, confirming both the importance of basic mathematics knowledge for academic achievement in engineering and the reliability of the chosen pre-posttest design. Other outcomes were quite unexpected and demanded deeper analyses. Students who had participated in additional face-to-face courses, for example, showed less learning gains than students who had participated in an e-tutoring version. It also could be observed that meta-cognitive variables failed to explain successful course participation. Reasons for these outcomes are discussed, suggesting reliability threats and interactions between students' preconditions and their learner behaviour. A significant and unmoderated impact on students' learning gains in the pre-course was found for the number of online test attempts, making this variable a reliable indicator of student engagement. The evaluations show that open learning designs with heterogeneous learner groups can deliver meaningful information, provided that limitations are considered and that external references, like academic grades, are available in order to establish consistency.

Keywords: learning analytics, pre-course, mathematics, formative e-assessment, STEM

1. Introduction

The educational backgrounds of students entering university are increasingly diverse, leading to a growing demand for preparatory and bridging courses – not only, but particularly in mathematics (Parker, 2005; Croft*, et al.*, 2009; Faulkner*, et al.*, 2014). A growing number of undergraduates lack basic mathematical skills and are not adequately prepared for the demands of a STEM (Science, Technology, Engineering, Mathematics) degree programme, an issue addressed since the 1990ies as the "mathematics problem" (Howson*, et al.*, 1995). Different reasons for the mathematics problem have been suggested, from abridged school curricula (Lawson, 2000) to a general increase in transfers to tertiary education (HEFCE, 2013) to a higher rate of students from non-traditional backgrounds (Faulkner*, et al.*, 2014). Today, nearly all technical faculties in Germany provide pre-courses in mathematics.

When addressing diverse groups of learners the implementation of web-based content may be beneficial; students are free to pace their learning and the course can be accessed by participants who not (yet) live near the campus. With many learner data stored online, the evaluation of these courses has become much easier, and learning analytics seem to offer countless possibilities for statistical analyses. But not all collected data may deliver meaningful results. In distance education, drop-out rates tend to be higher (Ashby*, et al.*, 2011), particularly in open access courses (Pappano, 2012) while answer rates are often lower (Cook*, et al.*, 2000; Fan and Yan, 2010). Web-based university pre-courses thus can be a difficult field for evaluation: access is free for all prospective students but only a section of each cohort participates. In this group, commitment can be very diverse and students often withdraw without giving feedback (Smith and Ferguson, 2005; Street, 2010; Gasiewski*, et al.*, 2012). Technical barriers and data privacy policies may also prohibit a connection between pre-university performance and further academic achievement.

Reference this paper as: Derr K, "Identifying Consistent Variables in a Heterogeneous Data Set: Evaluation of a Web-Based Pre-Course in Mathematics" *The Electronic Journal of e-Learning Volume 15 Issue 1 2017*, (pp82-93) available online at www.ejel.org

In this article different aspects of the evaluation of a web-based pre-course in mathematics are reported, with a special focus on the reliability and consistency of the collected data. The pre-course was not mandatory but students were encouraged to take the initial diagnostic self-test in order to identify knowledge gaps. These test results were also considered one important factor in the overall data model. The diagnostic feedback advised students to close existing gaps via self-study or by participating in additional face-to-face or e-tutoring courses. Reliability issues had to be considered regarding the open design and the resulting non-randomized groups of learners. It also had to be considered that some students did not participate at all; it was interesting to compare this group's academic achievement with the "experimental" group but, again, students' preconditions and interactions with other variables had to be taken into account.

In order to differentiate between successful and less successful pre-course participation a pre-posttest design was administered and had to be evaluated regarding its consistency. Finally, the informative value of different sets of metacognitive variables, from attitudes towards the subject to the use of learning strategies to measures of student engagement, was investigated regarding their consistency, their reliability, and their potential to predict learning gains in the pre-course. Analyses were based on data collected from four cohorts (2011-2014) enrolled at Baden-Wuerttemberg Cooperative State University Mannheim. Anonymised examination results from the first and final year of the degree programme were added as dependent variable in a multiple regression model. A significant relation to this measure of academic achievement also served as an indicator for each independent variable's reliability.

1.1 Approach

In the academic field, there has been a growing interest in "educational data mining" (Romero and Ventura, 2010). Based on predictive models, students at risk to fail a course can be identified at an early stage and interventions suggested (Campbell and Oblinger, 2007; Corrigan*, et al.*, 2015). Prior achievement, for example, measured by secondary school GPA (grade point average), has been found a valid predictor of tertiary GPA and of student retention. Students with a high level of domain-related prior knowledge will have it easier to acquire new knowledge, thus in technical degree programmes mathematics grades or mathematics placement test scores have repeatedly been found of particular importance for later study success (Budny*, et al.*, 1998; Zhang*, et al.*, 2004; Warwick, 2007; Ehrenberg, 2010; Faulkner*, et al.*, 2010; Kokkelenberg and Sinha, 2010).

A first approach to evaluate the effectiveness of the mathematics pre-course programme therefore was to confirm these relations with the collected data. It was hypothesized that students with good secondary school grades and a high level of prior knowledge in mathematics would show a higher level of academic achievement in engineering. In this basic model the impact of personal and demographic variables (age, gender, federal state) was analysed, as well, with the overall goal to identify students "at risk" to perform poorly, or to withdraw from the degree programme.

In a second step students' learning gains in the pre-course were to be measured. Thus a pre- and a posttest in mathematics was developed and administered to pre-course participants. Both tests were designed to be equally difficult, but consisted of different items as suggested for single group pre-posttest designs (Kane, 2013). The gain score, or difference between posttest and pretest results, thus could be interpreted as a measure of change in relation to a student's pretest result. It was expected that participation in the pre-course would positively affect gain scores of students who had showed poor pretest performance.

It then was investigated which factors most contributed to this gain score, with a focus on the "at risk" group. The impact of different pre-course elements and their combinations – self-study, e-tutoring, face-to-face – was one major interest. Considering the role of affective and metacognitive variables in the learning process, the influence of scales addressing these variables on learning gains of the "at risk" group was analysed, as well (Robbins*, et al.*, 2004; Richardson*, et al.*, 2012). For STEM subjects, Ackerman*, et al.* (2013) suggested a multiple regression model including cognitive and meta-cognitive variables. The authors reported a strong influence of mathematics placement tests (isolated R^2 = .21), but they also stressed the importance of students' self-concepts in mathematics (their self-confidence and attitudes towards the subject) and their ability to master and organize learning. It therefore was expected that positive attitudes towards the subject as well as an efficient use of learning strategies and a high level of student engagement would be correlated with learning gains. Finally, the impact of the collected pre-course variables on first year achievement was to be evaluated in order to confirm, or disprove, the effectiveness of the pre-course design.

2. Data collection and tool development

Data from five engineering courses (mechanical engineering, mechatronics, computer science, electrical engineering, and industrial engineering) were analysed and evaluated. In a multiphase research design (Creswell and Plano Clark, 2011; Richey and Klein, 2005) repeated evaluations of test results, group interviews, questionnaire data and statistical information were used to revise and successively improve the programme. Throughout the study the learning management system (Open Source LMS Moodle) was used to administer, evaluate, and optimize the different quantitative tools. The finally enacted modular design consisted of an e-learning environment covering the secondary school curriculum in mathematics, initiated and completed by a pre- and a posttest, plus supplementary face-to-face and e-tutoring courses. The first two pre-course evaluations had shown that students' learning preferences were quite diverse. While many students wanted to learn independently (and alone), others claimed to need additional support and missed face-to-face interaction. Considering the differences in participants' starting positions and their personal situation in the phase between school and university it was decided to modularize the programme, with different learning scenarios open for self-selection (Jackson and Johnson, 2013). Students now could sign up for weeklong on-campus courses or for an e-tutoring programme that lasted one month. All students had access to the same web-based learning material, but in the e-tutoring course the learning process was structured and monitored by mathematics lecturers. Every week students uploaded a completed exercise sheet and were encouraged to discuss mathematical problems with peers and e-tutors.

2.1 Educational background

From the university's administration students' secondary school GPA (leaving certificate) was collected. Other school related variables, like gap between secondary and tertiary education, type of secondary school, or the area / federal state where school was attended were collected from a web-based questionnaire.

2.2 Prior knowledge level in mathematics (pre-posttest design)

Domain related prior knowledge was measured by a diagnostic test. As placement tests in mathematics are not mandatory at German universities, no standardized items were available for the development of the pre-posttest design. Two item sets were developed, covering the secondary school syllabus and structured alongside the ten e-learning modules, and underwent a two-year revision process. The first cohort's test results served as a database for classical and probabilistic item analyses. An Item Response Theory (IRT) approach was chosen to model each item's difficulty level and identify extreme outliers (Hambleton and Swaminathan, 2010). In combination with traditional measures, like mean scores and discrimination index, the Rasch model estimates delivered information on each item's quality and contribution to the test. Items that did not fit the model were revised or replaced, and the analysis was repeated in the following year. After two revisions both tests delivered consistent results (Cronbach's α pretest = .91 and posttest = .85) and no more outlying items; since 2013 the pre-posttest design has remained unchanged. Pre-posttest similarity was established by comparing pretest results with posttest results of a control-group that had neither participated in the pretest nor in the pre-course. With a consistent pre-posttest design, learning gains in the pre-course could be measured by pre-posttest difference.

2.3 Affective and meta-cognitive aspects of learning

Two Likert scales were administered, one addressing students' attitudes towards mathematics and mathematics learning, and another referring to their use of learning strategies. For the mathematics attitude scale an item set developed for the "Trends in International Mathematics and Science Study TIMSS" was employed (Kadijevich, 2006; Mullis*, et al.*, 2012). In this inventory, students' liking of the subject (for example "I am interested in mathematics") and their self-confidence in learning mathematics are addressed (for example "I learn things quickly in mathematics"). For the learning strategies scale subscales of the LIST inventory were used (Schiefele and Wild, 1994), a German adaptation of the "Motivated Strategies for Learning Questionnaire MSLQ" (Pintrich*, et al.*, 1991). MSLQ is a well-established item battery designed to address students' use of learning strategies (for example "I have a regular place set aside for studying" from the subscale "Resource management strategies").

Both scales underwent a revision process: based on students' answers an exploratory analysis including inter-item and item-total correlations plus factor analysis was performed on the first version (27 items) and again after outlying items had been replaced or removed (21 items). The final factor analysis produced one common factor describing "attitude towards mathematics", covering liking mathematics (for example "I am interested in mathematics"; "I enjoy learning mathematics") and self-confidence in learning mathematics (for example "I learn things quickly in mathematics"; "Mathematics is harder for me than any other subject"). Internal consistency of this subscale was acceptable, with Cronbach's α = .82. The scale describing the use of learning strategies was less consistent, Cronbach's α was only .71, and only four items loaded on a common factor that could be described with "mastering the self-study process": these items were related to self-organization and time management (for example "I usually managed to keep to my schedule").

2.4 Effort

Students' effort and engagement in the learning process were represented by variables from two different sources: self-reported measures of effort were collected from the evaluation questionnaire (number of learning modules and invested learning time per week) and the LMS log files provided the number of page views per student and the number of completed test attempts. At the end of each learning module students were advised to monitor their progress by taking a self-test, consisting of 10 to 15 randomized items. The data analysis was to reveal which variable showed the strongest explanatory power in relation to learning gains in the pre-course (Macfadyen and Dawson, 2010).

2.5 Academic achievement: first and final year

Overall academic achievement was measured by cumulated GPA at the end of the degree programme; a second measure was the dichotomous variable "graduation / withdrawal". For the analysis of students' first year performance a variable was needed that strongly correlated with these measures. Data from the first participating cohort that started the degree programme in 2011 and graduated in 2014 (n = 660) were used for this analysis. Nearly all examinations significantly correlated with GPA, but Mathematics I was the first year exam with the strongest correlation (r = .62; n = 660; p < .01). The simple linear regression model using Mathematics I as a predictor of GPA was significant and explained 38% of the variance in GPA (n = 660; R^2 = .38; R^2 adj. = .37; F (14, 450) = 405.77; p < .01). Mathematics I was also significantly related to the dichotomous variable graduation / withdrawal: For every increase in grades in Mathematics I grades the odds of completing the degree programme were 14 times greater than the odds of withdrawing (p < .01). Thus the hypothesized relevance of mathematics performance for general academic achievement in engineering was confirmed and Mathematics I grades identified as an early indicator of study success in engineering.

2.6 Dataset final analysis

The data collected for the main analysis were based on the cohort that started the degree programme in 2014 and had access to the revised e-learning programme, consisting of a pretest in mathematics, an e-learning environment open for self-study, and optional e-tutoring and face-to-face courses. The following groups of variables were collected from the LMS:

Table 1: Collected variables

	# items
Preconditions when entering tertiary education (traditional predictors)	
Prior knowledge level in mathematics (pretest mean score)	77
Demographic and personal variables (final school grades, type of school, mathematics grades, gender, age, gap between school and university, federal state)	7
Pre-course participation and variables related to effective learning	
Type of course attended (self-study, plus e-tutoring or face-to-face)	1
Attitude scale	14
Use of learning strategies	7
Level of engagement	5
First year achievement	
Results Mathematics I exam	1

Of 722 first year students, 603 participated in the pretest and the majority answered the associated questionnaire. The diagnostic test feedback informed students about their test results per mathematical field and advised them to close existing knowledge gaps with the related learning material. (Note: the design of the diagnostic feedback was significantly improved by a Moodle plug-in developed by Dreier (2014) for his bachelor thesis in computer science). 42% of all pretest participants decided to enrol in either additional programme: 119 students visited a face-to-face course and 132 opted for the e-tutoring version. Attrition rate in the e-tutoring course was 14%, so that 113 students completed this course with a certificate. A group of 28 students attended both additional programmes (see table 2). 105 first year students did not participate in the pretest nor the pre-course, but nearly all first year students (n = 708; 98%) participated in the posttest that was taken at the university's computer labs during induction week. For the regression analysis, data from 613 students who took the first year examination Mathematics I six months later were available.

Table 2: Pre-course participation and first year students (2014 cohort)

	n
Pre-course participants	603
Questionnaire I: personal and attitude scales	593
Pretest	603
Self-study*	386
+ E-tutoring course	85
+ Face-to-face course	91
+ E-tutoring and face-to-face course	28
Questionnaire II: evaluation and learning strategies scales	200
Enrolled students	722
Posttest	708
Posttest only	105
First year performance (Mathematics I)	613

3. Results

3.1 Students' preconditions and academic achievement

Standard multiple regression analysis was employed to investigate the power of personal and demographic variables in predicting first year exam results in Mathematics I. In each of the single and multiple regression models, the two IVs final school grades and pretest results showed significant impact on the dependent variable Mathematics I. The type of secondary school attended was also found an important factor, suggesting significantly poorer MI performance for students from vocational schools, or with non-traditional backgrounds (Faulkner, *et al.*, 2014; van Soom and Donche, 2014). Mathematics grades at school were also related to Mathematics I, but this variable showed less powerful results in the multiple model than expected (Zhang, *et al.*, 2004; Ehrenberg, 2010; Faulkner, *et al.*, 2010). One reason might have been that the data, unlike final school grades, was based on self-reports.

Some interactions between variables were identified, for example between gender, age, and educational background. As female students are underrepresented in engineering courses (in this study the average rate was 12%) it is very difficult to separate the influence of gender on academic achievement in engineering (Ackerman, *et al.*, 2013). In the literature, investigations of the relation between gender and performance in mathematics or science have led to mixed results (Zhang, *et al.*, 2004; Xie and Shauman, 2005; Johnson and Kuennen, 2006; Richardson, *et al.*, 2012; Faulkner, *et al.*, 2014). In this study, female students and younger students, on average, had higher school achievement levels, leading to sometimes contradictory effects in single and multiple regression analyses. A descriptive analysis showed that women were often younger and more often had traditional educational backgrounds and very good secondary school grades than male students. After controlling for these interactions gender was unrelated to achievement, as well as age, the length of the gap between secondary and tertiary education, and the federal state in which secondary school was attended.

The complete model accounted for 33% of variance in first year performance (n = 465; R^2 = .33; R^2 adj. = .31; F (13, 451) = 16.67; p < .01). Comparing the two most consistent IVs, final school grades and pretest results, the latter was found the strongest predictor of first year performance. After the removal of pretest results from the multiple model, R^2 decreased from .33 to .22. When the IV final school grades was removed from the model, R^2 decreased to .30. Note that the predictive quality of the diagnostic pretest was considerably improved throughout the four years of the study, with only 25% of variance explained in 2011. Pretest results were also found significantly related to final GPA in a multiple regression with one complete cohort (first year students of 2011), and both variables were significantly related to student withdrawal in a logistic regression.

These outcomes mirror the literature on academic achievement in engineering, with very stable relations between school performance, prior knowledge level, and success in MINT-related subjects (Budny, *et al.*, 1998; Zhang, *et al.*, 2004; Kokkelenberg and Sinha, 2010; Faulkner, *et al.*, 2014; van Soom and Donche, 2014). Students with poor values in any of the achievement-related IVs, but particularly those with a poor pretest result, had a higher risk to perform poorly in Mathematics I.

3.2 Learning gains in the pre-course

After having verified the importance of prior knowledge in mathematics for study success in engineering factors influencing learning gains in the pre-course were analysed. In 2014, 603 students participated in both tests and achieved an average pretest score of 49.7 (*SD* = 15.9) and an average posttest score of 55.2 (*SD* = 17.5). By comparison, students who had not participated in the pretest achieved a posttest mean score of 47.3 (*SD* = 18.2). In both groups a large variance in test results could be observed. The average gain score (posttest minus pretest) for the 2014 cohort was 5.4 (median = 5.1), with a maximum value of 61.8 and a minimum of -37.5. Students with poor pretest results (mean score < 50), thus considered the "at risk" group, had an average gain score of 8.3 (median = 7.3; max. = 61.8; min. = -23.4).

3.2.1 Course type

The highest learning gains were achieved by students who had participated in both course types, e-tutoring and face-to-face with an average gain score of 9.1 (n = 28, pretest mean score = 44.2). The remaining 85 e-tutoring participants had a gain score of 6.7, in combination with a pretest result of 47.5. The poorest gains were achieved by students who had attended the face-to-face course, only (gain score = 3.5). This group also had the poorest pretest results, with a mean score of 43.7. Pre-and posttest results per course type are depicted in Table 3 and Figure 1. The 19 students who had withdrawn from the e-tutoring course failed to improve, as well (note that this group includes 6 students who later on attended a face-to-face course).

Table 3: Pre- and posttest results 2014: complete dataset in comparison to chosen pre-course type (*n* = 603) (*6 students participated in a face-to-face course, as well)

	participants both tests	self-study	face-to-face	e-tutoring	face-to-face + e-tutoring	e-tutoring withdrawal*
n	603	386	91	85	28	19
pretest (%)	49.7	52.4	43.6	47.5	44.2	38.0
posttest (%)	55.2	57.9	47.2	54.2	53.3	39.9
gain score	5.5	5.5	3.5	6.7	9.1	1.9

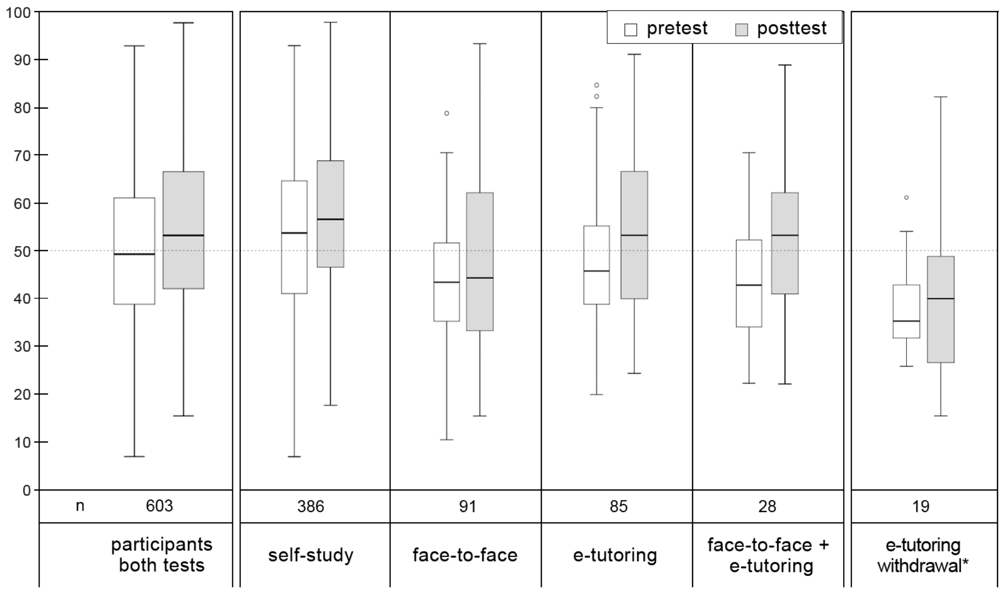

Figure 1: Pre- and posttest results 2014: complete dataset in comparison to chosen pre-course type (*n* = 603) (*6 students participated in a face-to-face course, as well)

Regarding the modular course programme, the strongest effects could be observed for the e-tutoring course, a one-month self-study programme supervised by mathematics lecturers. However, the significance of these results was limited, as students were not randomly assigned but had self-enrolled into the different course types (e-tutoring, face-to-face, self-study, or neither). An analysis of students' educational backgrounds suggested that students with higher "risk"-level, e.g. poorer school performance, or having attended a vocational school, more often chose the weekly face-to-face courses. Furthermore, variance in gains was high for all groups, so that it could be assumed that other factors had an influence on successful pre-course participation.

3.2.2 Attitude and learning strategies

It had been hypothesized that mathematics attitude items would correlate with each other, which they did, thus replicating existing results that suggest relations between mathematics liking and mathematics self-confidence (Parsons*, et al.*, 2009). Significant relations with pretest results were also found for nearly all attitude items, so that the presumption that a positive attitude towards mathematics would be related with a higher level of prior knowledge could be verified (Mullis*, et al.*, 2012). A critical point were the often skewed distributions: participants more often expressed positive attitudes towards mathematics, or felt reluctant to express negative attitudes, leading to small case numbers. For example, only 13% (n = 77) of first year students were on the negative side of the statement "I enjoy learning mathematics" (strongly disagree: n = 15; disagree: n = 62), whereas 63% agreed (n = 277) or strongly agreed (n = 89).

Not-normal distributions were also observed for the learning strategies scale. Four items addressing a proficient use of learning strategies were significantly related to each other, and to pretest results, indicating that students able to manage their learning process had a higher level of prior knowledge in mathematics, as well. However, these relations were never linear, so that these variables only allowed to differentiate between students who "strongly agreed" to an item like "I usually managed to keep to my schedule" (n = 43; pretest mean scores = 58.6) and the rest of the sample. Correlations between the attitude and the learning strategies scale were rather weak, as well. With regard to these non-linear patterns analyses of variance were performed for each single item in relation to learning gains in the pre-course. In this process it was found that both scales, students' attitudes towards mathematics and their use of learning strategies, were related to prior knowledge level, but were more or less unrelated to the variable gain score (posttest minus pretest). Thus students with deficits in basic mathematics knowledge only rarely showed a strong positive attitude or high efficiency in their use of learning strategies, but if so this was unrelated to learning gains.

3.2.3 Effort

It was expected that students who invested a lot of time and effort into the pre-course would achieve a higher gain score (Ackerman*, et al.*, 2013). Four different measures of effort were available for this analysis. In the evaluation (n = 205), students had answered how many hours per week they had studied. A first analysis suggested that students with more study time per week had poorer pretest results and higher learning gains. It also could be observed that with an increase in number of reported learning modules learning gains increased, as well. ANOVA estimations for these two items, however, were not significant, and showed a high variance in each subgroup's gain scores.

Two further variables were collected from the LMS log files. The number of page views per learning module did not deliver significant results. According to the database query, 83% of the pre-course participants (n = 603) had visited at least one page (note that page views were counted per login, so that the same page was only counted once per login session). The highest number of page views was 1585 (out of 684), but the majority of cases had no more than 200 page views (median = 121). Only 19 students had a page view count above 1000. An ordinal version of this variable was used for ANOVA, grouped to "no views" (n = 101), "1-10 views" (n = 83), "11-100 views" (n = 145), "101-200 views" (n = 77) and "200 and more views" (n = 197). Students with 1-10 page views showed poorer gains than the rest of the sample, but otherwise this item did not significantly explain achievement in the pre-course.

Finally, the number of self-tests per student were related to learning gains. Each learning module provided a final self-assessment, consisting of 10-15 randomized items (thus with each test attempt new items were presented and the number of attempts was unlimited). The highest number of test attempts was 83, but the majority of students took four tests (= median). Transformed to a five-step ordinal variable, with "no test attempts" (n = 296), "1-5 attempts" (n = 167), "6-10 attempts" (n = 55), "11-20 attempts" (n = 60), and "21 and more attempts" (*n* = 25), this variable significantly differentiated between higher / lower achievement in the pre-course. Students with no test attempts had the poorest learning gains (gain score = 3.8) and students with 21 and more attempts had an average gain score of 12.0.

These results strongly supported the view that study time or number of page views may be unreliable indicators of student engagement in e-learning environments (Samson, 2015). Macfadyen and Dawson (2010) reported weak relations between these measures and performance in an online biology course. They observed a good predictive power for number of tests completed and an even stronger impact of the total number of forum posts (an effect that could not be confirmed in this study due to low and irregular case numbers in discussion forums).

Table 4: Pre- and posttest results 2014: complete dataset in comparison to number of online self-test attempts (*n* = 603)

	participants both tests	number of test attempts				
		none	1 to 5	6 to 10	11 to 20	21 and more
n	603	296	167	55	60	25
pretest (%)	49.7	50.2	48.2	52.0	49.9	49.0
posttest (%)	55.2	54.0	53.7	57.2	60.4	61.0
gain score	5.5	3.8	5.6	5.3	10.5	12.0

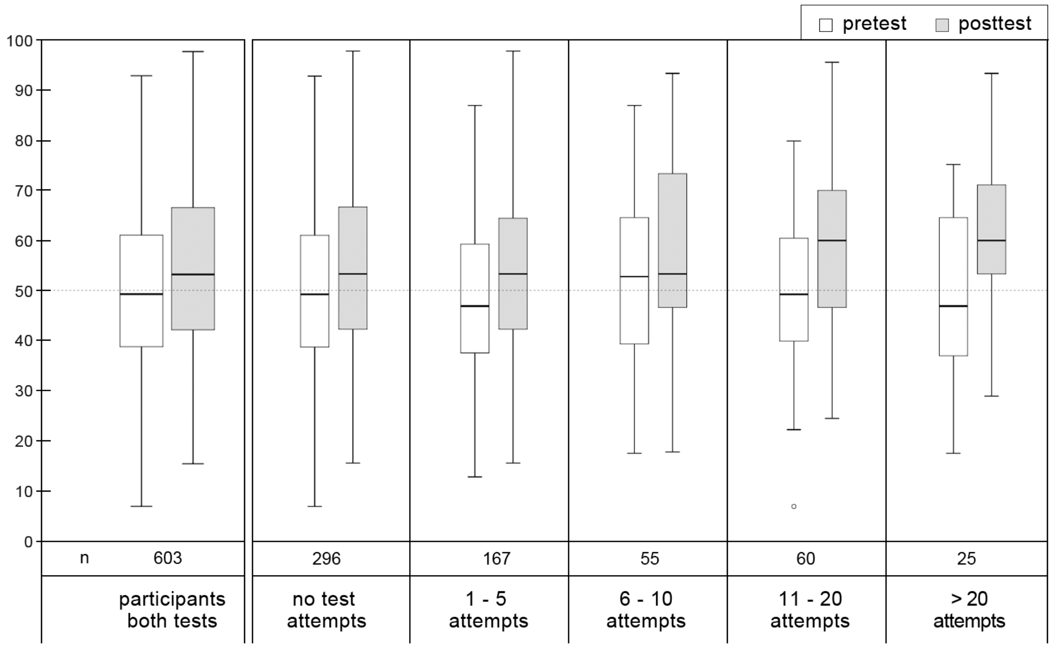

Figure 2: Pre- and posttest results 2014: complete dataset in comparison to number of online self-test attempts (*n* = 603)

3.3 Learning gains in the pre-course and academic achievement

In the final analysis, learning gains in the pre-course were related to first year academic achievement. Assuming a poor pretest performance being a risk factor, a high gain score was expected to reduce this risk. Accordingly, gain score as well as learner engagement, represented by number of test attempts, were expected to influence Mathematics I results. Thus in the final analyses these variables were added to the multiple regression as described in section 3.1. Table 5 gives a summary of the changes in variance explained (R^2) when pretest results (model 2), gain score (model 3), and effort (model 4) were added to the basic model (model 1). It can be seen that in the basic model 21% of variance in Mathematics I was accounted for. When diagnostic pretest scores were added, R^2 increased to .33. Finally, the variables gain score and number of test attempts led to a total variance explained of 36%.

Table 5: Summary of hierarchical regression analysis for variables predicting Mathematics I (n = 465). Model 1: students preconditions when entering university; model 2: plus pretest mean score; model 3: plus gain score (pre-course learning gains); model 4: plus measures of effort (number of page views, number of test attempts) (B: unstandardized regression coefficient; $SE\ B$: standard error; β: standardized regression coefficient; significance levels: $^*p < .05$; $^{**}p < .01$)

	model 1			model 2			model 3			model 4		
Predictor variable	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
gender	-.06	.13	-.21	.14	.12	.05	.17	.12	.06	.08	.12	.03
age	.03	.05	.52	.03	.05	.06	.04	.05	.07	.04	.04	.07
mathematics grades	.30	.15	.13*	.00	.15	.00	-.05	.14	-.02	-.28	.15	-.13
secondary school grades	.64	.11	.33**	.44	.10	.23**	.39	.10	.20**	.43	.10	.23**
pretest score				.03	.00	.40**	.03	.00	.48**	.03	.00	.43**
gain score							.01	.00	.18**	.01	.00	.14**
number of page views										.00	.00	-.01
number of test attempts										.02	.01	.11*
R^2 (R^2 adj.)	.21 (.19)			.33 (.31)			.35 (.33)			.36 (.34)		
F for change in R^2	10.03**			16.67**			17.44**			15.67**		

Gain score showed a significant impact on Mathematics I and also added to the variance explained (model 3: n = 465; R^2 = .35; R^2 adj. = .33; F (14, 450) = 17.44; p < .01). It should be stated that with a B coefficient of .014 the influence of this variable was not very strong; however, according to this model, a student with a gain score of 20 was predicted to achieve a Mathematics I exam that was .28 grades above that of a student with otherwise similar preconditions and a gain score of zero (note that test scores ranged from 0 to 100 and that Mathematics I grades ranged from 1 to 5). The number of test attempts, as well, showed a significant impact on Mathematics I results (model 4), confirming the importance of this variable.

Finally, it was investigated if the group of students who had not participated in the pretest or the pre-course programme (n = 105) differed in their Mathematics I results. The multiple model suggested poorer first year performance when a student had not taken the pretest, and this relation was significant, as well (p < .01). In the "at risk" group, the effect of *not* participating in the pretest led to a difference of -.5 in Mathematics I grades. The effectiveness of the pre-course thus could be established in the multiple model. As a limitation to this interpretation it should be considered that participation was voluntary, therefore groups were not randomized. It may be hypothesized that students who take the diagnostic pretest already show a higher interest in their degree programme, which might result in a better first year performance. Descriptive analyses suggested that the rate of "at risk" students in the posttest-only group was slightly higher, with more students having attended vocational schools and a higher rate of medium to poor school grades. These differences were not significant, with a high variance and a considerable number of very high performing students. Thus non-participation in the pre-course could not be described as a risk factor in itself, but students who were "at risk" in any of the predictive variables certainly would have benefitted from pre-course participation.

4. Discussion and conclusion

In this article different dimensions to the evaluation of a web-based pre-course in mathematics were summarized, relating students' preconditions when entering university, their learning activities in the pre-course and first year study success. In order to establish a consistent data model the impact of all collected variables on either learning gains in the pre-course or academic achievement was analysed. In this process, the relevance of "traditional" or performance-related variables for study success in engineering could be confirmed. Secondary school grades, for example, highly correlated with later study success, but the most consistent predictor of first year academic achievement were test results in a web-based diagnostic pretest in mathematics. In the multiple regression model this variable significantly influenced first year performance as well as cumulated GPA at the end of the degree programme, identifying poor pretest results as a dominant risk factor. These outcomes confirmed existing literature (Zhang*, et al.*, 2004; Ackerman*, et al.*, 2013; Faulkner*, et al.*, 2014) and indicated that the diagnostic pretest delivered consistent results. The model also showed a significant impact for the variable pre-course learning gains (pre-posttest difference); "at risk" students who were able to considerably improve throughout the pre-course showed better first year performance.

It then was investigated which factors most strongly supported successful pre-course participation. As the course was designed to address a heterogeneous group of learners and allowed self-enrolment into different modules some limitations regarding the interpretation of outcomes had to be considered. The highest pre-course learning gains, for example, were achieved by students who had combined two additional course programmes, e-tutoring and face-to-face, followed by e-tutoring-only participants. Students who had preferred to learn independently with the self-study programme were also able to improve considerably, whereas the face-to-face group had the least learning gains. Effects of course participation could even be linked to exam scores in Mathematics I, with significantly poorer results for the face-to-face group. Three interpretations of these unexpected results are suggested. First, reliability issues had to be taken into account as results were not based on randomized groups. Face-to-face courses appeared to be preferred by "at risk" students (non-traditional, poorer school grades, poorer pretest-results) and although the differences between this group and the e-tutoring group were not significant students' diverse preconditions may have added up and influenced the outcomes. Second, the face-to-face and e-tutoring course were difficult to compare regarding length, intensity, and concept. It may be hypothesized that the one-week face-to-face was too condensed to have a lasting effect on students with major knowledge gaps. The one-month e-tutoring course allowed for more practice, and with weekly tasks and a final certificate the learning process was monitored more strongly. A third reason might be that face-to-face participants felt less inclined to invest extra time into self-study once they had completed the course. For this group little or no online learning activity could be observed. Concluding, these results called for a revision of the face-to-face course concept, for example by expanding it to a four-week blended learning programme, giving students more time for individual practice.

Affective and metacognitive variables in this study were more or less unrelated to learning gains. Attitude towards mathematics, for example, strongly correlated with prior performance, suggesting that students with good grades also have a positive attitude towards the subject (Kadijevich, 2006; Mullis*, et al.*, 2012). A significant impact on students' learning gains in the pre-course, though, could not be observed. Even less related to learning gains was a scale addressing students' use of learning strategies (Pintrich*, et al.*, 1991; Schiefele and Wild, 1994). Distributions were often skewed and many items produced inconsistent results, being unrelated to prior knowledge level, learning gains, or the attitude scale. It is suggested that the weak impact of this scale might have been caused by a lack of representativeness. Only a third of the sample participated in the final evaluation survey and the learning strategies items in particular were often left unanswered, leading to even smaller case numbers. It also may be hypothesized that a superficial answer behaviour in the evaluation affected the quality of the scales (Spooren*, et al.*, 2013). It also has been reported that low-performing students are less likely to conscientiously answer questionnaires which also may have led to skewed answer patterns (Thiessen and Blasius, 2008). Summarizing, the impact of the meta-cognitive scales was disappointing. In order to better understand learner behaviour in an open e-learning environment a qualitative approach might be more beneficial.

One variable, however, led to consistent and unmoderated results: "at risk" students who had repeatedly engaged in online self-assessments achieved higher pre-course learning gains than "at risk" students who had not. Compared to other approaches, like invested time, number of learning modules, or number of page views, the number of test attempts thus was found the strongest indicator of effort, or student engagement. This outcome may be characteristic of the domain of mathematics, but similar observations were made by Macfadyen and Dawson (2010) for a web-based biology course. This variable also showed a significant relation to first year performance in the multiple model, suggesting that students with low prior knowledge level in mathematics were able to benefit from the pre-course, but only if they showed a strong learner engagement.

Concluding, it is stated that even volatile and inconsistent learning environments can produce valuable information, but the described limitations will have to be considered when interpreting the results. In order to identify consistent variables external references should be included in the model, in this case students' educational backgrounds and academic achievement. In this study, reliable and significant effects were mainly observed for straightforward quantitative measures, like pre- and posttest results. Further research might be needed into the reliability of affective and motivational items collected from web-based courses. Regarding the effectiveness of student learning, the role of practice should be investigated more deeply (Gibbs and Simpson, 2004; Pachman*, et al.*, 2013). Participants of the pre-course appeared to highly appreciate the possibility to practise online. By providing open question banks related to adaptive tests that provide item sets of different difficulty and complexity levels individual students' learning processes might be supported even better.

Acknowledgements

Support for this publication was provided by the German Federal Ministry of Education and Research (BMBF) in the context of the Federal "Quality pact for teaching" (ref. number 01PL12012). Responsibility for the content published in this article, including any opinions expressed therein, rests exclusively with the author. Test items and learning material are licensed under the creative commons attribution 3.0 unported and can be provided via www.optes.de.

References

Ackerman, P.L., Kanfer, R. and Beier, M.E. (2013). Trait Complex, Cognitive Ability, and Domain Knowledge Predictors of Baccalaureate Success, STEM Persistence and Gender Differences. *Journal of Educational Psychology*, 105(3), 911–927.

Ashby, J., Sadera, W.A. and Mcnary, S.W. (2011). Comparing student success between developmental math courses offered online, blended, and face-to-face. *Journal of Interactive Online Learning*, 10(3), 128–140.

Budny, D., LeBold, W. and Bjedow, G. (1998). Assessment of the Impact of Freshman Engineering Courses. *Journal of Engineering Education*, 87(4), 405–411.

Campbell, J.P. and Oblinger, D.G. (2007). Academic Analytics. *EDUCAUSE Learning Initiative*.

Cook, C., Heath, F. and Thompson, R.L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and psychological measurement*, 60(6), 821–836.

Corrigan, O., Glynn, M., KcKenna, A., Smeaton, A. and Smyth, S. (2015). Student Data: Data is Knowledge. Putting the Knowledge Back in the Students' Hands, 29-30 Oct, Hatfield, UK.

Creswell, J.W. and Plano Clark, V.L. (2011). *Designing and conducting mixed methods research*. Sage, Los Angeles, CA.

Croft, A.C., Harrison, M.C. and Robinson, C.L. (2009). Recruitment and retention of students – an integrated and holistic vision of mathematics support. *International Journal of Mathematical Education in Science and Technology*, 40(1), 109–125.

Dreier, O. (2014). Entwicklung eines Moodle Plug-Ins zur erweiterten Auswertung von Tests. Bachelor Thesis, Duale Hochschule Baden-Wuerttemberg, Mannheim.

Ehrenberg, R.G. (2010). Analyzing the factors that influence persistence rates in STEM field, majors: Introduction to the symposium. *Economics of Education Review*, 29, 888–891.

Fan, W. and Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26, 132–139.

Faulkner, F., Hannigan, A. and Fitzmaurice, O. (2014). The role of prior mathematical experience in predicting mathematics performance in higher education. *International Journal of Mathematical Education in Science and Technology*, 45(5), 648–667.

Faulkner, F., Hannigan, A. and Gill, O. (2010). Trends in the mathematical competency of university entrants in Ireland by leaving certificate mathematics grade. *Teaching mathematics and its applications*, 29(2), 76–93.

Gasiewski, J., Eagan, M., Garcia, G., Hurtado, S. and Chang, M. (2012). From Gatekeeping to Engagement: A Multicontextual, Mixed Method Study of Student Academic Engagement in Introductory STEM Courses. *Research in Higher education*, 53(2), 229–261.

Gibbs, G. and Simpson, C. (2004). Conditions under which assessment supports student learning. *Learning and Teaching in Higher Education*, 1(1), 3–31.

Hambleton, R.K. and Swaminathan, H. (2010). *Item response theory. Principles and applications*. Kluwer Academic Publishers, Boston.

HEFCE (2013). *Non-continuation rates at English HEIs. Trends for entrants 2005-06 to 2010-11.* Higher Education Council for England.

Howson, A.G., Barnard, A.D., Crighton, D.G., et al. (1995). Tackling the mathematics problem. Joint report by London Mathematical Society, Institute of Mathematics and its Applications, Royal Statistical Society.

Jackson, D.C. and Johnson, E.D. (2013). A hybrid model of mathematics support for science students emphasizing basic skills and discipline relevance. *International Journal of Mathematical Education in Science and Technology*, 44(6), 846–864.

Johnson, M. and Kuennen, E. (2006). Basic Math Skills and Performance in an Introductory Statistics Course. *Journal of Statistics Education*, 14(2).

Kadijevich, D. (2006). Developing Trustworthy Timss Background Measures: A Case Study On Mathematics Attitude. *The Teaching of Mathematics No.* 2, pp. 41–51.

Kane, M.T. (2013) Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73.

Kokkelenberg, E. and Sinha, E. (2010). Who succeeds in STEM studies? An analysis of Binghamton University undergraduate students. *Economics of Education Review*, 29, 935–946.

Lawson, D. (2000). Changes in the Competency of A Level Students. *Measuring the Mathematics Problem*, London, 6–9.

Macfadyen, L.P. and Dawson, S. (2010). Mining LMS data to develop an 'early warning system' for educators: A proof of concept. *Computers & Education*, 54(2), 588–599.

Mullis, I.V.S., Martin, M.O., Foy, P. and Arora, A. (2012). *TIMSS 2011 International Results in Mathematics*.

Pachman, M., Sweller, J. and Kalyuga, S. (2013). Levels of knowledge and deliberate practice. *Journal of Experimental Psychology: Applied*, 19(2), 108–119.

Pappano, L. (2012). The Year of the MOOC, *The New York Times, Education Life Supplement* Nov 4, p. 26.

Parker, M. (2005). Placement, Retention, and Success: A Longitudinal Study of Mathematics and Retention. *Journal of General Education*, 54(1), 22–40.

Parsons, S.J., Croft, T. and Harrison, M. (2009). Does students' confidence in their ability in mathematics matter?. *Teaching mathematics and its applications*, 28(2), 52–68.

Pintrich, P.R., Smith, D.A.F., Garcia, T. and McKeachie, W.J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor, Michigan.

Richardson, M., Abraham, C. and Bond, R. (2012). Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis. *Psychological Bulletin*, 138(2), 353–387.

Richey, R.C. and Klein, J.D. (2005). Developmental research methods: Creating knowledge from instructional design and development practice. *Journal of Computing in Higher Education*, 16(2), 23–38.

Robbins, S.B., Lauver, K., Le, H., Davis, D., Langley, R. and Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261–288.

Romero, C. and Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *Transactions on systems, man, and cybernetics, Part C: Applications and reviews*, 40(6), 601–618.

Samson, P.J. (2015). Can Student Engagement be Measured? and, if so, Does it Matter?. *Frontiers in Education Conference FIE*, El Paso, Texas, 21-24 Oct. 2015.

Schiefele, U. and Wild, K.P. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens, *Zeitschrift für Differentielle und Diagnostische Psychologie*, 15(4), 185–200.

Smith, G.G. and Ferguson, D. (2005), Student attrition in mathematics e-learning, *Australasian Journal of Educational Technology*, 21(3). 323–334.

Spooren, P., Brockx, B. and Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research*, 83(4), 598–642.

Street, H. (2010), Factors Influencing a Learner's Decision to Drop-Out or Persist in Higher Education Distance Learning. *Online Journal of Distance Learning Administration*, 13(4).

Thiessen, V. and Blasius, J. (2008). Mathematics achievement and mathematics learning strategies: Cognitive competencies and construct differentiation. *International Journal of Educational Research*, 47(6), 362–371.

van Soom, C. and Donche, V. (2014). Profiling First-Year Students in STEM Programs Based on Autonomous Motivation and Academic Self-Concept and Relationship with Academic Achievement. *PLoS one*, 9(11), e112489.

Warwick, J. (2007). A Pilot Study in the Application of the Analytic Hierarchy Process to Predict Student Performance in Mathematics. *PRIMUS*, 17(1), 181–194.

Xie, Y. and Shauman, K.A. (2005). *Women in science. Career processes and outcomes*. Harvard University Press, Cambridge, MA.

Zhang, G., Anderson, T.J., Ohland, M.W. and Thorndyke, B.R. (2004). Identifying factors influencing engineering student graduation: A longitudinal and crossinstitutional study. *Journal of Engineering Education*, 93, 313–320.