

Increasing performance by differentiated teaching? - Experimental evidence of the student benefits of digital differentiation

Citation for published version (APA):

Haelermans, C., Ghysels, J., & Prince, F. (2015). Increasing performance by differentiated teaching? - Experimental evidence of the student benefits of digital differentiation. *British Journal of Educational Technology*, 46(6), 1161. <https://doi.org/10.1111/bjet.12209>

Document status and date:

Published: 01/01/2015

DOI:

[10.1111/bjet.12209](https://doi.org/10.1111/bjet.12209)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Increasing performance by differentiated teaching? Experimental evidence of the student benefits of digital differentiation

Carla Haelermans, Joris Ghysels and Fernao Prince

Carla Haelermans is an assistant professor in Top Institute for Evidence-Based Education Research (TIER) at Maastricht University. She is also a guest researcher at the Centre for Innovations and Public Sector Efficiency Studies (IPSE Studies), Delft University of Technology. Joris Ghysels is an assistant professor in Top Institute for Evidence-Based Education Research (TIER) at Maastricht University. Fernao Prince is a teacher at Sophianum Gulpen. Address for correspondence: Dr Carla Haelermans, Top Institute for Evidence-Based Education Research (TIER), Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. Email: Carla.Haelermans@maastrichtuniversity.nl

Abstract

This paper explores the effect of digital differentiation on student performance using a randomized experiment. The experiment is conducted in a second year biology class among 115 prevocational students in the Netherlands. Differentiation allowed students in the treatment group to work at three different levels. The results show that there is a significant effect of digital differentiation on the posttest score. This effect is robust to adding covariates such as students' ability, grade repetition, age, gender, class and average neighborhood income. There are no differential effects when dividing students in three groups, by ability. The results imply that differentiation in large classrooms is possible and beneficial for all students, once done digitally.

Practitioner Notes

What is already known about this topic

- Unclear from literature whether ability grouping and differentiation works.
- More and more schools work with IT and implement it in their classes.

What this paper adds

- A description of an intervention to implement differentiation in the classroom.
- Evidence on the effect of digital differentiation among eighth grade students.

Implications for practice and/or policy

- It is beneficial for student performance to differentiate the level at which the student is taught.
- Differentiation in large classrooms is possible when using IT for implementation.

Introduction

Present-day education is forced to change to meet the high standards set by governments and other stakeholders. Change is necessary to cope with three present-day challenges, which we deal with in this paper. The first issue is the continuing call for differentiation in education, in order to meet the needs of both low-achieving and high-achieving students in the same classroom. Often, teaching was directed towards the “average” student, thereby neglecting the needs of high- and

low-achieving students (Fischer & Rose, 2001; Forsten, Grant & Hollas, 2002; Tomlinson & Kalbfleisch, 1998). Nowadays, more and more programs are developed for the gifted students and needed students, such as the No Child Left Behind Act (see, eg, Bui, Craig & Imberman, 2011; Davis, Engberg, Epple, Sieg & Zimmer, 2010). However, many educational systems still do not have programs like this. Differentiation in class also seems to increase student motivation (Tomlinson, 2004). However, differentiation in the traditional classroom setting is hard to achieve according to most teachers, as shown in the literature (Anderson, 2000; Grittner, 1975).

The second issue is the decreasing population numbers in many rural regions in many countries, among which is the Netherlands (Themagroep Krimp en Onderwijs, 2010). Families often move away from rural areas instead of settling there, and the number of children in these areas is rapidly decreasing. This has far-reaching implications for schools in these areas, eg, with respect to organization, personnel, quality and facilities (Themagroep Krimp en Onderwijs, 2010). Most of these implications are related to class size and the way classes are organized. Often, the easiest solution to decreasing student numbers is an increase in class size. However, this in turn might lead to lower quality and a higher workload for teachers. Larger classes also give less room to differentiate.

Lastly, schools feel the need to change towards a more digitally oriented curriculum, which could lead to increased student commitment, and in turn to increased student motivation and decreased dropout numbers (Cabus, 2013). Some schools already seek ways to complement traditional classroom teaching with digital learning, but many schools still teach in the traditional way.

Our study proposes a solution to all of the above-mentioned problems by initiating a randomized experiment in which the effect of digitally differentiated learning routes is analyzed. Digital learning routes meet the demand of students of more diverse education and more incorporation of digital learning materials in regular classes. Most methods offer digital learning materials, next to their traditional books, which makes it fairly easy to incorporate once sufficient computer rooms are available and/or students bring their own device. Differentiated learning routes meet the demand to teach students at their own level, even when they are in the same classroom. The combination of digital and differentiated learning routes provides teachers with the possibility to do all the above, even in large classrooms. It should be noted that we study these differentiated learning routes *within* the Dutch system of ability tracking and, hence, evaluate a fine-grained (within class) way of differentiating that deepens the between-classes differentiation that is the standard result of regular ability tracking.

This paper contributes to the literature in two ways: first, it is one of few studies that analyzes the effect of differentiation in a digital environment using an experimental setup and, as far as we know, the only study that was able to randomize at the individual level. Second, we show that digital differentiation leads to small positive results, even within the different levels of ability tracking that exist in the Netherlands.

The remainder of this paper is structured as follows: section 2 describes the related literature. Section 3 presents the context of the experiment, eg, a description of the field experiment and the content of the experiment. Section 4 presents the empirical model, methodology and the data. In Section 5, we discuss the baseline results and the extended regression results. Section 6 concludes the paper and discusses the findings.

Literature

The literature on differentiation is quite diverse, not least because it studies topics as varied as differentiated teaching within a classroom and structural differentiation by means of ability grouping at the class level ("ability tracking"), where the latter refers to an effort to form

homogeneous groups (classes) at some stage of the educational system. Correlational studies by Barta and Allen (1995), Reezigt, Houtveen and Van de Grift (2001) and Lou *et al* (1996) show that homogeneous groups (classes) are beneficial for high-achieving students but that heterogeneous groups are better for low- and medium-achieving students. This finding is underlined by the study on ability grouping by Kulik and Kulik (1982). However, in his best evidence synthesis, Slavin (1990) shows that there is no clear effect of ability grouping, regardless of the level of prior performance. This finding is confirmed by Betts and Shkolnik (2000). Hanushek and Woessman (2010) show that ability tracking leads to larger educational inequality, and the findings also suggest that ability tracking reduces educational outcomes. Brunello and Checchi (2007) also conclude that tracking increases inequality, but they find that there is a specialization effect that reduces the impact of parental outcomes on following further education.

However, in this study, we do not focus on ability tracking but on differentiated teaching within prevocational education. There are only a few experimental studies on the effect of individual differentiation in the traditional classroom. One of the few counterexamples is the study by Reis, McCoach, Little, Muller and Kaniskan (2010), who studied individual differentiation in reading instruction in primary education in an experimental setting. They found positive results.

Much as in the case of ability tracking and differentiation, which we discussed above, the results of research on the effect of the use of information technology (IT) for educational purposes depends strongly on its angle. Many of the previous evaluations of IT are wide in scope, as they evaluate increased budgets for IT either for schools or for households. They rely on the assumption that users have sufficient skills to use IT to their benefit and, hence, an increase in IT use does not need to be specified further to benefit educational outcomes. Yet, in practice, these general evaluations offer mixed results (eg, not significant: Brown & Liedholm, 2002; Coates *et al*, 2004; Goolsbee & Guryan, 2006; Navarro & Shoemaker, 1999; Terry, Lewer & Macy, 2003; positive: Kulik, 1994; Machin, McNally & Silva, 2007; Sosin, Blecha, Agarwal, Bartlett & Daniel, 2004; negative: Angrist & Lavy, 2002; Leuven, Lindahl, Oosterbeek & Webbink, 2007). As such, policy conclusions regarding increased budgets for IT that do not further specify its use are unclear.

A different strand of evaluations focuses on IT as a differentiation tool. IT is particularly suited to provide differentiation, as algorithms offer the opportunity to develop individual learning paths. For language, several experiments have been conducted, but none of them have found positive effects. Borman, Benson and Overman (2008), eg, conducted a randomized field trial to study the effects of a computer-based training on literacy skills and conclude that this program did not help students to improve their language skills. Similar conclusions are drawn by Rouse and Krueger (2004), Given, Wasserman, Chari, Beattie and Eden (2008) and Potocki, Ecalle and Magnan (2013). Evaluations of IT-based differentiation programs in math range from general teaching to remedying programs and cover both general pupil audiences and pupils with learning disabilities. Evaluation outcomes tend to be positive. Arroyo, Park Woolf, Royer, Tai and English (2010) show that a math-oriented, IT-based tutoring system leads to improved student performance for middle school students. Burns, Kanive and DeGrande (2012) show that a computer delivered math fact intervention results in gains for participating students. Furthermore, they show that significantly fewer of the students at risk for math difficulties beforehand were still at risk after the program. Banerjee, Cole, Duflo and Linden (2007) report on the positive outcomes of an experiment with an IT-based math remedying program, introduced in public schools of two cities in India, which illustrates that the benefits of IT differentiation are not confined to pupil audiences living in highly technologized societies. Pilli and Aksu (2013) also show a positive effect on some units of the math test.

Haelermans and Ghysels (2013) set up a field experiment and show that treatment students in the first year of secondary school who use an interactive online differentiated practice tool have a

significantly higher basic math performance than control group students. A counterexample reporting no significant effect relates to language learning (see Borman *et al.*, 2008). It should be noted that almost all of the above-discussed studies apply differentiation using computers, but hardly any of them study micro differentiation using computers in the classroom. Computers are used as remedying devices outside of regular classes, rather than incorporated in teaching to all students. A few studies do discuss differentiated instruction using technology (eg, Scalise, 2007; Smith & Throne, 2009). However, these studies do not study effects of digital differentiation at the micro level but mainly discuss principles and guidelines on how to implement technology-based differentiation in the classroom.

Context of the experiment

In the Netherlands, secondary education offers four different levels of education, which are concluded with a central examination in the last year. Students enter secondary school at age 12 and are directed into a specific level of education on the basis of a standardized primary school ability test and the recommendation given in elementary school. As such, the Dutch educational system is, internationally speaking, among the “early tracking” systems.

The first 3 years of secondary education are called lower secondary education. The last years of secondary education (the total number of years in secondary education depends on the level of education) are considered upper secondary education. The four different levels of secondary education are the following (Dutch abbreviations in brackets):¹

- practical prevocational school (4 years) (vmbo-bk);
- theoretical prevocational school (4 years) (vmbo-gt or, by its older acronym, mavo);
- higher general school education (5 years) (havo); and
- pre-academic education (6 years) (vwo).

Most schools do not focus on one particular track but offer a variety of levels of education.

The school under study offers all types of education. However, the experiment is conducted in the biology classes of the second year of prevocational education. Up to the second year, the school does not make a distinction between practical and theoretical prevocational school. The tracking *within* prevocational school only takes place after the second year. The school under study is a typical, average sized secondary school (about 1650 students in total) in a rural region with decreasing population numbers.

The field experiment

The main problem with determining the effect of digital differentiation is the potential correlation of characteristics of the students and unobservable factors with both the treatment and the outcome variables. In this study, we use exogenous variation in the treatment by means of a field experiment to ensure that this potential correlation is absent.

Figure 1 shows the timeline of the field experiment and shows that the field experiment consists of a preexperimental period and the experimental period, which in itself also consists of three periods. Randomization, as well as the first data collection, took place in the preexperimental period at the individual level. Stratified randomization at the student level was applied, to avoid assignment bias in the relatively small sample. Thus, randomization was effectuated in strata determined by class, gender, age and grade repetition. From the total of 115 students, which were divided over five classes, 57 students were assigned to the treatment group and 58 to the control group. Both groups practiced with the digital way of working for 1 week, so they could get accustomed to the digitalization. The field experiment consists of 12 weeks in total, distributed over three topics of each 4 weeks. Note that the treatment only affects the content that is discussed, not the skills needed to process the content or other aspects that could influence the

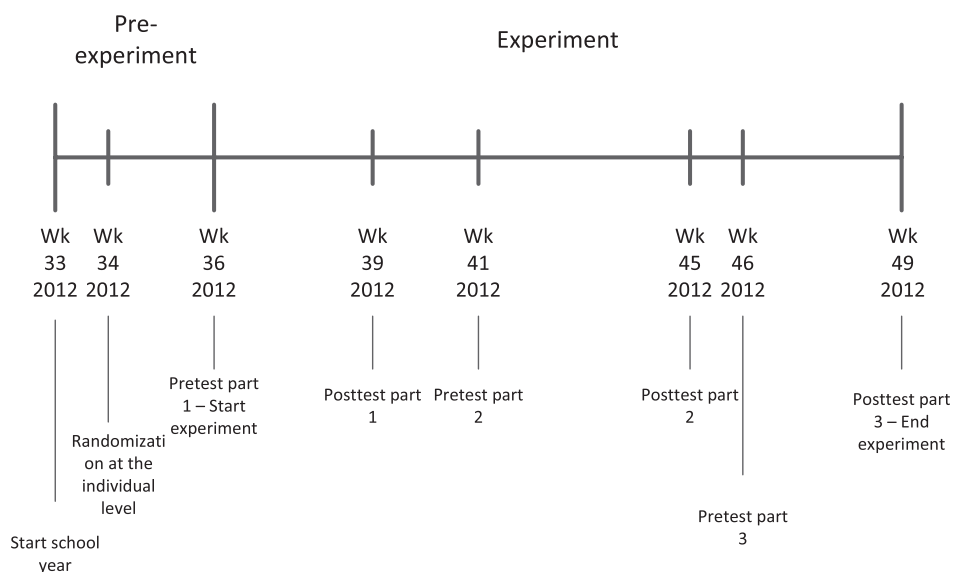


Figure 1: Overview of the field experiment

pretest of the next topic, because of changed study behavior. The first topic was “metabolism and respiration,” the second topic was “blood circulation,” and the third topic was “your health.” Because the school wanted students to be graded separately for all topics, we have a pre- and posttest for each of the three topics. The first pretest took place in week 36 and contained easy questions on the topic of the first experimental period. The first part of the experiment lasted 4 weeks, in which the treatment group followed differentiated learning routes at three different levels, based on weekly written small tests, whereas the control group also wrote weekly tests but followed only one learning route. In week 39, the first posttest was written. The second pretest, on the second topic, was written in week 41, with again 4 weeks of experiment and a posttest in week 45 (note the second period included 1 week of holiday). The third pretest was written in week 46, and the third posttest was written in week 49, which was the end of the experiment. The test consisted of questions on this topic that come from a test data bank with questions that have been used for years by this school for this course.

We have decided to take an aggregate measure of the tests of these three topics, for several reasons. First of all, we expect the effect to evolve over time. For some students, eg, students that have a long adjustment time, the 4-week period per topic may be too short to actually show a difference in performance. For other students, eg, those who adapt fast to a new situation, differentiation might show an increasing effect and, hence, strengthens the longer they work in the differentiated situation. The adjustment process might start over again for the student who is not comfortable with changes when a new topic starts after 4 weeks, whereas the comfortable student continues on the previously built differentiated path. Four weeks might be too short to rule out the option that the uncomfortable student even performs worse than the average control student, which seems to be the case for some students. In order to avoid the results of the uncomfortable students overtaking the longer term effect that might be there, we take the average of the three periods. We also ran separate regression analysis on the three topics, which gave about the same coefficients (with the largest coefficient in the third period, which seems to confirm our suspicions), but are just about not significant, most likely due to these type of students for which 4 weeks was not long enough to make a difference (results available upon request).

Furthermore, taking the three periods together allows for analysis of a larger research period, in which larger effects might appear, as students have had a differentiated learning route for a longer period and had a chance to get used to studying at the appropriate level. Keep in mind that the treatment only affects the content that is discussed, not the skills needed to process the content or other aspects that could influence the pretest of the next topic, because of changed study behavior. Lastly, there was no reason to expect the effect of differentiation to be dependent on the topic (also confirmed by the similar size of the treatment coefficient in the separated regressions), which would be the main reason for separating the three periods. As the duration of each topic within the experiment was 4 weeks for all three topics, we simply calculated the total pretest score² as the average of the score of all three pretests and the total posttest score as the average of all three posttest scores.

Digital differentiation

During the 12 weeks of the experiment, the students had two lessons (50 minutes per lesson) of biology each week. All classes were computerized. In the first week of every 4-week block, the pretest on that specific topic determined the track for the treatment group students. From the second class on, each student took a computerized short test in the first class each week to determine his or her level. For a treatment group student, this test determined in what track he or she was going to work during that week. For control students, there were no consequences to these tests, except the possibility to gain bonus points (which was the case of treatment group students as well), in order to motivate the students to take the tests seriously. There were three different tracks: the practical prevocational track, the theoretical prevocational track and the higher general track, which is one level higher in the Dutch tracking system. Note that it is possible for treatment group students to follow a different track each week, depending on their score on the test. The lessons at the practical prevocational track level were written in more simple language, using fewer words and less complicated sentences. Furthermore, the pace of the exercises would be a bit lower, which means these students studied the minimum amount of topics. The content of the basic topics did not differ between the groups. The theoretical prevocational track had some extra topics compared with the practical track and more difficult explanation and more challenging exercises. This was also the case for the higher general track, where more topics were discussed (in most cases slightly off topic but interesting to the students).

Control group students also made the short tests, but always followed the theoretical prevocational track, no matter how they performed on the test. Pre- and posttest were written at the theoretical prevocational track level. When studying for the posttests, treatment group students could choose at what level they wanted to study, after the course of 4 weeks all tracks were open to all students. However, it was recommended that they would study at the level they practiced in.

All students used digital instruction and digital assignments to study the biology course content, although all students also had their book. However, in practice, they hardly ever used the book. The digital instruction was a combination of digital instruction packages from the publisher, sections from the book that were digitalized by the teacher and additional digital material from the Internet (such as exercises, puzzles and short movies).

The teacher was present in class to answer questions and explain things when necessary. Every lesson, the teacher started with a short introduction. Because of the digitalization, the teacher had more time to devote to individual questions during the lessons. In a traditional classroom setting, there is not much time left for individual questions and explanations after the teacher has explained the topic to the class. It cannot be said whether there were more questions from treated than from control students, as they were in the same class, and it was hard to distinguish the groups within one class for the teacher, especially because conditions have to be as similar as

possible. The teacher was the same for the complete research population, which avoids teacher bias in the measurement of the effect.

Methodology and data

Methodology

To identify the effect (average treatment effect [ATE]) of digital differentiation on test scores, we use the method pioneered by Rosenbaum and Rubin (1983). We define the effect of the treatment as the average posttest score for the control students subtracted from the average treatment score of the treated students. We can do it this way because the random assignment of students to the treatment and the control group ensures the independence between the treatment and potential outcomes. We estimate the ATE using both simple *T*-statistics and linear regression. In the regression, we also take student ability and other student characteristics, such as age and gender, into account. The randomization also ensures that these observable student characteristics are independent of the treatment.

Data

Table 1 presents the descriptive statistics of the five second-year prevocational school classes. Table 1 shows that the 115 students were rather nicely divided over the five classes, with each having between 18 and 26 students. The average age was a little over 13, although age ranges from 12 to 15, mainly due to grade repetition. Two percent of the students has repeated a grade at some point in his or her school career, and 57 students (49%) are female. The average primary school ability test score is 530 (test is scored on a scale between 500 and 550), with a minimum of 516 and a maximum of 545. Note that the primary school ability test score is missing for four students. This is explained by the fact that this test is not compulsory, although the large majority of schools uses it. The average monthly income of the neighborhood in which the students are living (an indicator of socio-economic status [SES]) is €2126, but ranges from about €1500 to €3500 per month.

Lastly, Table 1 shows that the average score between 0 and 1 at the pretest is 0.71 (71%), whereas the average score at the posttest is 0.55. The fact that the posttest score is lower than the pretest score can easily be explained by the types of test that were used. The pretest consisted of very easy questions on the topic, based on knowledge students should have gathered in previous years. The posttest was a more complicated test, based on the biology content that was discussed during the experimental period. The posttest was perceived by the students as much more difficult than the pretest, and this shows in the scores. However, this is not a problem for our analysis as we only use the pretest as an indication of prior knowledge and student ability.

Table 1: Descriptive statistics

	Observations	Average	Standard deviation	Min	Max
Age	115	13.20	0.52	12	15
Female	115	0.49	0.50	0	1
Grade repetition	115	0.02	0.13	0	1
Primary school ability test total score	111	530.65	5.74	516	545
Average monthly income neighborhood	115	2107.12	319.34	1532	3358
Pretest	115	0.71	0.07	0.43	0.85
Posttest	115	0.55	0.08	0.39	0.83
Class 1	26				
Class 2	25				
Class 3	24				
Class 4	18				
Class 5	22				

Table 2 presents the observable characteristics of the treatment and control group, as well as the *T*-statistics/Mann–Whitney statistics on the differences between the groups. Apart from the randomization, these statistics give us full confidence with respect to the comparability of the treatment and control group.

Results

Baseline results

The first results we present are the simple *T*-statistics of the difference between treatment and control group at the pretest and at the posttest. Table 3 presents these results. In Table 3, we see that the treatment and control group are not significantly different at the pretest, whereas the treatment group scores 0.02 higher (only barely significant at the 10% level) at the posttest. The standardized version of the effect on the pretest points towards a small effect of 0.22 of a standard deviation.

The returns of digital differentiation

The next step is to analyze the returns of digital differentiation using regression analysis, which leads to more precise estimates due to including additional information on the individuals. The results of these analyses of the returns of digital differentiation are presented in Table 4. In Table 4, we present four models: model 1 is the basic model in which no covariates are included. In model 2, we include the average score for the pretest T0. In model 3, we add student characteristics and variables that account for student ability and past education, such as age, gender, grade repetition and the total scores for the standardized exit exam of primary education. Moreover, we add dummies reflecting to which class a pupil belonged so as to control for potential class effects. In the last model, we add the environmental variable average income of the neighborhood. This variable is taken from administrative records of the tax administration and reflects the average SES of all households within a particular neighborhood. As such, model 4 combines measures of student characteristics (age and gender), student ability (eg, grade repetition and primary school ability test) with indicators of the environment (eg, income) and current education (eg, class). It thus avoids overestimating the direct impact of the treatment.

The results presented in Table 4 show that the score at the posttest is around 0.03 higher for treatment students than for control students in all models. In models 3 and 4, the effect is significant at the 5% level. The coefficient of 0.03 corresponds to a small effect of 0.18 of a standard deviation. The standardized effect has decreased slightly with the inclusion of the covariates, compared with the *t*-test and the first regression model. In sum, the significance of the effect proves robust to adding different types of student specific information and to class and environmental information.

Differential effects of the returns of digital differentiation

Next to the average effect of digital differentiation, we are interested in whether there are differential effects between groups of students with a different ability. Therefore, in model 5 in Table 5, we add interaction terms to incorporate the possibility that the effect is mainly driven by the students at the lower level, the middle level or the higher level of differentiation, based on their pretest scores. The lowest scoring group is used as the reference category and is represented by the variable “Treatment” and the interaction terms of middle performing group times treatment variable and higher performing group times treatment variable are added to the regression. We remove the pretest variable in this specification due to multicollinearity.

Model 5 in Table 5 shows that the effect is fairly homogeneous across performance groups. In this specification, the base effect of differentiation, the effect for the lowest scoring group, is not significant and roughly of the same size as in the previous specifications. The estimates of students, who belong to the middle or highest performing on the pretest, show that they experience no statistically significant gain or loss in comparison with the reference category, the lowest

Table 2: T-statistics and Mann–Whitney statistics of treatment and control group

Variable	Control group			Treatment group			T-statistic
	Observations	Average	Standard deviation	Observations	Average	Standard deviation	
Age	58	13.18	0.51	57	13.22	0.53	−0.39
Female	58	0.53	0.50	57	0.46	0.50	0.85
Grade repetition	58	0.00	0.00	57	0.03	0.18	−1.44
Primary school ability test total score	56	531.30	5.30	57	530.00	6.13	1.20
Average monthly income neighborhood	58	2086.24	324.73	57	2128.37	315.20	−0.71
Variable		Rank sum	Expected		Rank sum	Expected	Mann–Whitney Z-score
Class	58	3476.5	3364	57	3193.5	3306	0.64
Place of residence	58	3368.5	3364	57	3301.5	3306	0.03

Please note there are no significant differences between the two groups.

Table 3: Baseline results: T-tests of the pretest and the posttest

Variable	Control group (n = 58)		Treatment group (n = 57)		T-statistic
	Average	Standard deviation	Average	Standard deviation	
Pretest	0.71	0.08	0.71	0.07	0.13
Posttest	0.54	0.09	0.56	0.09	-1.65*

*, significant at the 10% level; **, significant at the 5% level; ***, significant at the 1% level.

Table 4: The returns of digital differentiation

	Model 1		Model 2		Model 3		Model 4	
Number of observations	115		115		111		111	
R ²	0.02		0.15		0.22		0.22	
F-stat	(1,113)		(2,112)		(10,100)		(11,99)	
Prob > F	0.09		0.00		0.01		0.01	
Posttest score	Coef.	T-stat	Coef.	T-stat	Coef.	T-stat	Coef.	T-stat
Treatment	0.03	1.65*	0.03	1.81*	0.03	2.03**	0.03	2.01**
Pretest score			0.43	4.03***	0.35	2.97***	0.34	2.86***
Age					-0.03	-1.73*	-0.03	-1.64*
Female					-0.01	-0.42	-0.01	-0.41
Grade repetition					0.02	0.28	0.02	0.25
Primary school ability					0.00	2.23**	0.00	2.25**
test total score								
Class 2					0.02	0.83	0.02	0.81
Class 3					0.00	-0.16	-0.01	-0.21
Class 4					-0.03	-1.14	-0.03	-1.12
Class 5					0.00	-0.04	0.00	-0.04
Average monthly income							0.00	0.40
neighborhood								
Constant	0.54	47.02***	0.23	2.97***	-1.33	-1.43	-1.40	-1.47

Reference is class 1.

*, significant at the 10% level; **, significant at the 5% level; ***, significant at the 1% level.

performing group of students. We therefore conclude that the average effect from Table 4 is not driven by a specific group of students but that all groups of students are represented in this effect.

Conclusions and discussion

Conclusions

In this paper, we study the ATE of digital differentiation on student performance of second year prevocational students. A randomized experiment was carried out during 12 weeks among 115 secondary students. Treatment group students got weekly tests that determined at which of three levels the student would work in the upcoming week. Control group students also made the weekly tests but all followed the theoretical prevocational track. The results show that there is a significant effect of digital differentiation of 0.03 on the posttest score, which corresponds to a small effect of 0.18 of a standard deviation. This effect is robust to adding different types of student, class and environmental specific information.

Discussion and policy reflection

Our contribution to the literature is the randomized experimental design, which allows for causal analysis and for estimating the ATE of digital differentiation. We show that there is a small but

Table 5: The differential effects of the returns of digital differentiation

Model 5		
Number of observations		111
R ²		0.16
F-stat	(12,98)	1.56
Prob > F		0.05
Posttest score	Coef.	T-stat
Treatment	0.02	0.63
Middle performing students * treatment	0.03	0.92
High performing students * treatment	0.02	0.69
Age	−0.02	−1.76
Female	−0.01	−0.27
Grade repetition	0.01	0.07
Primary school ability test total score	0.00	2.25**
Class 2	0.03	1.17
Class 3	−0.00	−0.24
Class 4	−0.02	−0.56
Class 5	0.01	0.28
Average monthly income neighborhood	0.00	0.66
Constant	−1.24	−1.23

Reference is class 1. Variable “Treatment” represents low performing students.

*, significant at the 10% level; **, significant at the 5% level; ***, significant at the 1% level.

positive and significant effect of digital differentiation on student performance. These findings are in line with the findings of the comparable study by Reis *et al* (2010). The findings are also in line with the studies by Arroyo *et al* (2010) and Burns *et al* (2012), although those studies focus on individual differentiation for math, which is similar but not exactly the same. Our findings differ from the findings by, eg, Borman *et al* (2008) and Rouse and Krueger (2004), but given that those studies focused on language performance and not courses more similar to biology, we find it hard to really compare these results.

The small positive outcomes we report in this paper may motivate schools and teachers to implement similar schemes. Although the effects are small, it is important to keep in mind that it is only a 12-week intervention in one of many secondary school courses, where the treatment intensity is rather limited. Given these limitations, we believe that the small finding is relevant and important nonetheless.

Two aspects of the intervention may determine the feasibility of its adoption: the availability of computers and the development effort. The former is self-evident. To be able to differentiate individually in class, students need access to a personal computer during class. Depending on the digital situation in the school, this might be rather cheap or rather costly to implement. However, in the current time in which laptops and tablets are more and more common in education, it would be an option to apply this digital differentiation method with the use of those as well.

The second aspect, development effort, may sound evident as well but is less stringent than it may seem in the case of the intervention studied here. In effect, we compare the use of IT-based teaching with and without differentiation. If we assume that a teacher already masters the skills to implement digital teaching, the additional effort is limited. At setup, sufficient differentiation material needs to be found, which may be covered completely by an educational package,

although it may also require further searching and testing. In class, the differentiated material is likely to save the teacher instruction time, which he or she can use to tackle individual questions. All in all, the experiment shows that differentiation slightly improves biology learning in a classroom setting with digital learning and weekly testing. It should be reiterated that the effectiveness of the digital learning and weekly testing environment is not evaluated as such in this experiment. This remains for future research. Regarding the former, two research routes can be envisaged. First of all, an experiment may test whether student performance is higher when using digital learning routes, compared with traditional classroom teaching, regardless of the differentiation. Second, one could imagine that students, asking for diversification of learning materials, might prefer mixed teaching, in which half of the classes take place in the traditional classroom setting and half of the classes take place at the computer. It would still be possible to differentiate and take small tests to determine the level at which to differentiate, but it would also increase the diversity of the lessons, which in turn might improve students' focus and concentration. Moreover, replication of the differentiation experiment may also be warranted, because testing was restricted to the least academic track of the Dutch secondary education system. Whether this type of differentiation also would work in the more heterogeneous classes of countries that do not track students early in secondary education, eg, is not a priori clear, although the literature does not give counter-indications on this matter. Therefore, the discrete conclusion for now is that differentiation in a digital learning environment seems to be an effective way to improve learning outcomes. However, more research is necessary to see if this conclusion holds for different educational systems.

Notes

1. Eurydice is used for the terminology (Executive Agency of Education Audiovisual and Culture, 2009a, 2009b).
2. Note that there were no significant differences between treatment and control groups on the pretest for each of the three topics (topic 1: mean treatment (MT) = 0.70, mean control (MC) = 0.71, $t = 0.34$; topic 2: MT = 0.59, MC = 0.58, $t = -0.49$; topic 3: MT = 0.83, MC = 0.84, $t = 0.44$).

References

- Anderson, L. W. (2000). Why should reduced class size lead to increased student achievement? In M. C. Wang & J. D. Finn (Eds), *How small classes help teacher do their best* (pp. 3–24). Philadelphia: Temple University Center for Research in Human Development and Education.
- Angrist, J. D. & Lavy, V. (2002). New evidence on classroom computers and pupil learning. *Economic Journal*, 112, 735–765.
- Arroyo, I., Park Woolf, B., Royer, J. M., Tai, M. & English, S. (2010). Improving math learning through intelligent tutoring and basic skills training. In V. Aleven, J. Kay & J. Mostow (Eds), *ITS 2010, Part I, LNCS 6094* (pp. 423–432). Berlin Heidelberg: Springer-Verlag.
- Banerjee, A. V., Cole, S., Duflo, E. & Linden, L. (2007). Remedying education: evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, 122, 3, 1235–1264.
- Barta, J. J. & Allen, M. G. (1995). The dilemma of tracking and grouping in early childhood and middle grades: are we speaking the same language. In H. Pool & J. A. Page (Eds), *Beyond tracking. finding success in inclusive schools* (pp. 95–105). Bloomington/Indiana: Phi Delta Kappa Educational Foundation.
- Betts, J. R. & Shkolnik, J. I. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, 19, 1–15.
- Borman, G. D., Benson, J. G. & Overman, L. (2008). A randomized field trial of the fast forward language computer-based training program. *Educational Evaluation and Policy Analysis*, 31, 1, 82–106.
- Brown, B. W. & Liedholm, C. E. (2002). Can web courses replace the classroom in principles of microeconomics? *The American Economic Review*, 92, 2, 444–448.
- Brunello, G. & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, 22, 52, 781–861.
- Bui, S. A., Craig, S. G. & Imberman, S. A. (2011). *Is gifted education a bright idea? Assessing the impact of gifted and talented programs on achievement*. NBER Working paper 17089. Cambridge.

- Burns, M. K., Kanive, R. & DeGrande, M. (2012). Effect of a computer-delivered math fact intervention as a supplemental intervention for math in third and fourth grades. *Remedial and Special Education*, 33, 3, 184–191.
- Cabus, S. (2013). Does enhanced student commitment reduce school dropout? Evidence from two major dropout regions in the Netherlands. *Regional Studies*. doi: 10.1080/00343404.2013.799760.
- Coates, D., Humphreys, B. R., Kane, J., Vachris, M., Agarwal, R. & Day, E. (2004). “No significant distance” between face-to-face and online instruction: evidence from principles of economics. *Economics of Education Review*, 23, 5, 533–546.
- Davis, B., Engberg, J., Epple, D. N., Sieg, H. & Zimmer, R. (2010). *Evaluating the gifted program of an urban school district using a modified regression discontinuity design*. NBER Working paper 16141. Cambridge.
- Executive Agency of Education Audiovisual and Culture (2009a). *National summary sheets on educational system in Europe and ongoing reforms—The Netherlands*. Brussels: European Commission.
- Executive Agency of Education Audiovisual and Culture (2009b). *Organisation of the education system in the Netherlands*. Brussels: European Commission.
- Fischer, K. W. & Rose, L. T. (2001). Webs of skill: how students learn. *Educational Leadership*, 59, 3, 6–123.
- Forsten, C., Grant, J. & Hollas, B. (2002). *Differentiated instruction. Different strategies for different learners*. Peterborough: Crystal Springs Books.
- Given, B. K., Wasserman, J. D., Chari, S. A., Beattie, K. & Eden, G. F. (2008). A randomized, controlled study of computer-based intervention in middle school struggling readers. *Brain and Language*, 106, 83–97.
- Goolsbee, A. & Guryan, J. (2006). The impact of internet subsidies in public schools. *The Review of Economics and Statistics*, 88, 2, 336–347.
- Gritter, F. M. (1975). Individualized instruction: a historical perspective. *The Modern Language Journal*, 59, 7, 323–333.
- Haelermans, C. & Ghysels, J. (2013). *The effect of an online practice tool on math in secondary education—evidence from a randomized field experiment*. TIER Working Paper Series.
- Hanushek, E. A. & Woessman, L. (2010). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, 116, 150, C63–C76.
- Kulik, C.-L. C. & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: a meta-analysis of evaluation findings. *American Educational Research Journal*, 19, 3, 415–428.
- Kulik, J. A. (1994). Meta analysis study of finding on computer-based instruction. In E. L. Baker & H. F. O’Neil (Eds), *Technology assessment in education and training* (pp. 9–34). Hillsdale, NJ: Lawrence Erlbaum.
- Leuven, E., Lindahl, M., Oosterbeek, H. & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economics and Statistics*, 89, 4, 721–736.
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B. & d’Apollonia, S. (1996). Within-class grouping: a meta-analysis. *Review of Educational Research*, 66, 4, 423–458.
- Machin, S., McNally, S. & Silva, O. (2007). New technologies in schools: is there a payoff? *The Economic Journal*, 117, 1145–1167.
- Navarro, P. & Shoemaker, J. (1999). The power of cyberlearning: an empirical test. *Journal of Computing in Higher Education*, 1, 1, 29–54.
- Pilli, O. & Aksu, M. (2013). The effects of computer-assisted instruction on the achievement, attitudes and retention of fourth grade mathematics students in North Cyprus. *Computers and Education*, 61, 62–71.
- Potocki, A., Ecalte, J. & Magnan, A. (2013). Effects of computer-assisted comprehension training in less skilled comprehenders in second grade: a one-year follow-up study. *Computers and Education*, 61, 131–140.
- Reezigt, G. J., Houtveen, A. A. M. & Van de Grift, W. (2001). *Vormgeving en effecten van adaptief onderwijs*. Groningen: RUG/GION.
- Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M. & Kaniskan, R. B. (2010). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal*, 2011, 48, 2, 462–501.
- Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 1, 41–55. doi: 10.1093/biomet/70.1.41.
- Rouse, C. E. & Krueger, A. B. (2004). Putting computerized instruction to the test: a randomized evaluation of a “scientifically-based” reading program. *Economics of Education Review*, 23, 4, 323–338.
- Scalise, K. (2007). *Differentiated e-learning: five approaches through instructional technology*. Eugeno: University of Oregon.

- Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: a best-evidence synthesis. *Review of Educational Research*, 60, 3, 471–499.
- Smith, G. E. & Throne, S. (2009). *Differentiating Instruction with Technology in Middle School Classrooms*. International Society for Technology in Education.
- Sosin, K., Blecha, B. J., Agarwal, R., Bartlett, R. L. & Daniel, J. I. (2004). Efficiency in the use of technology in economic education: some preliminary results. *Recent Research in Economic Education*, 94, 2, 253–258.
- Terry, N., Lewer, J. & Macy, A. (2003). The efficacy of alternative instruction modes in economics. Retrieved January 25, 2014, from: <http://ssrn.com/abstract=392561>
- Themagroep Krimp en Onderwijs (2010). “Krimp en onderwijs” (“shrinkage and education”) *Positionpaper* Nationaal Netwerk Bevolkingsdaling.
- Tomlinson, C. A. (2004). Research evidence for differentiation. *School Administrator*, 61, 7, 28–35.
- Tomlinson, C. A. & Kalbfleisch, M. L. (1998). Teach me, teach my brain: a call for differentiated classrooms. *Educational Leadership*, 56, 3, 52–55.