Chapter 23
Randomized Controlled Trial of Kindergarten Students Using Literacy Technology

Haya Shamir, Erik Yoder, Kathryn Feehan and David Pocklington

Abstract Randomized controlled trials in education are necessary to keep pace with the evidence-based practices demanded by schools and the nation. In this study, kindergarten students in a school district in Indiana were randomized on the class level: The experimental condition consisted of students utilizing a computer-adaptive reading program for 15 minutes per day, 5 days per week over the course of the 2017–2018 school year; the control condition consisted of students receiving traditional, teacher-directed literacy instruction for the same amount of time. Thus, overall exposure to literacy instruction was the same for both groups. At the end of the school year, students randomized to the experimental condition outperformed their control counterparts on two different literacy assessments, indicating that early literacy instruction in the format of computer-adaptive reading programs may be beneficial to kindergarten students. Experimental students also outperformed their control counterparts across demographics, demonstrating a benefit from the computer-adaptive program for all students.

Keywords Early childhood · Literacy · Randomized controlled trial

23.1 Introduction

In today's schools, evidence-based practices are not only encouraged to be utilized by teachers, but are also expected [1]. Originating in the medical field, evidence-based practice is a term that has become engrained in education, the demand for it increasing parallel to the increase in standardized testing of students: All teachers are expected to utilize research-based, quality practices while meeting the demands of district- and state-mandated assessments [1, 2]. This involves utilizing the tools that have been found by rigorous research to be valid and reliable. Although many other factors contribute to students' academic success, our nation has a set of rigorous Common

H. Shamir (✉) · E. Yoder · K. Feehan · D. Pocklington
Waterford Research Institute, Sandy, UT, USA
e-mail: hayashamir@waterford.org

V. L. Uskov et al. (eds.), *Smart Education and e-Learning 2019*,
Smart Innovation, Systems and Technologies 144,
https://doi.org/10.1007/978-981-13-8260-4_23

Core State Standards for K-12 students to meet, and teachers need to ensure that students are prepared for these state standards, national standards, and standardized assessments [3]. These high expectations put pressure on teachers as well as students, yet educational research is not assessing curricula or assessments quickly enough to provide schools with the highest quality products as possible. If the educational sector demands such high standards for students and evidence-based practices from teachers, then educational research needs to keep pace with these demands.

The What Works Clearinghouse (WWC) of the Institute of Education Sciences provides the nation with evaluations of qualities of research studies, with the highest level of "Meets WWC Group Design Standards Without Reservations" being a rare and highly respected award [4]. This level is only achieved with a randomized controlled trial (RCT) in education: RCTs are considered the gold standard of educational research [5]. RCTs randomly assign participants to a treatment or control condition, theoretically eliminating any bias while simultaneously creating equivalent groups at the onset of the study. RCTs in educational research are quite difficult to achieve but the benefits to education outweigh the difficulties of running them [6], yet the amount of educational RCTs is lacking.

On the other hand, RCTs in educational research are seen as flawed design choices since the designs are perceived to control for all external variables [7, 8]. RCTs do not control for every issue that may arise in a study, but by accounting for issues such as self-selection bias they are extremely important to determine the highest quality educational products that can be effective for all students. This assists teachers in providing students with the best education possible. For example, an RCT found that literacy skills were improved in the experimental group after a storybook-based intervention was provided to prekindergarten students, but fidelity predicted improved outcomes in most cases [9]. One may consider it cruel to limit the control students to not receiving the experimental condition: Education is a right, and the question arises of whether research is limiting some students' opportunities to achieve academic success by prohibiting them some educational advantage [6].

Technology is a potential asset to education, an increasing role in schools that is proving to be as effective if not more so than traditional teaching methods. When easily incorporated into lesson plans, technology can enhance students' experiences while preparing them for the technology-driven world of today. Students need to become creative, problem-solving thinkers in order to keep pace with the ever-changing, globalized workforce. A meta-analysis found that individually played educational games create higher learning gains than other forms of educational virtual reality instruction [10]. Technology has the potential to provide students with the high-quality education they need, but proper evidence of the success of technology in schools is required [11].

While both RCTs and technology are potentially advantageous in education, RCTs incorporating technology could be the solution to providing evidence-based practices in schools. Technology has been found to improve students' emergent literacy skills [12]. Additionally, RCTs have demonstrated the benefits technology can have on teaching mathematics throughout elementary school: Prekindergarten students randomly assigned to use a technological mathematics tool learned significantly more

mathematics skills than their control counterparts [13]; an RCT replication study utilizing a mathematics intervention for first grade found that experimental students had significantly higher math scores following the intervention [14]; and an RCT study utilizing a digital mathematics tool in third grade classrooms of experimental schools while control schools followed traditional teaching methods found that experimental students improved in mathematics achievement and motivation [15]. Although these studies contribute greatly to the field of education and to students' futures, there are few studies addressing demographics in the realm of educational RCTs. More RCT studies analyzing students of varying and disadvantaged backgrounds are necessary to provide the nation with the most highly respect technological advances in education.

This study contributes to the field of educational research because it employs a rigorous RCT that investigates the impact of a computer-adaptive reading program on early literacy skills when utilized in kindergarten classrooms; the study also includes students of varying ethnicities and socioeconomic backgrounds. This study investigates the impact of randomly assigning kindergarten classes to use either a computer-adaptive reading program or to the control condition. The hypothesis is that students in classes receiving the computer-adaptive reading program will have higher end of year literacy scores than their control counterparts.

23.2 Methods

23.2.1 Participants

The initial sample for this study consisted of kindergarten students (N = 536) enrolled in a public school district in Indiana during the 2017–2018 school year. The experimental group consisted of kindergarten students who were randomly assigned to using Waterford Early Learning (WEL) during the 2017–2018 school year. The control group consisted of kindergarten students who were randomly assigned to traditional literacy instruction.

23.2.2 Materials

Waterford Early Learning. The program offers a comprehensive, computer-adaptive pre-reading and reading curriculum for prekindergarten through second-grade students. The software presents a wide range of multimedia-based activities in an adaptive sequence tailored to each student's initial placement and his or her individual rate of growth throughout the complete reading curriculum.

Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP). The MAP is a valid assessment intended to measure individual growth and

mastery for students in kindergarten through twelfth grade on a range of skills. The skills considered relevant to kindergarten students for this study are Reading Foundations, Literature and Nonfiction, Vocabulary, and Writing. Results are scored on a standardized Rasch Unit (RIT) scale.

Text Reading Comprehension (TRC). The TRC is a measure of reading performance that evaluates students' foundational reading skills and their ability to apply those skills to increasingly complex texts. Developmentally appropriate books are calibrated to benchmark levels and used to assess oral reading accuracy and comprehension.

23.2.3 Procedure

Twenty-two kindergarten classes across three schools participated in the study. Cluster randomization was used to assign classes to the experimental or control condition. A sequence of possible permutations with comparatively balanced distribution of assignments between conditions was generated for each school. A random number generator was used to select an assignment within each sequence for each school. This resulted in eleven classes in the experimental condition and eleven classes in the control condition.

Eleven kindergarten classes ($n = 273$) were randomly assigned to the experimental condition: Students in these classes were expected to use Waterford Early Learning (WEL) during the 2017–2018 school year for 15 minutes per day, 5 days per week. Eleven kindergarten classes ($n = 263$) were assigned to the control condition: Students received traditional literacy instruction for the same amount of time that the experimental group received CAI instruction. Thus, overall exposure to literacy instruction was the same for both groups.

For both literacy assessments, the analytic sample consisted of all randomized students with assessment data ($N = 430$) in the experimental ($n = 217$) and the control ($n = 213$) groups. Overall attrition, the percentage of the sample who left the study, remained within the tolerable level of expected bias as defined by the WWC at the overall attrition rate of 19.78%. Likewise, differential attrition, the difference in percentage points between the experimental and control rates of attrition, remained within the tolerable level of expected bias as defined by the WWC at the differential attrition of 1.50%. Combined, the overall and differential attrition remain a "tolerable threat of bias under both optimistic and cautious assumptions range" [4].

The MAP and TRC literacy assessments were administered to all students at the end of the school year. Results from the TRC were converted into Lexile Levels based on students' assigned book levels and performance levels.

23.3 Findings

23.3.1 MAP

Group Differences Using Independent Samples *t*-tests. Independent samples *t*-tests examining group differences in end of year scores were conducted (see Fig. 23.1).

RIT Score. Analysis of RIT end of year scores revealed a significant difference between groups $t(1, 428) = -2.07, p < 0.05$, due to higher end of year scores made by students who used Waterford ($M = 155.31$) than by control students ($M = 152.77$). Effect size ($d = 0.20$).

Reading Foundations. Analysis of Reading Foundations end of year scores revealed a significant difference between groups $t(1, 428) = -2.60, p < 0.01$, due to higher end of year scores made by students who used Waterford ($M = 158.76$) than by control students ($M = 155.15$). Effect size ($d = 0.25$).

Literature and Nonfiction. Analysis of Literature and Nonfiction end of year scores did not reveal a significant difference between groups $t(1, 428) = -1.45, p = 0.147$; however, students who used Waterford had higher end of year scores ($M = 154.04$) than control students ($M = 152.08$).

Vocabulary. Analysis of Vocabulary end of year scores did not reveal a significant difference between groups $t(1, 428) = -1.63, p = 0.104$; however, students who used Waterford had higher end of year scores ($M = 153.50$) than control students ($M = 151.19$).

Writing. Analysis of Writing end of year scores did not reveal a significant difference between groups $t(1, 420) = -1.75, p = 0.080$; however, students who used Waterford had higher end of year scores ($M = 154.98$) than control students ($M = 152.68$).

Group Differences Using ANOVAs—RIT Score. Two separate two-way ANOVAs were conducted to examine the effects of WEL and demographics on RIT end of year scores (see Fig. 23.2).

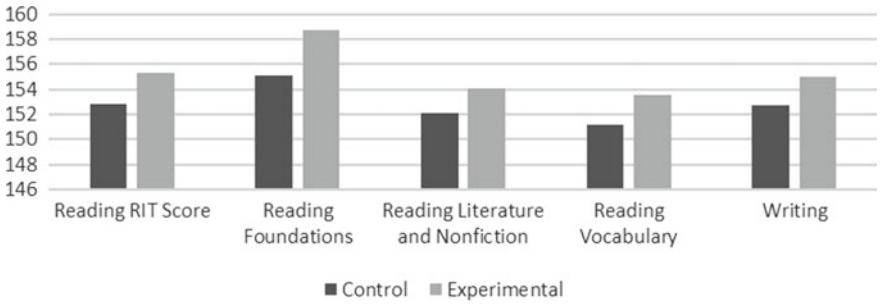

Fig. 23.1 MAP end of year scores by strand

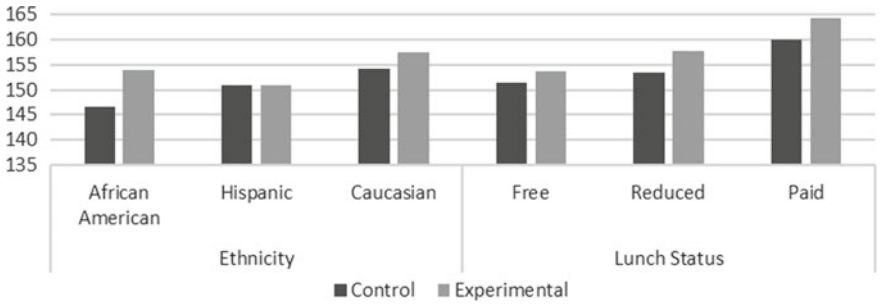

Fig. 23.2 MAP RIT scores by demographics

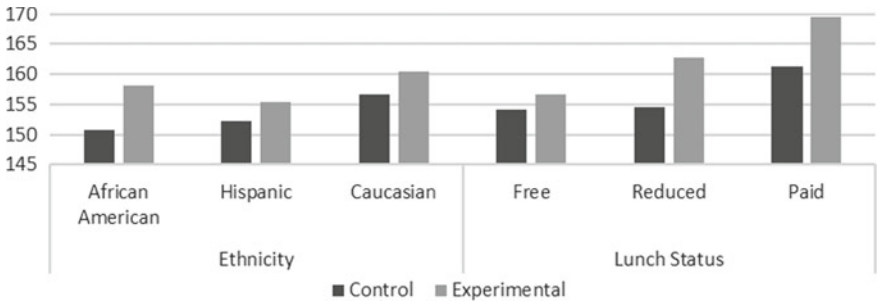

Fig. 23.3 MAP reading foundations scores by demographics

Ethnicity. There was no significant interaction between the effects of ethnicity and Waterford curriculum on RIT end of year scores, $F(3, 421) = 2.56$, $p = 0.054$. Simple effects analysis showed that for African American and Caucasian students, students in the experimental group significantly outperformed students in the control group. Hispanic students in the experimental group scored slightly higher than the control group, but the difference was not significant.

Lunch Status. There was no significant interaction between the effects of lunch status and Waterford curriculum on RIT end of year scores, $F(2, 422) = 0.48$, $p = 0.700$. Simple effects analysis showed that for students with free lunch, reduced lunch, and paid lunch, students in the experimental group scored slightly higher than the control group, but the difference was not significant.

Group Differences Using ANOVAs—Reading Foundations. Two separate two-way ANOVAs were conducted to examine the effects of WEL and demographics on Reading Foundations end of year scores (see Fig. 23.3).

Ethnicity. There was no significant interaction between the effects of ethnicity and Waterford curriculum on Reading Foundations end of year scores, $F(3, 421) = 0.98$, $p = 0.402$. Simple effects analysis showed that for African American and Caucasian students, students in the experimental group significantly outperformed students in the control group. Hispanic students in the experimental group scored slightly higher than the control group, but the difference was not significant.

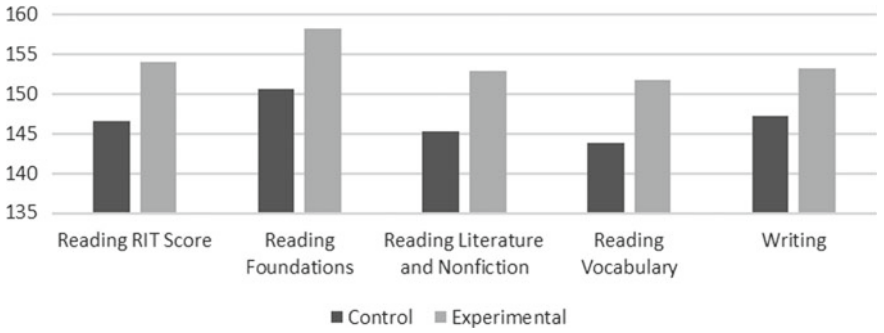

Fig. 23.4 MAP end of year scores by strand—African American students

Lunch Status. There was no significant interaction between the effects of lunch status and Waterford curriculum on Reading Foundations end of year scores, $F(2, 422) = 1.35$, $p = 0.258$. Simple effects analysis showed that for students with paid lunch, students in the experimental group significantly outperformed students in the control group. Students with free lunch and reduced lunch in the experimental group scored slightly higher than the control group, but the difference was not significant.

African American Analysis—Group Differences Using Independent Samples *t*-tests. Independent samples *t*-tests examining group differences in end of year scores were conducted for the subsample of African American students (see Fig. 23.4).

RIT Score. Analysis of RIT end of year scores revealed a significant difference between groups $t(1, 71) = -2.90$, $p < 0.05$, due to higher end of year scores made by students who used Waterford ($M = 154.00$) than by control students ($M = 146.59$). Effect size ($d = 0.68$).

Reading Foundations. Analysis of Reading Foundations end of year scores revealed a significant difference between groups $t(1, 71) = -2.58$, $p < 0.05$, due to higher end of year scores made by students who used Waterford ($M = 158.19$) than by control students ($M = 150.68$). Effect size ($d = 0.60$).

Literature and Nonfiction. Analysis of Literature and Nonfiction end of year scores revealed a significant difference between groups $t(1, 71) = -2.46$, $p < 0.05$, due to higher end of year scores made by students who used Waterford ($M = 152.92$) than by control students ($M = 145.35$). Effect size ($d = 0.58$).

Vocabulary. Analysis of Vocabulary end of year scores revealed a significant difference between groups $t(1, 71) = -2.83$, $p < 0.01$ due to higher end of year scores made by students who used Waterford ($M = 151.78$) than by control students ($M = 143.76$). Effect size ($d = 0.66$).

Writing. Analysis of Writing end of year scores did not reveal a significant difference between groups $t(1, 71) = -1.93$, $p = 0.057$; however, students who used Waterford had higher end of year scores ($M = 153.19$) than control students ($M = 147.19$).

23.3.2 TRC

The end of year TRC assessment assigned students to a book level based on their level of oral reading accuracy and comprehension. Book levels were converted to corresponding Lexile Level ranges, and a specific Lexile Level was calculated for each student based on their performance during the assessment. Students with less than 90% accuracy on their book level were given the minimum value of the corresponding Lexile Level range, students between 90% and 94% accuracy on their book level were given the mean value of the corresponding Lexile Level range, and students with at least 95% accuracy on their book level were given the maximum value of the corresponding Lexile Level range.

Group Differences Using Independent Samples *t*-tests. An independent samples *t*-test examining group differences in end of year Lexile Levels was conducted (see Fig. 23.5). Analysis of end of year Lexile Levels revealed a significant difference between groups $t(1, 428) = -3.45$, $p < 0.01$, due to higher end of year Lexile levels made by students who used Waterford ($M = 160.16$) than by control students ($M = 116.78$). Effect size ($d = 0.33$).

Group Differences by Demographics Using ANOVA. Two separate two-way ANOVAs were conducted to examine the effects of WEL and demographics on end of year Lexile Levels (see Fig. 23.6).

Ethnicity. There was no significant interaction between the effects of ethnicity and Waterford curriculum on end of year Lexile Levels, $F(4, 418) = 0.75$, $p = 0.560$. Simple effects analysis showed that for Hispanic and Caucasian students, students in the experimental group significantly outperformed students in the control group. African American students in the experimental group scored slightly higher than the control group, but the difference was not significant.

Lunch Status. There was no significant interaction between the effects of lunch status and Waterford curriculum on end of year Lexile Levels, $F(1, 418) = 0.49$, $p = 0.486$. Simple effects analysis showed that for students with free/reduced lunch and paid lunch, students in the experimental group significantly outperformed students in the control group.

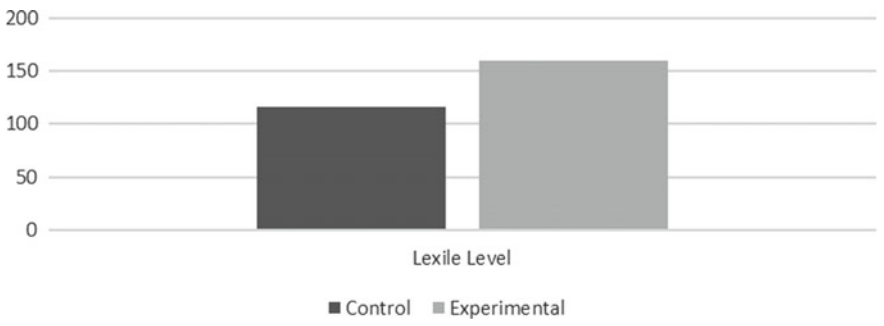

Fig. 23.5 TRC end of year Lexile levels

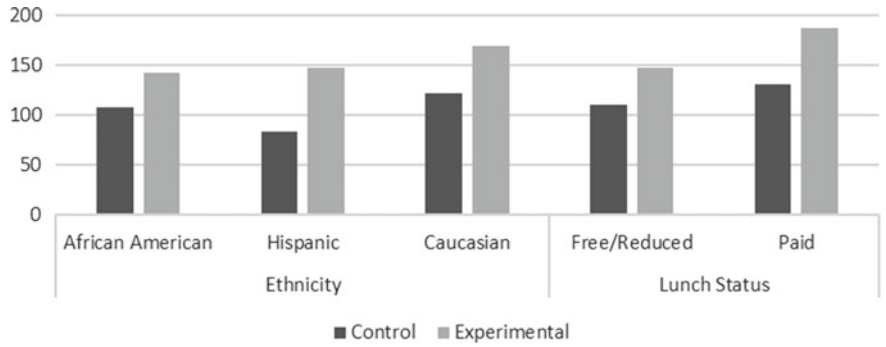

Fig. 23.6 TRC end of year Lexile levels

23.4 Conclusions

In this study, the hypothesis was upheld: Students who were randomly assigned to the experimental condition outperformed students randomly assigned to the control condition on end of year literacy assessments. Across ethnicities and socioeconomic backgrounds, students who used the computer-adaptive reading program outperformed their control counterparts. These effects were seen on two different reading assessments and across all literacy strands.

This study adds to the body of RCTs in educational research, providing evidence in favor of computer-adaptive reading programs in early elementary school classrooms. We can tentatively conclude that technology in education is beneficial to early education and early literacy skills: The kindergarten students who were randomly assigned to use the program gained essential early literacy skills to prepare them for their academic future. These findings are extremely important to the future of elementary school academics because of the high demand for research-based practices. In today's schools, teachers have high expectations for their school district, state, and country to uphold all students to proficiency standards, and these results indicate that computer-adaptive reading programs could assist students of all demographics in achieving necessary early literacy skills to meet expectations. Educational research can benefit from these results, providing more evidence in favor of technology in schools for individualized literacy assistance to all students.

Although this is an RCT, the study only included students from one grade in one school district. In the future, an RCT with various school districts would provide more conclusive evidence concerning the effectiveness of technology in education. Overall, this study is an RCT that contributes to the educational research sector but could be done in a more large-scale setting to amplify the conclusions drawn.

References

1. Masters, M.: The role of evidence in teaching and learning. In: Australian Council for Educational Research: Research Conference 2018, Melbourne, Australia (2018)
2. Devaney, L.: Survey: Teaching demands taking toll on educators. eSchool News: Daily Tech News & Innovation, https://www.eschoolnews.com/2016/05/13/survey-teaching-demands-taking-toll-on-educators/. Last accessed 1 Dec 2018
3. National Governors Association Center for Best Practices, Council of Chief State School Officers: Common Core State Standards. National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington, D.C. (2010)
4. What Works Clearinghouse: Standards handbook (Version 4.0). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf
5. Hutchison, D., Styles, B.: A guide to running randomised controlled trials for educational researchers. Slough: NFER (2010)
6. Bouguen, A., Gurgand, M.: Randomized controlled experiments in education: Report for the European Commision. EENE Analytical Report No. 11 (2012)
7. Ginsburg, A., Smith, M.S.: Do randomized controlled trials meet the "gold standard". American Enterprise Institute (2016)
8. Rowe, M., Oltmann, C.: Randomised controlled trials in educational research: Ontological and epistemological limitations. Afr. J. Health P. Educ. **8**(1), 6–8 (2016)
9. Bleses, D., Højen, A., Justice, L.M., Dale, P.S., Dybdal, L., Piasta, S.B., Markussen-Brown, J., Clausen, M., Haghish, E.F.: The effectiveness of a large-scale language and preliteracy intervention: The SPELL randomized controlled trial in Denmark. Child Dev. **89**(4), 342–363 (2018)
10. Merchant, Z., Goetz, E.T., Cifuentes, L., Keeney-Kennicutt, W., Davis, T.J.: Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: a meta-analysis. Comput. Educ. **70**, 29–40 (2014)
11. Bulman, G., Fairlie, R.W.: Technology and education: Computers, software, and the internet. In: Hanushek, E., Machin, S., Woessmann, L. (eds.) Handbook of the Economics of Education, vol. 5, pp. 239–280. Elsevier (2016)
12. Connor, C.M., Goldman, S.R., Fishman, B.: Technologies that support students' literacy development. In: Spector, J.M., Merrill, M.D., Elen, J., Bishop, M.J. (eds.) Handbook of Research on Educational Communications and Technology, pp. 591–604. Springer, New York (2014)
13. Clements, D.H., Sarama, J., Spitler, M.E., Lange, A.A., Wolfe, C.B.: Mathematics learned by young children in an intervention based on learning trajectories: a large-scale cluster randomized trial. J. Res. Math Educ. **42**(2), 127–166 (2011)
14. Gersten, R., Rolfhus, E., Clarke, B., Decker, L.E., Wilkins, C., Dimino, J.: Intervention for first graders with limited number knowledge: large-scale replication of a randomized controlled trial. Am. Educ. Res. J. **52**(3), 516–546 (2015)
15. Faber, J.M., Luyten, H., Visscher, A.J.: The effects of a digital formative assessment tool on mathematics achievement and student motivation: results of a randomized experiment. Comput. Educ. **106**, 83–96 (2017)