Chapter 9
Empirical Evidence for Benefit? Reviewing Quantitative Research on the Use of Digital Tools in Mathematics Education

Paul Drijvers

Abstract The benefit of using digital tools in education, and in mathematics education in particular, is subject to debate. To investigate this benefit, we focus on effect sizes on student achievement reported in reviews of experimental and quantitative studies. The results show significant positive effects with modest effect sizes. Possible causes for this are discussed and illustrated with one case study. We wonder if the review studies capture the subtlety of integrating digital tools in learning as much as qualitative studies do, and question their potential to address the "how" question. As a conclusion, a plea is made for replication studies and for studies that identify decisive factors through the combination of a methodologically rigorous design and a theoretical foundation in domain-specific theories from mathematics didactics.

Keywords Digital technology · Mathematics education · Effect size
Student achievement

9.1 Introduction

The benefit of using digital tools in education, and in mathematics education in particular, is subject to debate. For example, the header of a September 2015 BBC news item was "Computers 'do not improve' pupil results, says the OECD".[1] A Dutch news site[2] provided an even stronger claim: "Poorer school performance through increased computer use." Both items were based on a report by the

---

[1] 15 September 2015, http://www.bbc.com/news/business-34174796.
[2] 15 September 2015, http://nos.nl/artikel/2057772-slechtere-schoolprestaties-door-meer-computergebruik.html.

P. Drijvers (✉)
Freudenthal Institute, Utrecht University, Utrecht, The Netherlands
e-mail: p.drijvers@uu.nl

L. Ball et al. (eds.), *Uses of Technology in Primary and Secondary Mathematics Education*, ICME-13 Monographs, https://doi.org/10.1007/978-3-319-76575-4_9

Organisation for Economic Co-operation and Development (OECD) on student achievement and the use of computers, that just had been published (OECD, 2015). Indeed, the results of this study included negative correlations between mathematics performance and computer use in mathematics lessons and led to conclude that there is little evidence for a positive effect on student achievement:

> Despite considerable investments in computers, internet connections and software for educational use, there is little solid evidence that greater computer use among students leads to better scores in mathematics and reading. (OECD, 2015, p. 145)

Even if correlations do not imply causality, the "little solid evidence" in the above OECD quote at least challenges the research community. Other voices, however, point out the benefits of using digital technology in education. For the case of mathematics education, the National Council of Teachers of Mathematics claimed that we cannot and should not neglect digital tools: "Technology is an essential tool for learning mathematics in the 21st century" (NCTM, 2008, p. 1). This quote recognizes the potential of digital technology for mathematics teaching and learning, including a possibly changing focus in mathematics curricula towards conceptual understanding and higher order thinking skills. This potential is underpinned by research findings, such as the ones reported by Ronau et al.:

> Over the last four decades, research has led to consistent findings that digital technologies such as calculators and computer software improve student understanding and do no harm to student computational skills. (Ronau et al., 2014, p. 974)

Others (e.g., Hoyles & Lagrange, 2010; Hoyles & Noss, 2003) took a more nuanced stance, claiming that it is the *how* that determines the effect of ICT use on performance in mathematics education: how to design effective ICT environments and how to "exploit" them for student learning?

These different claims and opinions with respect to *if*, *how*, and *how much* to use digital tools in mathematics education raise several questions. What does empirical research really tell us about the effects on student performance of using digital technology in mathematics education? Does the answer depend on student grade, on the mathematical topic, on the type, size, scale and duration of the intervention? Do we see trends in research findings on these questions over the recent decades according to review studies? How can we explain the differences between studies? Is it possible anyway to answer such overarching questions through the review of empirical studies? What are the limitations of this approach? These questions form the core of this chapter, and will lead to considerations on the relationship between qualitative studies, addressed in more detail in Heid's chapter in this volume on the one hand, and quantitative studies and review studies on the other. A reflective stance is taken; as such, this chapter has an essay-like character rather than a traditional research paper format.

In this chapter, we will first revisit and synthesize the results of five important review studies on empirical, quantitative studies on the use of digital technology in mathematics education (Sect. 9.2). This section is central in the chapter. To illustrate the difficulty to find convincing evidence of the potential of digital tools in such (too?) general review studies, Sect. 9.3 describes one empirical study that was grounded in qualitative work and well-focused, but not successful in terms of student performance. Some possible causes are discussed. In the reflective Sect. 9.4, we reflect on the interpretation of effect sizes, the subtlety of using digital tools in mathematics education and some methodological issues. Finally, in the concluding Sect. 9.5 limitations of review studies are addressed, and a plea is made for an appropriate integration of qualitative and quantitative methods, and for methodologically rigorous studies grounded in theories on the learning of mathematics.

9.2 Revisiting Review Studies

9.2.1 Some Relevant Studies Before 2010

Of course, the question of the benefits of integrating digital tools in mathematics education is not new and has been investigated before. In this section, we briefly review early studies in the field, that is, studies that were published before 2010 that try to summarize research findings in the field. In one of the first synthesizing studies, Heid (1997) provided an overview of principles and issues of the integration of digital technology, and sketched the landscape of the different types of tools and their pedagogical potential. On the topic of using handheld graphing technology in particular, Burrill et al. (2002) reported on 43 studies and concluded that these devices can be important in helping students develop a better understanding of mathematical concepts; this conclusion, however, is not quantitatively underpinned. Ellington (2003, 2006) also focused on graphing calculators, which were indeed important in the implementation of digital tools in mathematics education at the end of the 20th century. Her review of 54 studies showed an improvement of students' operational skills and problem-solving skills when calculators are an integral part of testing and instruction. The effect sizes, however, were small—which is not uncommon in educational research. Lagrange, Artigue, Laborde and Trouche (2003) developed a multi-dimensional framework to review a corpus of 662 mostly qualitative research studies on the use of technology in mathematics education and to investigate the evolution of research in the field, to identify trends, without explicitly addressing learning outcomes. Kulik (2003) did address learning outcomes and reported an average effect size of $d = 0.38$ in 16 studies on the effectiveness of integrated learning systems in mathematics.[3]

---

[3] The effect sizes reported here are means to express the differences between two populations in terms of their pooled standard deviation. The most commonly used methods are Cohen's d and Hedge's g. The difference between the two is important for small sample sizes, but neglected in this paper as we do not want to get into measurement details too much. The d reported here means that the average difference between experimental group and the control group equals 0.38 of their pooled standard deviation, which is considered a weak to medium effect.

Two subsequent large-scale experimental studies by Dynarski et al. (2007) and Campuzano, Dynarski, Agodini and Rall (2009), however, concluded that the effects of the use of digital tools in grade 9 algebra courses was not statistically different from zero. For the use of computer algebra systems, Tokpah (2008) found significant positive effects with an average of $d = 0.38$ over 102 effect sizes.

Altogether, these early studies provided mixed findings on the effect of using digital tools in mathematics education and showed different degrees of quantitative evidence. Also, the dissemination of digital tools and the experience teachers and students had with their use in class were limited by that time. These considerations provide ample reason to look at more recent studies in more detail.

9.2.2 Five More Recent Review Studies

To further investigate more recent findings, we now focus on five review studies that provide information on the effect of using digital technology in mathematics education through reporting effect sizes.[4] The selection of these five studies is not based on a systematic database survey, but on an informal literature and Google Scholar search using terms such as review study, mathematics education, and digital technology. It is interesting to notice that the studies included in each of these review studies are very different and hardly show any overlap, due to different criteria and foci.

The first one is the study by Li and Ma (2010). It reviewed 46 studies on using five different types of computer technology (tutorials, communication media, exploratory environments, tools, and programming languages) on mathematics education in K–12 classrooms, reporting in total 85 effect sizes. The researchers found a statistically significant effect with a weighted average effect size of $d = 0.28$, which led them to report "… a moderate but significant positive effect of computer technology on mathematics achievement" (Li and Ma 2010, p. 232). The reported unweighted average effect size, $d = 0.71$, seems less appropriate as it does not take into account the number of students involved. Additional findings were that higher effect sizes were found in primary education compared to secondary, and in special education compared to general education. Also, effect sizes were bigger in studies that used a constructivist approach to teaching, and in studies that used non-standardized tests. Differences with respect to the five types of technology were not found.

The second review study by Rakes, Valentine, McGatha, and Ronau (2010) focused on algebra in particular. The authors included two studies that were also in the Li and Ma (2010) study, and found 109 effect sizes. The interventions were categorized; here we only report on the categories Technology tools (with calculators, graphing calculators, computer programs, and java applets as categories) and

---

[4] We addressed three of them in earlier publications (Drijvers, 2014, 2015).

Technology curricula, being computer-based curricula for use in onsite classes, online courses, and tutoring curricula. The average weighted effect sizes for these two categories were $d = 0.151$ and $d = 0.165$, respectively. Over all categories, the authors concluded that interventions focusing on conceptual understanding provide about twice as high effect sizes as the interventions focusing on procedural understanding. Also, they noted that interventions over a small period of time may have significant effect, and that the grain size differences in interventions (whole-school study versus single-teacher interventions) did not make a significant difference.

The third review study by Cheung and Slavin (2013) took into account 74 effect sizes from 45 elementary and 29 secondary studies on K–12 mathematics. The primary studies included one study that was also part of the Rakes et al. review; the secondary studies category included the two studies addressed in the previous paragraphs. The average effect size was $d = 0.16$. The authors' final conclusion refers to a modest difference: "Educational technology is making a modest difference in learning of mathematics. It is a help, but not a breakthrough." (Cheung & Slavin, 2013, p. 102). Some additional findings are worth mentioning. First, the overall effectiveness of educational technology did not improve over time. Second, like Li and Ma (2010), the authors found higher effect sizes in primary than in secondary education. Third, lower effect sizes were found in randomized experiments compared to quasi-experimental studies. Fourth and final, effect sizes in studies with a large number of students were smaller than in small-scale studies.

The fourth review study by Steenbergen-Hu and Cooper (2013) focused on the effectiveness of intelligent tutoring systems (ITS) on K–12 students' mathematical learning. The authors' corpus of studies had four studies in common with the Rakes et al. (2010) study. The 65 effect sizes included in their study ranged from $g = 0.01$ to $g = 0.09$. This led the authors to careful conclusions: "ITS had no negative and perhaps a very small positive effect on K–12 students' mathematical learning relative to regular classroom instruction" (Steenbergen-Hu & Cooper, 2013, p. 982). Additional findings were that the effects of the ICT interventions proved less big in cases of long interventions (more than one school year). Also, the general student population seemed to benefit more from the ITS use than their low achieving peers, which questions the potential of ITS for reducing achievement gaps.

The fifth and final study we address here is a meta-study carried out by Sokolowski, Li, and Willson (2015). The authors particularly investigated the use of exploratory computerized environments (ECEs) for grade 1–8 mathematics. The interventions focused on digital tools for supporting word problem solving and exploration. The average of the 24 effect sizes included was $g = 0.60$, which is a moderate effect size. Additional findings were that the effects were most positive in middle school grades (grades 6–8). Concerning the mathematical domain, the effect sizes tended to be slightly higher for geometry than for algebra. In terms of teaching styles, teacher-based support proved to be more effective than computer-based support, which led the authors to claim that in spite of the positive effects, "this finding does not diminish the importance of good teaching" (Sokolowski, Li, & Willson, 2015, p. 13).

Table 9.1 Effect sizes reported in five review studies

Study	Number of effect sizes	Average effect size	Global conclusion
Li and Ma (2010)	85	d = 0.28 (weighted)	Moderate significant positive effects
Rakes et al. (2010)	109	d range 0.151–0.165	Small but significant positive effects
Cheung and Slavin (2013)	74	d = 0.16	A positive, though modest effect
Steenbergen-Hu and Cooper (2013)	61	g range 0.01–0.09	No negative and perhaps a small positive effect
Sokolowski, Li, and Willson (2015)	24	g = 0.60	A moderate positive effect size

Table 9.1 summarizes the findings of the five review studies with respect to the effect sizes and their global conclusion. The overall image is that the use of digital technology in mathematics education can have a significant positive effect, with effect sizes ranging from small to moderate. The average of these (average!) effect sizes is about 0.2, and we notice quite some variation: comparing the value for g ranging from 0.01 to 0.09 in one study and being 0.6 in another, the results do not really converge. On the one hand, this is somewhat disappointing; on the other, the different studies are based on different sets of research studies with different foci. Meanwhile, we conclude that these studies do not provide an overwhelming evidence for the effectiveness of the use of digital tools in mathematics education.

Of course, this summary of review studies provides a highly (or even too?) aggregated view and neglects detailed differences. Can we learn more about decisive factors that explain these different effects? A first possible factor is *student age and student level*. Sokolowski, Li and Willson (2015) found the effects to be most positive in middle school grades (grades 6–8), whereas Steenbergen-Hu and Cooper found the highest effects in elementary school (grades K–5). Both Li and Ma (2010) and Cheung and Slavin (2013) reported higher effect sizes in primary education compared to secondary. The former also claimed that effects are higher in special education compared to general education. This is in line with the finding by Steenbergen-Hu and Cooper (2013), who concluded that the general student population seemed to benefit more from ITS use than low achieving students. In sum, evidence of benefit is larger in primary and lower secondary education, and it is not self-evident that digital tool use helps to bridge the gap between high and low achieving students. We can conjecture about the reasons for the latter point: if digital environments provide rich learning opportunities, it seems likely that high achieving students manage to better exploit these opportunities. As for grade level, we do not know why digital tools would work better for younger students; maybe other factors such as the availability of the tools and the mathematical sophistication needed play a role here?

This brings us to the second factor: the *mathematical domain*. Steenbergen-Hu and Cooper (2013) found bigger effect sizes for basic math than for algebra. The Rakes et al. (2010) study showed low effect sizes in the domain of algebra, whereas a review study by Chan and Leung (2014) reported a high effect size ($d = 1.02$) for the use of Dynamic Geometry Systems. These findings are in line with Sokolowski, Li, and Willson (2015), who reported effect sizes to be slightly higher for geometry than for algebra. Again, we wonder why this would be the case. Is using digital tools for geometry more natural, and are geometry tools more intuitively used than algebra tools that may require more syntax? These questions clearly need further investigation.

A third possible factor concerns *learning goals and teaching style*. Li and Ma (2010) found bigger effect sizes in studies that used a constructivist approach to teaching. More or less in line with this, Rakes et al. (2010) reported the largest effect sizes in studies on conceptual understanding rather than on procedural skill acquisition. Sokolowski, Li, and Willson (2015) found high effect sizes in studies explicitly focusing on word problem solving and exploration, and teacher-based support in these studies was more effective than computer-based support. Even if these findings are somewhat eclectic, they suggest that using digital tools can be effective in interventions focusing on higher-order learning goals, such as conceptual insight and problem solving, with a constructivist view on learning and with an important role for the teacher. These findings are interesting as they may challenge the view of digital tools mainly supporting skill acquisition with no important role for the teacher.

Possible external factors that might impact on learning effects are the intervention's *duration and sample size*. Rakes et al. (2010) showed that short interventions may have significant effect, and Steenbergen-Hu and Cooper (2013) claimed that interventions shorter than one school year are more effective than longer ones. It seems that short interventions do not necessarily lead to weaker effects. With respect to sample size, Cheung and Slavin (2013) found that effect sizes in studies with a large number of students were smaller than in small-scale studies. Steenbergen-Hu and Cooper (2013) reported higher effect sizes for studies with less than 200 participants. In contrast to this, Rakes et al. (2010) found that single-teacher interventions were not more effective than whole-school interventions. Apparently, the picture with respect to sample size remains unclear.

As a final factor, we briefly address the *development over time*. Over the last decades, digital tools for mathematics have become more sophisticated, ICT infrastructures have drastically improved both in schools and at home, and both teachers and students have become more familiar with using ICT in education. Therefore, one might expect the benefits for student achievement to increase over time. If we consider these review studies in more detail, however, we agree with Cheung and Slavin (2013) and with Steenbergen-Hu and Cooper (2013) that the effect sizes reported in the different research reports did not significantly increase over time. A possible explanation might be that there indeed is a positive development over time, but that it is compensated by other factors, such as more rigorous study designs and methods, and bigger sample sizes.

One might wonder if *publication bias* might play a role in the review studies addressed above. Would it be possible that actual effect sizes are smaller, due to the fact that studies that did not result in significant effects were not published? Most review studies took this into account. For example, both Steenbergen-Hu and Cooper (2013) and Sokolowski, Li, and Willson (2015) found little evidence that publication bias had impact on their findings.

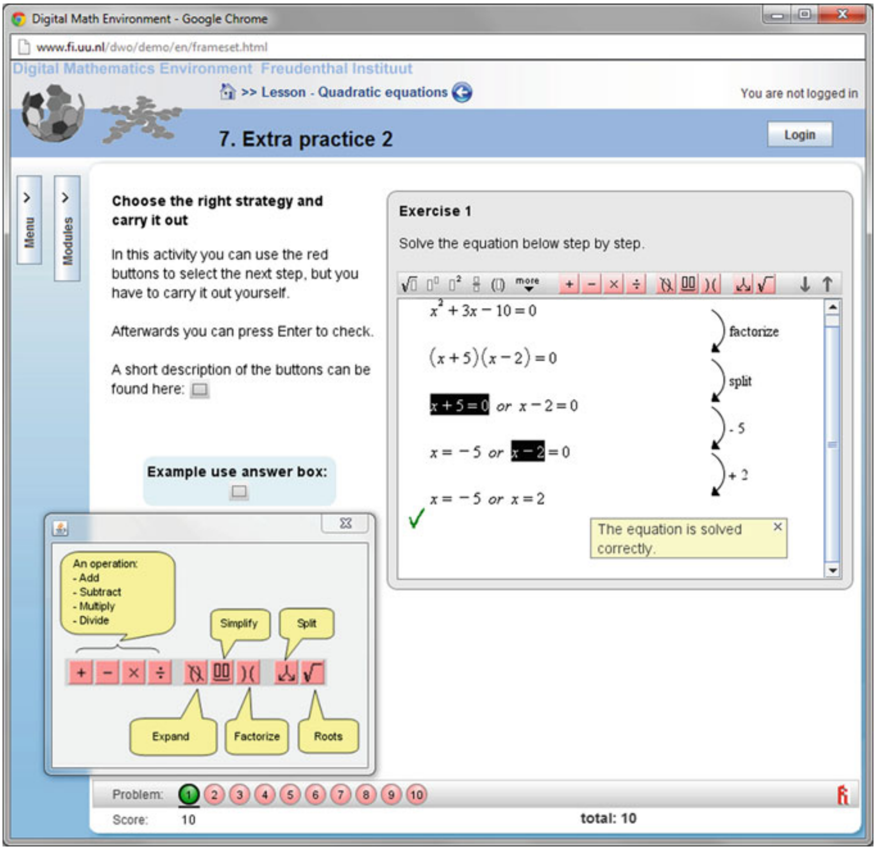
All in all, the review studies show that the use of digital technology in mathematics education can have a significant positive effect, with effect sizes ranging from small to moderate and with considerable variation in size. Benefits seem to be best for younger students (primary level or early secondary), better for geometry than for algebra, effective in interventions focusing on higher-order learning goals, and already beneficial in short interventions. Over the last decades, effect sizes do not increase and publication bias does not seem to play a role in this picture.

9.3 An Example: The Case of Applets for Algebra

The picture provided by review studies, however, is limited. Different types of interventions, students, mathematical domains and digital tools are merged into one global average effect size. Would this merging of studies with different perspectives explain the modest overall benefits in terms of student performance? In this section, we counterbalance the global picture by briefly presenting one single empirical study that reported no significant results. It illustrates that, in spite of a focus on one mathematical topic and one type of digital tool and the qualitative preparatory study, providing empirical evidence for the benefits of using digital tools is not straightforward. Some tentative explanations will be provided.

In the study (Drijvers, Doorman, Kirschner, Hoogveld, & Boon, 2014), two online algebra modules were used in 8th grade. The modules were designed in the Digital Math Environment, which proved to be successful in improving student achievement in algebra in grade 12 (Bokhove & Drijvers, 2012). Also, teachers had reported success while implementing the online materials in lower grades. Figure 9.1 shows a task from one of the modules.

The study had an experimental design, in which each of the involved teachers taught to two classes in parallel, each randomly assigned to the experimental condition of using the online modules, or to the control condition of regular teaching. Figure 9.2 shows the results of the pretest, the intermediate test (Post_Linear), the posttest (Post_Quadractic) and the two retention tests, all administered with paper-and-pen. In spite of the earlier positive experiences with these types of modules, the results show that the experimental condition did not lead to students outperforming their peers in the control condition. The experimental group did not catch up the small initial (and coincidental) lag; indeed, this gap became significantly larger in the final retention test.



Choose the right strategy and carry it out

In this activity you can use the red buttons to select the next step, but you have to carry it out yourself.

Afterwards you can press Enter to check.

A short description of the buttons can be found here:

Example use answer box:

Exercise 1

Solve the equation below step by step.

$x^2 + 3x - 10 = 0$

$(x+5)(x-2) = 0$

$x+5=0$ or $x-2=0$

$x=-5$ or $x-2=0$

$x=-5$ or $x=2$

The equation is solved correctly.

Fig. 9.1 Algebra task in the Digital Math Environment

As a possible explanation for these findings, the authors mention a spill-over effect. All participating teachers taught one control and one experimental class, and they may have picked up pedagogical ideas from the online intervention and used these in the control classes as well. Such a spill-over effect is well-documented in research literature (e.g., see Creemers, Kyriakides, & Sammons, 2010). A second possible explanation is that the work on the online tasks was not an adequate preparation for more complex tasks. Third, the feedback provided in the digital environment might have lacked quality, and, finally, the integration of paper-and-pen skills and digital practice might not have been optimal, so that transfer to the “traditional media” was hindered. In short, in spite of a careful experimental design and an environment that had proven to be useful in other settings, the researchers did not find a positive effect.

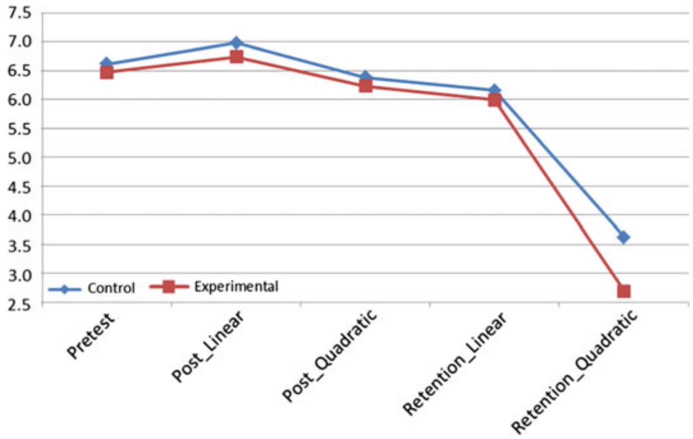


Fig. 9.2 Average grades for control and experimental group (N = 842)

9.4 Reflection

To reflect on the above findings from quantitative studies, we first discuss the interpretation of effect sizes and next address two other factors that may play a role: the too general claims made, which ignore the subtlety of using digital tools for learning, and the methodological weaknesses that some studies suffer from.

9.4.1 Interpreting Effect Sizes

First, let us notice that the results from experimental and quantitative studies are more positive than the correlational findings from the OECD (2015) study cited in this chapter's introduction. However, the effect sizes, with their overall average in the order of $d = 0.2$, are modest. How do we interpret them? Higgins et al. (2012) claimed that technology-based interventions produce just slightly lower effect sizes than other types of educational interventions not involving digital tools, thus suggesting that these results are not that disappointing. Slavin (2016) supported this stance, pointing out that the interpretation of an effect size mainly depends on two factors: the sample size and whether or not the students are assigned randomly to the different conditions. For a number of large scale studies with random assignment on different topics, Slavin found an average effect size of 0.11, suggesting that it is very optimistic to expect more. From this perspective, the reported effect sizes are not that low. In the meanwhile, the interpretation of effect sizes should be done with care and is subject to debate, as is the case for the interpretation of significant *p*-values.[5]

9.4.2 The Too General Claims that Ignore the Subtlety of Using Digital Tools for Learning

Would we not all agree that research findings such as "The use of paper-and-pen has a positive effect on student achievement" would be too general? Why, then, would we try to find evidence for similar claims on the use of ICT? It makes sense to assume that digital technology is not a panacea, and that its effectiveness will largely depend on particular implementations and situations. The following two quotes underline that the effect of ICT in mathematics education is a subtle matter and will depend to an important extent on the specific technological application, the educational setting and the orchestration by the teacher. It is the "how" that counts!

> The range of impact identified in these studies suggests that it is not whether technology is used (or not) which makes the difference, but how well the technology is used to support teaching and learning. There is no doubt that technology engages and motivates young people. However this benefit is only an advantage for learning if the activity is effectively aligned with what is to be learned. It is therefore the pedagogy of the application of technology in the classroom which is important: the how rather than the what. (Higgins, Xiao, & Katsipataki, 2012, p. 3)

> There have been several reviews of the benefits of ICT to student learning in mathematics that suggest positive effects from the use of digital technology. […] However, the type and extent of the gains are a function of how the technology is used in the teaching of mathematics. (Drijvers, Monaghan, Thomas, & Trouche, 2015, p. 15).

If we agree that the learning of mathematics is a complex domain and that we need to know more about the factors that determine the contribution of digital tools to it, it is important that research is grounded in theoretical knowledge from domain-specific mathematics pedagogy and from man-machine interaction. To mention just some possible perspectives, theories on reification (Sfard, 1991), on emergent modeling (Doorman, Drijvers, Gravemeijer, Boon, & Reed, 2012), or on instrumental genesis may offer such a theoretical basis (Drijvers, Kieran, & Mariotti, 2010). Educational research on the use of digital tools for mathematics education that is not based on domain-specific didactical knowledge may miss opportunities to discover decisive factors.

As an aside, we should note that didactical knowledge and practice may also change under the influence of digital technology. In fact, this is what the OECD mentions as a possible explanation for their surprising findings:

---

[5] For a current debate on *p*-values see http://www.statslife.org.uk/news/2116-academic-journal-bans-p-value-significance-test.

> … we have not become good enough at the kind of pedagogies that make the most of technology. […] Technology can amplify great teaching but great technology cannot replace poor teaching (OECD 2015, pp. 3–4).

In this line of reasoning, an important research question would be “What type of student achievement can be improved through which type of use of which kind of digital tools?” rather than the very general “Does the use of digital tools improve student achievement?”

9.4.3 Methodological Limitations

In this chapter, we limited ourselves to review studies that summarize the results from experimental studies. The body of such experimental studies shows some remarkable methodological characteristics. First, replication studies have hardly ever been carried out. Why is this the case? If replication studies had been done, would we encounter similar replication issues as in the field of cognitive and social psychology?[6] Do we manage to control relevant variables? Second, it is interesting to notice that smaller studies tend to report bigger effect sizes than larger ones and that the reported effect sizes do not seem to increase over time. This suggests that scaling up successful interventions identified in effective small-scale studies may not be so easy. As far as the trend over time is concerned, the criteria for publication and for inclusion in review studies seem to be getting higher, and this is indeed what we should strive for according to Ronau and colleagues, who in a recent study on the quality of 480 mathematics education technology dissertations argued for higher quality in both research reports and reviews:

> The mathematics education technology research community must in turn begin to demand greater quality in its published studies, through both how researchers write about their own studies and how they review the works of others. (Ronau et al. 2014, p. 1002)

A possible cause of the lack of positive trends in reported effect sizes, therefore, might be these higher methodological standards, which might filter out the studies that report high effect sizes. From a methodological point of view, more rigor in research methods to improve the quality of our results is welcomed of course.

---

[6] See, for example, http://www.theguardian.com/science/2015/aug/27/study-delivers-bleak-verdict-on-validity-of-psychology-experiment-results.

9.5 Conclusion

In the introduction, we raised the question of what empirical research really tells us about the effects on student performance of using digital technology in mathematics education. The literature review revealed mixed results. The OECD correlational study showed little evidence for benefit. Experimental studies, and their review studies in particular, reported significant positive effects, with average effect sizes ranging from small to moderate with considerable variation. Compared to effect sizes reported for other types of innovative interventions, the evidence for benefit is not overwhelming. Also, insight into factors that are decisive for the (lack of) positive benefit of the use of digital tools is limited. Younger students (primary level or early secondary) seem to benefit more, results are better for geometry than for algebra, interventions focusing on higher-order learning goals may be effective, and short interventions may be beneficial. Over the last decades, effect sizes do not increase and publication bias does not seem to play a role in this picture.

Of course, the above conclusion has some important limitations. First, review studies are based on studies that themselves are older, and one might wonder if the picture has changed over, say, the last five years. The fact that effect sizes so far have not been increasing, however, does not favor this argument. Second, we focus on experimental, quantitative studies and neglect qualitative studies and studies that follow a design research paradigm, whereas such studies can contribute to the body of knowledge, and in many cases take an in-depth view on student learning and are firmly grounded in theories from the field of mathematics didactics.[7] The study described in Sect. 9.3 shows that there can be many reasons why the effect of using of ICT in mathematics education may not show up. A third limitation of the type of review studies revisited is that these studies do not differentiate between educational levels, types of technology used, and other educational factors that may be decisive. Rather, they provide an overview without nuances, which may cause us to miss important insights in the phenomenon.

In spite of these limitations, the conclusion is that evidence for the benefit of using technology in mathematics education from experimental studies is modest and that evidence-based insights in factors that affect these benefits are limited. What we need on our research agendas, therefore, are studies (including replication studies) that focus on the identification of decisive factors that determine the eventual benefits in specific cases. Such studies should on the one hand be methodologically well designed according to the standards from educational science, and on the other hand be strongly based in sound theoretical foundations from domain-specific mathematics didactics, as to better address the "how"-question. In many cases, preliminary qualitative studies may show to be indispensable to set up learning arrangements that also will result in positive effects in experimental studies. To combine the best of both worlds is the challenge we are facing.

---

[7] The findings from qualitative studies are addressed in the chapter by Heid in this volume.

References

Bokhove, C., & Drijvers, P. (2012). Effects of a digital intervention on the development of algebraic expertise. *Computers & Education, 58*(1), 197–208.

Burrill, G., Allison, J., Breaux, G., Kastberg, S., Leatham, K., & Sanchez, W. (2002). *Handheld graphing technology at the secondary level: Research findings and implications for classroom practice*. Dallas, TX: Texas Instruments Corp. http://education.ti.com/research.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts—Executive summary* (NCEE 2009–4042). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Chan, K. K., & Leung, S. W. (2014). Dynamic geometry software improves mathematics achievement: Systematic review and meta-analysis. *Journal of Educational Computing Research, 51*(3), 311–325.

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K–12 classrooms: A meta-analysis. *Educational Research Review, 9,* 88–113.

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in educational effectiveness research*. London/New York: Routledge.

Doorman, M., Drijvers, P., Gravemeijer, K., Boon, P., & Reed, H. (2012). Tool use and the development of the function concept: From repeated calculations to functional thinking. *International Journal of Science and Mathematics Education, 10*(6), 1243–1267.

Drijvers, P. (2014). Digital technology in mathematics education: A reflective look into the mirror. In J. Roth & J. Ames (Eds.), *Beiträge zum Mathematikunterricht 2014* (pp. 21–28). Münster: WTM-Verlag.

Drijvers, P. (2015). *Denken over wiskunde, onderwijs en ICT*. Inaugurele rede. [Thinking about Mathematics, Education and ICT. Inaugural lecture.] Utrecht: Universiteit Utrecht.

Drijvers, P., Doorman, M., Kirschner, P., Hoogveld, B., & Boon, P. (2014). The effect of online tasks for algebra on student achievement in grade 8. *Technology, Knowledge and Learning, 19,* 1–18.

Drijvers, P., Kieran, C., & Mariotti, M. A. (2010). Integrating technology into mathematics education: Theoretical perspectives. In C. Hoyles & J.-B. Lagrange (Eds.), *Mathematics education and technology—Rethinking the terrain* (pp. 89–132). New York/Berlin: Springer.

Drijvers, P., Monaghan, J., Thomas, M., & Trouche, L. (2015). *Use of technology in secondary mathematics. Final report for the international baccalaureate*. http://www.ibo.org/globalassets/publications/ib-research/technologyindpmathematicsfinalreport.pdf.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., & Campuzano, L. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort. Report to Congress.* Publication NCEE 2007–4005. Washington, DC: U.S. Department of Education.

Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education, 34*(5), 433–463.

Ellington, A. J. (2006). The effects of non-CAS graphing calculators on student achievement and attitude levels in mathematics: A meta-analysis. *School Science and Mathematics, 106*(1), 16–26.

Heid, M. K. (1997). The technological revolution and the reform of school mathematics. *American Journal of Education, 106*(1), 5–61.

Higgins, S., Xiao, Z., & Katsipataki, M. (2012). *The impact of digital technology on learning: A summary for the education endowment foundation.* Report. Durham: Durham University.

Hoyles, C., & Lagrange, J.-B. (Eds.). (2010). Mathematics education and technology-rethinking the Terrain. In *The 17th ICMI Study*. Springer US.

Hoyles, C., & Noss, R. (2003). What can digital technologies take from and bring to research in mathematics education? In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. Leung

(Eds.), *Second international handbook of mathematics education* (Vol. 1, pp. 323–349). Dordrecht: Kluwer Academic Publishers.

Kulik, J. A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say*. Arlington, VA: SRI International. http://www.sri.com/policy/csted/reports/sandt/it/Kulik_ITinK-12_Main_Report.pdf.

Lagrange, J.-B., Artigue, M., Laborde, C., & Trouche, L. (2003). Technology and mathematics education: A multidimensional study of the evolution of research and innovation. In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 239–271). Dordrecht, Netherlands: Kluwer Academic Publishers.

Li, Q., & Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educational Psychology Review, 22,* 215–243.

National Council of Teachers of Mathematics. (2008). *Technology in teaching and learning mathematics. A position of the national council of teachers of mathematics*. http://www.nctm.org/Standards-and-Positions/Position-Statements/Technology-in-Teaching-and-Learning-Mathematics/.

OECD. (2015). *Students, computers and learning. Making the connection*. Paris: OECD Publishing. https://doi.org/10.1787/9789264239555-en. http://www.oecd.org/edu/students-computers-and-learning-9789264239555-en.htm.

Rakes, C. R., Valentine, J. C., McGatha, M. B., & Ronau, R. N. (2010). Methods of instructional improvement in Algebra: A systematic review and meta-analysis. *Review of Educational Research, 80*(3), 372–400.

Ronau, R. N., Rakes, C. R., Bush, S. B., Driskell, S. O., Niess, M. L., & Pugalee, D. K. (2014). A survey of mathematics education technology dissertation scope and quality: 1968–2009. *American Educational Research Journal, 51*(5), 974–1006.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics, 22*(1), 1–36.

Slavin, R. E. (2016). What is a large effect size? *Huffington Post*. http://www.huffingtonpost.com/robert-e-slavin/what-is-a-large-effect-si_b_9426372.html.

Sokolowski, A., Li, Y., & Willson, V. (2015). The effects of using exploratory computerized environments in grades 1–8 mathematics: A meta-analysis of research. *International Journal of STEM Education, 2*(1), 1–17.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*(4), 970–987.

Tokpah, C. L. (2008). *The effects of computer algebra systems on students' achievement in mathematics*. Doctoral dissertation. Kent, OH: Kent State University. https://etd.ohiolink.edu/!etd.send_file?accession=kent1215528446&disposition=inline.