KNOWLEDGE CENTRE
FOR EDUCATION

A SYSTEMATIC MAPPING OF THE EFFECTS OF ICT ON LEARNING OUTCOMES

KONRAD MORGAN, MADDY MORGAN, LOTTA JOHANSSON & ERIK RUUD

KNOWLEDGE CENTER FOR EDUCATION
Drammensveien 288, 0283 Oslo
Adress: Postboks 564, NO-1327 Lysaker

ISBN: 978-82-12-03560-7
Referance number: KSU 4/2016
Published: November 2016
Photo: Shutterstock

TITLE: Morgan, K., Morgan, M., Johansson, L. & Ruud, E. (2016). A systematic mapping of the effects of ICT on learning outcomes. Oslo. Knowledge Center for Education. www.kunnskapssenter.no

KNOWLEDGE CENTER FOR EDUCATION:
PHONE: +47 22 03 70 00
E-MAIL: kunnskapssenter@forskningsradet.no
WEBPAGE: www.kunnskapssenter.no
FACEBOOK: kunnskapssenter
TWITTER: kunnskapsrad

CONTENT

EXECUTIVE SUMMARY: ICT EFFECTS ON EDUCATIONAL OUTCOMES

Since the first use of computers in classrooms in the 1960's there has been significant interest from educational stakeholders in determining answers to some fundamental questions about how Information and Communications Technology (ICT) impacts educational outcomes. Specifically:

- Does ICT provide a positive influence on academic performance and if it does are there subjects or disciplines that are more strongly influenced or less strongly influenced than others?
- Does ICT improve the effectiveness of the learning process and if it does what aspects of ICT make the strongest improvements on learning?

This systematic mapping of research in the field shows that so far, these questions cannot be answered as clearly and consistently as policy makers and practitioners might hope. The systematic mapping provides a summary of rigorous empirical studies in the fields of educational ICT to determine the causal effect of the use of ICTs on students' learning outcomes.

Many of the 30 included studies are systematic reviews and meta-analyses. The total number of studies included in this mapping review exceeds 1900, spanning several decades. Studies were assessed on quality and relevance and categorized under three broad themes with subcategories:

While few studies document convincing effects of ICT on students' learning outcome, an analysis across studies shows a consistent, but small positive impact from the use of ICT in classroom settings. Although some research has reported large Effect Sizes (ES) (>> +2,0) from novel technology implementations, the more rigorous meta-analyses of large scale randomized control studies, consistently reports ES's in the range of +0,1 to +0,3. The most important finding being that the highest ES's from such comprehensive and rigorous analyses are associated with studies where ICT has been implemented as a planned part of a comprehensive teaching environment with clear goals, teaching plans, teaching materials, supporting technical resources, teacher training and development. In such a context the improvements associated with ICT in education are to be viewed as part of a broader improvement in the educational environment and not just as a single technology.

Educational technology	**Design features**	**Pedagogical aspects of teaching and learning with ICT**
Desktop Computer systems	Intelligent tutoring systems	Blended learning
Mobile device systems	System design features	Assessment and feedback
Game based systems		Educational psychology

Table 1: The themes and subcategories of the report.

1. INTRODUCTION

1.1 TECHNOLOGY AND EDUCATIONAL POLICY

Over the last decades, our daily lives are increasingly influenced by information and communication technology (ICT). However, whilst society at large has experienced extensive changes due to technological development, the OECD reports concerns that institutionalized education appears to be lagging behind in this development (OECD 2015[1]). Previously, ICT was regarded as a potential online resource for broadening access to higher education[2]. More recently, however, this development trend has been accompanied by a strong belief in technology's inherent potential for transformational educational practice and improved student learning.

ICTs encompass a set of devices, tools, modalities, programs, etc. expected to strengthen the educational context. At the same time ICT, and the employment of ICTs, represents a skill in itself. As digital skills become increasingly important in all domains of society, formal education represents an essential arena for developing a digitally native generation, equipped with these desirable competencies, and ready for the labor market (OECD 2015)[3]. Digital skills were promoted as one among five basic skills (along with reading, writing, arithmetic and oral skills) in the Norwegian National Curriculum for Knowledge Promotion in Primary and Secondary Education and Training from 2006. This challenges traditional educational provision and instructional practices, a development trend further emphasized with the launch of White Paper 28 (2015-2016) [4].

This mapping of research is undertaken to state whether ICTs contribute to improving students' learning outcome. The intention is to supplement the existing knowledge base about technology use in education by asking: What may – realistically – be expected from introducing digital technology in educational settings?

1.2 EXPECTATIONS OF ICT IN EDUCATION

These are among the most commonly mentioned expectations on the use of technology in education:

STRUCTURAL	
Democratic	Digital skills are essential in contemporary society and, as OECD conclude, also in the labour market. 21 century citizens must master technology, both with regards to digital literacy and the technology itself. Formal education has a key role in developing digitally literate population, preparing them for participation in the increasingly complex information society. Since classroom instruction demands presence, online education enhances access and may contribute to improved equality in education.
Cost-effective	Technology is ubiquitous, and assumed to offer cost-effective alternatives to traditional education. The prevailing question within this context is how much of the traditional education can be replaced, without compromising teaching quality, student performance or educational outcomes.

Table 2: Structural expectations of ICT in education.

1 OECD (2015), *Students, Computers and Learning: Making the Connection,* PISA, OECD Publishing. *http://dx.doi.org/10.1787/9789264239555-en*

2 Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. Teachers College Record, 115(3), 1-47.

3 OECD (2015), *OECD Skills Outlook 2015: Youth, Skills and Employability*, OECD Publishing. http://dx.doi.org/10.1787/9789264234178-en

4 Meld. St. nr. 28 (2015-2016) *Fag – fordypning – forståelse. En fornyelse av Kunnskapsløftet*

PEDAGOGICAL	
Increased possibilities	Initially, ICT was expected to deliver content, but expectations on its inherent potential have risen with the increasing complexity and capability of technology (Schmid et al. 2014). Educational technology is currently regarded as something that can add value to education in general; not as a replacement of existing provision. Ed Tech is often designed to make possible what has been impossible in traditional education, such as expanded classroom possibilities (e.g. simulation-programs for medical training), self-regulated learning (students may themselves via devices formulate e.g. learning criteria and assess each other), greater possibilities for personalized learning (led, tutored and assessed by a computer/other device).
Increased motivation and learning	Compared to traditional education, children, youths and adults are expected to be engaged by technology (Edutainment). ICTs are subsequently assumed to spur motivation for learning – thus influencing student attainment and completion rates. Wide ranges of game based learning software (often defined as serious games) and simulation based learning have been designed in order to address this expectation.

Table 3: Pedagogical expectations of ICT in education.

Many of the included studies refer to such policy expectations when stressing the importance of the research field. As they are designed and conducted to measure the causal effect of ICT based educational interventions on students' learning outcomes, they provide little or no information about *how* the use of technology in schools may support students' development of digital skills or teachers' digital competence.

1.3 EXPANSION OF ICT IN EDUCATION

While there are many ways to map the expansion of ICT in education, in the end it all comes down to provision and use. The provision of technology is an obvious prerequisite for the use of technology. According to an OECD report from 2015, 99% of the Norwegian students have access to a computer at home (OECD 2015)[5]. In Norwegian schools, there are on average three students for every two computers. However, and with all kinds of technological devices included, the provision of ICT in Norwegian schools is widespread, making Norwegian schools among the most technologically developed schools in the world (OECD 2015[6]; European Commission 2013[7]). The provision is, however, solely a prerequisite for the effective employment of technological devices, which appears to be a more complex topic.

A report from the European Commission (2013[8]), finds no correlations between the level of computer provision in school and the frequency of use. Too often, technological devices constitute unused classroom resources. Thus, provision of technology in classrooms is no guarantee of usage, and even less for effective usage. The EC report also finds that Norwegian teachers (especially at higher grades) disagree about the relevance and positive impact of the use of ICT, with regards to transversal skills, higher order thinking skills, achievement, motivation and collaborative work. Between 25-50 % of the Norwegian students in the 11th grade are being taught by teachers who are skeptical towards the educational potential in technology (European Commission 2013[9]). The paradox is therefore that while Norwegian schools are technologically well-equipped, Norwegian teachers are among the most technology-skeptical teachers in Europe. Based on the identified gap between provision and frequency of use, the report from the EU Commission expresses a need for a policy shift. Having met the goals for provision, the focus should be on developing teachers' competence in integrating ICT in their teaching practice. Additionally, the use of ICT in teaching needs the involvement and support of all stakeholders, also by policy and strategies, thus highlighting school leaders' active engagement. Several studies included in the mapping show the potential inherent in integrating technology in ordinary classroom instruction. For example,

5 OECD (2015), *Students, Computers and Learning: Making the Connection,* PISA, OECD Publishing. *http://dx.doi.org/10.1787/9789264239555-en*

6 ibid.

7 European Commission (2013). Survey of Schools: ICT in Education. Benchmarking access, use and attitudes to technology in Europe's schools. European Commission.

8 ibid.

9 ibid.

Bernard et al. (2014) investigated the effects of blended learning (combination between ordinary classroom instruction and online education) on learning outcomes, and find that the effects almost double when some kind of human interaction is included, thus highlighting the inherent potential in the social aspects of technology use in the classroom. Archer et al. (2014) found that training and support encourages increased use of technology and stresses the importance of training and supporting teachers on how to use technology in their classroom instruction.

1.4 EFFECT ON LEARNING OUTCOMES

According to the OECD, the investment of ICT has a weak; and sometimes even negative, correlation with student performance. Even in computer-specific tasks, such as digital reading, Norway scores just above the OECD average. Based on the results from PISA 2012 and supplementing research, the report suggests that the increased access to computers in itself does not lead to significant advances in learning. When positive effects are registered, it is restricted to certain outcomes as well as certain uses of computers (OECD 2015, p. 163[10]). This report finds that it is the context of use and not the digital tool in itself which determine successful educational outcomes.

1.5 WHY A SYSTEMATIC MAPPING IS NEEDED

This systematic mapping of the effects of ICT on students' learning outcome shows that the field is inherently heterogenic and pervaded by conflicting ideologies, influenced by many stakeholders and agendas. Obviously, the nature of the topic creates a field of research in constant flux, which makes it hard to study. This is, however, a strong argument for a mapping of research in the field and a summary of research findings.

As this is a systematic mapping of the *effects* of ICT on learning outcomes, the report includes studies with the potential to measure effects; either as single studies (designed as randomized controlled trails or quasi-experimental studies), or as meta-analyses and systematic reviews summarizing several single studies with these designs. Studies aiming to investigate causal effect can potentially answer *what* works-questions. Effect studies are not designed to answer questions on *why* and *how* something works/ does not work. Hence, this systematic mapping can potentially provide insights into the most effective teaching interventions using technology.

1.6 HOW THIS REPORT IS STRUCTURED

After a brief introduction describing the provision and use of technology in education, as well as the expectations associated with it, the method of the report is presented. The search is described, and characteristics of a systematic mapping explained. The effect sizes used in the report are explained throughout the report. The two following chapters present the results of the systematic mapping. The first of these chapters present studies focusing mainly on technological aspects of ICT in education, either as devices or software. In the second chapter, the focus is primarily on the pedagogical aspects of the use of technology, investigating how technology can enhance education practice and instruction, and thus advance learning outcomes. In the final chapter, overall features identified across the included studies are described and discussed, in light of expectations on the use of technology in education. Also, knowledge gaps are presented.

This mapping is conducted to investigate if and how ICT influences learning outcomes. All included studies use variables on learning outcome and student performance, but they define these variables differently (academic achievement, cognitive skills, physical skills, literacy etc.) and use various measures (e.g. tasks, tests, grades). In general, studies trying to investigate the impact of different kinds of interventions on learning outcomes struggle with documenting clear effects.

10 OECD (2015), *Students, Computers and Learning: Making the Connection,* PISA, OECD Publishing. *http://dx.doi.org/10.1787/9789264239555-en*

2. METHOD

2.1 SYSTEMATIC RESEARCH MAPPING

Systematic mapping is one among several formats for systematically collecting, assessing, analyzing and summarizing research. It can be used to describe the current state of knowledge for a particular topic or research field, but unlike a systematic review, a mapping does not synthesize the research findings. The format is particularly useful for policy-makers and practitioners, as it covers the breadth of a theme and gives an overview that is well suited to answer their questions.

Methodology for systematic mapping was originally developed by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)[11]. It has similarities with scoping studies that aim to rapidly map the concepts underpinning a research area and the main sources and types of evidence available.[12] Systematic maps and coping studies can be undertaken as stand-alone projects in their own right, especially where the area is complex or has not been reviewed comprehensively before.[13] Systematic mapping follows the same rigorous, objective and transparent processes as do systematic reviews, including extensive search strategies and the fact that there is always more than one person involved in each step of the mapping process. Study results are often not included in systematic maps as no synthesis of results is undertaken[14]. However, there are cases when the authors have included data relating to results in the mapping report[15], as this can be used to inform the synthesis step in a future systematic review or it is perceived as beneficial for the particular research theme. In this case, the mapping includes studies measuring the effect of educational technologies on learning outcomes, and effect sizes reported in the studies are included in the systematic mapping. The Norwegian Knowledge Centre for Education has conducted this mapping as a contribution to the ongoing debate about the effect of digital technology on students' learning outcome.

This systematic mapping follows procedures outlined for systematic reviews[16], and specific pre-defined screening criteria were used to assess studies for inclusion and exclusion.

2.2 TOPIC FOR THE RESEARCH MAPPING

This systematic mapping aims to document the effects of ICT on students learning outcomes, and thus contribute knowledge about what realistically may be expected from introducing digital technology in schools.

11 Peersman, G. (1996). *A descriptive mapping of health promotion studies in young people*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
Gough, D., Kiwan, D., Suttcliffe, K., Simpson, D., & Houghton, N. (2003). A systematic map and synthesis review of the effectiveness of personal development planning for improving student learning.
Gough, D., Olivier, S. and Thomas, J. (2012): An introduction to systematic reviews, p45-46. London: Sage publications.

12 Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology, 8, 19–32.

13 Mays, N., & Roberts, E. (2001). Synthesising research evidence. Studying the organisation and delivery of health services: Research methods. N. Fullop, P. Allen, A. Clarke and N. Black.

14 James, K. L., Randall, N. P., & Haddaway, N. R. (2016). A methodology for systematic mapping in environmental sciences. *Environmental Evidence*, *5*(1), 1.
Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: a systematic mapping study. *Educational Technology & Society*, *18*(3), 1-14.

15 Cruz, S., da Silva, F. Q., & Capretz, L. F. (2015). Forty years of research on personality in software engineering: A mapping study. *Computers in Human Behavior*, *46*, 94-113.
Randall, N. P., Donnison, L. M., Lewis, P. J., & James, K. L. (2015). How effective are on-farm mitigation measures for delivering an improved water environment? A systematic map. *Environmental Evidence*, *4*(1), 1.

16 Gough, D., Olivier, S. and Thomas, J. (2012): An introduction to systematic reviews, p 156. London: Sage publications.

2.3 DATABASES

Online searches were conducted in six databases on 18 January 2016: Education Resources Information Center (ERIC), Applied Social Sciences Index and Abstracts (ASSIA), International Bibliography of the Social Sciences (IBSS), ProQuest Education Journals (PQEJ), Scopus and Psycinfo.

Database searches were limited to peer-reviewed articles published after 1 January 2013, and the language is limited to English, Norwegian, Swedish or Danish.

2.4 SEARCH STRING

The search string (Attachment 1) was derived from the research topic and designed to find empirical studies related to the use of Information and Communications Technology in educational settings that had reported objective learning outcomes related to technology based education. The database searches were performed 18 January 2016 and provided 2649 studies.

2.5 SCREENING OF STUDIES FOR INCLUSION AND EXCLUSION

All the references were imported to the EPPI-Reviewer 4 software (ER4), developed by the EPPI Centre at the University College London. Following removal of 740 duplicate references, the remaining 1909 studies were screened for inclusion and exclusion in two steps by three independent researchers (Figure 1).

This systematic mapping follows acknowledged procedures outlined for systematic reviews[17], and specific pre-defined screening criteria were used to assess studies for inclusion and exclusion. The screening process involved two steps. In Step 1 the studies were assessed based on title and abstract, and in Step 2 assessments were based on full text (for screening criteria, see Appendix 2). A total of 1853 studies were excluded in Step 1, and 26 studies were excluded in Step 2 (Figure 1).

After the completion of these stages a total of 30 studies (of which 23 were systematic reviews and meta-analysis) were included in the report.

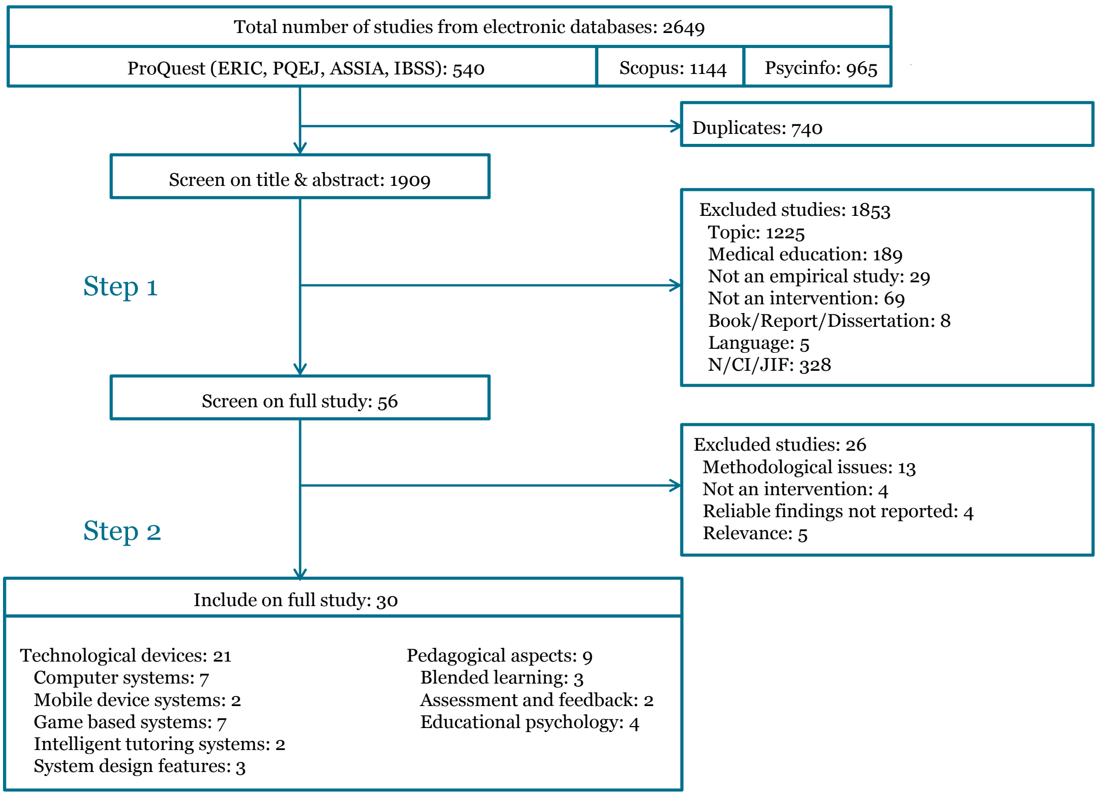

Figure 1: Flow diagram

17 Gough, D., Olivier, S. and Thomas, J. (2012): An introduction to systematic reviews, p 156. London: Sage publications.

2.6 MAPPING

The remaining 30 papers were then mapped into their respective types of investigation: Systematic review (2 studies), Cluster RCT (2 studies), Quasi experimental design (5 studies), or Meta-analysis (21 studies); And in 1) different types of educational technology (computer systems (7 studies), mobile device systems (2 studies) and game based systems (7 studies)), design features (5 studies), and 2) pedagogical aspects of teaching and learning with ICT (blended learning (3 studies), assessment and feedback (2 studies), educational psychology (4 studies).

Once these stages were completed the literature review took place.

2.7 EXPLANATIONS OF THE USE OF EFFECT SIZES

An ***effect size*** is a statistical technique for measuring the size of a difference between two groups, usually a control and an intervention within a social science context, such as a controlled comparison of a new technique in education (for a more thorough description, see appendix 3). The majority of the included studies (20) use Cohen's *d* to measure the size of the effect. In order to correct for small sample bias, d gets in many studies converted to an unbiased estimator denoted as g (Hedge's *g*) (see table 4).

In addition to the commonly used Cohen's d and Hedge's g, the measurements occur in the review are presented in table 5.

Small = An effect size of 0,2 is proposed as "small" and would probably not be noticeable in real world comparisons. Cohen suggested an example being the comparative heights of 15 and 16 year old students.

Medium = An effect size of 0,5 is proposed as "medium" and would probably be large enough to be noticed in real world comparisons. Cohen suggested an example being the heights of 13 year old and 18 year old students.

Large = Finally an effect size of 0,8 is proposed as "large" and would probably be easily perceivable. Cohen's example here was the intellectual difference between a college freshman and a PhD graduate.

Table 4: Description of effect sizes.

Measurement	Scale	Description
r	-1-+1	Correlation coefficient (r=0,10 small; r=0,30 medium; r=0,50 large)
η^2	0-1	Eta-squared (0,02 small; 0,13 medium; 0,26 large)
F	No range	F-statistics
ρ	-1-+1	Spearman's rank correlation coefficient

Table 5: Different measures used in the report.

3. EDUCATIONAL TECHNOLOGY

3.1 TECHNOLOGICAL DEVICES

In this chapter, focus is directed towards different kinds of technological devices and educational games. Due to its comparatively long history, much research has been conducted on computer based systems and their effect on different aspects of learning – from mathematics to literacy. Some of the included studies report from research conducted from as early as the 1970's. Although the use of desktop computers traditionally has been widespread, there has, over the last decades, been a shift towards a greater use of mobile devices.

Mobile devices have many advantages compared to ordinary computers. They are portable and individual, they can be context sensitive and socially connectable (Sung et al. (2015)) Adapted both for computers and for mobile devices, certain types of educational games, often called serious games, have been designed with the aim to enhance the students' motivation.[18]

3.1.1 DESKTOP COMPUTER SYSTEMS

Study	Method	Topic	Effect on academic achievement	Effect on cognitive & affective outcomes
Archer et al. (2014)	Meta-analytic review	Effectiveness of the use of technology in classrooms	g=0,18	-
Cheung & Slavin (2013)	Longitudinal meta-analysis	Effectiveness of educational technology applications in mathematics for K-12 students	g=0,15	-
Grgurovic et al. (2013)	Longitudinal meta-analysis	Effectiveness of computer technology-supported language learning	g=0,26	-
McEwan (2015)	Meta-analysis	Evaluation of the impact on educational interventions in language and mathematics	g=0,15	-
Schmid et al. (2014)	Meta-analysis	Effects of use of technology in postsecondary education	g=0,27	g=0,2
Takaci et al. (2015)	Quasi-experiment	Effect of collaborative learning using GeoGebra	$\eta^2 = 0,18$	-
Takacs et al. (2015)	Meta-analysis	Effects of technology on children's literacy development	g=0,17 (comprehension) g=0,20 (vocabulary)	-

Table 6: Effect sizes - Desktop computer systems

18 Note that the underlying capability of computing in education has changed dramatically over the review periods in included studies. Moving from numerical manipulations to higher order conceptual features – these changes may require separate analysis for each of the changes in the technology.

Included studies addressing desktop computer systems

In a tertiary meta-analytic review, **Archer et al. (2014)** re-assessed literacy learning outcomes presented in three previous meta-analyses. Students of different age groups and a variety of ICT and computer-assisted instruction interventions were included in the meta-analysis. The reported overall effect of educational technology on literacy learning outcomes was small (g=0,18). When specialized teacher education and support are included as a moderator[19] variable the ES's associated with technology in literacy learning outcomes can rise as high as 0,57. The study reported no significant differences in effect size as a function of implementation fidelity, or whether the interventions were delivered by researchers or teachers. The study by Archer et al. (2014) concludes that "the training and support of those conducting the interventions and attention to the fidelity of the intervention program, contribute to successful outcomes."

A longitudinal meta-analysis of 74 studies by **Cheung & Slavin (2013)** compared traditional (non ICT-based teaching) with ICT based educational technology. The study reported a consistent but small overall impact (g=0,15) on mathematics achievement in K-12 classrooms. Supplemental computer assisted instruction (blended learning) had the largest effect on mathematics achievement (g=0,18), while smaller effect sizes were reported with more traditional rote learning approaches such as comprehensive programs and computer-management learning (g=0,07-0,08). Analyzing the use of program intensity (frequency of intended use) as a moderator variable, the effect sizes for low, medium and high intensity were 0,03, 0,20 and 0,13, respectively. Furthermore, the effect size of studies with a high level of implementation[20] (g=0,26) was significantly greater than for studies of low and medium level of implementation (g=0,12). The study by Cheung & Slavin (2013) states that "Educational technology is making a modest difference in learning of mathematics".

19 The level of detail that can be assumed or detected from a meta analysis on other meta analyses is vague. More detailed moderator variables should be interpreted with caution as the error multipliers from multiple meta analyses may create statistical noise.

20 As technology has advanced, what is stated to be a high level of implementation five years ago might not be regarded the same way today.

A longitudinal meta-analysis of 37 studies by **Grgurovic et al. (2013)** showed that second/foreign language instruction supported by computer technology was at least as effective as conventional instruction without technology. Across the various conditions of technology use, the study reported a small but positive and statistically significant overall effect size of 0,26.

In a large scale meta-analysis comprising 77 randomized experiments, **McEwan (2015)** evaluated the impacts of different forms of educational interventions on language and mathematics outcomes in primary school (grade 1-8). The study reported that the impacts of ICT on educational outcomes (g=0,15) were comparable to impacts of increased teacher training (g=0,12) and smaller class sizes (g=0,12). These studies however, were directed at primary education in developing countries so these findings need to be replicated in a developed post-secondary setting.

In a detailed meta-analysis of the experimental literature of technology use in postsecondary education **Schmid et al. (2014)** reviewed 1105 studies featuring a broad variety of educational technologies and applications. The study reported the overall average effects of educational technology use on achievement and attitude outcomes, and found a positive association with improvements in academic performance (g=0,27) and student attitudes (g=0,20). In addition, the study explored moderator variables in an attempt to explain how technology use can lead to positive or negative effects. When more novel applications such as cognitive support tools (which are aiming to scaffold the active creation and negotiation of information) were involved, the effect sizes increased substantially in the 0,30-0,45 range, and equivalently, when search and retrieval tools (defined as tools that provide capabilities for knowledge seeking and retrieval, e. g. search engines, data bases etc.) were included the effect sizes increased even more in the 0,50-0,75 range. The introduction of communication tools to help students communicate among themselves and with teachers had less impact on effect sizes in the range 0,20-0,30. The overall message emerging from the study by Schmid et al. (2014) is that "learning is best supported when the student is engaged in active, meaningful exercises via technological tools that provide cognitive support".

In a quasi-experimental study including 180 students, **Takaci et al. (2015)** compared the effect of collaborative learning with or without the use of the GeoGebra algebra system on student calculus ability in higher education. GeoGebra is a novel mathematics software package that enables the students to check whether each step in the process of solving a task was correct. The study reported a medium positive effect size on student calculus ability (η^2=0,18)[21] for collaborative use of the software package over a two month test period. The effect size was reduced to η^2=0,10 for individual use of the package.

A meta-analysis of 43 studies by **Takacs et al. (2015)** was conducted on the effects of technology enhanced stories for young children's literacy development when compared to more traditional storybook reading. The use of technology to enhance children's story book reading comprehension and vocabulary was associated with improvements in comprehension (g=0,17) and expressive vocabulary (g=0,20). The average effect size for expressive vocabulary was heterogeneous with a significant effect for disadvantaged children (g=0,27) and a nonsignificant effect for non-disadvantaged children (g=0,05). Technology characteristics such as animated pictures and music were found to be beneficial while hotspots and sound effects were found distracting.

Summary: Desktop computer based systems

Findings from seven studies investigating the effect of computer based systems on academic achievement show small overall effect sizes in the range g=0,15-0,27 (Table 6). Schmid et al. (2014) also reported an overall effect size of 0,20 for student attitudes toward instruction in learning environments involving technology. Interestingly, three studies reported moderate to substantial increases in effect sizes of specific moderator variables: Archer et al. (2014) found substantial increases in effect size when specialized teacher education and support was included as a moderator variable (g=0,57); Cheung & Slavin found a significant increase in effect size of studies with a high level of implementation (g=0,26); and Schmid et al. (2014) reported that when educational technology included cognitive support tools, effect sizes increased substantially in the 0,30-0,45 range, and equivalently, when search and retrieval tools were included the effect sizes increased even more in the 0,50-0,75 range.

3.1.2 MOBILE DEVICE SYSTEMS

Included studies on mobile device systems

A detailed longitudinal meta-analysis by **Burston (2015)** summarized 20 years (1994-2014)[22] of research on learning outcomes using mobile assisted language learning (MALL) technology. Despite the publication of over 600 MALL studies over the past 20 years, no study has systematically evaluated the learning outcomes of MALL implementation projects. Over half of the MALL related studies focused on technological aspects of mobile devices, and did not involve MALL implementation projects, or learning gains were based on subjective teacher assessment or student self-evaluation. A number of other studies lacked statistically reliable learning outcome data due to short duration of projects or small number of participants involved. Yet other studies suffered from serious design shortcomings, thus leaving only 19 studies to reliably determine the learning outcomes of MALL applications.

Study	Method	Topic	Effect on academic achievement	Effect on cognitive & affective outcomes
Burston (2015)	Meta-analysis	Effects on learning outcome using mobile assisted language learning technology	-	-
Sung et al. (2015b)	Meta-analysis	Effectiveness of mobile devices in language learning	g=0,53	g=0,55

Table 7: Effect sizes - Mobile devices

21 Note that this is an eta-squared measure, operating with the scale 0-1.

22 The changes in mobile technology over the past 20 years have been significant. There are methodological challenges with merging findings using mobile technology from 20 years ago into single effect sizes.

Unfortunately, the remaining 19 studies reported such high variability of quality that the authors were unable to provide sufficient reliable results to estimate any effect size. Of the 19 studies, 15 reported unequivocal positive results, with those focusing on reading, listening and speaking without exception reported a positive advantage of MALL applications. The four remaining studies, all focusing on vocabulary, reported no significant differences.

A longitudinal meta-analysis by **Sung et al. (2015b)** investigated the contribution of mobile devices to language learning (MALL technology). The meta-analysis included 44 peer-reviewed journal articles and doctoral dissertations published from 1993 to 2013[23]. The study reported both achievement-related effect sizes (such as test scores) and affective/cognitive-related variables (such as motivation, engagement, attitude, satisfaction and preference). Overall effect sizes for achievement and affective/cognitive variables were g=0,53 and g=0,55, respectively, which suggest that MALL has a similar moderate effect on students' academic achievement and affective/cognitive variables in language learning. The study also conducted analyses for the effect of moderator variables on learning achievement. The mean effect sizes of learning stage differed significantly between categories, with the largest effect on adult MALL usage (g=0,95) followed by young children (g=0,51).

Furthermore, significant differences between various categories of hardware usage was reported between handheld devices (such as iPods, cell phones, digital pens and MP3 players) and laptop computer (such as laptops, tablet PCs and e-book readers), where handheld devices achieved a moderate-to-high effect size (g=0,73) as compared to no significant effect for laptop computers (g=0,15). Furthermore, interventions of 1-6 months had the largest effect size (g=0,77), followed by 2-4 weeks (g=0,62) and > 6 months (g=0,13). No significant effect size was found for interventions lasting only one week (g=0,23). The meta-analysis revealed that MALL instruction has produced a meaningful improvement in language learning.

Summary: mobile device systems

The included studies on mobile device systems both report from mobile-assisted language learning technology, which often are mobile device adapted versions of the former computer-assisted learning technology. Both Burston and Sung et al. (2015b) question the quality of research conducted within this field, but while Burston refuses to draw any conclusion of the effectiveness of mobile devices, Sung et al. (2015b) reports some quite significant effects. This highlights the need for further studies within this area. However, the effects on both academic achievement (0,53) and cognitive and affective outcomes (0,55) are rather convincing, and especially for adults (0,95) and although a bit less, also for young children (0,51).

23 See previous footnote.

3.1.3 GAME BASED SYSTEMS

Study	Method	Topic	Effect on academic achievement	Effect on cognitive & affective outcomes
Arnab et al. (2013)	Cluster RCT	Serious-games used to support Relationship and Sex Education	-	η^2=0,084
Miller et al. (2015)	Cluster RCT	Evaluation of a game-based physical activity program in primary school	Object control: d=0,96 Pedometer activity: d=1,02 (physical skills)	-
Riconscente (2013)	Experimental repeated measures crossover	Impact on learning when using touch interface games on iPads	15 % increase in test scores (above control)	10 % increase in test scores (above control)
Santos et al. (2014)	Meta-analysis	Effects of augmented reality learning experiences on K-12 students performances	d=0,56	-
Sung et al. (2015c)	Quasi experiment	Evaluation of students performance using a contextual decision-making game in health education	F=7,10 p=0,01	F=5,15 p=0.028 (problem-solving ability)
Wouters & Van Oostendorp (2013)	Meta-analysis	Role of instructional support in game-based learning	-	d=0,34
Wouters et al. (2013)	Meta-analysis	Cognitive and motivational effects of serious games	d=0.29	-

Table 8: Effect sizes – Game based systems

Included studies on games based systems

The study by **Arnab et al. (2013)** described the development of the digital game PR:EPARe (Positive Relationships: Elimination Coercion and Pressure in Adolescent Relationships) as a didactic approach to Relationships and Sex Education (RSE). Early efficacy testing of the game solution was validated in a cluster randomized controlled trial including 505 participants in school year 9 aged either 13 or 14 years. Data was collected as self-reported questionnaire measures. The study reported an intermediate effect in favor of the game (η^2=0,084)[24], and indicates that such serious game technology may have effective roles in training and remediating emotional responses and attitudes.

A cluster randomized controlled trial by **Miller et al. (2015)** investigated how serious games technology may impact sports attitudes and performance in primary school students. The study reported on the efficacy of the Professional

24 Note that this is an eta-squared measure, operating with the scale 0-1.

Learning for Understanding Games Education program. Students were assessed at baseline and 8-week follow-up for fundamental movement skill competency (FMS object control test), in-class physical activity and (pedometer steps) and perceived sports competence (self-reported profile). The study reported substantial increases for object control in students using the serious games practice (d=0,96) and in class activity (pedometer measure) (d=1,02). However, no difference was reported for perceived sports competence between the serious games practice group and the control.

An experimental repeated measures crossover study by **Riconscente (2013)** investigated the impact from the use of touch interface games on iPads to teach fractions to fourth grade students. The study reported on the efficacy of the fractions game Motion Math on fractions knowledge and attitudes. The data reported suggests that one week of exposure to Motion Math improved students' fraction test scores by an average of 15 %, and students' self-efficacy and liking of fractions each improved an average of 10 %. Both measures represented statistically significant increases compared to a control group. The game was designed to help children understand the relationship between fractions, proportions and percentages. The author suggested that one contributing factor to the positive impact was the instant feedback provided by the game, and that the entertainment value of the game provided children with the motivation necessary to persist in extensive practices.

A meta-analysis of 7 studies by **Santos et al. (2015)** was conducted to evaluate the effect of augmented reality learning experiences (ARLEs) on K-12 and university students' performance in various educational settings, including science and language classes. The included ARLE applications were intended to complement traditional curriculum materials, and included research papers must have at least a preliminary working ARLE prototype. The study found that ARLE applications showed a widely variable effect on student performance from a small negative effect (d=-0,28) to a large positive effect (d=1,00), with a mean moderate effect size of 0,56. The wide variability in ARLE effect sizes was ascribed to the many possible ways to use augmented reality, as well as, differences in experimental design of the studies. With such a wide variability of effects, there is a need for replication studies to clarify the findings.

The study by **Sung et al. (2015c)** described the development of a contextual digital game for improving students' learning performance in an elementary school health education course. A quasi experiment was conducted to evaluate the effects of the digital game on students' learning achievement, learning motivation and problem-solving ability. There were 52 students in both the experimental group and the control group. Students in the experimental group learned with the contextual digital game, while the students in the control group learned with the conventional e-book approach. The experimental results showed that the novel game-based learning prototype resulted in significant increases in the academic performance of highly motivated students, and more so than with the lower motivated students. Also, the game-based learning prototype improved the students' problem-solving competencies. Results of the statistical analysis were presented as F-statistics. F-values derive from an ANOVA test or a regression analysis to find out if the means between two populations are significantly different. No effect sizes are given in the study.

In a meta-analytic review including 29 studies, **Wouters & van Oostendorp (2013)** investigated the importance of instructional support in game-based learning, comparing studies with and without instructional support. In addition, a value-added approach was used, focusing on how specific game features facilitate learning and motivation. Wouters & van Oostendorp (2013) found that students that received instructional support in game-based learning outperformed the comparison group (d=0,34). Specifically, the meta-analysis reported learning improvements, in knowledge (d=0,33) and skills (d=0,64). The most effective features of instructional support facilitates the students to select relevant information (d=0,46), much more than features helping to organize and integrate information (d=0,14). In addition, instructional support that facilitates system interaction modality[25] (d=1,24), personalization (d=1,06), feedback (d=0,49), modeling[26] (d=0,46), reflection (d=0,29), and improves learning outcomes. Wouters & van Oostendorp (2013) reported on publication bias within this field, showing that the effect sizes in articles published in peer-reviewed journals (d=0,44)

25 Audio-channels for verbal explanations

26 Showing which kind of information should be used and how, in a specific situation.

were significantly higher than those reported in gray literature (proceedings: d=0,08; unpublished: d=0,14).

Among many researchers, serious games are emphasized as improving both cognitive processes and motivation among students. However, there is little evidence for this conclusion. **Wouters et al. (2013)** conducted a meta-analysis (39 studies) trying to shed light upon the effectiveness of serious games compared to conventional instruction methods on the cognitive dimensions of learning. Wouters et al. (2013) showed that serious games improve learning compared to conventional instruction (totally d=0,29), with regards to both knowledge (d=0,27) and cognitive skills (d=0,29). Even though serious games without any instruction seem to be somewhat effective (d=0,2), the most effective strategy is to combine serious games with instructional methods (d=0,41). Further, serious games lead to higher level of retention (d=0,36). However, surprisingly there is no statistically significant difference with regards to motivation.

Summary: Game based systems

Game based systems are reported to improve different aspects of learning, with regards to academic achievement, cognitive and affective outcomes, and physical skills. Wouters & van Oostendorp (2013) reported that the effect increases when certain kinds of instructional support are accompanying the games.

While six studies report on rather convincing improvements, one study (Wouters et al. 2013) found small effects. Small scale "one off" trials with extremely novel technologies are often associated with substantial reported impacts on learning. However, often there is a lack of replication with such studies and the findings may reflect the enthusiasm and novelty of the technology as opposed to significant evolutions in educational technologies. In addition, a publication bias might contribute to high effect sizes (Wouters & van Oostendorp 2013).

3.2 DESIGN FEATURES

The following section reports on five studies investigating different forms of software design and its effect on learning. Two studies report on intelligent tutoring systems which are a computer-assisted learning environment that aims to provide immediate and customized instruction or feedback to learners, usually without intervention from a human teacher. Intelligent tutoring systems have been developed for a number of academic subjects including mathematics, computer sciences, reading, writing and for training of specific skills, such as metacognitive skills. The other three studies investigate different kinds of instructional software scaffolds, designed to enhance learning.

3.2.1 INTELLIGENT TUTORING SYSTEMS

Study	Method	Topic	Effect on academic achievement	Effect on cognitive & affective outcomes
Wang et al. (2015)	Quasi experiment	Intelligent tutoring systems effect on basic computer skills	Improved learning effectiveness	-
Steenbergen-Hu & Cooper (2013)	Meta-analysis	Effectiveness of intelligent tutoring systems on mathematical learning	g=0,01-0,09	-

Table 9: Effect sizes – Intelligent Tutoring Systems

Included studies on intelligent tutoring systems

The quasi experimental study by **Wang et al. (2015)** compared the teaching of basic computer skills in higher education with or without the use of iTutor, a problem solving oriented intelligent tutoring system. 137 freshmen students from four classes were randomly assigned to an experimental group and a control group. The experimental group practiced the skills with iTutor and the control group did not use iTutor, but could access the same materials organized in the form of folders. The results indicate that students in the iTutor group experienced better learning effectiveness than those in the control group, and by comparison with the materials organized in folders, iTutor enabled students with any level of prior knowledge to experience more effective learning. Results of the statistical analysis are presented as F-statistics. F-values derive from an ANOVA test or a regression analysis to find out if the means between two populations are significantly different. No effect sizes are given in the study.

In a large scale meta-analysis on the impact of intelligent tutoring systems (ITS) on K-12 student mathematics learning, **Steenbergen-Hu & Cooper (2013)** reported small positive effect sizes ranging from 0,01 to 0,09. Most of the studies compared the effectiveness of ITS with that of regular classroom instruction, and it was concluded that ITS had no negative and perhaps a small positive effect on K-12 students' mathematics learning. Moderator analysis showed that shorter interventions with ITS (less than a calendar year) appear to provide the largest gains in math learning and that students with higher achievement levels benefitted the most from ITS interventions.

Summary: Intelligent tutoring systems

The studies investigating the use of intelligent tutoring systems in basic computer skills and mathematics show only small improvements in academic achievement. As Steenbergen-Hu & Cooper (2013) concluded, and compared to ordinary classroom instruction, there are however no negative effects to report from the use intelligent tutoring systems, despite the lack of human teachers. Thus, a possible potential for mass education with this system is indicated.

3.2.2 SYSTEM DESIGN FEATURES

Study	Method	Topic	Effect on academic achievement	Effect on cognitive & affective outcomes
McElhaney et al. (2015)	Meta-analysis	Dynamic visualisations in science curriculum	-	g=0,12
Merchant et al. (2014)	Meta-analysis	Effectiveness of virtual reality-based instruction on learning in higher education	Games: g=0,51 Simulations: g=0,41 Virtual worlds: g=0,41	-
Richter et al. (2016)	Meta-analysis	Text-picture signal relations in multimedia learning	-	r=0,17

Table 10: Effect sizes – System design features

Included studies on system design features

The meta-analysis of **McElhaney et al. (2015)** reviewed 47 independent comparisons between dynamic and static materials in science education, and 76 visual design comparisons that test the effect of specific instructional scaffolds. Dynamic visualizations for science education are defined as computer-based, animated representations of scientific phenomena. Each reported effect size was coded as either recall assessment (learners to learn of specific ideas such as names of structures) or interference assessment (ask learners to engage in inquiry and construct new knowledge). Using both types of assessment outcomes, the effect sizes ranged from -0,89 to 1,02, with a mean overall effect size marginally significant in favor of dynamic visualizations (g=0,12). To fully realize the potential of dynamic visualizations, instructional scaffolds are needed to help students use the dynamic visualizations to make sense of their own ideas. The most successful scaffolds include 1) Visual cues (use of arrows or colors to highlight salient features) (g=0,50); 2) Sequential conceptual representations (unique representations in the treatment condition occurring either before or after the visualization used in the control condition) (g=0,52); 3) Interactivity (learner control features such as play/pause controls or specifying input parameters) (g=0,45); and Inquiry prompts (sense-making or self-monitoring prompts) (g=0,26). Other instructional design features such as simultaneous conceptual representations (additional representations in the treatment condition occurring concurrently with the visualizations used in the control condition) and 3D-information (additional three-dimensional information present in the treatment condition) showed no or negligible benefits. The mean overall effect size was significant in favor of the refined instruction designs (g=0,35), almost three times as high as the impact of dynamic visualizations in general.

The meta-analysis by **Merchant et al. (2014)** examined the impact of technology based instruction in K-12 or higher education settings. The meta-analysis included 13 studies in the category of game-based instruction, 29 studies in that category of simulation-based instruction and 27 studies in the category of virtual worlds.[27] Analysis of the relationship between instructional technology use and learning outcome gains resulted in a moderate mean effect sizes of 0,51 for game-based instruction and 0,41 for both simulation-based instruction and virtual worlds, showing that games produce higher learning gains than simulations and virtual worlds. A moderator analysis was performed to highlight effect sizes of selected instructional design parameters. Key findings included that: For simulation studies, elaborate explanation type of feedback was more appropriate for declarative tasks (g=2,29) than visual cues type feedback (g=0,81). This is likely because students may need detailed instruction based on factual knowledge to complete a task. For procedural tasks, knowledge of correct response type of feedback was more appropriate (g=1,08) than visual cues (g=-0,06), indicating that when a task is procedural in nature, merely providing knowledge of correct response is sufficient to guide learners to complete the task. Furthermore, student performance is enhanced when they conduct game-based learning individually (g=0,72) rather than in a group (g=-0,004).[28]

A meta-analysis of 27 primary studies by **Richter et al. (2016)** investigated the role of signaling in multimedia on transfer and comprehension outcomes in K-12 and higher education settings. The signaling principle denotes how visual representations (e.g., color coding) are presented in learning materials to trigger broader recall. The study reported a small-to-medium overall effect size (r=0,17)[29] in favor of signaled as compared to non-signaled multimedia learning material. The signaling effect was significantly moderated only by domain-specific prior knowledge of the learners. Learners with low-level prior knowledge profited more from multimedia integration signaling (r=0,19) than high-level prior knowledge learners (r=-0,08). Although the effects were small, the findings indicate the effectiveness of the signaling principle in particular for learners with low prior knowledge.

27 There is no common agreed difference between these three types of system and there may be some overlap that might complicate interpretation of these findings.

28 This is contradictory to findings in other included studies within this review, e. g. Bernard et al. (2014).

29 Note that this is a correlation coefficient, operating with a scale from -1-+1.

Summary: System design features

The three studies investigating different kinds of learning software reported significant improvements on different aspects of learning. Interactive technology and elaborate explanation type of feedback within the software were proven to be effective, and Merchant et al. (2014) concluded that individual game-based learning is significantly more effective than group game-based learning. However, some negative effects are also identified, and the complexity in the effects both in regards to different aspects of the system features and to different learners, indicates the improbability of finding software suitable for all.

4. PEDAGOGICAL ASPECTS OF TEACHING AND LEARNING WITH ICT

4.1 BLENDED LEARNING

In recent decades, the political interest for online education has increased for several reasons, but foremost due to its cost-effectiveness and for providing learning opportunities that are independent of space and time (so called anytime anywhere learning). It offers a cost effective way to provide equal educational opportunities for a wide range of students from different social and economic backgrounds. Many studies have traditionally compared different forms of computer based distance education with ordinary classroom instruction, with rather inconclusive results. As many meta-analyses on the topic have concluded, online education does not seem to be more effective than classroom instruction, but on the other hand – not less effective either. Such lack of difference has legitimized a broad investment in online education in the US. The term "blended learning" or "hybrid learning" (hereafter just: blended learning) has developed as a result of the ambition to find more effective instructional conditions, trying to balance between online education and ordinary classroom instruction. Blended learning has subsequently been promoted as "the best of two worlds" (Means et al. 2013).

Based on a sub-collection (96 studies) of a meta-analysis (totally 674 studies), **Bernard et al. (2014)** investigated the effectiveness of blended learning compared to classroom instruction in higher education. Using achievement outcomes as the primary variable, Bernard et al. (2014) found that blended learning exceeds ordinary classroom conditions close to one-third of a standard deviation (g=0,334). In addition, the aim of the study was to more carefully outline the most important specific aspects of blended learning. Bernard et al. (2014) found that the kind of computer support used is of importance; cognitive support (g=0,59) (e.g. simulations and serious games) seems to be more effective than content/presentational support (g=0,24) (mere presentations of information). Furthermore, if the support is combined with one or more sources of interaction, between students, teachers and students, and/or students and the educational content, student academic achievement is even more enhanced. Of these variables, the single most important instructional feature seems to be the interaction between students (g=0,49).

Included studies on blended learning

Study	Method	Topic	Effect on academic achievement	Effect on cognitive & affective outcomes
Bernard et al. (2014)	Meta-analysis	Blended learning in higher education	g=0.33	-
Means et al. (2013)	Meta-analysis	Effectiveness of online and blended education	g=0.2	-
Spanjers et al. (2015)	Meta-analysis	Subjective and objective learning outcomes of blended learning	g=0.34	-

Table 11: Effect sizes – blended learning

Means et al. (2013) meta-analysis, reported from 45 studies focusing on online learning and blended learning and its effect on learning outcomes among different kinds of students, compared to ordinary classroom instruction. The result of the study resembles previously conducted meta-analyses, finding that students engaged in online learning, solely or partly, perform modestly better (g=0,2) than students solely engaged in ordinary classroom instruction. However, learning conditions including both online and face-to-face aspects (the so called blended learning) were significantly more effective (g=0,35) than ordinary instruction. Interestingly, Means et al. (2013) found positive and significant effect sizes for collaborative instruction (g=0,25) and especially for expository instruction (g=0,39). There were no differences to be found either across subjects or age-groups. As concluded in the study, the educational situation in which the blended learning takes place is often characterized by additional learning time, resources and possibilities for interactions between students. Thus, there are reasons to believe that these aspects influence the positive effects of blended learning.

Spanjers et al. (2015) conducted a meta-analysis (69 studies) examining the effectiveness of blended learning compared to ordinary face-to-face learning in relation to students learning outcome, satisfaction and time investment. As objective measures, post-tests, gains in test scores, course grades etc., were used. The subjective measures consisted of the students' self-assessment, perceived self-efficacy, subjective learning gains, confidence in ability etc. In addition, the time investment measure is also subjective, based on the students' perceived amount of work or effort devoted to the work and the appropriateness of the devoted time. Overall, the meta-analysis reported on small to medium effect-sizes in favor for blended learning (objective g=0,34, subjective g=0,27). However, the effect on satisfaction was inconsiderable small (g=0,11) and the investment evaluation was significantly negative (g=-1,04) (based on 4 studies). Thus, according to Spanjers et al. (2015), blended learning has an effect on students' objective learning outcomes, but that does not seem to be correlated with students' satisfaction. In addition, the students perceived blended learning to be more time demanding and less effective with regards to workload compared to ordinary face-to-face learning.

Summary: Blended learning

The three meta-analyses on the topic all conclude in favor for blended learning on learning outcome, although the effect sizes vary from small to relatively small (from 0.2-0.34). As Means et al. (2013) stated, the positive result of blended learning might be caused by an overall enhanced learning situation and increased resources. Both Bernard et al. (2014) and Means et al. (2013) stress that collaboration between students enhance the effect and thus also learning. In addition, the studies find that computer support focusing on cognitive aspects, such as simulations or serious games are more effective, as well as instructions that are expository. Spanjers et al. (2015) reported that the students experience blended learning as time consuming. This indicates a potential risk of lower students' satisfaction.

4.2 ASSESSMENT AND FEEDBACK

The importance of feedback is often emphasized in all kinds of educational contexts, addressed both by teachers and students in collaborative self-assessment. For technological devices, assessment and feedback have often been restricted to immediate responses and corrections. Recently, the possibility to utilize more formative feedback has been investigated, both addressed by instructional features of the software itself and by peers – via mobile devices.

Included studies on assessment and feedback

Study	Method	Topic	Effect on academic achievement	Effect on cognitive & affective outcomes
Van der Kleij et al. (2015)	Meta-analysis	Effects of feedback in a computer-based learning environment	g=0,49	-
Lai & Hwang (2015)	Quasi-experiment	Evaluation of an interactive peer-assessment criteria development approach	-	d=2,39

Table 12: Effect sizes – assessment and feedback

In a meta-analysis based on 40 studies, **van der Kleij et al. (2015)** investigated the effects of feedback on learning outcomes in computer-based learning environments. Shute's (2008)[30] categorization of feedback is used, distinguishing between knowledge of results (indication of whether the answer is correct or not, but does not reveal the correct answer), knowledge of correct results (reveal the location of the result, but not the correct answer) and elaborated feedback (includes many kinds of feedback, such as; additional information, hints and explanations of the correct answer). In addition, a distinction is made between immediate and delayed feedback. Computer-based educational programs are often characterized by immediate knowledge of results or knowledge of correct results. Van der Kleij et al. (2015) found *knowledge of the result* to be the least effective kind of feedback (g=0,05), whilst *knowledge of the correct result* was significantly more effective (g=0,32). However, elaborated feedback improved the students' feedback the most (g=0,49) and especially in higher-order tasks and in mathematics. The timing of the feedback did not seem to have any effect, not even on lower order learning.

In a quasi-experimental study (N=103), **Lai & Hwang (2015)** reported from an evaluation of the effectiveness of an interactive peer-assessment criteria development approach created with the aim to help students to develop abilities for self-assessment, learning from other peers work and making self-reflection of their own learning and progress through a mobile device.

The effectiveness was measured with regards to learning achievement, learning motivation, meta-cognitive awareness, and cognitive load. Lai and Hwang (2015) refer to studies highlighting the many advantages of peer-assessment, it leads to e. g.: improvements in learning, stimulation of meta-cognitive awareness and increased autonomy. However, peer-assessment can also be associated with problems. According to Lai and Hwang (2015), students often have difficulties fully understanding the assessment criteria formulated by teachers. The interactive peer-assessment criteria development approach seeks to solve this issue. The evaluation of the intervention showed significantly improved learning achievement in the experimental group (d=2,39)[31], as well as improved learning motivation and meta-cognitive awareness. However, the cognitive load did not increase compared to the control group. With regards to the learning process, this study showed the importance of integrating the students in the development of assessment criteria, and also that a mobile device can serve as an effective tool for realizing that ambition.

Summary: Assessment and feedback

Van der Kleij et al. (2015) found that elaborated feedback improved learning more than different forms of correctional feedback. Lai & Hwang (2015) reported on results in favor (2.39) for an application developed for students to define their own learning assessment criteria and for conducting peer-assessment in accordance to those criteria. The two studies indicate the potential in the usage of technology as assessment and feedback tools.

30 Shute, V. J. (2008). Focus on formative feedback. Review of Educational Research, 78, 153–189. doi:10.3102/0034654307313795

31 The reported effect size is very large, beyond those expected in social sciences

4.3 EDUCATIONAL PSYCHOLOGY, ICT AND LEARNING

In recent decades, there has been a rapid increase in possibilities for students to undertake higher education online. Online education has many advantages, since it both can be synchronistic (making simultaneous interaction possible despite geographical differences) and asynchronistic (as it is independent of time and space). However, online education students have to develop self-regulated learning abilities. Two of the articles described below investigated the conditions of the most effective aspects of self-regulated learning strategies, while the other three focus upon the potentiality for ICT to enhance affective learning and self-efficacy.

Study	Method	Topic	Effect on academic achievement	Effect on cognitive & affective outcomes
Broadbent & Poons (2015)	Systematic review	Self-regulated learning strategies in online education	r=0,13	-
Brydges et al. (2015)	Systematic review	Computer-based supervision for self-regulated learning strategies	Na	na
Lee et al. (2013)	Meta-analysis	Effects on affective and cognitive learning	-	g=0,42 (cognitive) g=0,18 (affective)
Gergenfurtner et al. (2013)	Meta-analysis	Effects on computer support on self-efficacy and transfer of training	-	ρ=0,31(before training)/ ρ=0,39 (after training)

Table 13: Effect sizes – assessment and feedback

Included studies on educational psychology, ICT and learning

Broadbent & Poons' (2015) systematic review (based on 12 studies) aimed to investigate the most effective self-regulated learning strategies in online education in higher education with regards to academic outcomes. The combined self-regulated learning strategies correlated positively with academic outcomes, but not very strongly (r=0,13)[32]. Different aspects of self-regulated learning strategies were investigated, showing that time management (r=0,14), metacognition (r=0,06), effort regulation (r=0,11) and critical thinking (0,07) were positively but weakly correlated with academic outcomes, whereas no correlation was found with regards to rehearsal, elaboration, and organization. Broadbent & Poons (2015) concluded that the self-regulated learning strategies that have showed to be effective in traditional education might not be as effective in online education. Thus, there is a possibility that other, currently unexplored, strategies might be more important and effective in online education.

In a systematic review and meta-analysis (32 studies), **Brydges et al. (2015)** investigated if supervision that aims to develop self-regulated learning is correlated with improvement in learning. The included studies in the review and meta-analysis reported from simulation-based training interventions designed to support students to develop self-regulated learning strategies. No reported results were statistically significant. The groups that did not receive supervision performed worst on the post-test (d=-0,34, p=0,09). Interventions supported by supervision did however show some, but small effects outcome on both the post-test (d=0,23 p=0,22) and the delayed retention test (d=0,44 p=0,067). The study shed light on the insufficiency of simulation-based digital training support alone, and highlights the importance of supervision also in these contexts. However, it is important to note that no results were statistically significant.

32 Note that this is a correlation coefficient, operating with a scale from -1-+1.

In a meta-analysis based on 58 experimental and quasi-experimental studies, **Lee et al. (2013)** summarized research published over the last 15 years[33] on how technology effects K-12 students cognitive and affective learning. With regards to cognitive outcome, a moderate effect (g=0,42) was identified, indicating that the use of technology can be beneficial, especially for K-8 students. With regards to cognitive outcomes, the following seems to be especially effective; software using tutorials (g=0,81), tasks that are either basic (g=0,88), project-based (g=1,39) or based upon inquiry/investigation (g=0,61) as well as letting the students cooperate in groups (3-5) per computer (g=1,08). When the teacher acts as a facilitator, the highest effect sizes are measured (g=0,62). The overall effects on the affective outcomes are small (g=0,18).

Finally, **Gergenfurtner et al. (2013)** conducted a meta-analysis (based on 29 studies) investigating the longitudinal development of the relationships between self-efficacy and transfer of training throughout the last 25 years[34], with regards to computer support, collaboration and time lag. Self-efficacy (Bandura 1977)[35] denotes the beliefs in one's capability to perform in accordance to specific requirements. Self-efficacy is considered to be a predictor of academic achievement, and so does the transfer of training, which is a description of the ability to use new knowledge. Gergenfurtner et al. (2013) found a small but positive relationship between self-efficacy and training, measures before (ρ=0,31)[36] and after training (ρ=0,39). Compared to no support, computer support strengthens the correlation between the belief in efficacy and training transfer (pre ρ=0,23 post ρ=0,31). However, this correlation appears only shortly after training, but less in post-tests. The most fruitful combination is computer support, without any collaboration with peers (pre ρ=0,37, post ρ=0,62).

Summary: Educational psychology, ICT and learning

These studies show that engagement in online education seems to demand other self-regulated learning strategies than traditional education. Broadbent & Poons (2015) stress the importance to identify online education-specific strategies. Brydges et al. (2015) investigated the importance of supervision for developing self-regulated learning strategies, but found no statistically significant effects. Lee et al. (2013) found that technology enhances cognitive learning more than affective learning, and Gergenfurtner et al. (2013) showed the importance of computer support when developing belief in efficacy among students.

33 Note that this is a long time with technology change.

34 See footnote 13.

35 Bandura, A (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review. 84* (2): 191–215. doi:10.1037/0033-295x.84.2.191. PMID 847061.

36 Note that this is Spearman's rank correlation coefficient, operating with a scale from -1-+1.

5. CONCLUSIONS

As shown in this report, there are a wide range of expectations assigned with the use of digital technology in education. Not only is the use of digital devices expected to increase the student's motivation to learn; this motivation is also expected to raise student achievement. Additionally, technology is expected to create new educational possibilities, to offer potentially cost-effective tools and balance societal inequalities. This mapping has been undertaken to establish if, how and to what degree ICT influences students' learning outcomes.

5.1 POSITIVE BUT SMALL IMPACT

In reviewing the included studies, it becomes clear that educational technology is not a single homogenous intervention but a broad variety of modalities, tools, and strategies for influencing and assisting teaching and learning. The different forms of educational technologies included in this mapping are, in addition, used in a variety of educational contexts, and with a wide range of goals. The inherent heterogeneity of the material therefore makes it difficult to draw clear conclusions concerning the effectiveness of ICT in education. However, some features appear across the studies.

Although the effects are small, the review shows a consistent positive impact from the use of ICT in classroom settings. Some studies report large ES's (>> 2,0) from novel technology implementations, but the more rigorous meta-analysis, focusing on large scale randomized control trials, consistently report ES's in the range of +0,1 to +0,3. However, when the technology is accompanied with some kind of instructional support, either embedded in the software or through teacher supervision, the effects seem to increase significantly. Obviously, it is not the technology in itself that promotes learning outcomes, but the design of the software and/or the pedagogical use of the device.

5.2 INSTRUCTION AND HUMAN SUPPORT

Some studies investigating the effect of technology partly or solely without any present teacher e.g. on intelligent tutoring systems and online education (as mentioned in Means et al. 2013, Steenbergen-Hu & Cooper 2013), show neither positive nor negative effects on learning outcomes, thereby concluding that technology can replace traditional classroom education without risking academic performance. However, as shown in the figure 2 below, this may not be conclusive.

The table presents studies using interactional

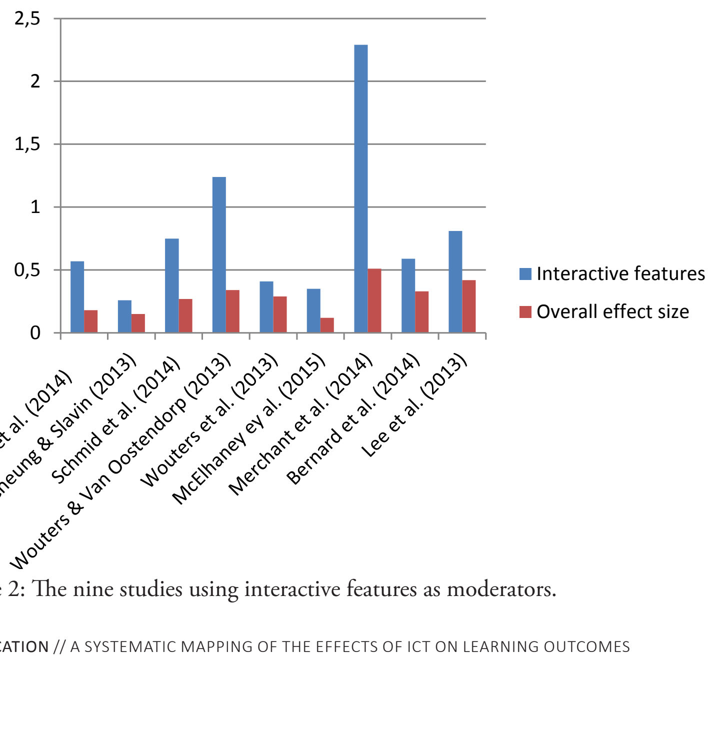

Figure 2: The nine studies using interactive features as moderators.

features (human support and/or ICT support tools) as moderators, indicating that the effect almost doubles when ICT is accompanied with different kinds of either technological or human support. Interactional features might be physical (teacher, peer) as non-physical (mediated by teachers, peers, or e.g. tutoring systems).

Several studies show that having a teacher physically present enhances learning with ICT (e.g. Archer et al. 2014, Bernard et al. 2014, Lee et al. 2013). Although some studies report contradictory findings (Gergenfurtner et al. 2013), numerous studies find that peer collaboration contribute to improved learning outcome (Lee et al. 2013, Bernard et al. 2014, Means et al. 2013). This indicates that interactional features (physical or non-physical) contribute to increased effect sizes, generally highlighting the importance of providing support in the use of technology. This also indicates the potential of educational technology as a supplement to ordinary education rather than a replacement. The effectiveness of ICT in education depends entirely on how well it assists teachers and students in achieving the educational goals. The highest ES's were associated with studies where ICT had been implemented as a planned part of a comprehensive teaching environment with clear goals, teaching plans, teaching materials, supporting technical resources, teacher training and development. Improvements associated with ICT in education should therefore not be ascribed to a single factor, but understood and interpreted contextually.

5.3 TEACHERS AND TECHNOLOGY

Technology in education can serve a multitude of purposes, from administrative to educational. However, and in accordance with findings in this mapping, the effectiveness of ICT in education is determined by the context in which it is introduced and employed. The quality of the instructional design appears to be the single most important aspect, and as described in the introduction, this depends on teachers' professional pedagogical and didactic competence, their room for maneuver and how school leaders and school owners support their work. All of these aspects influence the teachers' ability to effectively integrate ICT in their teaching practice. As indicated in the introduction, reports from the OECD and EU advise Norway to be less concerned with the provision of technology and more concerned with teachers' professional development and focus on how technology may support teachers' everyday instructional practice.

5.4 REASONS TO INTERPRET EFFECTS OF ICT ON LEARNING WITH CARE

There are several reasons why findings in this mapping of the effects of ICT on learning should be interpreted with caution.

- **Enthusiasm of novel technology:** Extremely novel technologies are often associated with substantial impacts on learning (so called Hawthorne effect). However, the findings may reflect the enthusiasm of the novelty of the technology, and might not be sustained in the long-term.
- **Development of technology**: Educational technology is constantly developing. Meta-analyses and systematic reviews investigating the effects of a specific modality, tool or software, through several decades might in fact not be studying the same thing, thus making it difficult to accumulate insights.
- **Publication bias:** Wouters & van Oostendorp (2013) report on a publication bias, indicating significantly higher effect sizes in published studies, than in gray literature (reports, unpublished papers, conference papers etc.).
- **Study heterogeneity:** Different measures, different learning environments, different teaching methods/approaches, different student demographics over time make any longer term comparisons and effect comparisons difficult.
- **A "noise" of effects:** There are wide variations of effects reported, both in this review and within the single studies, thus causing a "noise" of effects which make it difficult to interpret the significance of the results. Additionally, when combining, merging or summing the reported statistical effects from numerous studies the statistical errors become multiplied.
- **Overall intervention bias:** Educational technology interventions are often accompanied by a re-structuring of a whole educational setting, sometimes influenced by increased resources and increased time dedicated for learning and preparation, thus making it difficult to identify if the effect is caused by the technology or the re-structured educational context itself.

5.5 KNOWLEDGE GAPS

This mapping has revealed a need for more research on:

- How teachers experience, implement and learn about educational technology.
- Characteristics of teachers' work conditions that may hinder or promote successful implementation of digital technologies in schools.
- Leadership support when new technologies are being introduced.
- The impact of teachers' digital competence on how technology is used in education.
- A more systematic approach to educational technology studies, with common measures of ability or competence that can be shared across multiple educational settings.
- Better defined teacher competencies and training skill sets. This information is often missing when studies are reported making comparisons between studies problematic.

REFERENCES

Archer, K., Savage, R., Sanghera-Sidhu, S., Wood, E., Gottardo, A., & Chen, V. (2014). Examining the effectiveness of technology use in classrooms: A tertiary meta-analysis. *Computers & Education*, *78*, 140-149.

Arnab, S., Brown, K., Clarke, S., Dunwell, I., Lim, T., Suttie, N., ... & De Freitas, S. (2013). The development approach of a pedagogically-driven serious game to support Relationship and Sex Education (RSE) within a classroom setting. *Computers & Education*, *69*, 15-30.

Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, *26*(1), 87-122.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1-13.

Brydges, R., Manzone, J., Shanks, D., Hatala, R., Hamstra, S. J., Zendejas, B., & Cook, D. A. (2015). Self-regulated learning in simulation-based training: a systematic review and meta-analysis. *Medical education*, *49*(4), 368-378.

Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL*, *27*(01), 4-20.

Cheung, A. C., & 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.

Gegenfurtner, A., Veermans, K., & Vauras, M. (2013). Effects of computer support, collaboration, and time lag on performance self-efficacy and transfer of training: A longitudinal meta-analysis. *Educational Research Review*, *8*, 75-89.

Grgurović, M., Chapelle, C. A., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, *25*(02), 165-198.

Lai, C. L., & Hwang, G. J. (2015). An interactive peer-assessment criteria development approach to improving students' art design performance using handheld devices. *Computers & Education*, *85*, 149-159.

Lee, Y. H., Waxman, H., Wu, J. Y., Michko, G., & Lin, G. (2013). Revisit the Effect of Teaching and Learning with Technology. *Educational Technology & Society*, *16*(1), 133-146.

McElhaney, K. W., Chang, H. Y., Chiu, J. L., & Linn, M. C. (2015). Evidence for effective uses of dynamic visualisations in science curriculum materials. *Studies in Science Education*, *51*(1), 49-85.

McEwan, P. J. (2015). Improving Learning in Primary Schools of Developing Countries A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, *85*(3), 353-394.

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, *115*(3), 1-47.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, *70*, 29-40.

Miller, A., Christensen, E. M., Eather, N., Sproule, J., Annis-Brown, L., & Lubans, D. R. (2015). The PLUNGE randomized controlled trial: Evaluation of a games-based physical activity professional learning program in primary school physical education. *Preventive medicine*, *74*, 1-8.

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, *17*, 19-36.

Riconscente, M. M. (2013). Results from a controlled study of the iPad fractions game Motion Math. *Games and Culture*, *8*(4), 186-214.

Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2014). Augmented reality learning experiences: Survey of prototype design and evaluation. *IEEE Transactions on Learning Technologies*, *7*(1), 38-56.

Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A., ... & Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. *Computers & Education*, *72*, 271-291.

Spanjers, I. A., Könings, K. D., Leppink, J., Verstegen, D. M., de Jong, N., Czabanowska, K., & van Merriënboer, J. J. (2015). The promised land of blended learning: Quizzes as a moderator. *Educational Research Review*, *15*, 59-74.

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.

Sung, H. Y., Hwang, G. J., & Yen, Y. F. (2015c). Development of a contextual decision-making game for improving students' learning performance in a health education course. *Computers & Education*, *82*, 179-190.

Sung, Y. T., Chang, K. E., & Yang, J. M. (2015b). How effective are mobile devices for language learning? A meta-analysis. *Educational Research Review*, *16*, 68-84.

Takači, D., Stankov, G., & Milanovic, I. (2015). Efficiency of learning environment using GeoGebra when calculus contents are learned in collaborative groups. *Computers & Education*, *82*, 421-431.

Takacs, Z. K., Swart, E. K., & Bus, A. G. (2015). Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks a meta-analysis. *Review of educational research*, *85*(4), 698-739.

Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes A Meta-Analysis. *Review of Educational Research*, *85*(4), 475-511.

Wang, D., Han, H., Zhan, Z., Xu, J., Liu, Q., & Ren, G. (2015). A problem solving oriented intelligent tutoring system to improve students' acquisition of basic computer skills. *Computers & Education*, *81*, 102-112.

Wouters, P., & Van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, *60*(1), 412-425.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249.

APPENDIX 1: SEARCH STRING

Search string (ERIC, ASSIA, IBSS, PQEJ databases)

(TI,AB("1 to 1 computer" OR "blended learning" OR "CAI" OR "CAL" OR "CBL" OR "collaborative learning" OR "computer aided" OR "computer assisted instruction" OR "computer assisted learning" OR "computer based instruction" OR "computer based learning" OR "computer based teaching" OR "computer simulation*" OR "computer supported" OR "computer technology" OR "computer use" OR "computer-aided" OR "computer-assisted instruction" OR "computer-assisted learning" OR "computer-based instruction" OR "computer-based learning" OR "computer-based teaching" OR "computeri?ed instruction" OR "computers and learning" OR "computers in education" OR "computer-supported" OR "computing education" OR "digital learning" OR "digital technology" OR "educational technology" OR "effect* on learning" OR "e-learning" OR "electronic learning" OR "game*" OR "ICT*" OR "information communication technolog*" OR "innovative technology" OR "instructional technolog*" OR "intelligent tutoring system*" OR "interactive learning environment*" OR "interactive learning object*" OR "interactive simulation*" OR " interactive white board*" OR "learning effect*" OR "media in education" OR "mobile learning" OR "multimedia learning" OR "OLPC" OR "one laptop per child" OR "one to one computer" OR "one2one computer" OR "online learning" OR "online self study" OR "online self-study" OR "online study" OR "serious game*" OR "simulation based education" OR "simulation based teaching" OR "simulation-based education" OR "simulation-based teaching" OR "simulation" OR "social network" OR "supplemental CAI" OR "tablet*" OR "technology enhanced instruction" OR "technology enhanced learning" OR "technology use" OR "technology-enhanced instruction" OR "technology-enhanced learning" OR "TEL" OR "tutoring system*" OR "virtual learning" OR "virtual reality" OR "web-based instruction*" OR "web-based learning" OR "web-based training")) AND (TI,AB("academic achievement" OR "academic outcome*" OR "academic performance" OR "academic progress" OR "academic success" OR "achievement gain*" OR "basic skill*" OR "career readiness" OR "cognitive gain outcome*" OR "college readiness" OR "educational achievement" OR "educational benefit*" OR "educational improvement" OR "educational outcome*" OR "educational performance" OR "effect*" OR "effective learning" OR "enhancing learning" OR "graduat*" OR "knowledge acquisition" OR "learner outcome*" OR "learning outcome*" OR "mathematics achievement" OR "mathematics learning" OR "mathematics skills" OR "program* effect*" OR "reading outcome*" OR "reading skills" OR "science achievement" OR "student improvement" OR "student outcome*" OR "student* achievement*" OR "student* performance" OR "test score*" OR "treatment" OR "treatment effect*" OR "writing achievement" OR "writing skills")) AND (TI,AB("average treatment effect" OR "causal effect*" OR "control group" OR "difference-in-difference" OR "effect study" OR "instrumental variable*" OR "meta*" OR "PIRLS" OR "PISA" OR "propensity score" OR "propensity score matching" OR "quasi-experiment" OR "randomi?ed controlled trial*" OR "randomi?ed controlled stud*" OR "randomi?ed experiment" OR "regression discontinuity" OR "TIMSS" OR "treatment group"))

The same search string with custom syntax was used in the Scopus and Psycinfo databases.

APPENDIX 2: INCLUSION AND EXCLUSION

Step 1 – screening and quality assessment based on title and abstract using the following criteria:

EXCLUDE on topic (The study focus is not on ICT and learning)

EXCLUDE on medical education (The focus of this report is not on medical education)

EXCLUDE on not empirical (Study needs to be evidence based, not conceptual or philosophical only)

EXCLUDE on not an intervention (Study needs to report on an intervention, not contextual only)

EXCLUDE on book/report/dissertation (Study is not peer-reviewed)

EXCLUDE on language (Study is not written in English, Norwegian, Swedish or Danish)

EXCLUDE on N<50 (Studies with few participants are likely to have low validity, and conclusions based on the results may be uncertain. We have chosen to set a lower limit of 50 participants in the intervention group and 50 participants in the control group)[37]

EXCLUDE on Citation Index (Normally, articles cited by other researchers have high quality and relevance within a research field. We have chosen to exclude articles with lower citation index (CI) than average for the remaining articles)

EXCLUDE on Journal Impact Factor (Journal Impact Factor (JIF) is an index based on the average number of citations of articles published in a scientific journal, and is used as an impact-measure of a journal in the research field. We have chosen to exclude journals with lower JIF than average for the journals publishing the remaining articles)

INCLUDE based on title and abstract (Need to retrieve full report for full text screening)

Step 2 – screening and quality assessment based on full text using the following criteria:

EXCLUDE on methodological issues (The studies intervention does not meet high quality experimental conditions)

EXCLUDE on not an intervention (Study needs to report on an intervention, not contextual only)

EXCLUDE on findings not reported (The study does not report on an intervention with data or outcomes)

EXCLUDE on relevance (The study is not relevant for the report)

INCLUDE on full study (Include based on full text. Item ready for in-depth review)

---

37 Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology, 1*(1), 9-38.
Weisburd, D., & Gill, C. (2014). Block randomized trials at places: rethinking the limitations of small N experiments. *Journal of Quantitative Criminology, 30*(1), 97-112.

APPENDIX 3: EFFECT SIZES

What is an effect size?

Within the evidence based reviews conducted by the Knowledge Centre we make extensive use of a statistical measure called an ***effect size***. An ***effect size*** is a statistical technique for measuring the size of a difference between two groups, usually a control and an intervention within a social science context, such as a controlled comparison of a new technique in education. A graphical representation of two groups with an ***effect size*** difference of 1.0 is shown below.

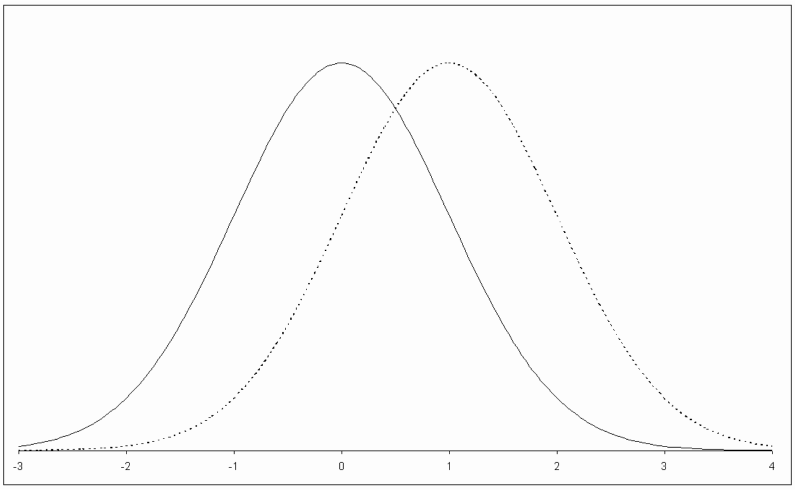

Figure 3: Graphical representation of an effect size of 1.0 between 2 groups (from Coe, 2002).

The power of this specific technique is that, unlike more traditional measures that focus on the ***statistical significance*** (or probability) of an outcome, an ***effect size*** shows the effectiveness of a specific intervention in comparison to either a control condition or another intervention (Coe 2002)[38].

In contrast ***statistical significance*** measures if an outcome did not occur by chance, this is often shown using the ***P*** or probability value (*for example P<0.05*). The weakness with such traditional ***statistical significance*** measures are that they are susceptible to bias from sample size, which can make very weak effects appear highly significant if a study has a large sample size and conversely very strong effects can appear non-significant if a study has a small sample size.

The ability to directly compare the strengths of the outcomes from interventions makes the ***effect size*** very suitable in determining which intervention is more effective in a given experimental comparison. Effect sizes also permit much easier comparisons of any replications for a study, showing quickly and easily if a proposed intervention shows a similar sized effect reported by earlier experiments.

38 Coe, R. (2002) "It's the Effect Size, Stupid. What effect size is and why it is important" Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England, 12-14 September 2002

How is it calculated?

The effect size is very easy to calculate, being the standardised mean difference between the two groups.

$$\text{Effect Size} = \frac{\text{[Mean of experimental group]} - \text{[Mean of control group]}}{\text{Standard Deviation}}$$

With normally distributed populations this calculation means that the effect size is also the "Z Score" of a standard normal population. So an effect size of +0.6 means that the score of the average person in the intervention group of an experiment is 0.6 of a standard deviation above that of the average person in the control group of an experiment (note that effect sizes can be either positive or negative depending on the positive or negative influence of an intervention). This allows researchers to combine the findings from similar studies and calculate a common effect size for multiple studies that share a common intervention and outcome measure. Combining effect sizes together from many experimental comparisons of similar interventions to estimate an overall effect size for a specific intervention is called ***meta-analysis***.

Cohen's Real World Effect Size Scale

In his 1969[39] paper that extolled many of the modern principles of using effect sizes Cohen proposed that effect sizes could be best understood by reflecting them into real world comparisons within a scale of Small, Medium and Large effect sizes. In our Knowledge Centre reviews we often adopt Cohen's proposed 3 item scale in our tabular summaries to permit rapid understanding of the strength of reported effect sizes.

Small = An effect size of 0,2 is proposed as "small" and would probably not be noticeable in real world comparisons. Cohen suggested an example being the comparative heights of 15 and 16 year old students.
Medium = An effect size of 0,5 is proposed as "medium" and would probably be large enough to be noticed in real world comparisons. Cohen suggested an example being the heights of 13 year old and 18 year old students.
Large = Finally an effect size of 0,8 is proposed as "large" and would probably be easily perceivable. Cohen's example here was the intellectual difference between a college freshman and a PhD graduate.

Table 15: Description of effect sizes.

There are some risks in adopting a simplified coding of effect sizes into small, medium and large (see Glass et al, 1981[40] for a detailed summary) but these risks are generally only of concern when taking such an effect size coding out of its context. Since in our reviews we present relatively coherent studies all within similar educational contexts we have chosen to use colour coding within our tabular presentations of effect sizes when displaying the comparative outcomes, in order to make the information more easily understood. Please note that full ***effect size*** details for each summarised study are provided in the more detailed text descriptions that follow the summary tables.

39 Cohen, J. (1969) Statistical Power Analysis for the Behavioral Sciences. NY: Academic Press.

40 Glass, G.V., McGaw, B. and Smith, M.L. (1981) Meta-Analysis in Social Research. London: Sage.

Cautions when using effect sizes

As we have shown ***effect size*** provides a valuable tool when understanding the strength of a causal effect from a specific intervention. However when using ***effect size*** we must always be careful that we are comparing similar interventions, settings and outcomes. This is especially important in educational research where we may have significant variations in student populations, test instruments, educators, and learning environments that may not be completely reflected in a summary of a study.

Assumptions when using alternative measures of effect-size

Within the research literature related to ***effect size*** you will sometimes find alternative measures of ***effect size*** reported other than the "standardised mean difference" technique that we have been describing.

For example, in many studies you will see the correlation "***r***" between two variables being used to calculate the square of the two values (shown as "***R2"***) which indicates the ***proportion of variance accounted by the independent variables*** (for a more detailed discussion see Thompson, 1999[41]). However when effect size is calculated from this "***proportion of variance accounted for***" method you should be aware that it suffers from a number of limitations, standard errors can be large and two studies with opposite results would report identical "***variance accounted for***" results (See Olejnik & Algina 2000[42] for more details). Good summaries of many of the different kinds of effect size measures that can be used and the relationships among them can be found in Snyder and Lawson (1993)[43], Rosenthal (1994)[44] and Kirk (1996)[45].

However such alternative ***effect size*** measures often hide a more complex issue, the possible confusion of measures of association with causal effect. As has been noted by Coe (2002)[46]:

"*The crucial difference between an effect size calculated from an experiment and one calculated from a correlation is in the causal nature of the claim that is being made for it. Moreover, the word 'effect' has an inherent implication of causality: talking about 'the effect of A on B' does suggest a causal relationship rather than just an association."* (Coe, 2002).

For this reason many statisticians recommend caution in using the term "effect" unless there is an explicit causal mechanism being described and instead to use the term "variance accounted for" or "strength of association" or cite the regression coefficient instead of calling it an ***effect size*** (see Fitz-Gibbon[47], 2002 and Coe, 2002 for a fuller discussion).

---

41 Thompson, B. (1999) 'Common methodology mistakes in educational research, revisited, along with a primer on both effect sizes and the bootstrap.' Invited address presented at the annual meeting of the American Educational Research Association, Montreal. [Accessed from http://acs.tamu.edu/~bbt6147/aeraad99.htm, January 2000]

42 Olejnik, S. and Algina, J. (2000) 'Measures of Effect Size for Comparative Studies: Applications, Interpretations and Limitations.' Contemporary Educational Psychology, 25, 241-286.

43 Snyder, P. and Lawson, S. (1993) 'Evaluating Results Using Corrected and Uncorrected Effect Size Estimates'. Journal of Experimental Education, 61, 4, 334-349.

44 Rosenthal, R. (1994) 'Parametric Measures of Effect Size' in H. Cooper and L.V. Hedges (Eds.), The Handbook of Research Synthesis. New York: Russell Sage Foundation.

45 Kirk, R.E. (1996) 'Practical Significance: A concept whose time has come'. Educational and Psychological Measurement, 56, 5, 746-759.

46 Coe, R. (2002) "It's the Effect Size, Stupid. What effect size is and why it is important" Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England, 12-14 September 2002

47 Fitz-Gibbon C.T. (2002) 'A Typology of Indicators for an Evaluation-Feedback Approach' in A.J.Visscher and R. Coe (Eds.) School Improvement Through Performance Feedback. Lisse: Swets and Zeitlinger.

Summary

An ***effect size*** is a measure of the size of the causal effect of an intervention within a controlled experimental study or quasi experimental evaluation. The interpretation of ***effect size*** is dependent on the assumption that the control and experimental (intervention) groups are normally distributed with the same standard deviations. Without these assumptions the interpretation of ***effect sizes*** can be problematic , for example, when a sample has a restricted range, does not come from a normal distribution, or if the measurement from which it was derived has unknown reliability.

Care must be therefore be taken in comparing or aggregating ***effect sizes*** based on different outcomes, different operationalisations of the same outcome, different treatments, levels of the same treatment, or measures derived from different populations.

KNOWLEDGE CENTRE
FOR EDUCATION

KNOWLEDGE CENTER FOR EDUCATION:
PHONE: +47 22 03 70 00
E-MAIL: kunnskapssenter@forskningsradet.no
WEBPAGE: www.kunnskapssenter.no
FACEBOOK: kunnskapssenter
TWITTER: kunnskapsrad

TIDLIGERE UTGIVELSER FRA KUNNSKAPSSENTER FOR UTDANNING:

Lillejord, S., Vågan, A., Johansson, L., Børte, K. & Ruud, E. (2016). H*vordan fysisk aktivitet i skolen kan fremme elevers helse, læringsmiljø og læringsutbytte. En systematisk kunnskapsoversikt.* Oslo. Kunnskapssenter for Utdanning. www.kunnskapssenter.no

Børte, K., Lillejord, S. & Johansson, L. (2016). *Evnerike elever og elever med stort læringspotensial: En forskningsoppsummering.* Oslo: Kunnskapssenter for Utdanning. www.kunnskapssenter.no.

Lillejord, S., Børte, K., Halvorsrud, K., Ruud, E., & Freyr, T. (2015). *Tiltak med positiv innvirkning på barns overgang fra barnehage til skole: En systematisk kunnskapsoversikt.* Oslo: Kunnskapssenter for utdanning, www.kunnskapssenter.no (fins på engelsk)

Lillejord, S., Halvorsrud, K., Ruud, E., Morgan, K., Freyr, T., Fischer-Griffiths, P., Eikeland, O. J., Hauge, T. E., Homme, A. D., & Manger, T. (2015). *Frafall i videregående opplæring: En systematisk kunnskapsoversikt.* Oslo: Kunnskapssenter for utdanning, www.kunnskapssenter.no

Lillejord, S., Ruud, E., Fischer-Griffiths, P., Børte, K., & Haukaas, A. (2014). F*orhold ved skolen med betydning for mobbing. Forskningsoppsummering.* Oslo: Kunnskapssenter for utdanning www.kunnskapssenter.no

Lillejord, S. & Børte, K. (2014). *Partnerskap i lærerutdanningen – en forskningskartlegging – KSU 3/2014.* Oslo: Kunnskapssenter for utdanning, www.kunnskapssenter.no

Lillejord, S., Børte, K., Ruud, E., Hauge, T. E., Hopfenbeck, T. N., Tolo, A., Fischer-Griffiths, P. & Smeby, J.-C. (2014). *Former for lærervurdering som kan ha positiv innvirkning på skolens kvalitet: En systematisk kunnskapsoversikt KSU 1/2014.* Oslo: Kunnskapssenter for utdanning, www.kunnskapssenter.no

KNOWLEDGE CENTRE
FOR EDUCATION