

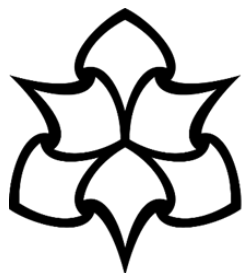


Education  
Endowment  
Foundation

# Using Digital Technology to Improve Learning: Evidence Review

December 2019

Professor Cathy Lewin  
Andrew Smith  
Professor Stephen Morris  
Elaine Craig  
(Manchester Metropolitan University)



**Manchester  
Metropolitan  
University**

Please cite this report as follows: Lewin, C., Smith, A., Morris, S. and Craig, E. (2019). *Using Digital Technology to Improve Learning: Evidence Review*. London: Education Endowment Foundation. The report is available from:

[https://educationendowmentfoundation.org.uk/public/files/Using\\_Digital\\_Technology\\_to\\_Improve\\_learning\\_Evidence\\_Review.pdf](https://educationendowmentfoundation.org.uk/public/files/Using_Digital_Technology_to_Improve_learning_Evidence_Review.pdf)

The EEF Guidance Report ***Using Digital Technology to Improve Learning*** is available at: <https://educationendowmentfoundation.org.uk/tools/guidance-reports/using-digital-technology-to-improve-learning/>

## Contents

Introduction.....	2
Methodology .....	3
Different ways that technology can be used to improve learning outcomes? ..	5
Digital technology characteristics that improve learning outcomes?.....	19
Evidence of the impact of digital technology in English schools .....	24
Conclusions .....	29
References .....	35
Appendices.....	40

## Introduction

The EEF review of the impact of digital technology on learning (Higgins et al., 2012) found positive benefits but noted that how technology is used (the pedagogy) is key and that future research should focus on identifying the specific conditions under which a positive impact is found. The OECD (2015) confirms that effects depend on context and specific uses with analyses of PISA data from 2012 suggesting that some use of technology can positively impact on attainment but that above average use can be 'associated with significantly poorer student performance' (p16). Digital technology can have a positive impact when it is embedded in a well-designed learning system (including factors such as the curriculum, pedagogy, teacher professional development and assessment) (US Office of Educational Technology, 2014), and when it supplements rather than replaces other forms of instruction (Higgins et al., 2012; Hattie & Yates, 2013). However, we still need to understand more about how and when digital technology should be used to support raising attainment.

There has been a plethora of research publications since the original review in 2012 but also a number of systematic reviews and meta-analyses. These studies examine specific uses of technology such as robotics/artificial intelligence, game based learning, augmented reality, virtual reality and mobile learning (e.g. Clark et al., 2016, Bano et al., 2018). In addition, some reviews focus on the use of technology in curriculum content areas such as mathematics, reading and science (e.g. Cheung & Slavin, 2013; D'Angelo et al., 2014). A synthesis of more recent evidence is timely, making an important contribution to the field and enabling teachers to make informed decisions about when and how to integrate digital technology in their pedagogy. To frame this research, we have adopted the typology of technology uses in education identified by Luckin and colleagues (2012), that of learning: from experts, with others, through making, through exploring, through inquiry, through practising, from assessment, and in and across settings. That is, the typology focuses on the pedagogy rather than the technology.

The review undertaken in 2012 (Higgins et al.) concluded that positive benefits of technology use for supporting learning have been consistently identified. As well as supporting innovative pedagogies, technologies can make teaching and learning more efficient and/or offer supplementary practice. Typically, the overall effect size was reported as being between 0.3 and 0.4 (small to medium), just slightly below average educational interventions, although the effect sizes ranged from -0.03 to 1.05. However, the landscape is complex. The heterogeneity of the research conducted to date makes it "difficult to identify clear and specific implications for educational practice in schools" (Higgins et al., 2012, p3). That is, there is great variety in terms of the types of technology used, the context including the subject area, the pedagogy employed, implementation and support, and the research methodology adopted. Indeed, how the technology is used is crucial.

The review also noted that an issue unique to technology in education is the continual and rapid change; in contrast it takes time for evidence of impact of new technologies or specific approaches to be gathered and published. Furthermore, many meta-analyses include studies under the broad heading of technology use (e.g. mobile learning), often including different education stages and curricula foci, which makes it difficult to interpret. Finally, the focus of research shifts to new technologies which means that studies are not conducted with well-established and embedded learning technologies beyond its introduction.

Key themes from the analyses of 48 studies that synthesised primary research on the use of technology to support teaching and learning (Higgins et al., 2012) were identified as:

- Collaborative use (pairs or small groups) is usually more effective than individual use.
- It is difficult to draw conclusions about the ideal duration and intensity of technology use as findings are not consistent. However, it seems that short, focussed, and regular use (e.g. 3 x per week, over 1 term) can be effective.

- Technology can offer a means for lower ability, special educational needs and disadvantaged young people to access intensive support and catch up with their peers.
- “[T]echnology is best used as a supplement to normal teaching rather than as a replacement for it.” (p.14)
- Greater gains in attainment are found in the domains of maths and science; in literacy there are more examples of technology being more effective for writing than for reading.
- Professional development for teachers is a key success factor in interventions.

Higgins and colleagues (2012) noted that the link between learning technology and attainment may not be causal and may not be linear; there may be a limit to the amount of technology use that can be beneficial. Furthermore, the relative impact of learning technology compared to other pedagogical innovations is not usually investigated; the impact of technology innovations is similar to other educational innovations.

## Methodology

A detailed account of our methodology can be found in Appendix C. Consistent with previous work in this area (Higgins et al., 2012), we sought to identify meta-analyses which summarise the effects of the use of digital technology in schools, and their effects on attainment. Indeed our review sought to identify relevant literature published since the previous work was completed in order to provide an update to that synthesis. We therefore searched for papers published since 2012 using a very similar approach to Higgins et al. (2012). Following our rapid assessment of meta-analyses, we conducted an evidence review (REA) of individual studies which described research undertaken in English second rapid schools which has been published since 2012.

Meta-analyses synthesise the results from a number of primary studies, and they therefore offer a thorough methodology for synthesising knowledge within a particular area. By systematically searching for meta-analyses our review therefore offers a comprehensive overview of international research into the effect of technology on attainment in schools.

The terms *systematic review* and *meta-analysis* are used somewhat interchangeably in this document. Systematic review refers to the overall process of searching for and selecting evidence, analysing it and reaching conclusions about the body of evidence; meta-analysis is one particular quantitative method of synthesising the findings from the evidence included within a systematic review.

In particular, we sought to identify meta-analyses which contained studies representing levels 4 and 5 on the Maryland Scientific Methods scale. Such studies allow for strong inferences to be made between the intervention and the observed effects, as:

- i) they either match the intervention and comparison groups (level 4), thereby controlling for between group variables which might otherwise explain the effect, rather than the intervention, or
- ii) randomise the allocation of the intervention (level 5), thereby creating groups which are statistically matched on *observed* and *unobserved* variables.

As randomisation creates equivalency on unobserved variables, it allows for stronger inferences to be made about the causes of observed effects.

Table 1: Maryland Scientific Methods scale.

Standard	Comparison	Description	Examples of methods
Level 1	No comparison	Attainment rate is reported for intervention group only	Before and after study
Level 2	Comparison with predicted rate	Actual and expected attainment rates of intervention group are compared	Expected local attainment rates compared with national attainment rates
Level 3	Unmatched comparison group	Attainment rate of intervention group is compared with attainment rate of an unmatched comparison group	Comparison of mean levels of attainment
Level 4	Well-matched comparison group	Attainment rate of intervention group is compared with attainment rate of a comparison group matched on static (and dynamic) factors e.g. progress, gender	Propensity score matching; regression discontinuity
Level 5	Randomised control trial (RCT)	Attainment rates are compared between intervention and control groups that have been created through random assignment	Randomisation

### The meta-analyses included in this review: effect sizes

Throughout this document we report overall summary effect sizes of the included meta-analyses. Effect sizes represent the relationship between two variables, and as described in this document represent the effect of the intervention (a particular pedagogical use of technology) on attainment. Meta-analyses summarise the effect sizes of the individual studies which constitute them, such that each meta-analysis arrives at an overall summary effect size. A positive overall effect size suggests that the intervention caused an increase in attainment, whereas a negative effect size suggests the opposite was true. We have included effect sizes as reported in the reviewed meta-analyses. Authors of the reviewed meta-analyses have calculated effect sizes by using various statistical methods (commonly Cohen's *d* or Hedges' *g*).

Meta-analyses also often report various summary effect sizes for different subgroups in the study (for example, primary and secondary pupils, girls and boys). Throughout this document, we report summary effect sizes where appropriate.

We have used the acronyms *FEM* (fixed-effects model) and *REM* (random-effects model) throughout this document. These denote the methods by which the meta-analyses have calculated overall and summary effect sizes. A fixed-effects model assumes homogeneity of effect (Borenstein et al., 2009), i.e. that all the included studies are functionally equivalent. By contrast, a random-effects model is more appropriate for use when meta-analysts are synthesising data from a number of studies performed by other researchers as it does not assume a true effect size which is consistent across all the included studies. REM therefore calculates wider confidence intervals around the effect sizes.

When trying to understand the magnitude and practical significance of effect sizes there are a range of considerations, and various interpretations have been offered by statisticians and researchers (Higgins, Kokotsaki and Coe, 2012). Cohen (1969, *ibid*) describes an effect of 0.2 as small, 0.5 as medium and 0.8 as large. We have used these boundaries in our report to offer some interpretation of the various effect sizes reported.

### The meta-analyses included in this review: quality and certainty of the findings

We analysed each meta-analysis included in our evidence review to make a judgement about the quality of the systematic review process of each, and the certainty of the effect estimate of each. To rate quality we used the AMSTAR2 appraisal tool, and to rate certainty of the effect estimates we used the GRADE framework. For more details of these tools, please see Appendix C.

Each study was rated as low, medium or high *quality*, representing the quality of the systematic review process based on a judgement wholly informed by our reading of the published article. Due to

variability in how systematic reviews are written (including the different lengths of the reviews and author choices about where to include more or less detail), it was sometimes difficult to make a judgement with certainty. As a result, some systematic reviews may have been downrated due to a lack of information available to evidence a specific quality-related criterion. This is an important point to note when interpreting quality ratings for specific systematic reviews included in this review, and also when considering the body of evidence as a whole.

*Certainty of effect* means how certain we can be that the effect sizes identified by the included meta-analyses are as stated. For example, if a meta-analysis concludes that the use of a specific type of technology leads to an increase in attainment, how confident can we be in the claim about the direction of the effect (i.e. that the intervention resulted in an increase, rather than a decrease, in attainment)? Moreover, where the meta-analysis states the magnitude of the effect, how confident can we be in the precision of that stated magnitude as a *true effect* which would generalise to the population of interest (in this instance, school children in the United Kingdom).

Taking five criteria into account (detailed in Appendix C), we made an overall judgement about the certainty of effect of each of the meta-analyses included in our review:

- *High*. We are very confident that the true effect lies close to that of the estimate of the effect.
- *Moderate*. We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- *Low*. Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
- *Very Low*. We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

After discussion, we decided not to include any systematic reviews which we rated as having a *Very Low* certainty of effect. This resulted in the exclusion of eight reviews from further analysis.

### Other literature taken into consideration

Seven recent reports from the What Works Clearinghouse have been included. Two additional reports identified initially have been disregarded because one reported on a web-based program that is no longer available for purchase and the other presented no high quality evidence.

We have also drawn on Hattie's latest updated list of factors related to student achievement (Hattie, 2017). We have provided some of these effect sizes to make comparisons with the synthesis we present below.

## Different ways that technology can be used to improve learning outcomes?

We have organised the evidence in two primary ways: by domain areas (linked to the key themes of interest for the EEF) and by drawing on the typology of technology uses identified by Luckin and colleagues (2012). Studies may incorporate different domain areas and multiple pedagogical approaches; therefore, the themes identified below are not mutually exclusive. In addition to the meta-analyses that we have identified in our review, we have included findings from wider literature such as the What Works Clearinghouse that supplement the findings. Evaluation reports from the EEF individual studies of technology use in England are reviewed below (see What evidence is there about how English schools are currently using digital technology to improve learning?).

## Learning in Literacy and Language (including second language learning)

(13 articles: Abrami, Borokhovski & Lysenko, 2015; Archer et al., 2014; Castillo-Manzano et al., 2016; Chen, Tseng & Hsiao, 2018; Cheung & Slavin, 2012; Gerard et al., 2015; Kunkel, 2015; Lin, 2014; Sung, Chang & Liu, 2016; Sung, Chang & Yang, 2015; Takacs, Swart & Bus, 2015; Wouters et al., 2013; Zheng et al., 2016)

A total of 13 meta-analyses met the inclusion and quality criteria. Five (Abrami et al, 2015; Archer et al., 2014; Cheung & Slavin, 2012; Kunkel, 2015; Takacs et al., 2015) of the meta-analyses focus specifically on literacy and language interventions whereas three (Gerard et al., 2015; Sung et al., 2016; Zheng et al., 2016) report on a range of domains which included literacy and language. A further five meta-analyses focus on second language learning (Castillo-Manzano et al., 2016; Chen et al., 2018; Lin, 2014; Sung et al., 2015; Wouters et al., 2013).

As with previous research, the majority of these studies report a small, positive overall effect of technology use on attainment in literacy and language when compared to traditional classroom practices.

Five meta-analyses report on the impact of technology on reading (primarily in primary education) with effect sizes ranging from 0.1 to 0.23 in primary education and from 0.03 (non-significant) to 0.31 in secondary education. This compares with Hattie's (2017) effect size for the use of technology in reading of 0.29. Notably, two of these meta-analyses (Cheung & Slavin, 2012; Kunkel, 2015) report effect sizes for the impact of technology on reading attainment in the early years. The interventions tested include a game-based application for primary education (Abrami et al., 2015), electronic story books (Takacs et al., 2015) and one-to-one laptop programs (Zheng et al., 2016). Cheung and Slavin (2012) and Kunkel (2015) reviewed a large number of studies designed to support the teaching of reading, some of which are intended to provide supplementary practice and others designed to be integrated with other classroom activities.

Only one study (Zheng et al., 2016) reports specifically on the impact of technology on writing with an overall effect size of 0.20 for both primary and secondary education (combined). This study focuses on one-to-one laptop programs without providing details about the specific ways in which the laptops were used to support writing. This compares with Hattie's (2017) effect size for the use of technology in writing of 0.42.

Five meta-analyses report on language and literacy more generally (including reading and writing) with overall effect sizes ranging from 0.15 to 0.473 in both primary and secondary education. However, some of these effect sizes include other domains (Archer et al., 2014; Gerard et al., 2015) or adults (Sung et al., 2016; Wouters et al., 2013), which could account for them being slightly higher than for studies that focus exclusively on reading and writing interventions in school contexts. The types of technology use include drill and practice, intelligent tutoring systems, mobile learning, serious games and one-to-one laptops.

In relation to second language learning, the evidence suggests there is a positive effect of technology on attainment with overall effect sizes ranging from 0.29 to 0.873 for primary education and from 0.48 (non-significant) to 0.873 for secondary education. Four of these five meta-analyses include adults in addition to primary and secondary students, but separate effect sizes were calculated for adults and children in all cases. Furthermore, the effect size reported by Castillo-Manzano and colleagues (2016) for primary and secondary school students (0.48) was not statistically significant (by our standards rather than the authors), which can be attributable to the small number of effect sizes (6 of 53) relating to school-aged students. Only one of these effect sizes related to language learning. This meta-analysis concerned the use of audience response systems (or clickers) for assessment. Therefore, it is not possible to comment on whether the use of audience response systems is beneficial in language learning as the evidence is insufficient; we have included the study here for completeness.



The large effect size reported by Chen et al. (2018) included 6 studies at primary level and one at secondary level which explored the use of game-based learning for second language vocabulary learning.

Two of these studies present different evidence in relation to the impact of the use of mobile technology for different types of language learning (Lin, 2014; Sung et al., 2015). Lin (2014) suggests that computer mediated communication is significantly more effective for 'foreign language learning' (0.610) than for second language learning (where the language is commonly used outside the classroom) (-0.038). Sung and colleagues (2015) conclude that mobile learning is more effective for second/foreign language (0.837) than for first language learning (0.181).

Table 2: Effect sizes relating to language and literacy (P: Primary, S: Secondary)

Focus	Author	Overall	P/ S	P	S	Early Years	Notes
Reading	Abrami et al.	0.179		0.179			
	Cheung & Slavin, 2012 <sup>1</sup>	0.16		0.10	0.31	0.15	
	Kunkel	0.21 0.13		0.23 0.12	0.03 (NS) 0.18	0.50 0.20	Post-test designs Gain designs
	Takacs et al. <sup>2</sup>	0.17 0.20		0.17 0.20			Story comprehension Expressive vocabulary
	Zheng et al. <sup>3</sup> 2016	0.16	0.12 (NS)				
Writing	Zheng et al. 2016	0.16	0.20				
Literacy and Language	Archer et al. <sup>4</sup>	0.181	0.181				
	Gerard et al. <sup>5</sup>	0.34 0.27	0.34 0.27				Automated, adapted guidance Enhanced automatic guidance
	Sung et al., 2016 <sup>6</sup>	0.523	0.473				
	Wouters et al. 2013 <sup>7</sup>	0.29	0.66				Serious games
	Zheng et al. 2016	0.16	0.15				
Second language learning	Castillo-Manzano et al. <sup>8</sup>	0.288	0.48 (non-significant)				
	Chen et al. <sup>9</sup>	1.027	0.873				
	Cheung & Slavin, 2012 <sup>10</sup>	0.29		0.29			
	Lin <sup>11</sup>	0.441		0.615	0.426		
	Sung et al. 2015 <sup>12</sup>	0.351		0.508	0.488		

Chauhan (2017) did not meet the stringent inclusion criteria applied in this review as the meta-analysis included 54 (of 122) studies that did not have a control group, and without a counterfactual there is a high degree of uncertainty around the effect sizes. In common with three of the studies reviewed above, Chauhan (2017) reports on the impact of technology on attainment in primary

<sup>1</sup>Test for difference by educational level suggests that the effect at secondary level is greater than at primary or in the early years.

<sup>2</sup> 24 of 43 studies were conducted in the US.

<sup>3</sup> Effect size for reading calculated for primary and secondary combined. Overall effect size includes all domains.

<sup>4</sup> The majority of these studies focused on reading. Effect size calculated for primary and secondary combined.

<sup>5</sup> Overall gain reported for all domains/educational levels. 19% of studies were categorised as 'literacy and social studies' for automated guidance and 29% as 'language arts' for enhanced guidance. No statistically significant difference by domain or educational level – breakdown of effect sizes for sub-groups not provided.

<sup>6</sup> Overall effect size includes all domains/educational levels. Effect size for language arts reported for all participants, including adults. Domain was not a statistically significant moderator. There was a statistically significant difference by educational level with primary  $g=0.636$ , secondary  $g=0.451$  across all domains. However, post-hoc analyses to identify where the differences lie were not conducted.

<sup>7</sup> Overall effect size includes all domains and educational levels. Domain of 'language' includes first and second language learning.

<sup>8</sup> Overall effect size includes school and university. Only 6 of 53 effect sizes related to school students with only one study in primary education on English language learning and the remaining five from secondary education on science. An analysis of differences between school and home was conducted but the authors only tested if effect sizes were significantly different from zero and not from each other.

<sup>9</sup> Overall effect size includes adults. 6 studies at primary level, 1 at secondary level, 3 involve adults. Effect size calculated for primary and secondary combined. No statistically significant difference between educational level.

<sup>10</sup> Effect size for primary based on 3 studies only which focused on English as an Additional Language.

<sup>11</sup> Overall effect size includes adults. 59 studies of which 10 are at secondary level and 3 are at primary level. No statistically significant difference between educational level.

<sup>12</sup> Overall effect size includes adults. 31 of 44 studies involve primary and secondary school students. Statistically significant difference by educational level but this could be attributed to the 'mixed age' sub-group.

schools for a range of domains which includes literacy and language. Chauhan<sup>13</sup> reported an overall effect size for all domains of 0.546 and an effect size for literacy and language of 0.448. This compares to Hattie's (2017) effect sizes of 0.44 for technology in primary schools, 0.42 for the use of technology in writing and 0.29 for the use of technology in reading.

Abrami and colleagues (2015) concluded that the use of the ABRACADABRA reading system (a fully digital approach to teaching reading through interactive stories and games) has a positive impact on: phonics (0.189), phonemic awareness (0.324), listening comprehension (0.381) and vocabulary knowledge (0.138). No impact was identified on reading fluency and comprehension. A further study, commissioned by the EEF (McNally, Ruiz-Valenzuela and Rolfe, 2018), using a measure of phonics, literal comprehension and reading for meaning, suggests a positive effect of ABRACADABRA on literacy (0.138) although there was a positive impact for an offline equivalent (0.231), suggesting that the programme rather than the delivery method (technology or paper-based) accounted for the positive impact.

Evidence from the What Works Clearinghouse confirms that the picture is very complex. It is not the technology per se that makes the difference; rather it is the pedagogy and the context, as noted above. The reports summarise the evidence by providing a judgement about the strength of the evidence. Reviews are considered to offer evidence of a **positive effect** if they draw on two or more statistically significant studies, at least one of which is considered to be high quality, with no studies reporting a statistically significant negative impact. Reviews are considered to offer evidence of a **potentially positive effect** if one study is statistically significant, no studies report a statistically significant negative impact, and there are at least the same number of studies reporting a positive effect than there are an indeterminate effect. In addition, the six studies include three that are technology-based and three where technology use is integrated with other resources and optional rather than mandatory.

---

<sup>13</sup> From moderator analyses by domain; author only tested if effect sizes were significantly different from zero and not from each other. Studies identified include second language learning. Overall effect size includes all domains.

Table 3: Summary of evidence from the What Works Clearinghouse (n=number of studies considered)

Study	Educational Stage	Format	General literacy	Comprehension	Reading fluency	General reading	Alphabetics
Accelerated Reader	primary	Technology based		Inconsistent evidence (n=2)			
Achieve3000	primary secondary	Technology based	Potentially positive (n=2)	Potentially positive (n=2)			
Fast ForWard	primary	Technology based		Inconsistent evidence (n=4)	No discernible effect (n=1)		Positive effect (n=5)
Prentice Hall/Pearson Literature	secondary	Integrated package, optional online components; not clear if technology used in studies	No discernible effect (n=2)	No discernible effect (n=3)			
Read Naturally	primary	Integrated package, optional software resource; only 3 of 5 studies tested the software version		No discernible effect (n=2 software, n=2 audio CDs/texts)	Inconsistent evidence (n=2 software, n=2 audio CDs/texts)	Potentially positive effect (n=1 software although only in 1 of 3 schools, n=1 audio CDs/texts)	No discernible effect (n=1 software, n=1 audio CDs/texts)
Read 180	primary, secondary	Integrated package, technology standard but only as a small part (10-20 minutes of a 45-90 minute session)	Positive effect (n=6)	Positive effect (n=6)	Potentially positive (n=2)		No discernible effect (n=1)

## Learning in Mathematics

11 articles: Belland et al., 2016; Cheung & Slavin, 2013; D'Angelo et al., 2014; Gerard et al., 2015; Kim, Belland and Walker, 2017; Kulik & Fletcher, 2016; Sokolowski, Li & Wilson, 2015; Steenbergen-Hu & Cooper, 2013; Sung, Chang & Liu, 2016; Wouters et al., 2013; Zheng et al., 2016).

A total of 11 meta-analyses met the inclusion and quality criteria. Four (Cheung & Slavin, 2013; Kulik & Fletcher, 2016; Sokolowski et al., 2015; Steenbergen-Hu & Cooper, 2013) of the meta-analyses focus specifically on mathematics interventions whereas seven (Belland et al., 2016; D'Angelo et al., 2014; Gerard et al., 2015; Kim et al., 2017; Sung et al., 2016; Wouters et al., 2013; Zheng et al., 2016) report on a range of domains which included mathematics.

The evidence for learning in mathematics is broadly positive, with effect sizes suggesting a small to medium impact. These findings compare to Hattie's (2017) effect size for the use of technology in mathematics of 0.33. The reported effect sizes vary from 0.15 to 0.61 for primary education and from -0.1 (non-significant) to 0.65 for secondary education. The non-significant impact of technology in mathematics reported by D'Angelo and colleagues (2014) for school-aged students is based on only 4 of 96 effect sizes, so should be viewed with caution. Similarly, Kim and colleagues' (2017) report on the impact of scaffolding in STEM subjects but only 5 of 47 outcomes relate to mathematics. Finally, only 3 effect sizes relate to primary level education in Steenbergen-Hu and Cooper's meta-analysis (2013). After discounting these effect sizes, there remains some variability in those reported.

The types of technology use include drill and practice, simulations, mobile learning, game-based learning and pedagogical scaffolding or intelligent tutoring systems. Scaffolding is defined as support or guidance from a more capable other in order to complete a task which could not be undertaken independently. Intelligent tutoring systems, which often include scaffolding, adapt in response to a learner's progress. The evidence suggests that simulations, scaffolding and intelligent tutoring systems lead to the greatest impact on mathematics attainment.

*Table 4: Effect sizes relating to mathematics (P: Primary, S: Secondary); highlighted estimates should be viewed with caution*

Author	Overall	P / S	P	S	Notes
Belland et al. <sup>14</sup> ,	0.46		0.55	0.37 0.48	Lower secondary Upper secondary
Cheung & Slavin, 2013 <sup>15</sup>	0.16		0.17	0.14	
D'Angelo et al. <sup>16</sup>	0.62 0.49	-0.15 (NS)			Simulations compared to none Enhanced simulations compared to simple
Gerard et al. <sup>17</sup>	0.34 0.27	0.34 0.27			Automated, adapted guidance Enhanced automatic guidance
Kim et al. <sup>18</sup>	0.385	0.425			
Kulik & Fletcher <sup>19</sup>	0.49	0.40			
Sokolowski et al. <sup>20</sup>	0.60		0.61 0.41	0.65	Lower primary Upper primary Lower secondary
Steenbergen-Hu & Cooper <sup>21</sup>	0.09 (NS) 0.02 (NS)		0.25 (NS) 0.42 (NS)	0.08 (NS) -0.1 (NS)	Unadjusted effect sizes (no pre-test) Adjusted effect sizes
Sung et al., 2016 <sup>22</sup>	0.523	0.377			
Wouters et al. <sup>23</sup>	0.29	0.17			Serious games
Zheng et al. <sup>24</sup>	0.16	0.17			

Chauhan (2017) did not meet the stringent inclusion criteria applied in this review as the meta-analysis included 54 (of 122) studies that did not have a control group, and without a counterfactual there is a high degree of uncertainty around the effect sizes. In common with seven of the studies reviewed above, Chauhan (2017) reports on a range of domains which includes mathematics.

<sup>14</sup> Overall impact includes all educational levels. For STEM only; no analyses by domain and unclear what proportion are mathematics focused. No statistically significant difference between educational levels.

<sup>15</sup> Overall impact includes all educational levels. No statistically significant difference between educational levels. 39 of 74 studies were published before 2000, although no statistical differences by publication period. 79% of studies were categorised as drill and practice (CAI).

<sup>16</sup> Overall impact includes all domains and educational levels. Primary and secondary based on only 4 of 96 effect sizes that focus on achievement in mathematics. Educational level for these 4 effect sizes is not clear. 21 of 59 studies were conducted in North America and 18 were conducted in Europe. Moderator analyses only performed on data relating to science education.

<sup>17</sup> Overall gain reported. 44% of studies were categorised as mathematics. No statistically significant difference by domain or educational level – breakdown of effect sizes for sub-groups not provided.

<sup>18</sup> Overall impact includes all domains. Only 5 of 47 outcomes relate to mathematics. The ages of the participants are not identified. It is likely that they include adults and the proportions of primary and secondary aged students are not known.

<sup>19</sup> Overall impact includes all domains. Effect size for mathematics includes all ages. School level learners:  $g=0.44$ . Adult learners:  $g=0.78$ . Impact for other domains:  $g=0.72$ .

<sup>20</sup> Overall impact includes all educational levels. Of 24 studies, 7 from Taiwan, 5 from US and 4 from Turkey. No statistical tests of differences conducted but effect sizes presented.

<sup>21</sup> Overall impact includes all educational levels. 10 different ITS that are popular in the US. REM only presented here; authors also calculated FEM.

<sup>22</sup> Overall impact includes all domains/educational levels. Effect size reported for all participants; includes adults. Domain was not a statistically significant moderator. There was a statistically significant difference by educational level with primary:  $g=0.636$  and secondary:  $g=0.451$  across all domains. However, post-hoc analyses to identify where the differences lie were not conducted.

<sup>23</sup> Overall impact includes all domains and educational levels.

<sup>24</sup> Overall impact includes all domains. Effect size for mathematics calculated for primary and secondary combined.

Chauhan<sup>25</sup> reported an overall effect size for all domains of 0.546 and an effect size for mathematics of 0.469.

Evidence from the What Works Clearinghouse is limited. Only one study (University of Chicago School Mathematics Project, UCSMP) involves mathematics and technology; in this programme the technology is integrated with a textbook and other resources for the teacher. The measures include general maths achievement and algebra. However, the number of eligible studies included in the review is small and studies themselves involved small numbers of students; the review acknowledges that the extent of the evidence is limited. There is no detail about the technology used but it includes graphing calculators, geometry systems and spreadsheets. The review concludes that the algebra unit of UCSMP has potentially positive effects on both general mathematics achievement and algebra. In addition, the use of multiple units of UCSMP has a potentially positive effect on general mathematics achievement.

## Learning in Science

9 articles: Belland et al., 2016; Castillo-Manzano et al., 2016; D'Angelo et al., 2014; Gerard et al., 2015; Kim, Belland and Walker, 2017; Sung, Chang & Liu, 2016; Wouters et al., 2013; Zheng et al., 2016; Zheng et al., 2018.

A total of 9 meta-analyses met the inclusion and quality criteria. All of these studies report on a range of domains which included science; that is, there are no meta-analyses that focus specifically on the domain of science. Only two of the nine studies report effect sizes that relate just to science in primary and/or secondary education without including other domains or older students (D'Angelo et al., 2014; Zheng et al., 2018).

The effect sizes range from 0.11 (non-significant) to 1.42 for primary education and from 0.11 (non-significant) to 1.47 for secondary education. As explained below, the large effect sizes for both primary and secondary education should be viewed with caution. More reliable top-end estimates are 0.82 for primary education and 0.64 for secondary education. In comparison, Hattie's (2017) effect size for the use of technology in science is 0.23. The evidence includes the effects of technology in other domains and there is some variability in the effect sizes reported.

The effect size reported by Castillo-Manzano and colleagues (2016) for primary and secondary school students (0.48) was not statistically significant (by our standards rather than the authors), partly attributable to the small number of effect sizes (6 of 53) relating to school-aged students. This meta-analysis concerned the use of audience response systems (or clickers) for assessment.

D'Angelo and colleagues' (2014) study relates to the use of simulations in science. Whilst the effect sizes calculated for each educational stage are significantly different from zero, they are not significantly different from each other. Furthermore, the large effect size at primary level (1.42) is only based on three studies meaning it could be inflated. Even so, the overall effect size for simulations in science education in schools is medium and positive, suggesting that this particular use of technology in science can be worthwhile.

Zheng and colleagues' (2016) review concerns 1-1 laptop provision. The effect size reported for science is based on 3 effect sizes from 2 studies and so should be viewed with caution. Zheng and colleagues' (2018) review focused on mobile learning and inquiry-based learning. With regards to educational level the differences between higher education, secondary and primary levels was not statistically significant. Furthermore, the large effect size for secondary education (1.47) is only based

---

<sup>25</sup> Overall impact includes all domains. From moderator analyses by domain; author only tested if effect sizes were significantly different from zero and not from each other.

on two studies and should be viewed with caution. Nevertheless, the evidence suggests that combining mobile learning and inquiry-based learning can be beneficial.

Considering the types of technology use, the most effective uses are simulations (D'Angelo et al., 2014) and mobile learning, both in general (Sung et al., 2016) and when combined with inquiry-based learning (Zheng et al., 2018). Scaffolding is also effective and one-to-one laptop programs can have a small, positive impact on science attainment.

*Table 5: Effect sizes relating to science (P: Primary, S: Secondary);*

Author	Overall	P / S	P	S	Notes
Belland et al. <sup>26</sup> ,	0.46		0.55	0.37 0.48	Lower secondary Upper secondary
Castillo-Manzano et al. <sup>27</sup>	0.288	0.48 (NS)			
D'Angelo et al. <sup>28</sup>	0.62		1.42	0.64 0.63	Lower secondary Upper secondary
Gerard et al. <sup>29</sup>	0.34 0.27	0.34 0.27			Automated, adapted guidance Enhanced automatic guidance
Kim et al. <sup>30</sup>	0.385	0.146			
Sung et al., 2016 <sup>31</sup>	0.523	0.564			
Wouter et al., 2013 <sup>32</sup>	0.29	0.11 (NS)			Serious games
Zheng et al., 2016 <sup>33</sup>	0.16	0.25			
Zheng et al., 2018 <sup>34</sup>	0.91		0.82	1.47	

Chauhan (2017) did not meet the stringent inclusion criteria applied in this review as the meta-analysis included 54 (of 122) studies that did not have a control group, and without a counterfactual there is a high degree of uncertainty around the effect sizes. In common with the studies reviewed above, Chauhan (2017) reports on a range of domains which includes science, with a specific focus on primary education. Chauhan<sup>35</sup> reported an overall effect size for all domains of 0.546 and an effect size for science of 0.727 and for science and technology of 0.435.

### Learning from experts/with others and learning through practice

12 articles: Abrami, Borokhovski & Lysenko, 2015; Belland et al., 2016; Cheung & Slavin, 2013; Cheung & Slavin, 2012; Clark, Tanner-Smith & Killingsworth, 2016; Gerard et al., 2015; Kim, Belland

<sup>26</sup> Overall impact includes all educational levels. For STEM only; no analyses by domain and unclear what proportion are science focused. No statistically significant difference between educational level.

<sup>27</sup> Overall effect size includes school and university. Only 6 of 53 effect sizes related to school students from one study in primary education on English language learning and three studies from secondary education on science. Separate analyses by subject area were not conducted.

<sup>28</sup> Primarily focuses on science. 21 of 59 studies were conducted in North America and 18 were conducted in Europe. Moderator analyses only performed on data relating to science education, including educational level. No statistically significant difference by educational level.

<sup>29</sup> Overall gain reported. 37% of studies were categorised as science. No statistically significant difference by domain or educational level – breakdown of effect sizes for sub-groups not provided.

<sup>30</sup> Overall effect size includes all domains. 25 of 47 outcomes relate to science. The ages of the participants are not identified. It is likely that they include adults and the proportions of primary and secondary aged students are not known.

<sup>31</sup> Overall effect size includes all domains/educational levels. Effect size reported for all participants; includes adults. Domain was not a statistically significant moderator. There was a statistically significant difference by educational level with primary:  $g=0.636$ , secondary:  $g=0.451$  across all domains. However, post-hoc analyses to identify where the differences lie were not conducted.

<sup>32</sup> Overall effect size includes all domains and educational levels.

<sup>33</sup> Effect size calculated for primary and secondary combined. Overall impact includes all domains.

<sup>34</sup> Overall effect size includes all domains/educational levels. Domain was a statistically significant moderator but educational level was not. Effect sizes for primary and secondary include all domains with over two thirds (68%) conducted in natural sciences.

<sup>35</sup> Overall impact includes all domains. From moderator analyses by domain; authors only tested if effect sizes were significantly different from zero and not from each other.

& Walker, 2017; Kulik & Fletcher, 2016; Kunkel, 2015; Means et al., 2013; Steenbergen-Hu & Cooper, 2013; Zheng, 2016).

We have categorised meta-analyses that focus on scaffolding, intelligent tutoring systems, computer assisted instruction (CAI) or online learning as 'learning from experts', conceptualising the expertise as being embedded within the systems. Scaffolding is defined as support or guidance from a more capable other in order to complete a task which could not be undertaken independently. Intelligent tutoring systems are defined by Steenbergen-Hu & Cooper (2013, p970) as "self-paced, learner-led, highly adaptive, and interactive learning environments" which adapt to a learner's individual needs. CAI is an umbrella term that is used for a variety of systems. Cheung & Slavin (2012, p201) define their use of this term as systems that "provide additional instruction at students' assessed levels of need to supplement traditional classroom." Kunkel (2015, p.8) defines CAI as "instruction [that] is provided via the application rather than a teacher, volunteer, or classroom aid, and the application is both reliant on and responsive to the learner's actions." CAI is often associated with software designed to provide 'drill and practice' in specific skills.

Four articles focus specifically on scaffolding (Belland et al., 2016; Gerard et al., 2015; Kim et al., 2017; Zheng, 2016) whilst in two studies scaffolding is an aspect of the technology under consideration (Abrami et al., 2015; Clark et al., 2016). Two studies focus specifically on intelligent tutoring systems which also incorporate scaffolding (Kulik & Fletcher, 2016; Steenbergen-Hu & Cooper, 2013). Three studies include analyses of CAI (Cheung & Slavin, 2012; Cheung & Slavin, 2013; Kunkel, 2015).

The majority of the evidence suggests a small, positive effect of systems incorporating computer-based scaffolding when compared to traditional classroom practices. The effect sizes range from 0.179 to 0.55 for primary education and from 0.08 (non-significant) to 0.841 for secondary education. In comparison, Hattie's (2017) effect size for intelligent tutoring systems is 0.48.

Two studies report moderate effect sizes for scaffolding, one in relation to primary education (Belland et al., 2016) and one in relation to secondary education (Zheng, 2016). Two studies that focus on the use of intelligent tutoring systems (ITS) in mathematics compared to conventional instruction provide contradictory evidence. One (Kulik & Fletcher, 2016) suggests a small to moderate effect size (0.44) whilst the other (Steenbergen-Hu & Cooper, 2013) suggests no difference between ITS use and regular classroom instruction.

One study (Lin, 2014) specifically focuses on learning through communication, reviewing the evidence on the impact of computer mediated communication (CMC) and second language learning. This meta-analysis includes studies involving adults with 10 of 59 studies conducted in secondary schools and 3 conducted in primary schools. The author reports moderate, positive effect sizes for the impact of CMC on language learning in schools. Another study (Means et al., 2013) focuses on online and blended learning, a broad categorisation that includes communication and content, as well as online assessment. Means and colleagues compared online and blended learning to traditional classroom instruction. However, they noted that most of the 45 studies that they identified were modest in scope and furthermore, only seven of these studies were conducted with school-aged students. The studies focused on different domains with the most popular being medicine or health care.



Table 6: Effect sizes relating to learning from experts, others and through practice (P: Primary, S: Secondary)

Focus	Author	Overall	P / S	P	S	Notes
Scaffolding	Abrami et al. <sup>36</sup>	0.179		0.179		
Scaffolding	Belland et al. <sup>37</sup>	0.46		0.55	0.37 0.48	Lower secondary Upper secondary
Scaffolding	Gerard et al. <sup>38</sup>	0.34 0.27	0.34 0.27			Automated, adapted guidance Enhanced automatic guidance
Scaffolding	Kim et al. <sup>39</sup>	0.385	0.385			
Scaffolding	Zheng <sup>40</sup>	0.438		0.277	0.841	
Intelligent tutoring system	Kulik & Fletcher <sup>41</sup>	0.66	0.44			
Intelligent tutoring system	Steenbergen-Hu & Cooper <sup>42</sup>	0.09 (NS) 0.01 (NS)		0.25 (NS) 0.42 (NS)	0.08 (NS) -0.1 (NS)	Unadjusted effect sizes (no pre-test) Adjusted effect sizes
CAI	Cheung & Slavin, 2012 <sup>43</sup>	0.16	0.11			
CAI	Cheung & Slavin, 2013 <sup>44</sup>	0.16	0.19			
CAI	Kunkel	0.21 0.13		0.23 0.12	0.03 (NS) 0.18	Post-test designs Gain designs
Computer mediated communication	Lin <sup>45</sup>	0.441		0.615	0.426	
Online and blended learning	Means et al. <sup>46</sup>	0.2		0.1664 (NS)	0.1664 (NS)	

Some of the moderator analyses conducted in these studies provide further insight on who benefits from scaffolding and the types of scaffolding that are beneficial.

Belland and colleagues conducted multiple analyses on specific scaffolding feedback (e.g. whether or not it adapted according to user performance) but there was no evidence to suggest that any of them had a greater effect than others. In contrast, Kim and colleagues (2017) reported that conceptual scaffolding (providing hints and tools to identify what to consider when problem-solving) ( $g=0.126$ ) is not as effective as metacognitive scaffolding (reflecting on the learning process and considering possible solutions) ( $g=0.384$ ) and strategic scaffolding (helping students to locate and evaluate information for problem-solving and guiding the selection of an approach) ( $g=0.345$ ).

<sup>36</sup> Some scaffolding built in; feedback and suggestions provided following incorrect answers.

<sup>37</sup> For STEM only; no analyses by domain and unclear what proportions are mathematics and science focused. No statistically significant difference between educational level.

<sup>38</sup> Overall gain reported. 37% of studies were categorised as science. No statistically significant difference by domain or educational level – breakdown of effect sizes for sub-groups not provided.

<sup>39</sup> Only 5 of 47 outcomes relate to mathematics. The ages of the participants are not identified. It is likely that they include adults and the proportions of primary and secondary aged students are not known.

<sup>40</sup> Four of 29 studies involved primary school students. Eight of 29 studies involved secondary aged students. The remaining 17 studies were conducted in higher education contexts. No statistically significant difference by educational level. No statistically significant difference by domain (natural science, social science, medicinal science).

<sup>41</sup> Overall impact for all domains. Study reported the median effect size. Adult learners:  $g=0.78$ .

<sup>42</sup> 10 different ITS that are popular in the US. No difference overall (primary and secondary); unadjusted (no pre-test):  $g=0.09$  (non-significant), adjusted:  $g=0.01$ .

<sup>43</sup> Effect size for supplemental learning or CAI. 67% of studies were categorised as supplemental (CAI). Applies to all educational levels. Test for difference by educational level suggests that the effect at secondary level is greater than at primary or in the early years.

<sup>44</sup> Effect size for primary/secondary for supplemental CAI only. Over 70% of studies were categorised as supplemental CAI. Applies to all educational levels. 39 of 74 studies were published before 2000, although no statistical differences by publication period.

<sup>45</sup> 59 studies of which 10 are at secondary level and 3 are at primary level. No statistically significant difference between educational level.

<sup>46</sup> Only 7 of 45 studies were conducted with school students. Overall effect size includes university students for whom online learning had an impact (0.309).

Kunkel's analyses suggest that a positive effect of CAI for reading instruction is only evident when the comparisons are made with no treatment or non-reading CAI. That is, the young people involved get more learning time and this could account for any gains made. When comparisons were made with teacher-led interventions targeting reading (that is, similar activities) there was no difference in effect.

In relation to CMC and second language learning, Kim (2014) suggests that communicating with peers (0.495) or native speakers (0.487) are more effective than communicating with teachers (0.031, non-significant).

Although analyses did not include comparisons with traditional instruction and so this study has not been included in the main body of this review, Wouters & van Oostendorp (2013) compared game-based learning with instructional support to that without. Domains included biology and mathematics; participants included primary, secondary and university students. They concluded that game-based learning with instructional support is more effective than without any support (0.34). There were no differences by educational level. There was a greater effect on skills (0.62) than there was on knowledge (0.33) or in-game performance (0.19). Similarly, Clark and colleagues (2016) reported that games with scaffolding (such as personalised support, adapting to needs or interests) are more effective than games without (0.41)<sup>47</sup>. Notably, teacher provided scaffolding was the most effective (0.58).

### Learning through exploring

5 articles: Chen, Tseng & Hsiao, 2018; Clark, Tanner-Smith & Killingsworth, 2016; D'Angelo et al., 2014; Sokolowski, Li & Wilson, 2015; Wouters et al., 2013.

This section synthesises the research in relation to two specific uses of technology: simulations and game-based learning. Luckin et al. (2012) include these two kind of technology in their categorisation, noting that very little research exists in this area suggesting that 'technology-supported exploration is underused and undervalued within educational settings' (p29).

Two meta-analyses concern learning through exploring. D'Angelo and colleagues (2014) identified 59 studies that focus on the use of simulations in primary and secondary education. The majority relate to the domain of science with a very small number conducted in the domain of mathematics. Moderator analyses including educational level were only conducted for studies in the domain of science. Whilst the effect sizes calculated for each educational stage are significantly different from zero, they are not significantly different from each other. Furthermore, the large effect size at primary level is only based on three studies meaning it could be inflated. Even so, the overall medium effect size for simulations in science education in schools is medium and positive, suggesting that this particular use of technology in science can be worthwhile.

Sokolowski and colleagues (2015) focused on a slightly broader category of technology use which they defined as 'exploratory computerised environments'. This includes simulations but also dynamic geometry software. They identified 24 studies involving students in primary and secondary mathematics education.

The evidence from these two studies suggest that simulations are beneficial for science and mathematics with moderate, positive effect sizes.

Three meta-analyses relate to games-based learning. Chen and colleagues (2018) focused on studies of game-based learning for second language vocabulary learning, including those involving adult learners. Clark and colleagues (2016) included studies on game-based learning in general, but

---

<sup>47</sup> This is not summarised in the table above as it is a comparison of two formats of game. No breakdown by educational level but primary and secondary only; the authors state that majority of participants are from secondary school.

in school contexts only. Wouters and colleagues (2013) focused on serious games (defined as those with educational objectives), identifying 77 effect sizes the majority of which were generated from studies conducted in school contexts. Most of these studies were undertaken in the domain of biology or mathematics, with 11 effect sizes relating to language learning. The evidence suggests a small, positive effect of game-based learning on attainment.

*Table 7: Effect sizes relating to learning through exploring*

Focus	Author	Overall	Primary	Secondary	Notes
Simulation	D'Angelo et al. <sup>48</sup>	0.62	1.42	0.64 0.63	Lower secondary Upper secondary
Simulation	Sokolowski et al. <sup>49</sup>	0.65	0.61 0.41	0.65	Lower primary Lower secondary Upper primary
Game-based learning	Chen et al. <sup>50</sup>	1.027	0.873	0.873	
Game-based learning	Clark et al. <sup>51</sup>	0.33 0.35	0.33 0.35	0.33 0.35	All learning outcomes Cognitive learning outcomes
Game-based learning (serious games)	Wouters et al. <sup>52</sup>	0.29	0.30	0.33	

Moderator analyses conducted by some authors provide more insight into the ways in which games can be used to greatest effect.

D'Angelo and colleagues (2014) also compared enhanced simulations with standard simulations. Their evidence suggests that simulations that are very structured ( $g=0.88$ ) are more effective than those that have some structure ( $g=0.42$ ). Structure is defined as the degree of flexibility (and control) the simulation allows the student.

Chen et al. (2018) reported that adventure games (1.867) involving critical thinking and problem solving have a greater impact on attainment than non-adventure games (0.705) which predominantly focus on drill and practice. However, two of the three studies involving adventure games were conducted with adults only so this should be viewed with caution.

Clark et al. (2016) undertook many moderator analyses mainly relating to games mechanics. Their meta-analysis includes a diverse range of game-based learning technologies, which makes it difficult to draw conclusions.

## Learning in and across settings

3 articles: Sung, Chang & Yang, 2015; Sung, Chang & Liu, 2016; Zheng et al., 2018.

This section focuses on meta-analyses of mobile learning, which supports learning across multiple contexts and can enhance both individual and collaborative activity.

Sung and colleagues (2015) conducted a meta-analysis of studies investigating the use of mobile learning in language learning including school-aged students and adults. In a later review, Sung and colleagues (2016) conducted a meta-analysis of studies focused on mobile learning across three

<sup>48</sup> Overall effect for mathematics and science. Primary and secondary school effect sizes based on 92 effect sizes that focus on achievement in science only. 21 of 59 studies were conducted in North America and 18 were conducted in Europe. Moderator analyses only performed on data relating to science education.

<sup>49</sup> Of 24 studies, 7 from Taiwan, 5 from US and 4 from Turkey. No statistical tests of differences conducted but effect sizes presented.

<sup>50</sup> Includes six studies at primary level, one study at secondary level, three studies with adults. Effect size reported here for primary and secondary only which were combined in moderator analyses. No statistically significant difference by educational level.

<sup>51</sup> No breakdown by educational level; authors state that majority of participants are from secondary school.

<sup>52</sup> Overall effect for language learning, mathematics and science. No statistically significant difference by educational level.

domains (mathematics, science, literacy and language) again including school-aged students and adults. Zheng et al. (2018) conducted a meta-analysis of the impact of mobile learning and inquiry based learning in the domains of medical science, natural science (including mathematics and geography) and social science (including politics, education and psychology). With regards to educational level, the differences between higher education, secondary and primary levels were not statistically significant. Furthermore, the large effect size for secondary education (1.47) is only based on two studies and should be viewed with caution.

All three meta-analyses include studies involving adults. Overall, the evidence suggests that learning in and across settings through mobile learning has a moderate impact on attainment.

*Table 8: Effect sizes relating to learning in and across settings*

Author	Overall	Primary	Secondary	Notes
Sung et al. 2015 <sup>53</sup>	0.531	0.508	0.488	
Sung et al., 2016 <sup>54</sup>	0.523	0.636	0.451	
Zheng et al., 2018 <sup>55</sup>	0.91	0.82	1.47	

## Learning from assessment

1 article: Castillo-Manzano et al., 2016.

Castillo-Manzano and colleagues conducted a review of audience response systems (ARS) but most of these relate to university contexts rather than to schools. Of the 33 studies that they identified, only one concerned primary aged students (focused on language learning) and three related to secondary aged students (all focused on science). This partly explains the non-significant effect of ARS on learning in school contexts (by our standards rather than the authors). Hattie reports an effect size of 0.22 for 'clickers' (another term for audience response systems). There is insufficient evidence to comment on this further. Technology can support assessment in many different ways. A systematic review of the evidence available would be beneficial.

*Table 9: Effect sizes relating to learning from assessment*

Author	Overall	Primary/Secondary	Notes
Castillo-Manzano et al. <sup>56</sup>	0.288	0.48 (NS)	

<sup>53</sup> 31 of 44 studies involve primary and secondary school students. Statistically significant difference by level but this could be attributed to the 'mixed age' sub-group.

<sup>54</sup> Effect size reported for all participants; includes adults. Domain was not a statistically significant moderator. There was a statistically significant difference by educational level with primary:  $g=0.636$  and secondary:  $g=0.451$  across all domains. However, post-hoc analyses to identify where the differences lie were not conducted.

<sup>55</sup> Overall effect size includes adults (29% of studies) and all domains. Educational level not a statistically significant moderator.

<sup>56</sup> Only 6 of 53 effect sizes related to school students from one study in primary education on English language learning and three studies from secondary education on science. An analysis of differences between school and home was conducted but the authors only tested if effect sizes were significantly different from zero and not from each other. Overall effect size includes all educational stages (e.g. university).

# Digital technology characteristics that improve learning outcomes?

## Pedagogical organisation

A small number of meta-analyses consider the location in which learning takes place (formal settings such as the classroom, informal settings such as outdoors, and mixed settings). One review suggests that using technology in informal settings (0.768) is more effective than using it in formal settings (0.430) (Sung et al., 2016). Although not meeting the inclusion criteria for a high quality meta-analysis (44% of 122 studies did not have a control group), Chauhan (2017) also reported that using technology in an informal setting (0.700) is more effective than in a formal setting (0.534). However, two studies suggest that there is no statistically significant difference in effect by setting (Sung et al., 2015; Zheng et al., 2018). Therefore, following the categorisations proposed by the What Works Clearinghouse, the evidence suggests that technology is more beneficial in informal settings than in formal settings.

The way in which technology is integrated into teaching and learning can be important. Two broad categories are referred to: integrated learning where technology is embedded within traditional classroom instruction and supplemental uses of technology (i.e. additional practice). These are not mutually exclusive and can serve different purposes (e.g. targeting struggling readers). One meta-analysis suggests that integrated learning (0.28) is more effective than using technology as a supplemental activity in reading instruction (0.11) (Cheung & Slavin, 2012). However, Kunkel (2015), who also reviewed studies of technology use for reading instruction, suggests that the evidence on the type of use is inconclusive. Cheung and Slavin (2013) suggest the opposite case in mathematics with supplemental use (0.19) being more effective than computer based learning which tailors the content according to student progress (0.09) or integrated learning (0.06, non-significant). Wouters et al. (2013) report that serious games are more effective when integrated with other instructional approaches ( $d=0.41$ ) than when they are not integrated ( $d=0.20$ ). Thus, the way in which technology is integrated may be dependent on subject area and the type of technology used. Certainly, technology can be beneficial as a means of engaging in additional practice. Yet, it is necessary to ensure that the learner can make connections between the focus of technology use and the curriculum aims. That is, that its use should be integrated in teaching and learning rather than standalone.

The teacher's role can be important. Sokolowski and colleagues (2015) reviewed studies that focused on software supporting exploration and problem solving in mathematics. They reported an effect size of 0.75 for teacher-guided use of the software. In contrast, they reported an effect size of 0.56 for students who used the software independently without teacher support. However, no analysis was undertaken to determine whether or not these two effect sizes were statistically significantly different. Given that the confidence levels for these two effect size estimates overlap, it is possible that they are not. Therefore, it is not possible to state that teacher-guided use of digital technology is more effective than leaving students to work independently. Common sense suggests that teacher-guided use of software would be more beneficial than independent use, but this demands staffing whereas independent use may be more cost-effective, enabling staff to focus on other students who would benefit from the direct support of an adult.

Two studies looked at teaching methods (Sung et al., 2015; Sung et al., 2016). These studies both found that computer-assisted testing in mobile learning (2015: 1.769, 2016: 0.656) and self-directed study in mobile learning (2015: 0.581, 2016: 0.440) were more effective than other teaching methods,

noting that cooperative learning<sup>57</sup> with mobile technologies was not effective (2015: 0.268, 2016: 0.261). D'Angelo and colleagues (2014) report no difference in effectiveness between those studies involving collaboration<sup>58</sup> between students and those that do not in science education.

Two studies reported on student grouping in game-based learning. Clark et al. (2016) identified a statistically significant difference by player grouping but this could be attributed to the small number of effect size estimates in all but one of the subgroups. The analysis suggested that the only configuration to have a positive effect is playing as a single player in non-competitive play (0.45). Playing competitively and/or collaboratively did not have an impact on learning outcomes. In contrast, Wouters and colleagues (2013) suggest that playing games as a group, usually pairs, has a bigger impact on learning (0.66) than playing individually (0.22). D'Angelo and colleagues focused on simulations for STEM learning and found no statistically significant difference in impact when learning took place as individuals (0.5), in pairs (0.8) or as groups (0.88). It is difficult to draw any conclusions from these somewhat mixed findings.

One study identified that communicating with peers (0.495) or native speakers (0.487) is more effective than communicating with teachers which has no (positive or negative) impact on language learning (0.031, non-significant) (Lin, 2014).

## Technology

The technologies used in these studies vary from electronic books to hand held devices; in some cases the meta-analyses include a broad range of technologies. Cheung & Slavin (2013) make the important point that recent and currently used digital technologies have not been subjected to high-quality, large-scale trials.

Some studies note specific points as follows:

- Chen and colleagues (2018) noted that adventure games were more effective (1.867) than non-adventure games (0.705) in relation to vocabulary learning in a second language.
- Sung and colleagues (2015) noted (unsurprisingly) that handheld device are more effective (0.729) than laptops (0.151, non-significant) in relation to language learning and mobile learning.
- Takacs and colleagues (2015) make some observations regarding the differences between multimedia features (using sound and animation) and interactive features (hotspots, games, questions). They found that multimedia only technology enhanced storybooks had a positive impact on story comprehension ( $g=0.39$ ) whereas technology enhanced books with interactive features had no impact. They also noted that disadvantaged students who used multimedia only technology enhanced storybooks benefited ( $g=0.66$ ) whereas there was no beneficial effect for those who used similar stories but with interactive features.
- D'Angelo and colleagues (2014) noted the positive impact of modified simulations (instructional enhancement such as dynamic representations, meta-cognitive support or extended feedback) for STEM learning as compared to non-modified simulations (0.49). Furthermore, modified simulations are more effective when designed for science (0.53) than for maths (0.26, non-significant). Finally, modified simulations that are more structured allowing students less flexibility and control (0.88) are more effective than those that are less structured (0.42).

---

<sup>57</sup> Sung and colleagues use this term interchangeably with collaboration. They do not define its use.

<sup>58</sup> Collaboration treated as separate from working in groups. Collaboration specifically used to refer to the amount of collaboration required by the simulation.



In addition, Chauhan (2017, 44% of included studies did not have a control group) reported that there was no difference between software designed to meet an educational objective and general purpose software for learning in literacy, mathematics and science.

## Implementation

Abrami and colleagues (2015) state that four out of the nine studies that they reviewed reported that high quality implementation contributed to the positive effects. Similarly, Cheung and Slavin (2012) noted that six of the 40 studies that they reviewed associated low implementation quality with no impact on learning. In their later review (2013) they report that high quality implementations (0.26) are more effective than medium (0.12) or low quality implementations (0.12). Kulik and Fletcher (2016) noted that only four of the 50 studies they identified reported on implementation quality, noting that the median effect size for higher quality implementations was 0.44 and for weaker implementations was 0.01.

## Intervention length

Evidence relating to intervention length is mixed and is likely to depend on the context and specific technology uses. The evidence suggests that there is an optimum intervention length, with those being very short (less than 2 weeks) having very little impact and those that are too long being less effective. What counts as being too long differs (e.g. 6 months, 1 year).

Cheung and Slavin (2012) noted in their review of the impact of technology on reading that high intensity use (more than 15 minutes per day) is marginally<sup>59</sup> more effective (0.19) than low intensity use (less than 15 minutes per day) (0.11). In a subsequent review of the impact of technology on mathematics, Cheung and Slavin (2013) also found that high intensity use (30-75 minutes a week, 0.20; more than 75 minutes per week, 0.15) is significantly more effective than low intensity use (less than 30 minutes a week, 0.06, NS). However, D'Angelo and colleagues (2014) found that there was no difference in impact by the length of the sessions.

Sung and colleagues (2015) noted that the optimum intervention length is between one month and six months (0.722) with more than six months being less effective (0.130) and one week or less having no impact. Similarly, Kunkel states that longer and more intensive use (at least 30 minutes daily, for at least 5 months) is less effective than a shorter duration. Steenbergen-Hu and Cooper (2013), in a review of literature relating to intelligent tutoring systems<sup>60</sup>, reported that interventions that last for less than one year (0.26; 0.29<sup>61</sup>) are more effective than ones that last for more than one year (-0.01, non-significant; -0.08, non-significant). Sung and colleagues (2016) found no difference in intervention length but the effect size for studies running for six months or more was not significantly different from 0. Sokolowski and colleagues (2015) reported that shorter interventions (less than 2 weeks) (0.47) are relatively less effective than those lasting 2-5 weeks (0.63) or longer than five weeks (0.62). However, no analysis was undertaken to determine whether or not these effect sizes were statistically significantly different.

Wouters et al. (2013) suggested that playing a serious game for one session only makes no difference to learning (0.10, non-significant) whereas playing over multiple sessions has a moderate impact on learning (0.54). Clark and colleagues (2016) report that games played over multiple sessions had a small to medium positive impact on learning (0.44) whereas games played over one session did not seem to have any impact on learning (perhaps unsurprisingly) (0.08). But there was

---

<sup>59</sup> Statistical analysis suggests that the two effect sizes are significantly different at  $p < 0.08$ .

<sup>60</sup> Defined as learning environments that are self-paced, learner-led, highly adaptive and interactive, that respond to learners' needs depending on performance.

<sup>61</sup> Studies were categorised into two groups, adjusted and unadjusted, depending on whether or not they had taken into account pre-test scores. Hence, two sets of effect sizes.

no association between duration of game play (average 8 hours) and effect size magnitude. D'Angelo and colleagues (2014) in a study of simulations for STEM learning report that the number of sessions can have a differential impact. They found that interventions with 4-6 sessions (0.91) were significantly more effective than those that took place over 1 session (0.49), 2-3 sessions (0.39) or 11+ sessions (0.44).

In contrast, two studies reported no difference by intervention length (Gerard et al., 2015; Sung et al., 2016). Belland et al., (2016) did not do any specific analyses on intervention length but reported that interventions were considerably shorter than 1 year. Chauhan's study (2017), whilst lower quality, also reported no difference by intervention length.

## Educational Stage

Some studies which had a broad coverage of educational stages and domains undertook moderator analyses by educational stage. Sung and colleagues (2016) reported that impact of mobile learning is greater for primary students (0.636) than for secondary students (0.451). The difference by educational stage was statistically significant but they did not undertake specific analyses in order to elaborate on this. Another study by Sung and colleagues (2015) also reported a higher effect size for primary students (0.508) than secondary students (0.488). However, it should be noted that this difference could be related to a 'mixed' subgroup which only had 3 effect sizes and was not significantly different from zero.

Similarly, Cheung and Slavin (2013) reported a slightly higher effect size for primary school students (0.17) than that for secondary school students (0.14) for the impact of technology in mathematics. This difference was not statistically significant but the authors noted that the pattern is supported by other reviews. Kunkel<sup>62</sup> (2016) found that computer assisted instruction for reading was beneficial for the early years (0.50, 0.20) and primary aged children (0.23, 0.12) whereas the evidence in relation to secondary children was mixed with there being no difference in one data set and a small, positive effect in the other data set (0.18).

In contrast, Cheung and Slavin (2012) concluded that technologies to support reading are more effective at secondary (0.31) than elementary (0.10) and kindergarten (0.15). However only two of the 18 secondary studies were randomised, and the studies were dominated by two specific interventions, READ 180 and Accelerated Reader.

Several studies suggest there are no differences by educational stage (Belland et al., 2016; D'Angelo et al., 2014; Gerard et al., 2015; Lin, 2014; Steenbergen-Hu & Cooper, 2013; Zheng, 2016; Zheng et al., 2018). Of the six of these seven studies that report effect sizes for primary and secondary students, in half the cases the effect size for primary students was greater than that for secondary students, and in the other half vice-versa.

## Other Characteristics

Evidence suggests that lower ability students benefit more from technology use than higher ability students. Although Abrami and colleagues (2015) did not do any moderator analyses by student ability, they noted that three of the nine studies that they reviewed provided evidence to suggest that software for teaching reading (ABRACADABRA) makes more of a difference for low ability readers than high ability readers. Cheung and Slavin (2012) conducted moderator analysis on ability level and identified greater impact on low (0.37) and middle ability (0.27) students than on high ability students (0.08, not significant) for literacy learning. Gerard and colleagues (2015) reported similar findings with automated, adaptive guidance having a greater impact on learning than typical instruction for low

---

<sup>62</sup> Kunkel analysed the studies identified in three different data sets according to the study design.



( $g=0.59$ ) and medium ability ( $g=0.35$ ) than for high ability ( $g=0.14$ , NS). Similar results were also obtained in the analysis of studies that compared enhanced and simple feedback. Lin (2014) reported that computer mediated communication has more impact on language learning for beginners (0.782) than intermediate learners (0.393).

However, this finding is not necessarily the case for all uses of technology. Steenbergen-Hu and Cooper (2013) identified three studies that focused on the impact of intelligent tutoring systems in mathematics for low achieving students. In comparisons with general students, their complex analyses<sup>63</sup> suggest that low achievers may not benefit from this type of technology suggesting that students need to 'have sufficient prior knowledge, self-regulation skills, learning motivation, and familiarity with computers' (p984). Given the small sample size and complex analyses this finding needs to be viewed with caution. Similarly, Belland and colleagues (2016) report that scaffolding is more effective for traditional students ( $g=0.48$ ) than it is for low ability students ( $g=0.28$ ); this difference is statistically significant. In contrast, Gerard and colleagues (2015) report that automated, adaptive guidance is more effective than traditional classroom instruction for low ability students ( $g=0.59$ ) and medium ability students ( $g=0.35$ ) whereas it is less so for high ability students ( $g=0.14$ ). A similar pattern was identified in the comparisons of enhanced feedback versus simple feedback.

As well as ability, socio-economic status (SES) can account for differences in impact. Takacs and colleagues (2015) considered the differences between advantaged and disadvantaged students, a broad category including SES, learning disability and children with special needs. They reported that disadvantaged students benefited more than advantaged students from multimedia story books in relation to story comprehension. Similarly, disadvantaged students (0.27) were the only ones to benefit from technology enhanced story books in relation to developing expressive vocabulary, although the difference between disadvantaged and advantaged students was not statistically significant. McNally and colleagues (2018) reporting on the large-scale trials of ABRACADABRA in England, undertook analyses to identify the specific impact on attainment for children with low SES (using the proxy of eligibility for FSM). The analyses suggest that both ABRA and the offline equivalent had a higher impact for FSM students than for all students, which was small to moderate (ICT: 0.138; FSM, ICT: 0.368; non-ICT: 0.231; FSM, non-ICT: 0.396).

In contrast, Cheung and Slavin (2013) reported no statistically significant difference by SES when students use technology for mathematics learning (low SES<sup>64</sup>: 0.12; high SES: 0.25).

## Domain

Sokolowski and colleagues (2015) in their review of evidence in relation to mathematics exploratory environments, considered the content domain (calculated for all educational stages combined). They noted that technology is commonly used in algebra and geometry, the latter attributable to the visualisation opportunities that technology provides harnessed through specialised geometry applications. They reported that such environments are effective for arithmetic and algebra (0.61) but the evidence for geometry is less clear (0.67) because the confidence intervals [-0.07, 0.79] suggest that the effect may not significantly differ from 0 (a statistical test for this was not conducted). Similarly, Steenbergen-Hu and Cooper (2013) report that there is no difference in effectiveness when using an ITS to teach basic maths or algebra<sup>65</sup>, although the effect sizes were very small and the confidence intervals included zero.

---

<sup>63</sup> Analyses used both FEM and REM. Also, studies were categorised into those with unadjusted effect sizes (no account of data such as pre-test scores) and unadjusted effect sizes. Evidence suggests a negative impact on low achieving students, but analyses only related to 3 studies and most of the effect sizes reported were not significantly different from zero.

<sup>64</sup> 40% or more students eligible for free school meals.

<sup>65</sup> A difference was reported for FEM with adjusted effect sizes (studies that took account of baseline measurement differences between control and intervention groups). However, given the range of studies FEM is unlikely to be appropriate and therefore we have focused on results generated from a REM.

Studies that compare technology use in different domains and find no difference include Sung and colleagues (2016) review of mobile learning in the language arts, maths and science domains, and Zheng's (2016) review of scaffolding in natural science, social science and medical science domains.

In contrast, Wouters and colleagues (2013) suggest that serious games improve learning in all domains except biology and engineering. He notes that serious games are particularly effective in language learning (0.66) and significantly more so than mathematics (0.17). Similarly, Zheng's (2018) study of mobile devices and inquiry based learning suggests statistically significant differences by domain (medical science 1.82, natural science 0.93, social science 0.57). D'Angelo and colleagues (2014) reviewing the use of simulations in STEM subjects, reported that there was no effect in mathematics compared to a moderate positive effect in science (0.67). However, only 4 studies were identified that focused on mathematics (compared to 42 studies focusing on science) which could partially account for this finding so it should be viewed with caution.

## Variation

Archer and colleagues (2015) raised an important point about the variation in study outcomes. Meta-analyses of course take into account all effect sizes from the studies that meet the selection criteria. Of the 38 studies identified by Archer and colleagues, 8 reported negative impacts although only one of these was statistically significant. The variation is attributed to differences in the design and/or delivery of the intervention. Similarly, Kulik and Fletcher (2016) noted the variability in individual studies with some reporting large effects and some reporting none.

## Evidence of the impact of digital technology in English schools

To understand the strength of evidence about the effectiveness of digital technologies for learning in English schools, we carried out a modified Rapid Evidence Assessment.

Rapid Evidence Assessments (REAs) are a form of systematic review and are undertaken over a shorter period than a traditional systematic review (approximately three months, rather than 12 months). In this instance, the REA was undertaken in an even shorter period of time (approximately 14 days) as it was an adjunct to the initial review of systematic reviews.

REAs and systematic reviews systematically search for, evaluate and synthesise evidence about a specific intervention, and they are used to help policy makers to understand the impact of an intervention. Where possible an REA or systematic review will include a statistical meta-analysis of individual studies, in order to provide a clear indication of the likely impact (effect size) of the intervention. In this instance we chose not to quantitatively synthesise (i.e. by meta-analysis) the findings of the retrieved studies, due to the large amount of heterogeneity amongst the various technologies employed, contexts, applications, and outcome measures in the included studies.

We adapted our systematic search methodology from that used in the preceding search for systematic reviews (see Appendix C), to search for instances of evaluations of the use of technology in English schools, published between 2012 and 2018.

We found a small number of studies which did not employ research designs which could be classified as at levels 4 or 5 on the Maryland scale, and we chose to include these in this analysis due to the small overall number of studies which have been conducted in schools in England since 2012.

After completing the systematic search and reviewing the returned documents we included 31 documents in our final set for further analysis. Of these, 11 were Education Endowment Foundation

summaries which described other documents in the included set and are not referenced here as the detail is provided in the main evaluation reports (which are cited).

The included documents therefore described 19 different interventions or technologies evaluated in English schools. Of these, nine technologies/interventions pertaining to literacy and language were evaluated, six pertaining to mathematics, one to science, and four to cross-curricular or whole-school interventions. Where an intervention or technology was designed to develop skills, concepts and knowledge primarily employed within one specific subject we chose to classify it accordingly. Other technologies have a more general focus (e.g. developing metacognitive skills; communications to parents) but their outcomes were measured in terms of specific subjects. We have classified these as being cross-curricular.

Sixteen of the evaluations used a randomised design, and were therefore representative of levels 4 and 5 on the Maryland scale. Three used other quantitative designs (Maryland scale levels 1-3), and one used a qualitative approach.

The high degree of heterogeneity between the studies makes it difficult to derive consistent findings which can form the basis of recommendations for policy makers or practitioners. In line with the findings of the international systematic reviews, there is evidence that the use of technology to enhance learning can be positive across a range of subjects. Gains are often small, resulting on average in around two to three months' accelerated learning for children who use technology. There are some examples of where children eligible for FSM show greater gains than non-FSM eligible children, although these findings are not consistent across all of the evaluations.

A small proportion of the process evaluations highlight some of the implementation factors which may affect the efficacy of the use of digital technology in the classroom. These include children's access to the technology (i.e. if an internet connection is required at home), availability of space in schools for small group and one-to-one booster sessions, and the time taken out of other teaching for the use of technology. This latter point is an important consideration for schools and teachers when considering whether and how different types of technology can be integrated into regular scheduled lessons. In addition, staff training is highlighted in a number of the studies as being a factor which is important to the effective implementation of the technology.

A specific consideration for those who are responsible for selecting technologies is the ratio of benefits to costs. The work undertaken by the Education Endowment Foundation and summarised in project overviews represents a consistent approach across a number of different technologies, and highlights that evidence about the effectiveness of particular technologies (i.e. their effect on learning) can be considered in the context of their cost. For example, some technologies may exhibit smaller effects than others, but may have a much lower direct cost and be easier to implement.

We now summarise the evidence in relation to three domains and also cross-curricular use of technology.

## Learning in Literacy and Language

A total of eight studies of specific technologies or interventions met the inclusion criteria (ABRA; Accelerated Reader; Comprehension Booster programme; GraphoGame Rime; Improving Numeracy and Literacy; Trainertext; Tutoring with Alphie; Units of Sound).

Further details of each intervention can be found in Appendix B.

Overall the evaluations report mixed results, in terms of the impact of the various technologies on progress in learning. There is evidence that some evaluations found positive impacts of technology. *Abacadabra*, a 20 week online game-based literacy programme targeting phonic fluency and comprehension, was found to have an impact on attainment of between two and five months.

Similarly, Year 7 pupils who took part in the Accelerated Reader whole group reading management programme were found to make three months' additional progress compared with children not taking part in the programme. In both examples, the learning gains were found to be greater for children who were eligible for free school meals. Struggling readers who used Comprehension Booster over a course of six weeks, made nine months' additional progress in reading accuracy and comprehension, when compared to small groups of similar readers engaged in phonics activities. Trainertext offers practice for word decoding supported by visual mnemonics and resulted in improved decoding (12 months' additional progress) alongside improvements in other measures.

However, four studies found no significant positive effect for the specific reading technology. GraphoGame Rime was an example of such an evaluation, but this finding may be explained by the activities undertaken by the control group which the associated process evaluation found to be of a similar nature to those offered by the GraphoGame Rime computer game. Other instances in which no positive effect was found by the evaluation were explained in terms of the number of schools or children that dropped out of the study (e.g. Units of Sound), or incomplete data during analysis (Tutoring with Alphie).

An interesting finding from the evaluation of Abracadabra concerned the nature of the intervention where an online and paper-based equivalent of the same intervention were randomised to pupils. Children in the paper-based condition made greater learning gains than those in the computerised condition, which suggests that the literacy programme (rather than delivery mechanism) was instrumental in making the difference. This may also be the case in the GraphoGame Rime example outlined above; that a structured reading intervention may yield learning gains irrespective of the way in which the learning is delivered. The evaluation of Abracadabra also concluded that the intervention was more effective when the small groups engaging with the resources were of similar ability rather than mixed ability.

Nonetheless, it is clear from the evidence included that there is a place for learning technologies to deliver targeted learning interventions with specific groups of pupils. However, one potential drawback of employing technologies in this manner is the logistical challenge of how the additional learning is scheduled and resourced alongside the routine curriculum. This is similar to interventions which prioritise additional learning and which are not based on the use of digital technology. For example, Gorard, Siddiqui and See (Switch-On Reading; 2015) observe that pupils were missing other lessons on a daily basis in order to receive the intervention. Furthermore, additional staff may be required to provide access to the technology and support for its use; a number of examples reviewed made use of teaching assistants to fulfil this role.

Some process evaluations (e.g. Worth et al., 2018) pertaining to the reviewed technologies highlighted the role of staff training in the implementation of the technology. They also noted the requirement for the technology to be implemented with a high degree of fidelity, in terms of how it was designed to be used, and the conditions under which it would have the best chance of success. In addition there were in some instances technical problems which meant that schools found the technology difficult to use, and therefore were unable to proceed with the evaluation as planned.

It is clear therefore that schools need to carefully consider how technologies designed to enhance reading are implemented, and given scarce resources, which groups of pupils will be prioritised to receive the intervention.

## Learning in Mathematics

A total of nine articles met the inclusion criteria, describing six specific technologies or interventions (Affordable Maths Tuition; Improving Numeracy and Literacy; Maths 3-5, Maths 4-6 (onebillion); Maths tablet intervention (onebillion); MathsFlip; Touch Arithmetic).

In a similar manner to learning in literacy, the findings from evaluations of learning in mathematics were mixed. Four interventions led to a small, positive impact on pupils' learning. An evaluation of Improving Numeracy and Literacy (computer games focusing on logical principles) found that pupils who participated in the intervention made an estimated three months' additional progress when compared with pupils who were in the control group. There was a similar but smaller effect for pupils eligible for FSM, but this effect was not statistically significant, and therefore it is not possible to conclude that it did not occur by chance. Similar small effects resulted from the use of maths tablet apps Maths 3-5/Maths 4-6 (factual knowledge and basic conceptual understanding) and four numeracy apps (targeting core mathematical concepts: number, shape, space and measure). In the case of the former (a randomised controlled trial), apps developing factual knowledge and conceptual understanding were found to be associated with three to four months' learning acceleration for children in the treatment group, compared with those who did not use the maths apps. In another evaluation, children in schools randomised to receive the MathsFlip intervention (blended learning with technology used outside the lesson, focusing on core concepts) made one month's additional progress compared with children who did not receive the intervention, with a slightly higher impact for children eligible for FSM. Two evaluations however found no effect on pupil performance of specific interventions designed to aid learning in maths. Affordable Maths Tuition, a one to one tutoring programme via the internet, was found to have no effect on progress for children receiving the additional tuition. Touch Arithmetic (computer-assisted assessment of problem-solving steps) was no different to pencil and paper methods.

Process evaluations and researcher observations may explain some of the differential effects observed. For example, in the case of MathsFlip, teachers' lack of familiarity with the technology was reported, in addition to the required independence of pupils and support from parents needed for effective use of the technology. These examples of implementation factors are consistent with other findings detailed in this report which cite implementation factors as instrumental in the effectiveness of the use of technology to aid learning. To some extent they may be clues to understanding of why a particular technology was not found to have the anticipated effect when evaluated.

## Learning in Science

One article met the inclusion criteria, describing one specific technology (Talk Factory).

The evaluation of Talk Factory used a qualitative research design, and therefore we are unable to provide an estimate of the impact of the technology on attainment. Talk Factory is used to support exploratory talk and hypothesis formulation during science lesson plenaries. The authors of the report conclude that "TF can be a potentially effective tool for eliciting a change in dialogic practices during plenaries in the science classroom" (Kerawalla, L., Petrou, M., & Scanlon, E., 2013). Specifically, teachers using Talk Factory highlighted prompting students for more responses which explained their reasoning.

## Learning across the curriculum and whole school interventions

A total of seven articles met the inclusion criteria, describing four specific technologies or interventions (Extended examination answers; Learner Response System; ReflectED; Texting Parents).

Three of the technologies or interventions were evaluated using randomised controlled trials. Texting Parents made use of school communication systems to send text messages to parents about children's learning (e.g. progress, homework, exams). Small positive effects were found for mathematics and English attainment (one month's additional progress in each subject) although the finding in relation to English may have been affected by bias attributed to missing data. There was no additional learning progress effect in science. A surprising finding was the small effect of the intervention in terms of reducing absenteeism. This latter finding was of particular interest, as none of

the texts sent related to children's absence. This finding may be explained by parents taking an increased interest in their children's school life in general through participation in the intervention.

Another use of technology evaluated using a randomised controlled trial was ReflectED. This intervention comprised a programme of lessons designed to help primary aged children to develop their metacognitive skills. Although the lessons did not specifically make use of technology as central to the intervention, children made weekly reflections on their learning and recorded them using the Evernote application ([www.evernote.com](http://www.evernote.com)) on iPads. Evaluation data indicated that children who participated in ReflectED made four months' additional progress in maths, and two months' less progress in English, compared with children who did not receive the ReflectED intervention. However, statistical analysis suggests that we cannot be confident that these results would generalise to other schools given this was an efficacy study. Another evaluation based on a randomised controlled trial investigated the use of a Learner Response System, which featured handheld devices to allow children to answer multiple choice questions in real-time for immediate teacher feedback. This evaluation did not find a positive effect of the intervention for maths or reading, and in some instances found a small negative effect in maths for one of the cohorts (the group which used the Learner Response System for two years, as opposed to the group which used it for one year).

One study (Charman, 2014) used a non-randomised design which compared groups of students writing exam answers either using a screen or paper, in an attempt to understand whether there would be any systematic differences between the two response modes. This small study found that response mode did not significantly affect the marks achieved by the students, although there were slight variations in writing style; students using a computer wrote more, whereas students writing on paper used denser language which was more varied (Charman, 2014). The sample sizes were small (24 used a computer, 22 wrote on paper) and no tests of statistical significance were undertaken.



## Conclusions

Broadly speaking, technology can be beneficial for pupils but it depends on a range of factors including the context, the subject area, the content, the pedagogy, access to technology, training/support, the length of the intervention and how it is integrated with other classroom teaching. It is not sufficient to say that technology will make a positive impact on learning. Additionally, many reviews include limited evidence of technology that is currently in use instead drawing conclusions based on outdated practices and technologies. However, there are a number of high-quality digital interventions currently available and supported by robust evidence of a positive impact on learning. In some cases, the evidence is primarily from overseas, particularly the US. But there are also some recent, robust evaluations that have been commissioned by the EEF that will be of interest to school leaders and classroom teachers.

### The complex landscape of evidence

Undertaking a review of evidence of the impact of digital technology on attainment in schools is not an easy task. The landscape of available evidence is complex. Focusing entirely on meta-analyses for the review of international evidence whilst providing a clear synthesis of knowledge necessarily means a reliance on aggregated data and a lack of detail regarding the exact nature of the interventions that have been evaluated. Where possible, specific uses of digital technology that have been shown to be beneficial have been highlighted.

The purpose of this review was to update the first synthesis of evidence published by the Education Endowment Foundation (Higgins et al., 2012). As described above, a rigorous procedure was followed to identify meta-analyses published from 2012 to 2018, with 24 such reviews identified that met the required inclusion and quality criteria. Each of these meta-analyses is summarised in Appendix A below. Each has been rated for quality using the AMSTAR tool and certainty of effect size using the GRADE framework. All of the studies selected are rated as high or medium in relation to the quality assessment. None of the studies are rated as high in relation to the certainty of findings. Eleven of the selected meta-analyses are rated as low in relation to the certainty of findings for reasons such as the inclusion of non-randomised studies, few studies from school contexts, and diversity of the reviewed studies.

The variability both within and between the identified meta-analyses is considerable. There is great diversity both in relation to the foci of the meta-analyses identified but also in relation to the individual studies selected within the reviews. This makes it difficult to suggest specific recommendations for teachers in schools in England. The meta-analyses vary from one that focuses on a specific application (ABRACADABRA, reading instruction software) to another that considers the impact of any technology on learning in primary education. Many of the studies include a range of domains and educational levels beyond school (university, adult learning). Many cover a broad timespan with two meta-analyses identifying individual studies from the 1960s/1970s onwards and over half including studies from the 1990s. There is also an inevitable time lag given that meta-analysis synthesises existing published reports that may be published months or even years after the data collection has taken place. This is even more critical in the field of digital technology given the rapid changes in adoption and use that can take place. As a result, there is limited evidence that relates to technologies that are currently used in classrooms. The rapid evidence assessment undertaken as part of this review of evidence addresses this through synthesising the evidence from 19 different interventions that have been tested in English schools since 2012.

There are a range of other design issues that need to be considered. Some meta-analyses rely on different forms of outcome measures with some developed by teachers or researchers (likely to result in higher effect sizes) and others using standardised assessments. There is very little information about teacher training provided in any of the meta-analyses reviewed despite this being a crucial for

effective use of technology in the classroom. This does not necessarily mean that training did not take place in individual studies but that it is difficult for the review authors to quantify this and/or make judgements about the quality of CPD provision. There is also a bias towards studies conducted in North America. Interventions conducted in other countries may not transfer so readily into other contexts.

### What does the evidence say in relation to domain areas?

In relation to subject area, there is slightly stronger evidence of a positive impact on learning in mathematics than in literacy. Typically, interventions targeting mathematics lead to three to four months' additional progress whereas interventions targeting literacy lead to three months' additional progress. There is good evidence that some mathematics and some literacy programmes can make an impact on pupils' learning. In relation to science the evidence to date is less secure as all reviews include other subject areas; however, the evidence suggests that typically pupils make five months' additional progress. It should be noted that in all subject areas some reviews report no or very little impact. That is, the findings of all reviews are not consistent. It really depends on the specific intervention, the context and the rationale for using technology.

The evidence from meta-analyses suggests that digital technology has a very small, positive impact on language and literacy. Types of technology use vary and include drill and practice, intelligent tutoring systems, game-based learning and electronic story books. Many of the interventions included in the reviews provide opportunities for practicing literacy skills (such as phonics and phonemic awareness) either in addition to or as part of normal classroom teaching.

Recent studies in English schools of specific interventions offer a mixed picture. There is evidence that Abracadabra (an online literacy programme focusing on phonic fluency and comprehension) and Accelerated Reader (reading management and monitoring) have a small, positive impact on learning. Trainertext (visual mnemonics for decoding) and Comprehension Booster (designing to help struggling readers to improve comprehension) had moderate to large impacts on learning but both these studies involved small sample sizes and the Comprehension Booster study was not randomised. A cross-curricular study that focused on improving parental engagement through texting (Texting Parents) suggests a positive impact on English but this could have been due to bias attributed to missing data. Seven other studies did not provide evidence of a positive impact but this could be attributed to issues with the study design (e.g. school dropout, control group engaging in similar offline activities, lack of fidelity).

It should be noted that two meta-analyses and one study suggest that there is no benefit of technology use when compared to control groups that engage in similar or equivalent offline activities in literacy (Kunkel, 2015; McNally, Ruiz-valenzuela & Rolfe, 2018; Worth et al., 2018). Kunkel (2015) found no difference when comparing CAI use with teacher-led reading interventions. In these cases, the way in which the instruction was delivered to students (via technology or traditional means) did not make any difference. Similarly, the evaluation of Abracadabra also suggests that it is the content that leads to literacy improvement rather than the delivery mechanism. However, technology can offer an alternative and effective means of delivering and practicing language and literacy content and skills with little or no adult support. Schools need to carefully consider how technologies designed to enhance reading are implemented, and given scarce resources, which groups of pupils will be prioritised to receive the intervention.

The use of technology to support foreign and second language learning is more beneficial, the evidence suggesting a moderate, positive impact on learning. In particular, the evidence suggests that computer mediated communication, mobile learning and game-based approaches are effective for supporting second language learning. That is not to say that other technologies are not effective; these technologies happen to be the focus of more recent meta-analyses. In relation to computer



mediated communication, communicating with peers or native speakers is more effective than communicating with teachers.

In relation to mathematics, there is slightly more variation in the effect sizes reported. Typically, there is a small to moderate, positive impact on learning in mathematics. Greater impact is achieved when simulations, scaffolding and/or intelligent tutoring systems are used to support mathematics education. Only one study reports on specific mathematics topics, suggesting that technology is particularly beneficial for teaching algebra. Studies undertaken in England also suggest a small to moderate impact on learning of a variety of applications focusing on key concepts including games (Improving Literacy and Numeracy), apps for tablets (Maths 3-5, Maths 4-6), a flipped learning approach (Maths Flip) and a cross-curricular project focused on parental engagement (Texting Parents). However, two studies undertaken in England reported no effect on attainment (Affordable Maths Tuition, Touch Arithmetic).

The evidence suggests a small to moderate, positive impact on science, although 7 of the 9 identified meta-analyses report effect sizes that include other domains or older students. Notably, there are no meta-analyses of sufficient quality level published since 2012 that focus exclusively on uses of technology in science education. Simulations and mobile learning seem to be the most beneficial uses of technology in this discipline. Only one single study was identified that was conducted in England but it was qualitative. The study evaluated Talk Factory, an application to support exploratory talk and hypothesis formation. It was suggested that Talk Factory leads to a change in dialogic practices.

### What does the evidence say in relation to Luckin et al.'s typology?

*Learning from experts/with others and learning through practice* (e.g. scaffolding, intelligent tutoring systems) has a small positive effect on learning. Learning from experts can be more effective in mathematics and science education than in other domain areas, although one meta-analysis suggests that there is no impact of intelligent tutoring systems in mathematics when compared to conventional teaching. Using computer mediated communication has a moderate, positive impact on second language learning.

In relation to *learning through exploring*, the evidence suggests that simulations and other exploratory tools such as dynamic geometry environments have a moderate, positive impact on mathematics and science learning. More structured simulations, which are less flexible and give learners less control, are more effective than those that are less so. Game-based learning has a small, positive impact on learning. The evidence suggests that game-based learning may be more effective (leading to a moderate impact) on second language learning. Scaffolding can increase the effectiveness of games based learning.

*Learning in and across settings* through mobile learning has moderate, positive impact on learning across many domains.

There is very little evidence on the use of technology in relation to *learning from assessment*. Sung and colleagues (2015) report that computer-assisted testing in technology-enabled language learning is particularly effective (1.769). One study from England suggests little difference between writing exam papers online or on paper. Given the potential (efficiency) benefits of computer-based assessment this is a useful finding. However, this was a very small study and no statistical analyses were undertaken. One meta-analysis and another study conducted in England focused on learner response systems (also known as clickers) and found no evidence of impact on mathematics, reading language learning or science. It would be beneficial to conduct a meta-analysis or systematic review of recent evidence for computer-based assessment in school contexts.

## What does the evidence say in relation to use of technology in general?

Technology can provide a medium through which pupils can practice skills and consolidate their knowledge, either working independently or with the support of a teacher or a teaching assistant. Interventions targeted at enabling pupils to catch up with their peers often involve withdrawal from lessons and take place outside the classroom. Typically, these activities are supervised by teaching assistants. However, technology can also supplement class teaching outside lesson time by enabling pupils to continue their learning at lunchtime, after school or at home. Indeed, technology use can have greater impact when used in informal settings rather than formal settings. As well as being motivational and with learning activities designed for specific contexts such as museums, learning outside the classroom can extend the time spent learning. Evidence suggests that supplementing standard teaching rather than replacing it with technology interventions can have the most positive impact (Higgins et al., 2012; Kulik, 2003; Means et al., 2013; Outhwaite et al., 2017; Rutten et al., 2012). In sum, technology can provide access to additional resources and opportunities for additional learning time, often recording progress and providing feedback to pupils so that they can progress independently. However, the way in which this is organised needs careful consideration in relation to when and where this takes place, especially if technology is to be used during lesson times potentially meaning that something else will be missed.

Embedding instructional support into technology, such as scaffolding and feedback, can be more beneficial than that without (Morgan et al., 2016), leading to between four and six months' additional progress. Scaffolding is defined as guidance from a more capable other to complete a task that could not be undertaken independently. Embedded instructional support can enhance the effectiveness of simulations for STEM learning (D'Angelo, 2014), problem-based learning in STEM (Kim et al., 2017), game-based learning (Wouters & van Oostendorp, 2013; Clark et al., 2016) and technology interventions in literacy, maths and science (Belland et al., 2016; Gerard et al., 2015). Evidence relating to scaffolding and ability is mixed. It has been shown to be more effective for low and medium ability pupils than it is for high ability pupils (Gerard et al., 2015). However, other research contradicts this, suggesting that it is more effective for traditional students than for low ability students (Belland et al., 2016). Some studies suggest that technology such as intelligent tutoring system can be at least as effective as teacher-led instruction (Steenbergen-Hu & Cooper, 2013) and in some cases lead to greater progress than if not used (Kulik & Fletcher, 2016), most notably in mathematics. Intelligent tutoring systems are defined as 'self-paced, learner-led, highly adaptive, and interactive learning environments' (p.970) that respond to a learner's needs based on data gathered during interactions as a human tutor would do.

Unsurprisingly, higher quality implementations that are used exactly as intended tend to be more effective than those that are of lower quality (Cheung & Slavin, 2013; Abrami et al., 2015; Kulik & Fletcher, 2016). All staff involved should make sure that they know how the intervention is to be used and ensure that it is used for the advised length of time. Training for staff and also for students is essential.

There is some evidence that shorter interventions (1 to 6 months) tend to be more effective than longer ones (Steenbergen-Hu & Cooper, 2013; Sung et al., 2015) but again what counts as being too long varies (e.g. 6 months, 1 year). Interventions that are particularly short (a single session or less than 2 weeks) are not effective (Wouters et al., 2013; D'Angelo et al., 2014). The intensity of use can also affect the outcome, with high intensity use (more than 15 minutes daily) being more effective than low intensity use (Cheung & Slavin, 2012; 2013). Some studies however report no difference by intervention length (Gerard et al., 2015; Sung et al., 2016).

Some evidence exists that technology interventions can be more beneficial for low and medium ability pupils than for high ability pupils (Cheung & Slavin, 2012; Gerard et al., 2015). However, intelligent tutoring systems may not benefit low ability students (Steenbergen-Hu & Cooper, 2013). In addition,

disadvantaged students may benefit more than advantaged students from technology interventions (Takacs et al., 2015; McNally et al., 2015). However, in contrast another study revealed no difference between disadvantaged and advantaged pupils (Cheung & Slavin, 2013).

In relation to the original review in 2012 (Higgins et al., 2012), our review confirms many of the findings reported:

- Most studies report positive benefits of technology use across a range of subjects although effect sizes are often relatively small.
- The gains made in mathematics and science tend to be higher than the gains made in literacy and language.
- As noted above, the diversity observed in studies means that it is “difficult to identify clear and specific implications for educational practice in schools” (Higgins et al., 2012, p15).
- Rapid change of digital technology is an issue because meta-analyses inevitably do not always include evidence grounded in current practice.
- Broad interpretations of digital technology use makes it difficult to interpret the evidence.
- Longer interventions (typically more than 6 months) are less effective than shorter interventions but interventions may also be too short to make a difference (less than two weeks).
- Digital technology can be more effective for low and medium ability students than for higher ability students.

In contrast to the original review (Higgins et al., 2012), we have found limited evidence on the impact of collaborative learning with digital technology. This could be because insufficient information about collaboration is presented in the individual studies identified for meta-analyses. Alternatively, it may be that there has been less interest in computer supported collaborative learning in more recent years. Of the studies that did mention collaborative learning, one study reported no effect whilst two reported a positive impact.

There is very little evidence in the identified meta-analyses of the impact of digital technology on attainment in the early years with only two studies reporting effect sizes. It would be beneficial to consider evidence from rigorous qualitative studies to explore these further.

There is mixed evidence relating to whether or not technology should be used individually, in pairs or in groups.

There is no substantive evidence of a greater impact of technology in primary or secondary contexts.

Technology can benefit disadvantaged students (low SES, learning disability, special educational needs) more so than non-disadvantaged students. Evidence supporting this emanates from both the identified meta-analyses and the rapid evidence assessment of studies undertaken in England.

There is huge variability in the evidence. Teachers should carefully assess if the programme is right for their pupils and consider all the implementation variables. As positive impact has often been related to additional learning time or the content rather than the delivery mechanism, careful consideration needs to be given to the costs and the potential benefits when compared to more traditional alternatives. There is a lack of studies on contemporary classroom uses. Furthermore, there is very little (if any) information on the models of change underpinning technology interventions.

When investing in technology in a school, consideration needs to be given to the following:

- is the right equipment available and what features best meet the needs of the pupils;
- how can technology be used to enhance and support learning, rather than replace current teaching;
- is there appropriate space within or outside the classroom for pupils to use the technology;
- when should the programme be implemented and what will the pupils miss;
- which pupils (all, low ability, disadvantaged etc) should be targeted to receive the intervention;
- should an adult be on hand to offer support or can the pupil use the technology independently;
- if an adult should be on hand, should that be a teaching assistant or a teacher;
- how will delivery of the intervention be monitored to ensure that it is used as intended;
- how long should the intervention be delivered for;
- what training is required for teachers and teaching assistants, and for students.

The implications of this review are that schools should consider all the evidence available for a specific intervention before investing in technology-based programmes. Specific factors to focus on include: potential impact, cost (including what technology use replaces) and levels of supervision required. For example, technology that can be used at home or in school without the support of the teacher or a teaching assistant could be more cost-effective in relation to interventions that depend on adult support.

## References

- Abrami, P. Borohkovski, E., & Lysenko, L. (2015) 'The Effects of ABRACADABRA on Reading Outcomes: A Meta-Analysis of Applied Field Research'. *Journal of Interactive Learning Research*. 26(4), pp. 337-367.
- Adesina, A., Stone, R., Batmaz, F., & Jones, I. (2014) 'Touch Arithmetic: A Process-Based Computer-Aided Assessment Approach For Capture Of Problem Solving Steps In The Context Of Elementary Mathematics'. Loughborough University. [online] [Accessed 1<sup>st</sup> August 2018]  
[https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/18443/3/2014-06-20\\_TouchArithmetic\\_Paper\\_Final.pdf](https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/18443/3/2014-06-20_TouchArithmetic_Paper_Final.pdf)
- 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, pp. 140-149.
- Batdi, V. (2015) 'A Meta-Analytic Study Concerning the Effect of Computer-Based Teaching on Academic Success in Turkey'. *Kuram ve Uygulamada Egitim Bilimleri*. 15(5), pp. 1271-1286.
- Belland, B. R., Walker, A. E., Kim, N. J. & Lefler, M. (2016) 'Synthesizing Results from Empirical Research on Computer-Based Scaffolding in STEM Education: A Meta-Analysis'. *Review of Educational Research*. 87(2), pp. 309-344.
- Biggart, A., O'Hare, L. & Millar, S. (2015) *Tutoring with Alphie Evaluation Report and Executive Summary*. [online] [Accessed 4<sup>th</sup> December 2018]  
[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Tutoring\\_with\\_Alphie.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Tutoring_with_Alphie.pdf)
- Borenstein, M., Hedges, L.V., Higgins, J.P. and Rothstein, H.R. (2011) *Introduction to meta-analysis*. John Wiley & Sons.
- Castillo-Manzano, J. I. Castro-Nuno, M., Lopez-Valpuesta, L., Sanz-Diaz, M., T., & Yniguez, R. (2016) 'Measuring the Effect of ARS on Academic Performance: A Global Meta-Analysis'. *Computers & Education*. 96, pp. 109-121.
- Chan, K. K. & Leung, S. W. (2014) 'Dynamic Geometry Software Improves Mathematical Achievement: Systematic Review and Meta-Analysis'. *Journal of Educational Computing Research*. 51(3), pp. 311-325.
- Charman, M. (2013) 'Linguistic analysis of Extended Examination Answers: Differences Between On-Screen and Paper-Based, High and Low Scoring Answers'. *BJET*. Doi: 10.1111
- Chauhan, S. (2017) 'A Meta-Analysis of the Impact of Technology on Learning Effectiveness Of Elementary Students'. *Computers & Education*. 105, pp. 14-30.
- Chen, M.-H., Tseng, W.-T. & Hsiao, T.-Y. (2018) 'The Effectiveness of Digital Game-Based Vocabulary Learning: A Framework-Based View of Meta-Analysis'. *British Journal of Educational Technology*. 49, p. 69-84.
- Cheung, A. C. K., Slavin, R. E., Cheung, A. C. & Slavin, R. E. (2012) 'How Features of Educational Technology Applications Affect Student Reading Outcomes: A Meta-analysis'. *Educational Research Review*. 7(3), pp. 198-215.

- Cheung, A. C. K., Salvin, R. (2013) 'The Effectiveness of Educational Technology Applications for Enhancing Mathematics Achievement in K-12 Classrooms: A Meta-Analysis'. *Educational Research Review*. 6(9), pp. 88-113.
- Chiu, Y.-H. (2013) 'Computer-Assisted Second Language Vocabulary Instruction: A Meta-Analysis'. *British Journal of Educational Technology*. 44(2), pp. 52-56.
- Clark, D. B., Tanner-Smith, E. E. & Killingsworth, S. S. (2016) 'Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis'. *Review of Educational Research*. 86(1), pp. 79-122.
- D'Angelo, C., Rutstein, D., Harris, C., Bernard, R., Borokovski, E., Haertel, G. (2014) *Simulations For STEM Learning: Systematic Review and Meta-Analysis*. Menlo Park CA: SRI International, Issue March. [online] [Accessed 3<sup>rd</sup> August 2018]  
<https://www.sri.com/sites/default/files/publications/simulations-for-stem-learning-executive-summary.pdf>
- Darabi, A. & Suryavnashi, R. (2013) 'Effectiveness of Online Discussion Strategies: A Meta-Analysis'. *American Journal of Distance Education*. 27(4), pp. 228-241.
- Gerard, L., Matuk, C., McElhaney, K. & Linn, M., C. (2015) 'Automated, Adaptive Guidance for K-12 Education'. *Educational Research Review*. 15, pp. 41-58.
- Gorard, S., Siddiqui, N. & See, B.H. (2015) *Accelerated Reader Evaluation Report and Executive Summary*. [online] [Accessed 3<sup>rd</sup> December 2018]  
[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/EEF\\_Project\\_Report\\_AcceleratedReader\\_1.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_AcceleratedReader_1.pdf)
- Grade (no date) *The GRADE Working Group*. [online] [Accessed 3<sup>rd</sup> August 2018]  
<http://www.gradeworkinggroup.org/>
- Higgins, S., Xiao, Z., Katsipataki, M. (2012) *The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation*. [online] [Accessed 1<sup>st</sup> August 2018]  
[https://v1.educationendowmentfoundation.org.uk/uploads/pdf/The\\_Impact\\_of\\_Digital\\_Technologies\\_on\\_Learning\\_\(2012\).pdf](https://v1.educationendowmentfoundation.org.uk/uploads/pdf/The_Impact_of_Digital_Technologies_on_Learning_(2012).pdf)
- Horne, J., K. (2017) 'Reading Comprehension: A Computerized Intervention with Primary-age Poor Readers'. *Dyslexia*. 23(2), pp. 119-140.
- Hultcrantz, M., et al. (2017) 'The GRADE Working Group Clarifies the Construct of Certainty of Evidence'. *J Clin. Epidemiol.* 87, pp. 4-13.
- Kerawalla, L., Petrou, M., Scanlon, E. (2012) 'Talk Factory: Supporting 'Exploratory Talk' Around an Interactive Whiteboard in Primary School Science Plenaries'. *Technology, Pedagogy and Education*. 22(1), pp. 89-102.
- Kim, N. J., Belland, B. R. & Walker, A. E. (2017) 'Effectiveness of Computer-Based Scaffolding in The Context of Problem-Based Learning for Stem Education: Bayesian Meta-Analysis'. *Educational Psychology Review*. 29(7).
- Kulik, J. A. & Fletcher, J. D. (2016) 'Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review'. *Review of Educational Research*. 86(1), pp. 42-78.
- Kunkel, A. (2015) *The Effects of Computer-Assisted Instruction in Reading: A Meta-Analysis*. PhD University of Minnesota. [online] [Accessed 1<sup>st</sup> August 2018]  
[https://conservancy.umn.edu/bitstream/handle/11299/175221/Kunkel\\_umn\\_0130E\\_16099.pdf?sequence=1&isAllowed=y](https://conservancy.umn.edu/bitstream/handle/11299/175221/Kunkel_umn_0130E_16099.pdf?sequence=1&isAllowed=y)

- Lin, H. (2014) 'Establishing an Empirical Link Between Computer-Mediated Communication (CMC) and SLA: A Meta-Analysis of the Research'. *Language Learning & Technology*. 18(3), pp. 120-147.
- McNally, S., Ruiz-Valenzuela, J. & Rolfe, H. (2018) *ABRA : Online Reading Support Evaluation Report and Executive Summary*. Addendum added March 2018. [online] [Accessed 3<sup>rd</sup> December, 2018] [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/ABRA\\_with\\_addendum.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/ABRA_with_addendum.pdf)
- 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).
- Miller, S. Davidson, J., Yohanis, J., Sloan, S., Gildea, A. & Thurston, A. (2017) *Texting Parents Evaluation Report and Executive Summary*. [online] [Accessed 2<sup>nd</sup> August 2018] [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Texting\\_Parents.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Texting_Parents.pdf)
- Motteram, G., Choudry, S., Kalambouka, A., Hutcheson, G-H. & Barton, A. (2016) *ReflectED Evaluation Report and Executive Summary*. [online] [Accessed 2<sup>nd</sup> August 2018] [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/EEF\\_Project\\_Report\\_ReflectED.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_ReflectED.pdf)
- Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2017). 'Closing The Gap: Efficacy Of A Tablet Intervention to Support The Development of Early Mathematical Skills In UK Primary School Children.' *APA PsycNET*. 108, pp. 43-58.
- Rudd, P. Berenice, A., Aguilera, V., Elliott, L. & Chambers, B. (2017) *MathsFlip : Flipped Learning*. [online] [Accessed 1<sup>st</sup> August 2018] [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Flipped\\_Learning.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Flipped_Learning.pdf)
- Savelsbergh, E. R., Prins, G., T., Ritbergen, C., Fechner, S., Vaessen, B., E., Draijer, J., M. & Bakket, A. (2016) 'Effects of Innovative Science and Mathematics Teaching on Student Attitudes and Achievement: A Meta-Analytic Study'. *Educational Research Review*. 19, pp. 158-172.
- Shea, B., J., et al., (2017) 'AMSTAR 2: A Critical Appraisal Tool for Systematic Reviews that Include Randomised or Non-Randomised Studies of Healthcare Interventions or Both'. *BMJ*. 4008.
- Sheard, M., Chambers, B. & Elliott, L. (2015) *Units of Sound Evaluation Report and Executive Summary*. [online] [Accessed 3<sup>rd</sup> December 2018] [https://educationendowmentfoundation.org.uk/public/files/Support/Campaigns/Evaluation\\_Reports/EEF\\_Project\\_Report\\_UnitsOfSound.pdf](https://educationendowmentfoundation.org.uk/public/files/Support/Campaigns/Evaluation_Reports/EEF_Project_Report_UnitsOfSound.pdf)
- Siddiqui, N., Gorard, S., & See, B., H. (2015) 'Accelerated Reader as A Literacy Catch-Up Intervention During Primary to Secondary School Transition Phase'. *Educational Review*. [online] [Accessed 3<sup>rd</sup> August 2018] <http://dro.dur.ac.uk/16393/>
- 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), pp. 970-987.
- Sung, Y.-T., Chang, K.-E. & Liu, T.-C. (2016) 'The Effects of Integrating Mobile Devices with Teaching and Learning on Students' Learning Performance: A Meta-Analysis and Research Synthesis'. *Computers & Education*. 94, pp. 252-275.
- Sung, Y.-T., Chang, K.-E. & Yang, J.-M. (2015) 'How Effective are Mobile Devices for Language Learning? A Meta-Analysis'. *Educational Research Review*. 16, p. 68-75.



- 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), pp. 698-739.
- Tingir, S., Cavlazoglu, o., Koklu, O., Intepe-Tingir, S. (2017) 'Effects of Mobile Devices on K–12 Students' Achievement: A Meta-Analysis'. *Journal of Computer Assisted Learning*. 33(4), pp. 355-369.
- Torgerson, C., Ainsworth, H., Buckley, H., Hampden-Thompson, G., Hewitt, C., Humphry, D., Jefferson, L., Mitchell, N. & Torgerson, D. (2016) *Affordable Online Maths Tuition Evaluation Report and Executive Summary*. [online] [Accessed 3<sup>rd</sup> December, 2018]  
[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Affordable\\_Maths.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Affordable_Maths.pdf)
- U.S. Department of Education (2010) *What Works Clearing House: Accelerated Reader*. [online] [Accessed 1<sup>st</sup> August 2018]  
[https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc\\_acceleratedreader\\_061416.pdf](https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_acceleratedreader_061416.pdf)
- U.S. Department of Education (2013) *What Works Clearing House: Intervention Report: Fast ForWord*. [online] [Accessed 1<sup>st</sup> August 2018] <https://ies.ed.gov/ncee/wwc/EvidenceSnapshot/172>
- U.S. Department of Education (2013) *What Works Clearing House: Intervention Report: Read Naturally*. [online] [Accessed 1<sup>st</sup> August 2018]  
[https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc\\_read\\_naturally\\_072710.pdf](https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_read_naturally_072710.pdf)
- U.S. Department of Education (2014) *What Works Clearing House: Intervention Report: Academy of Reading*. [online] [Accessed 1<sup>st</sup> August 2018]  
[https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc\\_academy\\_reading\\_121614.pdf](https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_academy_reading_121614.pdf)
- U.S. Department of Education (2016) *What Works Clearing House: Intervention Report*. [online] [Accessed 1<sup>st</sup> August 2018]  
[https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc\\_ucsmmp\\_052416.pdf](https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_ucsmmp_052416.pdf)
- U.S. Department of Education (2016) *What Works Clearing House: Intervention Report: University of Chicago School Mathematics Project (UCSMP)*. [online] [Accessed 1<sup>st</sup> August 2018]  
<https://files.eric.ed.gov/fulltext/ED565883.pdf>
- U.S. Department of Education (2017) *What Works Clearing House: Intervention Report: Odyssey @ Math*. [online] [Accessed 1<sup>st</sup> August 2018]  
[https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc\\_odysseymath\\_013117.pdf](https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_odysseymath_013117.pdf)
- U.S. Department of Education (2017) *What Works Clearing House: Intervention Report: Prentice Hall/Pearson Literature* [online] [Accessed 1<sup>st</sup> August 2018]  
<https://files.eric.ed.gov/fulltext/ED577575.pdf>
- Wiggins, M., Sawtell, M. & Jerrim, J. (2017) *EFF: Learner Response System Evaluation Report and Executive Summary*. [online] [Accessed 1<sup>st</sup> August 2018]  
[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Learner\\_Response\\_System.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Learner_Response_System.pdf)
- Willson, V. (2015) 'The Effects of Using Exploratory Computerized Environments in Grades 1 To 8 Mathematics: A Meta-Analysis of Research'. *International Journal of STEM Education*. 2(1), p. 8.
- Worth, J, Nelson, J., Harland, J., Bernardinelli, D. & Styles, B. (2018) *Graphogame Rime Evaluation Report and Executive Summary*. [online] [Accessed 4<sup>th</sup> December 2018]  
[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/GraphoGame\\_Rime.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/GraphoGame_Rime.pdf)



- Worth, J., Sizmur, J., Ager, R. & Styles, B. (2015) Improving Numeracy and Literacy Evaluation Report and Executive Summary. [online] [Accessed 3rd December 2018]  
[https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Campaigns/Evaluation\\_Reports/EEF\\_Project\\_Report\\_ImprovingNumeracyAndLiteracyInKeyStage1.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Campaigns/Evaluation_Reports/EEF_Project_Report_ImprovingNumeracyAndLiteracyInKeyStage1.pdf)
- 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), pp. 249-265.
- Wouters, P. & Van Oostendorp, H. (2013) 'A Meta-Analytic Review of the Role of Instructional Support In Game-Based Learning'. *Computers & Education*, 60(1), pp. 412-425.
- Young, J. & Young, J. (2017) 'Technology-Enhanced Mathematics Instruction: A Second-Order Meta-Analysis of 30 Years Of Research'. *Educational Research Review*. Volume 22, pp. 19-33.
- Zheng, B., Warschauer, M., Lin, C.-H. & Chang, C. (2016) 'Learning in One-to-One Laptop Environments: A Meta-Analysis and Research Synthesis'. *Review of Educational Research*. 86(4), pp. 1052-1084.
- Zheng, L. (2016) 'The Effectiveness of Self-Regulated Learning Scaffolds on Academic Performance in Computer-Based Learning Environments: A Meta-Analysis'. *Asia Pacific Education Review*. 17(2), pp. 187-202.
- Zheng, L., Li, X., Tian, L. & Cui, P. (2018) 'The Effectiveness of Integrating Mobile Devices with Inquiry-Based Learning on Students' Learning Achievements: A Meta-Analysis'. *International Journal of Mobile Learning and Organisation*. 12(1), p. 77.

# Appendices

## Appendix A: Systematic reviews (international)

### Meta-analyses

*Abrami, Borokhovski and Lysenko (2015)*

Keywords: Reading, Learning from experts/others, Primary

*Abrami, Borokhovski and Lysenko (2015)* conducted a meta-analysis of nine RCTs and quasi-experimental studies published between 2009 and 2015, generating 65 effect sizes (k), that were conducted in Canada, Australia, Kenya and Hong Kong (2008-2014). The studies focused on a specific application, ABRACADABRA or ABRA for short, which provides game-based activities linked to interactive stories, and is designed for primary aged children (aged 4-9 years). An element of scaffolding has been incorporated with feedback and suggestions provided to students who do not provide correct answers. It was developed by researchers at Concordia University in Canada, and is underpinned by research on effective literacy pedagogies. Professional development materials are embedded in the software in a bid to keep set-up costs to a minimum. The software is free. It is recommended that students use ABRA for at least two hours per week, over at least 13 weeks.

The meta-analysis suggests a small positive effect overall (FEM:  $g = 0.170$ , REM:  $g = 0.179$ ,  $k = 65$ ). More of the activities in ABRA focus on phonics ( $g = 0.189$ ,  $k = 19$ ) and phonemic awareness ( $g = 0.324$ ,  $k = 20$ ), explaining a positive impact on these specific literacy skills, in addition to a positive impact on listening comprehension ( $g = 0.381$ ,  $k = 6$ ) and vocabulary knowledge ( $g = 0.138$ ,  $k = 15$ ). ABRA has no impact (positive or negative) on reading fluency and comprehension.

No moderator analyses were conducted as the findings from the individual studies were homogenous but evidence from individual studies drawn on for the meta-analyses were considered by the authors. ABRA is designed to be integrated into classroom practice rather than as a supplementary, standalone intervention. Four of the nine studies reported that high-quality implementation contributed to the positive effects. Three of the nine studies provide evidence that children with lower levels of reading ability made higher gains than other students.

Due to the inclusion of non-randomised studies and a study sample which does not include UK school children, our confidence in generalising the observed effects to UK schools is limited.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Low

*Archer and colleagues (2014)*

Keywords: Reading/literacy instruction, primary, secondary

*Archer and colleagues (2014)* set out to address the issue that evidence of the impact of technology on literacy instruction was inconclusive. They identified three meta-analyses, two of which focused on reading and one of which focused on literacy instruction, and re-analysed the available evidence (a small number of publications were no longer available). They identified 38 studies conducted between 1970 and 2007, which focused on interventions with students in the age range 5-16 years. They concluded that overall, ICT interventions lead to a small, positive effect ( $g = 0.181$ ).

They also identified four moderators: original review, training and support, implementation fidelity and who delivered the intervention (teacher or researcher). There was no difference in impact by original review or who delivered the intervention. Whilst Archer and colleagues reported findings in relation to teacher training and support, and fidelity, we have concluded that their judgement was based on the quality of information reported in each study (that is, how much detail and the quality of the process

undertaken e.g. if and how intervention fidelity was measured) rather than a judgement about the quality of the outcome (for example, whether intervention fidelity was high or low).

One of the important points raised by the authors concerns the variation in study outcomes. Of the 38 studies, eight reported negative impacts although only one of these was statistically significant. The variation is attributed to differences in the design and/or delivery of the intervention.

The heterogeneity of the effect sizes of the included studies (some of which is unexplained) means that our confidence that the effect estimate is limited when considering a UK pupil population.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Low

#### *Belland and colleagues (2016)*

Keywords: mathematics, science, primary, secondary

*Belland and colleagues (2016)* conducted a meta-analysis of studies published between 1993 and 2014 that investigated the impact of scaffolding on achievement in STEM subjects. They identified 144 experimental studies (including quasi-experimental designs), which reported on 333 outcomes (k). The participants included school-aged children and adults. There was an overall positive effect on cognitive outcomes in STEM disciplines ( $g=0.46$ ). Results by educational stage were: primary ( $g=0.55$ ,  $k=28$ ), middle ( $g=0.37$ ,  $k=108$ ) and secondary ( $g=0.48$ ,  $k=53$ ). There was a statistically significant difference between the effectiveness of scaffolding for traditional students ( $g=0.48$ ,  $k=279$ ) and underperforming students ( $g=0.28$ ,  $k=8$ ). This suggests that scaffolding is less effective for underperforming students. However, this applies to all age ranges. No analyses were conducted on subject area. No information is given about the intervention duration but the authors comment that the treatments were 'considerably shorter than one year'.

The authors conducted various analyses on specific features of the scaffolding that took place (e.g. whether or not it adapted according to user performance) but there was no evidence to suggest that any of them had a greater effect than others.

We are moderately confident in the effect estimates but note that the majority of the included studies did not take place within the UK education system, and hence similar effects might not be observed among UK children.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

#### *Castillo-Manzano and colleagues (2016)*

Keywords: language learning, science, primary, secondary, university

*Castillo-Manzano and colleagues (2016)* investigated studies of audience response systems (ARS). 33 papers published between 2008 and 2012 were identified, reporting 53 effect sizes (k). 20 of these studies included data from the US. The majority of the studies were undertaken in university level education; only six effect sizes concerned school aged learners from one study of primary aged children (English language learning) and three studies (five effect sizes) of secondary aged children (science). The overall effect size was small and positive (FEM:  $g=0.099$ , REM:  $g=0.288$ ,  $k=53$ ). The effect size of ARS use to support learning for non-university contexts was not statistically significantly different from zero at  $p<0.05$  (FEM:  $g=0.55$ , REM:  $g=0.48$ , NS,  $k=6$ ), and so we cannot be certain that the intervention made an impact, whereas a small effect was identified university contexts (FEM:  $g=0.091$ , REM:  $g=0.216$ ,  $k=46$ ). It should be noted that the authors use a non-standard cut-off of

$p < 0.1$  and so claim in their review that the effect size for non-university contexts is statistically significantly different from zero.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
High	Low

*Chen, Tseng and Hsiao, 2018*

Keywords: language learning, learning through play, primary, secondary, adult

*Chen and colleagues (2018)* conducted a meta-analysis of the impact of game-based learning (GBL) on vocabulary learning in a second language. They identified 10 studies ( $k$ ) published between 2003 and 2014, involving 635 participants. Six studies involved primary school children, one study involved secondary school children and three studies involved adults. The overall impact of GBL on vocabulary learning compared to traditional approaches was large and positive ( $d=1.022$ ,  $k=10$ ).

Moderator analyses suggest a statistically significant difference between adventure games ( $d=1.867$ ,  $k=3$ ) and non-adventure games ( $d=0.705$ ,  $k=7$ ). The authors argue that adventure games are more challenging and motivating than non-adventure games, involving deeper learning and engagement. However, two of the three studies that tested the impact of adventure games concerned adults rather than children. No significant difference was identified for age (children, adults). The effect size for children only was reported to be large ( $d=0.873$ ,  $k=7$ ).

It should be noted that the analyses in this paper are based on a small number of studies; this somewhat limits confidence in the observed effect.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

*Cheung and Slavin (2012)*

Keywords: reading, primary, secondary

*Cheung and Slavin (2012)* focused on 84 studies ( $k$ ) (1980-2010) of technology intervention designed to improve reading levels for students aged 5-16 years, which involved over 60,000 students. Interventions had to take place for a minimum of 12 weeks. In common with other studies, they reported a small, positive effect of technologies designed to support reading development when compared with traditional classroom instruction ( $ES=0.16$ ). They reported differences in impact according to the type of technology application, learner characteristics and the intervention methodology.

Cheung and Slavin suggest there was a marginal difference according to the type of intervention. Interventions designed to be supplemental ( $ES=0.11$ ,  $k=56$ ), primarily individualised computer assisted instruction packages, were less effective than those designed to be integrated with non-computer activities which can be described as an integrated instructional model ( $ES=0.28$ ,  $k=18$ ). These latter studies primarily focused on a small number of models: READ 180 and Voyager Passport both designed for struggling readers. However, these studies were not randomised; randomised studies tend to produce smaller effect sizes than non-randomised studies. A small number of studies were categorised as computer managed learning ( $ES=0.19$ ,  $k=4$ ) or innovative technology applications ( $ES=0.18$ ,  $k=6$ ).

Interventions were more effective at secondary ( $ES=0.31$ ,  $k=18$ ) than elementary ( $ES=0.10$ ,  $k=59$ ) and kindergarten levels ( $ES=0.15$ ,  $k=8$ ). However, only two of the 18 secondary studies were randomised, and the studies were dominated by two specific interventions, READ 180 and Accelerated Reader. Of 13 studies reporting student ability levels, there was a more positive impact

on low ability students (ES=0.37, k=12) and middle ability students (ES=0.27, k=8) than on high ability students (ES=0.08, k=9), possibly indicating ceiling effects. The authors found no differences in impact according to SES. Only a few studies reported on gender and/or race making it difficult to draw any conclusions about differential impact. Three studies focused on English as an Additional Language. There was a positive impact of technology use for this specific group (ES=0.29, k=3).

In relation to the intervention methodology, high intensity use (more than 75 minutes per week, ES=0.19) was marginally more effective than low intensity use (less than 75 minutes per week, ES=0.11). Of the 40 studies that reported on implementation quality, the six that judged it to be low reported no effect of the intervention. However, this could be attributed to a reporting bias; reports of no impact are more likely to comment on implementation quality.

Certainty in the observed effects is limited due to the inclusion of a majority of non-randomised studies. The authors conclude that there is a need for more randomised studies of technology interventions to develop reading.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
High	Low

*Cheung and Slavin (2013)*

Keywords: mathematics, primary, secondary

Cheung and Slavin (2013) reported on 74 studies (k) (1960-2011) involving 56,886 students, 45 of which were conducted with primary school students and 29 of which were conducted with secondary school students. It should be noted that 21 of the studies were published in the 1980s or earlier and 18 of the studies were published in the 1990s, although there were no statistically significant differences between studies conducted then and the 35 studies published since 2000. 64% of the studies were categorised as quasi-experimental.

Cheung and Slavin categorised the studies identified into three groups. Supplemental (drill and practice) which formed the majority of selected studies (79%), computer-based learning (assessment and monitoring progress) and comprehensive or integrated learning interventions. There was an overall small, positive effect of technology use on mathematics attainment (ES=0.16, k=74). Supplemental uses were more effective (ES=0.19) than computer-based learning interventions (ES=0.09) or integrated approaches (ES=0.06, NS).

The effect size for primary school students was higher (ES=0.17, k=45) than that for secondary school students (ES=0.14, k=29). This difference was not statistically significant but the authors noted that the pattern is supported by other reviews. There was no statistically significant difference between students from low SES backgrounds (ES=0.12, k=41) and students from high SES backgrounds (ES=0.25, k=10).

The frequency of use had a differential impact. Programs used for more than 30 minutes per week (30-75 minutes per week: ES=0.20, k=10; >75 minutes per week: ES=0.15, k=32) are more effective than those used for less than 30 minutes a week (ES=0.06, NS, k=29). The quality of the implementation (e.g. fidelity) also made a difference, with high quality implementations (ES=0.26, k=6) being more effective than medium (ES=0.12, k= 32) or low level implementations (ES=0.12, k=5).

The authors note that recent technologies and apps that are more typically used in mathematics classrooms have not yet been subjected to high quality, large-scale trials. They suggest a need for such studies to be carried out with commonly used classroom interventions designed to impact on mathematics achievement.

Certainty in the observed effects is limited due to the inclusion of a majority of non-randomised studies.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
High	Low

*Clark and colleagues (2016)*

Keywords: learning from experts/with others, learning through play, primary, secondary, university

*Clark and colleagues (2016)* reviewed the evidence relating to the impact of digital games on learning compared to no use of games, and also augmented games compared to standard games. Selected studies included primary, secondary and university aged students (not older than 25 years), and were published between 2000 and 2012. No breakdown is given by educational stage but the authors state that the majority of participants are at secondary school, with ages ranging from 5-21 years. 69 studies (n) from 68 journal articles were identified, with 173 effective sizes (k) from 55 studies measuring cognitive outcomes. The authors report that overall digital games have more impact on learning than comparisons that do not involve digital games ( $g=0.33$ ,  $n=57$ ). The impact on cognitive learning outcomes was reported as  $g=0.35$  ( $k=173$ ,  $n=55$ ). In addition, augmented games have greater impact than standard games ( $g=0.34$ ,  $k=40$ ,  $n=20$ ). In particular, games with scaffolding (e.g., personalised support, adapting to needs or interest) are superior when compared to those without ( $g=0.41$ ,  $k=20$ ,  $n=9$ ).

There was a significant difference between games played over multiple sessions and games played only once. Games played over multiple sessions had a small to medium positive impact on learning ( $g=0.44$ ,  $k=166$ ,  $n=40$ ). Games played for one session only did not seem to have any impact on learning (unsurprisingly) ( $g=0.08$ , NS,  $k=43$ ,  $n=17$ ). However, there was no evidence of an association between duration of game playing (average length was six hours) and effect size. Although there was a significant difference between player configuration sub-groups, this could be attributed to low numbers of effect sizes in all but one of the sub-groups.

Moderator analyses primarily concerned games mechanics with effects varying across the different characteristics. This suggests that game-based learning technologies are diverse and it is more difficult to draw general recommendations from any analyses. There was a significant difference between levels of scaffolding provided, with higher levels of scaffolding leading to greater learning games. Notably, teacher provided scaffolding was most effective ( $g=0.58$ ,  $k=8$ ,  $n=4$ ).

The observed effect sizes overall demonstrate a narrow confidence interval. However, the confidence interval is wide for the quasi-experimental studies included. Consequently, we are moderately confident in the effect estimate.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

*D'Angelo and colleagues (2014)*

Keywords: mathematics, science, primary, secondary

*D'Angelo and colleagues (2014)* report on the effect of simulations on STEM attainment in primary and secondary level education. 49 of the 59 studies, published between 1991 and 2012, concern science and most studies (47) focus on secondary level rather than primary level. The studies are primarily from North America (21) and Europe (18) with six from Asia (14 studies did not identify the geographical location). The authors identified 128 effective sizes (k), of which 96 relate to achievement. They report that teaching with simulations is more effective than teaching without

simulations ( $g=0.62$ ,  $k=46$ ). Similarly, modified simulations (with additional features such as scaffolding) are more effective than non-modified simulations ( $g=0.49$ ,  $k=50$ ).

Moderator analyses suggest that the impact of simulations as compared to traditional instruction is effective in science ( $g=0.67$ ,  $k=42$ ) but not in mathematics ( $-0.15$ , NS,  $k=4$ ). However, only four effect sizes in mathematics were included in the analyses, so it is difficult to reach firm conclusions about the comparability of the two subjects, despite the authors' analysis suggesting that there was a statistically significant difference in impact by domain. There were no differences in effectiveness by school grade levels, although there were only three effect sizes for primary school level (analyses for science only). There were no differences in effectiveness between studies employing independent study, those involving pairs and those involving small groups (analyses for science only). There were no differences in effectiveness for those studies involving collaboration and those not (analyses for science only). There were statistically significant differences by the number of sessions run, but not by session duration or overall duration. In relation to the number of sessions, 4-6 sessions ( $g=0.9$ ,  $k=9$ ) were more effective than 1 session ( $g=0.49$ ,  $k=11$ ), 2-3 sessions ( $g=0.39$ ,  $k=3$ ) or 11+ sessions ( $g=0.44$ ,  $k=8$ ).

With regards to studies that compared enhanced simulations (modified) with standard simulations (non-modified), enhanced simulations were more effective in science ( $g=0.53$ ,  $k=40$ ) than in mathematics ( $g=0.26$ ,  $k=10$ , NS). There were no differences in effectiveness: by school grade levels; between studies employing independent study, those involving pairs and those involving small groups; and for those studies involving collaboration and those not. There were no statistically significant differences by the number of sessions run session duration or overall duration. Simulations that were very structured were more effective ( $g=0.88$ ,  $k=9$ ) than those that were not ( $g=0.42$ ,  $k=22$ ).

Quality (AMSTAR 2)	Certainty of effect (GRADE)
High	Moderate

#### *Gerard and colleagues (2015)*

Keywords: literacy, mathematics, science, primary, secondary

*Gerard and colleagues (2015)* focus on the pedagogical strategy of scaffolding, conducting a meta-analysis of 41 studies reporting 57 different effect sizes ( $k$ ), conducted between 2000 and 2014. The studies were mainly conducted with students aged 10 to 14 (85%) with the remainder conducted with students aged 14 to 18. Automated, adaptive guidance was compared to traditional instruction. The domains included the broad sub-group of literacy and social studies (19% of studies), mathematics (44% of studies) and science (37% of studies). The overall effect size suggests that automated, adapted guidance leads to improved outcomes when compared to traditional instruction ( $g=0.34$ ,  $k=24$ ). The authors also compared enhanced automatic guidance with simple automatic guidance, reporting that enhanced guidance was more effective ( $g=0.27$ ,  $k=29$ ).

Only one of 13 moderators were statistically significant, that of ability (measured as prior knowledge). Automated, adaptive guidance was more effective for students of low ability ( $g=0.59$ ,  $k=5$ ) or medium ability ( $g=0.35$ ,  $k=15$ ) compared to those with high ability ( $g=0.14$ ,  $k=4$ ). For studies where the guidance features were optional, low and medium ability students made more use of the help features than high ability students did. In the analysis of the enhanced and simple forms of feedback, low and medium ability students benefited more from the enhanced feedback

Notably, analysis of the duration of the intervention did not suggest any differences although this was classified as one lesson of 50 minutes, or more than one lesson of 50 minutes. There was no effect of domain, suggesting that these findings are applicable to literacy, mathematics and science.



Again it was noted that the studies included in the meta-analysis were diverse in terms of the control conditions and contexts. Furthermore, a large proportion of non-randomised studies were included, thus limiting confidence in the observed effects.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

*Kim, Belland & Walker (2017)*

Keywords: mathematics, science,

*Kim and colleagues (2017)* also focused on the impact of computer-based scaffolding in the STEM subject areas, with a particular focus on problem based learning (PBL). They identified 21 studies with 47 outcomes published between 1990 and 2015, only five of which outcomes relate to mathematics education. They adopt a Bayesian approach to meta-analysis which accounts for small study effects commonly being larger than large study effects. Overall, there was a positive effect ( $g=0.385$ ) on learning in STEM subjects, although the confidence interval was wide. Analysis by subject area suggests a positive impact on mathematics ( $g=0.425$ ,  $k=5$ ) with a smaller positive impact on science ( $g=0.146$ ,  $k=25$ ).

They conducted specific analyses of scaffolding characteristics. Unlike Belland and colleagues, who reported no differences, Kim and colleagues suggest that conceptual scaffolding (directly related to the content) is not as effective ( $g=0.126$ ,  $k=32$ ) as metacognitive scaffolding ( $g=0.384$ ,  $k=4$ ) and strategic scaffolding ( $g=0.345$ ,  $k=2$ ). But problem based learning actually does not require conceptual scaffolding given its focus on open-ended, unstructured tasks. The authors suggest that their finding which counters that of Belland et al (2016) is attributed to this specific focus on PBL.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

*Kulik and Fletcher (2016)*

Keywords: mathematics, primary, secondary, university

*Kulik and Fletcher (2016)* reviewed the impact of intelligent tutoring systems on learning, reporting a positive impact. They identified 50 studies ( $n$ ) with 63 effect sizes ( $k$ ), published between 1985 and 2013, 39 of which were conducted in the US and 18 of which were conducted in the domain of mathematics. The studies included school children but also students at university. Students using intelligent tutoring systems outperformed those who did not (REM:  $g=0.49$ ,  $k=63$ ). The authors noted the variability in individual studies with some reporting large effects and some reporting none. They also noted variability according to outcome measure, with those using standardised assessments reporting lower effect sizes. Kulik and Fletcher reported a moderate impact in the domain of mathematics ( $ES=0.40$ ,  $n=18$ ) as compared to other domains which were categorised as single group ( $ES=0.72$ ,  $n=32$ ). However, further analyses suggest that the effect size varies according to the assessment format. Similarly there was a moderate impact for school level studies ( $ES=0.44$ ,  $n=23$ ) although the effect size for adults was higher ( $ES=0.75$ ,  $n=27$ ).

Four studies reported on implementation quality. The median ES for the stronger implementations was 0.44 and the median ES for weaker implementations was -0.01.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

*Kunkel (2015)*

Keywords: reading, primary, secondary

Kunkel's (2015) PhD dissertation focused on computer assisted instruction (CAI) from pre-school to high school, primarily involving studies from the US (89%). She included 61 studies (2000-2013), involving 38,940 students, from which she extracted 532 effect sizes (k), 101 being independent. She analysed them in three data sets according to study design. The way in which the interventions were organised were evenly split between supplemental use, integration with classroom instruction, and mixed or not specified. Most interventions were run over 3-5 sessions per week, with each session being up to 30 minutes in length. That is, (relatively) little and often. The majority of interventions targeted phonemic/phonological awareness (62%). They were predominantly 'linear' with students progressing through all activities in order rather than being adaptive and responding to students' ability level. The CAI was compared to control groups either receiving no treatment or using a non-reading CAI, resulting in small, positive effect for reading CAI. Moderator analyses included school level (or grade), specific reading skills (e.g. phonemic awareness) and study characteristics.

The mean effects for students receiving reading CAI were small, positive, and statistically significant when compared to control groups receiving no treatment or non-reading CAI: CAI vs no treatment (post-test) ( $g=0.21$ ,  $k=263$ ), CAI vs no treatment (gain) ( $g=0.13$ ,  $k=217$ ). When CAI was compared to teacher-led control activities (post-test) there was no difference ( $k=52$ ). We can be moderately confident in these effect sizes; although this meta-analysis includes a proportion of non-randomised studies from outside the UK, the effect these factors have on the certainty of effect is mitigated by the large sample size.

CAI appears to be beneficial for pre-school/kindergarten children (aged 4 to 6 years) when compared to no treatment ( $g=0.50$ ,  $k=20$ ;  $g=0.20$ ,  $k=2$ ) and for primary school children (aged 6 to 11 years) ( $g=0.23$ ,  $k=50$ ;  $g=0.12$ ,  $k=14$ ). Evidence relating to secondary school is mixed, there being no difference in one data set ( $k=9$ ) and a positive effect in the other data set ( $g=0.18$ ,  $k=10$ ).

Longer duration (5 times a week, more than 30 minutes per session, more than 50 hours overall) was less effective than shorter duration. That is shorter duration interventions should be prioritised. Adaptive programs in data set A were more effective than linear programs.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
High	Moderate

*Lin (2014)*

Keywords: learning from experts/with others (CMC), language learning, primary, secondary, adult

Lin (2014) focused on computer mediated communication (CMC) and second language learning, identifying 59 studies (k) published between 2000 and 2012, that involved 3562 participants. Of these, 10 were conducted in secondary schools and three were conducted in primary schools. They reported a positive, small to medium effect overall ( $g=0.441$ ,  $k=59$ ) with analyses by educational level suggesting that there was a medium effect for secondary aged students ( $g=0.615$ ,  $k=10$ ) and a small to medium effect for primary aged students ( $g=0.426$ ,  $k=3$ ), although there were no significant differences by educational level. It should be noted that there was a wide confidence interval around the effect size for high school students, and that this effects the overall certainty of the effect.

Moderator analyses suggest that CMC is more effective for learning a foreign language (i.e. classroom-based study only) ( $g=0.610$ ,  $k=43$ ) than for learning a second language (where the language is commonly used outside the classroom) ( $g=-0.038$ , NS,  $k=11$ ). CMC is also more effective for beginning learners ( $g=0.782$ ,  $k=8$ ) than for intermediate ( $g=0.393$ ,  $k=7$ ) and advanced learners ( $g=0.403$ ,  $k=4$ ). Interestingly, communicating with peers ( $g=0.495$ ,  $k=47$ ) or native speakers ( $0.487$ ,

k=4) is more effective than communicating with teachers (0.031, NS, k=5). There were no significant differences between the CMC mode (synchronous, asynchronous) or the CMC modality (text, voice).

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Low

*Means and colleagues (2013)*

Keywords: primary, secondary, university, adult

*Means and colleagues (2013)* identified 45 studies, published from 1996 to 2008, with 50 independent effect sizes (k), that compared online learning (100% or substantial amount of learning conducted via technology) or blended learning (at least 25% of learning conducted via technology) with traditional learning approaches. 27 effect sizes were obtained for studies of online learning and 23 effect sizes were obtained for studies of blended learning. The authors noted that most studies were modest in scope, although the number of participants varied from 16 to 1857. The mean effect size for all 50 contrasts was small (g=0.20, k=50). Blended learning (g=0.35, k=23) was significantly different to online learning (g=0.05, NS, k=27). Blended learning leads to higher gains in outcome scores when compared to traditional approaches, whereas online learning is no different to traditional approaches. The authors argue that, typically, blended learning approaches are designed to increase the amount of time a student spends engaged in learning. Evidence to date suggests that there is no benefit in replacing traditional classroom teaching with online learning. However, when students are unable to attend school for whatever reason then it could be a valuable alternative.

Moderator analyses suggest that the pedagogical approach has a differential impact. Independent uses do not differ from traditional approaches (g=0.05, NS, k=17) whereas teacher-directed uses (g=0.386, k=8) and collaborative uses (g=0.249, k=22) are more effective. There were no significant differences by treatment duration (<1 month, >1 month). Interventions in which the curriculum and/or instruction were different to the control condition (g=0.402, k=17) were more effective than when they were similar (g=0.130, k=29). This is perhaps common sense. If someone is taught something new and different then they will perform better in learning outcome measures. If something is taught but in different ways then there is only a small difference in performance.

However, only seven of 45 studies were conducted with school aged students, with a small non-significant effect (g=0.1164, NS, k=7) being identified between students who used online or blended learning and those who did not. This could suggest that school aged students do not benefit from these learning approaches for a variety of reasons (or may be an artefact of the small number of studies included). For example, school aged students may not be self-disciplined enough or sufficiently motivated to engage in learning outside the classroom. In comparison, a statistically significant effect was found for university students, who benefitted from using online or blended learning (g=0.309, k=21).

We need to view the general findings reported by Means and her colleagues with caution given that only a few studies collected data from school-based interventions. Furthermore, these studies were not predominantly of a UK sample, and hence our confidence in the certainty of the effect size is reduced.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Low

*Sokolowski and colleagues (2015)*

Keywords: mathematics, primary

*Sokolowski and colleagues (2015)* identified 24 studies (2000-2013, with 17 conducted from 2008 onwards) involving 4256 students which focused on exploratory computerised environments to support word problem solving and exploration in mathematics from primary to lower secondary school level. Exploratory learning is defined as 'gathering and investigating information' (citing Flum & Kaplan, 2006). This includes simulations and specific mathematics software such as dynamic geometry environments. The searches also included the term 'problem solving'. They identified a moderate effect ( $ES=0.60$ ) on attainment compared to traditional approaches. The studies were dominated by those from Taiwan (seven studies), US (five studies) and Turkey (four studies). The authors noted that the interventions were more effective when students were practising (putting knowledge to work) rather than learning new concepts.

It should be noted that the sub-groups in the moderator analyses are not statistically tested against one another (i.e. with a Q test), and that there are wide confidence intervals around some. The authors report sub-group effect sizes for educational level: lower primary ( $ES=0.61$ ), upper primary ( $ES=0.41$ ) and lower secondary ( $ES=0.65$ ). They also report sub-group effect sizes for intervention length: short ( $ES=0.47$ ), medium ( $ES=0.63$ ) and long ( $ES=0.62$ ).

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Low

*Steenbergen-Hu and Cooper (2013)*

Keywords: mathematics, primary, secondary

*Steenbergen-Hu and Cooper (2013)* identified 26 papers that were published from 1997-2010, reporting on 34 individual samples and 65 effect sizes. However, 47 of these effect sizes did not take into account any pre-test measure; these were referred to by the authors as unadjusted, and the remaining 18 were referred to as adjusted. The studies concern the use of intelligent tutoring systems (ITS) in mathematics in primary and secondary schools. Most of the studies compared ITS use to regular classroom instruction (31 of 34 independent samples, 61 effect sizes) but the rest compared it to human tutoring or homework practice. The analyses focus on the comparisons with regular classroom use only. The studies concern ten different ITS systems that are popular in the US. The authors reported a small, positive statistically significant overall effect of ITS use when using a fixed-effects model (FEM: 0.05,  $p=0.005$ ) and a non-significant effect of ITS use using a random-effects model (REM: 0.09,  $p=0.136$ ). When focusing on the adjusted overall effect sizes only (17), neither were significantly different to zero (FEM: 0.01; REM: 0.01). Given the likely variation in the true effect size across the analysed studies, we suggest emphasising analyses done under the random-effects model, and would therefore suggest accepting the small non-significant effect sizes of 0.09 (unadjusted) and 0.01 (adjusted) as the primary outcomes of interest.

The authors noted a high degree of heterogeneity between the studies. Moderator analyses suggest that interventions that lasted less than one year ( $k=11$ ; FEM: 0.23; REM: 0.26) are more effective than those that last for more than one year ( $k=15$ ; FEM: 0.02, NS; REM: -0.1, NS). In addition, the authors suggest that ITS use has a detrimental effect on low achieving students. However, this should be viewed with caution given the kind of studies that were included (a large number without pre-tests) and that only two studies included data on low achieving students.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
High	Moderate

*Sung, Chang and Liu (2016)*

Keywords: learning in and across settings, literacy and language, mathematics, science, primary, secondary

Sung, Chang and Liu focused on mobile learning, identifying 110 articles with 419 effect sizes relating to 18,749 participants, published between 1993 and 2013. Two effect sizes were unusually high and excluded from the analyses. The overall effect of mobile learning was moderate ( $g=0.523$ ). 39 studies took place in primary schools whilst 20 studies took place in secondary schools. 41 studies were conducted in the domain of language arts, 12 in mathematics and 27 in science.

There was a statistically significant differential effect by learning stage with the authors reporting effect sizes for primary ( $g=0.636$ ,  $k=39$ ) and secondary ( $g=0.451$ ,  $k=20$ ). The effectiveness also differed by location with a statistically significant difference between formal settings ( $g=0.430$ ,  $k=60$ ), informal settings ( $g=0.768$ ,  $k=21$ ), mixed ( $g=0.550$ ,  $k=25$ ). The confidence intervals for the effect of informal settings do not overlap with the other sub-groups. Therefore, the authors conclude that mobile learning is more effective in informal contexts than in formal or mixed settings. The authors also report a statistically significant difference by teaching method. Notably, games-based learning ( $k=4$ ) and co-operative learning ( $k=9$ ) were not effective compared to teaching that did not use mobile learning. However, this is based on a small number of comparisons so may be attributed to lack of power. The authors also discuss many reasons for this finding, including noting that the specific studies that involved co-operative learning were short interventions.

There was no difference by intervention length, although the authors noted that the effect size for studies of greater than six months was not significantly different to zero. However, this was only based on 9 effect sizes. There was no difference by domain. The effect sizes reported were: language arts ( $g=0.473$ ,  $k=39$ ), mathematics ( $g=0.337$ ,  $k=12$ ) and science ( $g=0.565$ ,  $k=27$ ).

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

*Sung, Chang and Yang (2015)*

Keywords: language learning, learning in and across settings, primary, secondary, adults

*Sung, Chang and Yang (2015)* conducted a meta-analysis of studies concerned with mobile learning and language learning. They identified 44 articles published between 1993 and 2013, involving 9154 participants. Of these, 31 relate to primary and secondary education. The overall mean effect size was 0.531 for achievement.

Moderator analyses focused on a variety of variables. There was a statistically significant difference for learning stage with the effect sizes reported as moderate for primary school students ( $g=0.508$ ,  $k=18$ ) and approaching moderate for secondary school students ( $g=0.488$ ,  $k=12$ ). However, it should be noted that this difference could relate to a 'mixed' subgroup which only had three effect sizes and was not significantly different from zero. Handheld devices are effective ( $g=0.729$ ,  $k=31$ ) whereas laptops (arguably less mobile) are not ( $g=0.151$ , NS,  $k=12$ ). There was a significant difference for teaching method. Computer-assisted testing ( $g=1.769$ ,  $k=1$ ) was superior to all other teaching approaches. However, this was based on one effect size only. Notably, mobile learning with co-operative learning was not effective, although only seven effect sizes related to co-operative learning. There was a significant difference by intervention duration with studies run between one month and six months ( $g=0.722$ ,  $k=17$ ) more effective than those run for longer than six months ( $g=0.130$ ,  $k=6$ ). Finally, interventions designed to support second language learning are more effective ( $g=0.837$ ,  $k=25$ ) than those designed to support first language learning ( $g=0.181$ ,  $k=18$ ). Interventions of one week or less had no effect.

There were no differences by setting (formal, informal, mixed), learning skill (e.g. reading, listening) or language.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

*Takacs and colleagues (2015)*

Keywords: reading, primary

Takacs and colleagues (2015) looked at studies comparing technology enhanced story books with traditional story books. The authors identified 43 studies with 57 effects, published between 1980 and 2014. 24 studies were conducted in the US, three in the UK, 11 in the Netherlands and five in Israel. Students were aged 3 years to 10 years old. Technology use had a small, positive impact on reading in relation to story comprehension ( $g=0.17$ ) and expressive vocabulary (using vocabulary to communicate meaning as opposed to understanding communication from others) ( $g=0.20$ ). Notably, both disadvantaged children (13 contrasts) and non-disadvantaged children (five contrasts) improved their expressive vocabulary when using technology enhanced story books (disadvantaged  $g=0.27$ ; non-disadvantaged  $g=0.05$ ). However, the difference between the three groups (overall, disadvantaged and non-disadvantaged) was non-significant, which may suggest that factors other than the status of the children explain the observed effect sizes.

Some features of interactive story books (multimedia features categorised as animation, sound and music) were identified as being beneficial and some were identified as detrimental (interactive features such as hotspots, games and questions, even when designed to support comprehension). Stories with only multimedia features had a positive effect on story comprehension ( $g=0.39$ ) whereas there was no effect for stories with interactive features or both multimedia and interactive features. Again, there was a more substantial benefit with regards to story comprehension for disadvantaged students who used multimedia only stories ( $g=0.66$ ,  $k=7$ ) whereas there was no effect for stories including interactive features.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

*Wouters, van Nimwegen, van Oostendorp and van Der Spek (2013)*

Keywords: learning through play, primary, secondary,

Wouters and colleagues (2013) focus on serious games, which they define as games with educational aims and objectives (as opposed to games designed purely for entertainment). They identified 39 studies published between 1990 and 2012, with 5547 participants. These included 77 effect sizes concerning learning outcomes, 39 of which concerned either primary school students ( $k=8$ ) or secondary school students ( $k=31$ ). Serious games were found to be more effective for learning than teaching that did not involve serious games ( $d=0.29$ ,  $k=77$ ,  $n=5547$ ). However the confidence interval was relatively wide, spanning both low and medium effect sizes, thus reducing the certainty of the effect.

Moderator analyses revealed that playing serious games has more impact on learning language ( $d=0.66$ ,  $k=11$ ) than on learning mathematics ( $d=0.17$ ,  $k=16$ ) or biology ( $0.11$ , NS,  $k=28$ ). However, it was noted by the authors that these analyses may have been confounded by other moderator variables. No differences were identified in relation to age level, with similar beneficial effects for primary school children ( $d=0.30$ ,  $k=8$ ) and secondary school children ( $d=0.33$ ,  $k=31$ ).

In terms of pedagogical organisation, playing a serious game for one session only makes no difference to learning ( $d=0.10$ , NS,  $k=47$ ) whereas playing over multiple sessions has a moderate



impact on learning ( $d=0.54$ ,  $k=30$ ). Learners benefit most when the games playing is integrated with other instructional approaches ( $d=0.41$ ,  $k=29$ ) than when the games are not integrated ( $d=0.2$ ,  $k=48$ ). Playing games as a group ( $d=0.66$ ,  $k=13$ ) has a bigger impact on learning than playing individually ( $d=0.22$ ,  $k=63$ ); in most of the studies included in the analyses the groups comprised pairs.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Low

*Zheng (2016)*

Keywords: Learning from experts/with others, primary, secondary, university

*Zheng (2016)* investigated the effects of self-regulated learning (SRL) scaffolds in computer-based learning on attainment. She identified 29 articles ( $k$ ) published between 2004 and 2015, involving 2648 participants, from primary, secondary and higher education contexts. She reported a positive overall effect on learning ( $ES=0.438$ ,  $k=29$ ).

There was no difference by scaffolding mechanism (prompts/hints, concept map, worked example, integrated SRL tool) or scaffold function (conceptual, strategic, metacognitive, mixed). There was a statistically significant difference between delivery forms with direct SRL (explicit) having no impact ( $ES=0.354$ , NS,  $k=1$ ) whereas indirect SRL (tacit) had a positive impact ( $ES=0.414$ ,  $k=27$ ). Similarly, domain general scaffolds were effective ( $ES=0.43$ ,  $k=27$ ) whereas domain specific scaffolds were not ( $ES=0.102$ , NS,  $k=3$ ). Providing multiple scaffolds (i.e. more than one simultaneously) is more effective ( $ES=0.577$ ,  $k=18$ ) than providing single scaffolds ( $ES=0.211$ ,  $k=11$ ).

Four studies involved 506 primary school students ( $ES=0.277$ ) and eight studies involved 698 secondary school students ( $ES=0.841$ ). It should be noted that the confidence interval around the effect size for secondary school students is particularly wide (0.109-1.574), which lowers our certainty in the estimate of the effect. The remaining 17 studies were conducted in higher education contexts. However, there was no significant difference in impact on learning by educational stage. There was no significant difference in impact by learning domain (categorised as natural science, social, science and medicinal science which was the focus of only one study).

With regards to the pedagogical organisation, there was a significant difference between interventions that involved students working in groups ( $ES=0.116$ ,  $k=5$ ) and those that involved students working individually ( $ES=0.496$ ,  $k=24$ ). Providing scaffolding at an individual level is more effective than providing it in groups. There was no difference by intervention length although the author had six different sub-groups from 'less than one hour' to '4-24 weeks' meaning that some sub-groups had very few effect sizes associated with them.

Findings revealed that self-regulated learning scaffolds in computer-based learning environments generally produced a significantly positive effect on academic performance ( $ES = 0.438$ ). It is also suggested that both domain-general and domain-specific scaffolds can support the entire process of self-regulated learning since they demonstrated substantial effects on academic performance.

There was no significant difference between different durations of treatment, but effect sizes for the longest duration (nine weeks+) are negative.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Low



Zheng, Li, Tian and Cui (2018)

Keywords: science, primary, secondary

Zheng and colleagues (2018) conducted a meta-analysis of 34 studies published between 2007 and 2016. They investigated the impact on learning of mobile devices used to support *inquiry-based learning*. They define inquiry-based learning as: “includ[ing] questioning, generating hypotheses, designing and planning, experimenting, collecting and analysing data, explaining results, drawing conclusions, and communicating findings” (Zheng et al., 2018, p.78). Mobile devices were categorised as mobile phones, PDAs and tablet PCs. The interventions in the studies identified were compared to ‘traditional learning’; no other details are provided and it is not clear if the control groups were engaged in inquiry-based learning without technology or another form of learning altogether. The 34 studies included higher education students (29%, 10 studies) with only 6% from high school (two studies) and 65% from primary school (22 studies).

The overall effect of mobile devices to support inquiry-based learning was 0.91 (REM). Moderator analyses were undertaken but only one was statistically significant: domain, with medical science (ES=1.82; k=3), natural science (ES=0.93; k=23) and social science (ES=0.57; k=8). No analyses were conducted to see where the differences lay but the confidence interval for the medical science effect size did not overlap with the confidence intervals for natural science or social science.

The non-significant moderators included: hardware types, software types.

There was no significant difference by context: formal (ES=1.37; k=7), mixed (ES=0.76; k=10) and informal (ES=0.82; k=17). Similarly, there was no difference by grouping: group (ES=1.01; k=11), individual (ES=0.78; k=20) and mixed (ES=1.43; k=3).

There was no significant difference by treatment length. Interestingly, 19 of the 34 studies had a duration of less than one day: less than one day (ES=0.82; k=19), 2-4 weeks (ES=1.09; k=7), 5-8 weeks (ES=1.19; k=5) and 9-24 weeks (ES=0.82; k=3).

There was no significant difference by educational stage: higher education (ES=0.99; k=10), high school (ES=1.47; k=2) and primary (ES=0.82; k=22).

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Low

Zheng, Warschauer, Lin and Chang (2016)

Keywords: language and literacy, mathematics, science, primary, secondary

Zheng and colleagues (2016) conducted a meta-analysis of 10 studies (n) generating 67 effect sizes (k) published between 2001 and 2015 which focused on the effect of one-to-one laptop schemes in primary and secondary schools. The overall effect of one-to-one laptop use compared to no such technology being present was positive but small (d=0.16, n=10, k=67). Separate analyses were conducted for the different domains but not educational stage. The evidence suggests that the use of one-to-one laptops has a positive impact in the domain areas of ‘English language arts’ (d=0.15, n=6, k=19), writing (d=0.20, n=3, k=11), mathematics (d=0.16, n=7, k=21), and science (d=0.25, n=2, k=3).

There was no effect on reading (d=0.12, NS, n=4, k=13).

No moderator analyses were undertaken.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

## Other studies

### *Chauhan (2017)*

Keywords: literacy and language, mathematics, science, primary

*Chauhan (2017)* gathered 122 articles that focus on the impact of technology on attainment at primary school, which were published between 2000 and 2016 and reporting 212 effect sizes and involving 32,096 participants. It should be noted that this selection may include studies that are not truly experimental or quasi-experimental; the authors noted that 54 studies did not have a control group but were considered to be 'experimental' because pre-tests and post-tests had been conducted. Studies were conducted in 26 different countries, with 41% (n=50) conducted in Taiwan, 12% (n=15) conducted in Turkey and 8% (n=10) conducted in each of the US and The Netherlands. The absence of any studies conducted in the UK is one factor meaning that we have little confidence that the effect estimates would be observed amongst UK school children. There was a positive, moderate effect of technology on learning in primary education when compared to teaching without technology ( $g=0.546$ ,  $k=212$ ).

In the moderator analyses, no statistical analyses were undertaken for differences between sub-groups. It was reported that there was a positive and moderate impact in science ( $g=0.727$ ,  $k=42$ ) and small to moderate impacts in mathematics ( $g=0.469$ ,  $k=41$ ), language ( $g=0.448$ ,  $k=77$ ) and science and technology ( $g=0.435$ ,  $k=15$ ). The type of software used was similarly effective: learning oriented ( $g=0.565$ ,  $k=155$ ), general ( $g=0.488$ ,  $k=57$ ). Effects were also similar, irrespective of intervention length: <1 week ( $g=0.704$ ,  $k=45$ ), 1-4 weeks ( $g=0.546$ ,  $k=50$ ), 1-6 months ( $g=0.483$ ,  $k=90$ ), >6 months ( $g=0.616$ ,  $k=18$ ). The effect sizes for learning environment were reported as: formal ( $g=0.534$ ,  $k=167$ ), informal ( $g=0.700$ ,  $k=21$ ), mixed ( $g=0.430$ ,  $k=24$ ).

Due to the inclusion of experimental designs without a control group, we have very little confidence that the effect estimate would be observed in the UK. Therefore, we have not included this study in our main review but as an additional study with lower quality evidence.

### *What Works Clearinghouse: University of Chicago School Mathematics Project (2016)*

Keywords: mathematics, secondary

University of Chicago School Mathematics Project (UCSMP) centres on problem-solving, real-world problems and the use of technology. That is, technology resources are an 'essential component' of the programme but it also includes a textbook and other resources for the teacher. The programme includes several different courses (e.g. algebra, geometry) which have been studied separately and cumulatively.

"UCSMP Algebra was found to have potentially positive effects on general mathematics achievement and algebra for secondary students. The cumulative effect of multiple UCSMP courses was found to have potentially positive effects on general mathematics achievement for secondary students." (p.2). One study reported a positive impact of the UCSMP Algebra course on general mathematics achievement. Use of the same course had a significant impact on a measure of algebra knowledge in one study and no impact in another study. One study reported a positive impact of the cumulative effect of UCSMP courses on general mathematics achievement. The extent of the evidence in all three cases was judged to be small.

There is no further information about the use of technology in the programme other than it includes graphing calculators, geometry systems and spreadsheets.

*What Works Clearinghouse: Read Naturally (2013)*

Keywords: reading, primary

Read Naturally is a supplementary reading programme that includes text, audio CDs and computer software. The underpinning strategy is: ‘modeling of story reading, repeated reading of text for developing oral reading fluency, and systematic monitoring of student progress by teachers and the students themselves. Students work at their reading level, progress through the program at their own rate, and work (for the most part) on an independent basis’ (WWC, 2013, p.1). The programme is designed to be used flexibly; the use of the software is optional.

WWC identified five studies (four experimental, one quasi-experimental). However, only two of these studies tested the software version of Read Naturally and one further study tested the software version in one of three participating schools. Four different outcome measures were adopted: alphabets, general reading achievement, comprehension and reading fluency. There were no effects of Read Naturally on alphabets (including word efficiency/identification and phonetic decoding) and comprehension. Evidence was mixed for reading fluency with one study reporting a positive impact and three reporting no difference. Notably, the study reporting a positive impact was one that focused on the software version of Read Naturally. Evidence was potentially positive for general reading achievement with one study reporting a positive impact and one study reporting no difference. The study reporting a positive impact only tested the software version in one of three participating schools.

*What Works Clearinghouse: Read 180 (2016)*

Keywords: reading, primary, secondary

Read 180 is a reading programme for struggling readers that includes “individualised instruction using an adaptive computer application”. Sessions last between 45 and 90 minutes and include whole-class instruction. The use of technology is one of three small-group activities alongside a 20-minute whole-group teaching session and a plenary discussion.

WWC identified nine studies (six experimental, three quasi-experimental). Of the six studies that reported a measure of comprehension, three reported a positive impact and three reported no difference. This was judged to be strong evidence of a positive effect. Similar results were obtained for the six studies that reported on general literacy achievement. Two studies reported on a measure of fluency, with one suggesting a positive impact and the other reporting no difference. This was judged as potentially positive. Two studies reported on alphabets (spelling) but both reported no difference.

*What Works Clearinghouse: Prentice Hall/Pearson Literature (2017)*

Keywords: literacy, secondary

This is a programme of study that targets reading, vocabulary, literary analysis and writing skills in secondary education. It is based on a student text book containing text and poetry, and can be accompanied by online components that offer features such as additional practice and assessment.

WWC identified three studies (all experimental). There were no discernible effects of the programme on general literacy achievement (two studies) or comprehension (three studies). However, it is also not clear whether or not the online components were used in these studies.

*What Works Clearinghouse: Accelerated Reader – Beginning Reading (2016)*

Keywords: literacy, primary

Accelerated Reader is a supplementary reading programme that records books read and assesses comprehension through a short quiz about the book and its vocabulary. Given that the readers in this study are beginners, it is expected that they will read the books with adult or peer support. Students are awarded points based on quiz outcomes, the number of words read and the reading level of the book.

WWC identified two studies (both experimental) that reported mixed effects in relation to measures of comprehension, with one study suggesting a positive impact and the other suggesting a negative impact (the effect size was greater than -0.25 but this was not statistically significant). There were no discernible effects in relation to reading fluency.

*What Works Clearinghouse: Fast ForWard – Beginning Reading (2013)*

Keywords: literacy, primary

Fast ForWard is a computer-based programme consisting of three strands designed to focus on the development of reading skills and cognitive skills (e.g. memory, processing, sequencing). Students are presented with activities or games, and the programme adapts the content presented based on their responses. The outcomes measured were alphabets (phonics, phonological awareness and letter knowledge), reading fluency and comprehension.

WWC identified nine studies (seven experimental, two quasi-experimental). Of the five studies reporting a measure of alphabets, two reported a positive effect and the remaining three reported no difference. This was judged to be strong evidence of a positive effect. Of the four studies that measured comprehension, two reported a positive effect, one reported a negative effect and one reported no difference. The evidence was judged to be mixed or inconsistent. Only one study reported on reading fluency and found no discernible effect.

*What Works Clearinghouse: Achieve3000 (2018)*

Keywords: literacy, primary, secondary

'Achieve3000 is a supplemental online literacy program that provides nonfiction reading content to students in grades pre K–12 and focuses on building phonemic awareness, phonics, fluency, reading comprehension, vocabulary, and writing skills.' (p1) The teacher selects an article for the class to engage with and the software adapts the reading level and activities depending on their ability level, assessing students in relation to comprehension and vocabulary.

WWC identified three studies (one experimental, two quasi-experimental). Two studies measured comprehension with one reporting a substantive positive effect (not significant but greater than 0.25) and one reporting no discernible effect. The evidence was judged to suggest potentially positive effects. Two studies measured general literacy achievement, with one positive effect and one with no discernible impact. Again, the evidence was judged to suggest potentially positive effects.

*Wouters and van Oostendorp (2013)*

Keywords: game-based learning, learning from experts, primary secondary

Wouters and van Oostendorp (2013) conducted a meta-analytic review of instructional support in game-based learning. They identified 29 studies, from 1990 to 2012, with 107 effect sizes (k) involving 3675 participants. Domains included biology and mathematics. Game-based learning (GBL) with instructional support is more effective than GBL without any support (d=0.34). This includes knowledge, skills and in-game performance.

Moderator analyses indicate a statistically significant difference in effects by learning outcome with more impact on skills (d=0.62) than on knowledge (d=0.33) or in-game performance (d=0.19). Instructional support in GBL was only more effective than GBL without instructional support at primary level (d=0.19) and university level (d=0.41). The impact at secondary level was not statistically significant. However, there were no statistically significant differences by age level. There was a positive impact of GBL with instructional support in the domains of biology (d=0.59, k=35) and mathematics (d=0.40, k=11) when compared to GBL without any support. However, this applies to all age ranges and mathematics was not significantly different to other domain areas.

Collaboration (as an instructional support type) was shown to have a small effect on learning through GBL (d=0.14, k=14) but modality (auditory rather than written presentation) (d=1.24, k= 10) and personalisation (d=1.06, k=4) were more effective.

Quality (AMSTAR 2)	Certainty of effect (GRADE)
Medium	Moderate

## Appendix B: Individual studies (English schools)

### Evaluations of technologies or interventions designed to enhance learning in Literacy and Language

#### ABRA

*What it is:* A 20-week online literacy programme focusing on phonic fluency and comprehension activities. The study also evaluated the effect of a paper-based version of the ABRA tool.

*Described in:* McNally, S., Ruiz-valenzuela, J., and Rolfe, H. (2018)

*Design:* Randomised controlled trial

*Impact:* Positive effects were found for online and paper versions of ABRA (between two and five additional months' progress), with larger impacts for students eligible for FSM and for the paper-based version of the system. A delayed post-test of one year after the intervention had completed suggested that the positive effect of ABRA was only sustained for the paper version.

#### Accelerated Reader

*What it is:* an internet-based reading management and monitoring programme for use with groups. It aims to foster independent reading by assessing reading age and suggesting books in line with pupils' needs and interests.

*Described in:* Gorard, S., Siddiqui, N., & See, B. H. (2015); Siddiqui, N., Gorard, S., & See, B. H. (2015).

*Design:* Randomised controlled trial

*Impact:* Children in the intervention group made an additional three months' progress in reading compared with those in the control group. There was a larger effect for pupils eligible for FSM, although there is less confidence in the certainty of this observed effect.

#### Comprehension Booster

*What it is:* A computer programme that provides low-achieving pupils with interactive reading practice using a variety of texts, and opportunities to acquire new vocabulary and learn the thinking skills required for understanding text.

*Described in:* Horne, J. K. (2017).

*Design:* Non-randomised.

*Impact:* Although the study involved a small number of students (19 intervention, 19 control), increases in reading accuracy (ES=0.72) and comprehension (ES=0.78) were found for the intervention group, with children receiving two booster sessions per week showing greater gains than those only taking part in one booster session per week. Control groups continued with their normal activities, which given that they had been identified as having special needs, generally meant being withdrawn from the classroom and working independently or in small groups with a teaching assistant, with a focus on phonics.

### GraphoGame Rime

*What it is:* A computer game designed to teach pupils to read by developing their phonological awareness and phonic skills.

*Described in:* Worth, J., Nelson, J., Harland, J., Bernardinelli, D., & Styles, B. (2018).

*Design:* Randomised controlled trial

*Impact:* The study concluded that there was no evidence that GraphoGame Rime resulted in improvements in pupils' attainment in reading. This finding was consistent for FSM and non-FSM children.

### Improving Numeracy and Literacy

*What it is:* Improving Numeracy and Literacy uses computer games designed to improve spelling and reading comprehension. These were incorporated with teacher training and the use of additional teaching materials.

*Described in:* Worth, J., Sizmur, J., Ager, R., & Styles, B. (2015).

*Design:* Randomised controlled trial

*Impact:* There was no evidence that the Literacy and Morphemes intervention had a positive impact on children's attainment. There is a high degree of certainty around this finding.

### Trainertext

*What it is:* This system uses visual mnemonics that appear above the text to supply relevant clues. This assists children in decoding the words without other help.

*Described in:* Messer, D., & Nash, G. (2018).

*Design:* Randomised controlled trial (waiting list control group).

*Impact:* The experimental group (45 students) had significantly higher standardised scores after ten months than the waiting list control group, with regard to decoding (ES=0.97), phonological awareness (ES=0.27), naming speed (ES=0.80), phonological short-term memory (ES=0.69) and executive loaded working memory (ES=0.65). Furthermore, the longer 10-month intervention group demonstrated greater gains than the six month intervention group. The control group (33 students) carried on engaging in traditional interventions for struggling readers, which were overseen by teaching assistants.

### Tutoring with Alphie

*What it is:* A computer-assisted programme based on collaborative learning that aims to improve the literacy skills of struggling readers.

*Described in:* Biggart, A., O'Hare, L. & Millar, S. (2015).



*Design:* Randomised controlled trial

*Impact:* It is apparent based on the evaluation that pupils who received Tutoring with Alphie showed a small positive improvement in reading comprehension, compared to pupils in the comparison group. There is low confidence in the certainty of the effect size however due to a high drop-out rate during the trial.

#### Units of Sound

*What it is:* A computer-based programme designed to help struggling readers with their reading and spelling skills. Students work in small groups, supervised by a teacher, and the programme develops skills from basic phonics through to adult reading levels.

*Described in:* Sheard, M., Chambers, B., & Elliott, L. (2015).

*Design:* Randomised controlled trial

*Impact:* There was a high drop-out rate during the evaluation. This means that there is low confidence in the certainty of the results, which found that the control group made greater learning gains than the treatment group.

## Evaluations of technologies or interventions designed to enhance learning in Mathematics

#### Affordable Maths Tuition

*What it is:* One to one tutoring programme. Maths graduates in India and Sri Lanka provide tutoring using the internet.

*Described in:* Torgerson, A, C., Ainsworth, H., Buckley, H., Hampden-Thompson, G., Hewitt, C., Humphry, D., Jefferson, L., Mitchell, N. & Torgerson, D. (2016).

*Design:* Randomised controlled trial.

*Impact:* There was no evidence that the intervention had an impact on Key Stage 2 maths.

#### Improving Numeracy and Literacy

*What it is:* Computer games (supported by teacher training and materials) to develop pupils' understanding of logical principles associated with mathematics.

*Described in:* Worth, J., Sizmur, J., Ager, R., & Styles, B. (2015).

*Design:* Randomised controlled trial.

*Impact:* Pupils who participated in the intervention made an estimated three months' additional progress when compared with pupils who were in the control group. There was a similar but smaller effect for pupils eligible for FSM. However this effect was not statistically significant and may therefore be a chance finding (i.e. may not have been caused by the intervention).

## Maths 3-5; Maths 4-6

*What it is:* Tablet based applications which focus on factual knowledge and basic conceptual understanding. Used during classroom time for 30 mins per day for 12 weeks.

*Described in:* Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J. (2018).

*Design:* Randomised controlled trial.

*Impact:* Use of the apps amongst children in the treatment group was found to be associated with three to four months' additional learning progress compared with those in the control group.

## MathsFlip

*What it is:* A blended learning program in which children undertake technology enabled learning before a lesson, in order to assimilate foundational concepts and knowledge. This is then built on during the lesson.

*Described in:* Rudd, P., Berenice, A., Aguilera, V., Elliott, L., & Chambers, B. (2017).

*Design:* Randomised controlled trial.

*Impact:* There was a small effect for children in the intervention group. They made an additional one month's progress compared with those in the control group. This impact was slightly higher for children eligible for FSM.

## Maths tablet intervention

*What it is:* Four applications, part of the onebillion package, which feature a virtual teacher who scaffolds children's learning, focusing on core mathematical concepts.

*Described in:* Outhwaite, L. A., Gulliford, A., & Pitchford, N. J. (2017).

*Design:* Non-randomised studies (4).

*Impact:* Large and significant learning gains were observed for children in the early years who used the four apps from the onebillion package focusing on number, shape, space and measure. This was conducted in a single, large primary school and involved relatively small sample, with three of the four studies not having age-matched control groups. These design aspects could account for inflated learning gains. In the first three studies, the children used the apps as a supplement to other numeracy activities. In the fourth study, the use of the apps was compared to It is now being evaluated through a randomised controlled trial funded by the Education Endowment Foundation.

## Touch Arithmetic

*What it is:* The Multi-Touch Arithmetic Tool (MuTAT). Captures and allows analysis of children's steps in solving mathematical problems.

*Described in:* Adesina, A., Stone, R., Batmaz, F., & Jones, I. (2014).

*Design:* Observational study.

*Impact:* Unknown. Study suggests that the MuTAT and pencil-and-paper methods were similar in terms of students' accuracy and efficiency when solving problems, and that the MuTAT provided deep insight into problem solving strategies.

## Evaluations of technologies or interventions designed to enhance learning in Science

### Talk Factory

*What it is:* Software designed to support exploratory talk and hypothesis formulation during science lesson plenaries.

*Described in:* Kerawalla, L., Petrou, M., & Scanlon, E. (2012).

*Design:* Qualitative

*Impact:* Unknown, due to research design. The authors conclude that Talk Factory "can be a potentially effective tool for eliciting a change in dialogic practices during plenaries in the science classroom".

## Evaluations of technologies or interventions designed to enhance learning across the curriculum, and whole school interventions

### Extended examination answers using a word processor

*What it is:* Comparison of response mode in long-format examination answers (i.e. writing exam answers on screen or on paper).

*Described in:* Charman, M. (2014).

*Design:* Non-randomised, comparison of groups.

*Impact:* No significant impact on exam marks achieved by students. Students who used a computer to write an extended answer to an exam question wrote more text whereas students who used paper wrote denser but less varied text. The sample sizes were small (24 students used a computer, 22 students wrote on paper) so no tests of statistical significance were undertaken.

### Learner Response System

*What it is:* A set of electronic handheld devices which allow pupils to respond to questions during lessons. Teachers are able to see an instant summary of responses and can provide real-time feedback.

*Described in:* Wiggins, M., Sawtell, M. & Jerrim, J. (2017).

*Design:* Randomised controlled trial.

*Impact:* No positive effect of the intervention for Maths or Reading, and some instances of a small negative effect in Maths for one of the cohorts (the group which used the Learner Response System for two years, as opposed to the group which used it for one year).

## ReflectED

*What it is:* A programme of lessons designed to help primary aged children to develop their metacognitive skills. Technology is not central to the delivery of lessons, but children made weekly reflections on their learning and recorded them using the Evernote application ([www.evernote.com](http://www.evernote.com)) on iPads.

*Described in:* Motteram, G., Choudry, S., Kalambouka, A., Hutcheson, G., & Barton, A. (2016).

*Design:* Randomised controlled trial

*Impact:* Children who participated in ReflectED made four months' additional progress in maths, and two months' less progress in English, compared to children who did not receive the ReflectED intervention. Statistical analysis suggests however that this finding cannot be confidently generalised to other schools, principally due to the study being underpowered (Mottram et al. 2016).

## Texting Parents

*What it is:* Texting Parents made use of school communication systems to send text messages to parents about children's learning (e.g. progress, homework, exams).

*Described in:* Miller, S., Davison, J., Yohanis, J., Sloan, S., Gildea, A., & Thurston, A. (2017).

*Design:* Randomised controlled trial

*Impact:* Small positive effects were found for Maths and English attainment (one month's additional progress in each subject) but the result for English may have arisen due to bias attributable to missing data. There was no additional learning progress effect in Science. There was a small effect of the intervention in terms of reducing absenteeism, although this was not specifically targeted by the intervention.

## Appendix C: Detailed methodology

We carried out a rigorous rapid review of the extant literature (published from 2012-2018) describing the impact of digital technologies on school children's attainment, with a specific focus on outcomes in mathematics, English and science. In line with previous work in this area (Higgins et al., 2012), our search was for systematic reviews of experimental and quasi-experimental studies.

### Search strategy

A 3-step search strategy using search keywords was adopted to identify systematic reviews:

1. electronic databases were searched for published studies;
2. governmental and organisational websites were searched for grey literature;
3. appropriate journals were hand searched.

We searched the following electronic databases:

FirstSearch

EBSCO

JSTOR

Web of Science

Science Direct

ProQuest Dissertations

ASSIA

ERIC

PsychInfo

PsychArticles

Scopus

Sociological Abstracts

Ethos

Gale

Ingenta

Connect

Proquest

Several governmental agencies and organisations associated with educational research were searched for reports and other grey literature:

What Works Clearing House (WWC)  
Best Evidence Encyclopaedia (BEE)  
Evidence for Policy and Practice Information (EPPI) Centre  
Washington State Institute for Public Policy  
The Danish Clearinghouse for Educational Research  
The Norwegian Knowledge Centre for Education  
The Campbell Collaboration  
The Cochrane Database  
The Education Endowment Foundation  
International Initiative for Impact Evaluation

We also hand searched the following journals looking for systematic reviews published between 2012 and 2018:

The American Education Research Association's Review of Educational Research  
The European Association for Research on Learning and Instruction's Education Research Review  
The British Education Research Association's Review of Education  
The British Education Research Association's British Educational Research Journal  
Computers & Education  
British Journal of Educational Technology

### Key words/terms

An initial search string based on population, intervention, outcome and study design was trialled and modified in order to maximise the probability of returning the greatest number of relevant results.

The final search string was as follows, with some adaptations necessary to accommodate the search engines of specific databases and websites. The string was used to search on titles and abstracts.

(app OR apps OR tablet\* OR iPad\* OR "artificial intelligence" OR "AI" OR robotic\* OR gami\* OR "digital learning resource" OR "educational resource" OR personali\*ed OR "learning analytic" OR "augmented reality" OR "mixed reality" OR "virtual learning environment" OR VLE OR "learning platform" OR "learning management system" OR makerspace OR "maker movement" OR making OR software OR quiz OR clicker OR "3D printing" OR "3D modelling" OR "interactive whiteboard" OR "interactive tabletop" OR "gesture-based" OR "gesture recognition" OR internet OR "Massive Open Online Course" OR "MOOC" OR "Web 2.0" OR "social media" OR "social learning platform" OR blog OR wiki OR "social networking" OR "social bookmarking" OR forum OR chat OR "video conference" OR "computer supported collaborative learning" OR CSCL OR "technology-supported collaborative learning" OR "e-textbook" OR "e-book" OR ebook OR "learning tools" OR "learning tool" OR digital OR computer OR "information technology" OR "information and communication technology" OR "ICT"

OR technology OR "technology enhanced learning" OR "TEL" OR "technology-enhanced stories" OR "technology-enhanced story" OR "electronic storybooks" OR "electronic storybook" OR "virtual-reality instruction" OR "computer-based training" OR "computer-based teaching" OR "computer-based learning" OR "computer-based instruction" OR "computer-assisted training" OR "computer-assisted teaching" OR "computer-assisted learning" OR "computer-assisted instruction" OR "CAI" OR "multimedia training" OR "multimedia teaching" OR "multimedia learning" OR "multimedia instruction" OR "intelligent tutoring systems" OR "intelligent tutoring system" OR e-learning OR "online learning" OR "virtual learning" OR "computer game" OR "computer gaming" OR simulat\* OR "serious games" OR "serious games" OR game OR word-processing OR mobile OR device\* OR laptop\*) AND (school\* OR pupil\* OR student\* OR elementary OR "K-12" OR "PK-12" OR "K-16" OR class\* OR child\* OR learner\* OR "primary school" OR "primary schools" OR "secondary school" OR "secondary schools" OR "high school" OR "high schools") **AND** (learning OR academic OR performance OR success OR outcomes OR effectiv\* OR achievement OR reading OR comprehension OR writing OR math\* OR science OR STEM OR teaching) **AND** (meta-analysis OR meta-analy\*ed meta-analytic OR "systematic review" OR "research synthesis" OR synthesi\*ed OR synthesis OR REA OR "rapid evidence assessment" OR "research review")

### Selection of studies

All references which were retrieved through the search process were imported into Mendeley, a specialist tool which assists with the screening of records for inclusion in reviews. Mendeley is also used to remove duplicates at the point of importation.

The titles and abstracts of retrieved studies were screened to identify studies that meet the exclusion criteria:

**Language:** is not English.

**Year of publication:** is before 2012.

**Type of paper:** is not a systematic review.

**Design of constituent studies:** constituent studies within the systematic review are not Maryland Scale level 4 or 5.

**Subjects:** did not include school children.

**Measure:** study did not include a measure of attainment.

**Use of technology:** technology was used to address a Special Educational Need only.

**Sample size:** contains less than 50 participants in the total sample.

References were screened by one reviewer and a second reviewer screened a proportion (20%) at random. Mendeley facilitates the logging of disagreements between reviewers so that they may be resolved through discussion and the involvement of a third reviewer if necessary.

The full texts of studies which were not initially excluded were then further screened.



Table 1: Maryland Scientific Methods Scale

Standard	Comparison	Description	Examples of methods
Level 1	No comparison	Attainment rate is reported for intervention group only	Before and after study
Level 2	Comparison with predicted rate	Actual and expected attainment rates of intervention group are compared	Expected local attainment rates compared with national attainment rates
Level 3	Unmatched comparison group	Attainment rate of intervention group is compared with attainment rate of an unmatched comparison group	Comparison of mean levels of attainment
Level 4	Well-matched comparison group	Attainment rate of intervention group is compared with attainment rate of a comparison group matched on static (and dynamic) factors e.g. progress, gender	Propensity score matching; regression discontinuity
Level 5	Randomised control trial (RCT)	Attainment rates are compared between intervention and control groups that have been created through random assignment	Randomisation

### The systematic reviews included in this review: quality and certainty of the findings

We analysed each meta-analysis included in our evidence review to make a judgement about the quality of the systematic review process of each, and the certainty of the effect estimate of each. To rate quality we used the AMSTAR2 appraisal tool, and to rate certainty of the effect estimates we used the GRADE framework.

AMSTAR 2 (Shea et al., 2017) is a critical appraisal tool developed to evaluate systematic reviews of randomised and non-randomised trials. The tool rates individual systematic reviews according to 16 criteria which rate specific aspects of the systematic review process. In line with Shea et al. (2017, p. 8) who noted that that 'users sometimes make modifications to the instrument', we chose to base our quality rating on the 10 AMSTAR 2 criteria which we deemed most important and relevant to this review and the included studies.

Each study was rated as low, medium or high *quality*, representing the quality of the systematic review process based on a judgement wholly informed by our reading of the published article. Due to variability in how systematic reviews are written (including the different lengths of the reviews and author choices about where to include more or less detail), it was sometimes difficult to make a judgements with certainty. As a result, some systematic reviews may have been downrated due to a lack of information available to evidence a specific quality-related criterion. This is an important point to note when interpreting quality ratings for specific systematic reviews included in this review, and also when considering the body of evidence as a whole.

The *GRADE* approach (GRADE Working Group, 2018) provides a framework for evaluating the certainty of the effect estimates of individual meta-analyses and of a body of evidence (i.e. multiple meta-analyses), and for making recommendations about the overall strength of the evidence base. Certainty refers specifically to how confident we can be that the "true effect lies on one side of a specified threshold or within a chosen range" (Hultcranz et al., 2017). In the context of our evidence review, this means how certain we can be that the effect sizes identified by the included systematic reviews are as stated. For example, if a systematic review concludes that the use of a specific type of technology leads an increase in attainment, how confident can we be in the claim about the direction of the effect (i.e. that the intervention resulted in an increase in attainment, rather than a decrease)? Moreover, where the meta-analysis states the magnitude of the effect, how confident can we be in the precision of that stated magnitude as a *true effect* which would generalise to the population of interest (in this instance, school children in the United Kingdom).

A systematic review's certainty of effect is likely to be downrated for one or more of five reasons:

1. There is a *risk of bias* inherent in the studies included in the systematic review. This may represent limitations in the study design (e.g. the inclusion of non-randomised studies) or in the execution of the studies. A large number of systematic reviews in this review were downrated as they included a large proportion of non-randomised studies.
2. There is unexplained heterogeneity in the results (*inconsistency*).
3. The evidence exhibits *indirectness*, for example by not directly addressing populations or interventions of interest, or by not measuring outcomes of interest. In the current review a number of included systematic reviews were downrated as their constituent studies included an overwhelming majority of non-UK children.
4. *Imprecision*, which may be evidenced by a wide confidence interval around the reported effect size, thereby implying uncertainty.
5. *Publication bias*, which can be estimated by various methods, and which refers to bias which may result from the inclusion of certain types of article and the exclusion of others. For example, many systematic reviews include only primary studies which have been published in academic journals (and which are therefore likely to report statistically significant findings).

Taking the five criteria into account, we made an overall judgement about the certainty of effect of each of the meta-analyses included in our review:

- *High*. We are very confident that the true effect lies close to that of the estimate of the effect.
- *Moderate*. We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- *Low*. Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
- *Very Low*. We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

After discussion we decided not to include any systematic reviews which we rated as being very low certainty of effect. This resulted in the exclusion of eight reviews from further analysis.

## Data extraction

We analysed the systematic reviews to identify the differential impact of types of technology and use, and other contextual factors such as age, socio-economic status and length of intervention, taking account of any reported moderator analyses. To enable this, the following information was extracted from the full texts of the eligible studies:

- Study details (first author; year of publication; published/unpublished)
- PICOTS (population, intervention; comparator; outcome; timing; setting)
- Research design of constituent studies where possible
- Results (effect type; effect size; 95% CI)

A narrative synthesis of the findings of this scoping review has been produced and presented in relation to previous findings (Higgins et al., 2012).