Effectiveness of online and blended learning from schools

Topping, Keith J.; Douglas, Walter; Robertson, Derek; Ferguson, Nancy

Published in:
Review of Education

DOI:
10.1002/rev3.3353

Publication date:
2022

Received: 21 July 2021 | Accepted: 20 April 2022

DOI: 10.1002/rev3.3353

SYNTHESIS

Review of Education BERA

Effectiveness of online and blended learning from schools: A systematic review

Keith J. Topping[1] | **Walter Douglas**[2] | **Derek Robertson**[1] | **Nancy Ferguson**[3]

[1]University of Dundee, Dundee, UK
[2]The Kelvin Centre, Glasgow, UK
[3]North Lanarkshire Council, Motherwell, UK

Correspondence
Keith J. Topping, Education, University of Dundee, Dundee DD1 4HN, UK.
Email: k.j.topping@dundee.ac.uk

Funding information
This work was not financially supported by any organization.

Abstract

This systematic analysis examines effectiveness research on online and blended learning from schools, particularly relevant during the Covid-19 pandemic, and also educational games, computer-supported cooperative learning (CSCL) and computer-assisted instruction (CAI), largely used in schools but with potential for outside school. Eight research databases were searched. Studies which were non-school, before 2000, not in English, without data and duplicates were removed, leaving 1355 studies: online 7%, blended 13%, CSCL 7%, games 26% and CAI 47%. Overall, digital technology was more effective (better) than regular instruction in 85% of studies, 8% the same and 3% worse. Blended learning was considerably better than online learning. CAI was the most effective, with games and CSCL coming after blended learning, but of course CAI was not searched for and these were not widely used outside of schools. Primary and early years/kindergarten were most effective (87% better) and secondary/high next (80%). Although science and mathematics were the most popular subjects, English as a foreign language interventions were the most effective, then writing and STEM, thinking, arts/music, humanities,

health and science, reading and maths, foreign languages and English in that order. Overall, females did better than males. 'Low ability' children and second language learners did especially well. Disadvantaged and special educational needs/disabled students did slightly worse. Self-efficacy was highly related to better outcomes. The limitations/strengths of the research were discussed and linked back to previous literature, a critical analysis made, and detailed implications for practitioners, policy makers and future researchers outlined. Digital technology's main advantage may be the possibility for enhanced task flexibility and learner autonomy, encouraging greater self-regulation. However, this may not be an advantage for all students.

KEYWORDS
blended, effectiveness, online, schools

Context and implications

Rationale for this study

During the covid pandemic, schools struggled with delivering online and blended learning. This study examines whether, in the best of circumstances, online and blended learning can be effective.

Why the new findings matter

The positive findings matter because schools need to know if online and blended learning are more or less effective than regular instruction.

Implications

Implications for teachers are that online and blended learning may be substituted for regular instruction provided this can be done efficiently.
Implications for school managers are that teachers should be encouraged, empowered and perhaps even required to develop the digital aspect of their teaching.
Implications for researchers are that further broad studies are necessary which include computer supported collaborative learning, educational games and computer assisted instruction delivered into homes, with analysis of study quality.
Implications for policymakers are that many teachers need intensive professional development about digital technology.

INTRODUCTION

This systematic review of the research literature on online and blended learning from schools starts by outlining recent perspectives on emergency remote learning, as occurred during the Covid-19 pandemic. We give aims for the study and explore the original contribution of this paper. However, much of the literature then reviewed relates to the period before the pandemic, with implications for action after the pandemic. Technology is not just digital, it is also social and political, shaping and being shaped by changing social practices (Feenberg, 2019). The pandemic increased the speed of growth of technology, but this review is also about how teachers come to engage with and implement technology, especially outside of school, which involves many adaptations and processes beyond the technology itself. Post-pandemic we hope to return to a calmer and more orderly state, but with a larger number of teachers much more experienced (and hopefully energised) in digital interventions outside school.

Recent perspectives

There have been several recent reviews of online and blended learning, particularly discussing 'emergency remote learning' as during the pandemic. Some of these discuss emergency situations (Barbour et al., 2020; Crompton et al., 2021; Hodges et al., 2020), some the management of school closures, changes in pedagogy and preparation for teachers (Bond, 2020a; Bond et al., 2021; Short et al., 2021; Viner et al., 2020), some vulnerable students and those with mental health issues (Drane et al., 2020; Nearchou et al., 2020) and some the implications for policy (Darmody et al., 2020).

Emergencies

Turning to emergencies first, none of the 'reviews' offer a systematic review of outcome data. Hodges et al. (2020) and Barbour et al., (2020) are at best narrative reviews of highly selected papers, often drawing more on higher education than schools, and making recommendations which go well beyond the evidence. Crompton et al. (2021) certainly address schools and conduct a proper systematic analysis, but only categorise their 60 included papers and do not report on outcomes. Nonetheless, all three studies have interesting things to say.

Hodges et al. (2020) described emergency remote learning as an attempt 'not to re-create a robust educational ecosystem, but rather to provide temporary access to instruction and instructional supports in a manner that is quick to set up and reliably available during emergency or crisis' (p. 13). This highlights the difference between the emergency response and a more organised longer-term programme of dealing with remote learning. Of course, there are many kinds of emergencies. For example, in Afghanistan, where education was disrupted by conflict and violence and schools themselves were targets, in order to take children off the streets and keep them safe, radio education and DVDs were used. Support personnel pre-emergency typically supported a small pool of staff interested in teaching online, rather than having to support virtually all staff at very different levels of expertise.

Hodges et al. (2020) cite Means et al. (2014), who identified nine dimensions of online learning, each with numerous options: modality (fully online, blended over 50% online, blended 25%–50% online, web-enabled face-to-face), pacing (self-paced—open-entry, open-exit, class-paced, class-paced with some self-paced), student-teacher ratio (<35:1; 36–99:1; 100–999:1, etc.), pedagogy (expository, practice, exploratory, collaborative),

teacher role online (active instruction, small presence, none), student role online (listen or read, complete problems and answer questions, explore simulation and other resources, collaborate with peers), online communication synchrony (asynchronous only, synchronous only, a blend of both), the source of feedback (automated, teacher, peers) and the role of online assessments (determine if student is ready for new content, indicate how to support the student, provide student or teacher with information about learning state, identify students at risk of failure, input to grade).

Barbour et al. (2020) reported most provinces across Canada offered some combination of face-to-face, blended and/or online instruction, including pre-existing online learning programs. Yet all of these tended to be described as 'online learning', ignoring the differences between instructional methods. Lessons from the pandemic needed to be retained, to help meet future issues due to natural disasters such as wildfires, hurricanes, earthquakes and the polar vortex. Much more planning was needed in teacher preparation, infrastructure, education policy and resources to be able to maintain quality instructional continuity during a crisis.

Of course, pandemics are not new. During the Spanish flu pandemic, the telephone was used for educating high school students. In 1948 the polio epidemic in New Zealand closed all schools, and the Correspondence School (Te Aho o Te Kura Pounamu) used regular mailing to send lessons to every household and educational radio to broadcast lessons. In Hong Kong, online learning was used when schools had to close due to the SARS outbreak. Barbour (2010) noted the planning required for remote emergency learning, commenting that in Singapore online and blended learning was so pervasive that teaching in online and virtual environments was a required course in teacher education programmes and schools were annually closed for week-long periods to prepare the K-12 system for forced closures. Schools that did not have adequate staffing to support all teachers should create communities of practice to foster peer-to-peer mentoring and encourage teachers to begin curating and using free resources offered by third-party organisations.

Students and families had contended with the availability of technology (e.g., sharing a limited number of devices while isolating with their family), ethical issues (e.g., Zoom privacy), and personal issues (e.g., taking care of loved ones or not being able to do so). Teachers needed to take time to test the learning experience on various mobile devices. Some students only used a mobile phone. However, some online activities and learning management system experiences might not work as well or behave differently on a phone. Teachers and students also needed health and wellness support. Teachers were expected to support students, but many of them felt anxiety and were stressed and overwhelmed. To support students with severe technology challenges, teachers could create a free Google Voice phone number and allow students to read aloud their handwritten work as a voicemail message. Some students managed the transition to remote learning quickly, or had previous experience taking online courses, and these students could be encouraged to be peer mentors and provide support for other students.

A proper systematic analysis of K-12 from 2010 to 2020 was offered by Crompton et al. (2021), but again was only of emergency responses with no comment on outcomes. Sixty articles from 48 countries were finally included, but only 13 focused on a specific discipline—most on maths, then literacy and science. Many studies covered all of K-12, nearly as many were secondary and only a few were elementary. Studies were mostly from high economic countries, with many fewer from middle and low. Emergency situations had included earthquakes, tsunamis, nuclear disasters, hurricanes, typhoons, Ebola outbreaks and cyclones. Many caused disruption to learning, as educational buildings were damaged, and/or schools were forced to close. Emergencies, often involving poverty and political turmoil, were more common in low-income countries than high income countries. However, the Covid-19 pandemic, unlike previous natural disasters,

involved 94% of the world's student population, and was longer in duration than some other emergencies.

Students in low-income countries were less likely to have Internet access. Students and teachers may also need access to hardware, such as laptops and mobile devices, remembering that emergency situations often come hand-in-hand with other financial hardships. The availability of software and applicability to language and other contextualised needs needed to be addressed, while meeting the needs of all learners. A large number of teachers had not had professional development in teaching online and many felt unprepared to implement it. Lack of teacher technology skills and knowledge were especially prevalent in low and low-middle income countries, such as sub-Saharan Africa, with only 64% of primary and 50% of secondary teachers receiving even minimum training. However, there were strategies to promote student learning and engagement, such as virtual nagging by text or phone and providing phone-based assessments.

Management of school closures, changes in pedagogy and preparation for teachers

Other reviews focused more on the details of emergency responding in schools. Viner et al. (2020) undertook a systematic review, searching three databases to identify the effectiveness of school closures and other school social distancing practices during coronavirus outbreaks. They included 16 of 616 identified papers. Despite the fact that school closures were deployed rapidly in many countries, there were no data on the relative contribution of school closures to transmission control. Data from the SARS outbreak in mainland China, Hong Kong and Singapore suggested that school closures did not contribute to the control of the epidemic. Modelling studies of Covid-19 predicted that school closures alone would prevent only 2%–4% of deaths, much less than other social distancing interventions.

A 'living rapid systematic review' of emergency remote learning was offered by Bond (2020a). Again, this paper did not summarise outcomes, but categorised studies into different bands and gave recommendations. A synthesis of 89 studies in K-12 from 70 different countries reported mainly online surveys (67%, $n = 60$, followed by interviews 30% and focus groups 11%), with a focus on student engagement. An initial 1032 items were reduced to 90 (71 of which were journal papers). There was no quality assessment. A narrative synthesis of findings was given. The majority of articles were about the secondary schooling, followed by primary schools and kindergarten. Qualitative methods were employed by 46 studies (52%), while 39 studies (44%) used quantitative methods and four (4%) used mixed methods.

Teachers should design activities with interaction to decrease feelings of isolation and boost engagement, engage in professional networks, and keep open communication with families. Linked with the need for social presence and interaction, opportunities to collaborate with peers were mentioned in 26% of studies. In order to facilitate this, some teachers used breakout rooms in Zoom for small group discussions, as well as asynchronous collaboration on discussion forums such as Edmodo and Google Slides. Some studies reported that online group work was not particularly effective, however pair assignments had worked quite well.

Open and frequent communication was needed between schools and families. Using platforms with collaborative features and push notifications could assist information dissemination, as well as promote interactivity between students, alongside providing parents with greater insight into learning activities. Parents and students needed further guidance in how to use new technologies. Online information sessions organised between teachers, or even between schools, could help reduce the burden on individual teachers to provide information and advice to families.

Two thirds of teachers found guidance for remote learning by official sources unhelpful, and 80% were further dissatisfied with how officials responded to feedback about this. Popular methods included asking students to take photos or make videos, explaining their thinking, carrying out independent projects, and rewriting activities to include home-based ingredients or equipment. From a research perspective, the lack of theoretical guidance was increasingly discussed. Framing research in a strong theoretical base assisted with the interpretation of data and also with redefining the view of a field.

A proper systematic review was reported by Bond et al. (2021) of emergency remote learning in secondary schools, but again results were reported as narrative synthesis and outcomes were not summarised, although what worked well and what worked less well were discussed. The Organisation for Economic Co-operation and Development estimated that school closures occurred in 188 countries, which impacted the education of over 1.7 billion children and their families (OECD, 2020). The review yielded 6274 records, which following quality assessment were reduced to 81 papers from 38 different countries, across all continents (Asia 42% studies, Europe 26%, North America 15%, Africa 6%, Middle East 6%, Oceania 4%, South and Central America 1%). Most studies were conducted in Indonesia (23%, $n = 19$), followed by the United States (14%, $n = 11$), China (5%, $n = 4$), and Slovakia, Austria and Hong Kong (4%, $n = 3$, each). Thirty-seven per cent of studies were from low or lower-middle income countries, and 63% from upper-middle income or high-income countries. Very few studies explored the perceptions and experiences of parents (6%) or school leaders (5%).

Despite confusing and sometimes contradictory government guidance, almost the same number of studies reported incidences of engagement and/or increased motivation (80%, $n = 65$) as the number of studies that reported incidences of disengagement and/or demotivation (81%, $n = 66$). Learning gains were reported in 18 studies (22). Some teachers reported that some students were more motivated to learn and complete schoolwork than prior to the pandemic, citing an increased ability to study and problem-solve independently, as well as a greater sense of responsibility. Students reported that teachers were motivating who checked in with them via email or phone, gave them timely feedback, used collaborative tools and made their expectations clear.

Several studies reported on hesitancy amongst teachers to facilitate peer collaboration, but approaches that facilitated peer collaboration reported largely positive effects. Teachers used live synchronous lessons for inquiry-based group work and experiments. Synchronous learning activities with interaction and support was found to keep students engaged, while formal and informal social interaction between peers helped well-being. Some teachers had students revise their work after live feedback from peers. Many studies reported facilitating discussion between students, and it was found that encouraging students to respond to the questions or work of peers alleviated pressure from teachers and encouraged social interaction. Some teachers prioritised social engagement before learning goals due to students' limited social interaction.

Considering the role of parents, parental involvement and support contributed to student learning, although issues of equity impacted the extent to which they could engage with their children's learning, alongside gaps in family content knowledge and technological skills. Some families failed to see the benefits of remote teaching and therefore did not encourage children to join online sessions. In other cases, inequalities in access to remote learning were aggravated by increased demands on girls to complete household chores or parents' fears that they could misuse newly found internet access. A small number of students felt their parents lacked the knowledge to support them with their learning. Teachers communicated with parents through WhatsApp or by email and phone, directly or through the school LMS. In remote areas lacking infrastructure and facilities to provide remote learning, parents collected lesson materials from schools and delivered their children's assignments. Students

welcomed their parents' help with online learning and cited the importance of parents for securing access to online resources, providing a suitable home-learning environment, reducing household chores and providing emotional support.

Nine studies focused specifically on vulnerable students, including migrants and refugees and students from lower socio-economic backgrounds. Five studies explicitly included evidence about students with special educational needs and disabilities.

Overall findings revealed that self-regulation and understanding were the most frequently reported indicators of student engagement, with online assessment tools, learning management systems with collaborative tools, live synchronous lessons with peer and teacher interaction, and teacher-made videos considered particularly engaging. Absence, social isolation, poor conduct and task completion were the most frequently reported indicators of disengagement, characterised by poor attendance in live lessons, a lack of opportunities to seek help with challenges and difficulties facilitating peer collaboration. In live lessons some students kept their cameras turned off, while others then did not engage further.

Although there were many reports that assessment online was particularly challenging, 21 different types of online assessments strategies were identified, with online quizzes and formative online feedback the most frequent. Several studies reported that online quizzes were beneficial for teachers, as they were easily corrected, could be used to rehearse recommended elements, could help maintain student interest in their learning, and allowed teachers to easily check student understanding of concepts. Live marking or recorded feedback and assessment were found to be particularly beneficial, as providing feedback during live lessons was sometimes challenging. Peer collaboration was facilitated through peer assessment, inquiry-based group work and experiments, aided by the use of collaborative software.

Short et al. (2021) identified peer-reviewed journal articles from 2007–2019 about preparing teachers to teach in K-12 blended environments. Initial searches yielded 272 hits, reduced to 88 articles, selected according to author impact by citation count and number of publications, journal impact by number of publications. Content analysis was used to identify patterns and this method could be termed a 'mapping review'. Qualitative approaches were employed in 52/88 articles. Five articles were literature reviews, and 11 presented models and theories.

Three methods were used to prepare teachers for blended teaching. The first was to have university coursework focused on the topic. Second, blended learning was used as a teaching method and evaluations were made of teachers' plans for using it in the classroom. Lastly, 21 articles focused on the impact of professional development on blended learning readiness, sometimes delivered through online courses. In-service teachers generally had a positive perception of blended teaching, recognising its benefits to both themselves and their students. There were various instruments to measure teachers' blended teaching readiness.

Vulnerable students and mental health

While disadvantaged students may be less likely to have computers and internet connections at home, schools, governments and charities have sought to overcome this, at least to some extent (Berson et al., 2022; Vincent-Lancrin et al., 2022). Nonetheless, the gap between advantaged and disadvantaged students seems to have widened. Indeed, the Education Endowment Foundation (2020a) estimated the median effect of school closures in the UK to have widened the gap between advantaged and disadvantaged by 36%. Kuhfeld et al. (2020) in the USA estimated a 35% loss in reading and a 50% or greater loss in mathematics. Liu et al. (2020) investigated the somatic symptoms of primary school students in China and found the pandemic had had mental health consequences. In India, Maity et al. (2020)

reported that students from less well-off backgrounds could not afford hardware, software or internet connection, and indeed some villages lacked electricity.

A scoping review of papers from 2001 to 2020, which focused on emergency remote learning and gave findings in narrative form, was offered by Drane et al. (2020). In 2020 UNESCO reported 191 countries had instigated nationwide school closures, affecting 1.5 billion students (91.3%) worldwide. Young people were often assumed to be digitally savvy, but their technology use at home was typically for personal use and not for learning purposes. Despite the discourse around the 'net generation' or 'digital natives', in reality these were largely myths.

Social vulnerability referred to the low resilience of communities when confronted by external stressors such as the effects of Covid-19. It involved varying levels of access to resources such as information, knowledge and technology, in order to prepare for, cope with and recover from external stressors. Factors such as overcrowding, poor health, difficulties with community safety, and unemployment might impact more significantly on learners within low socio-economic status communities. Households experiencing financial hardship may be restricted to accessing the Internet solely via mobile-only plans. These typically had lower download limits and additional costs could be accrued.

Schools were also very heterogeneous, with significant differences in context and socio-economic status, varying by region. For students in an online environment, immediate and constructive teacher feedback was vital to stabilise their engagement with education. This feedback may be absent for those learners with limited or no access to computers or the Internet, and were isolated in their homes, restricting their ability to 'see' their teachers. Students from disadvantaged backgrounds were reported as being more likely to experience markers of disengagement, such as daily absence, disruptive behaviour, and poor school connectedness. School connectedness was a protective factor for many students and associated with a reduction in risk-taking behaviours, as well as increases in school attendance and academic achievement. As students lost school connectedness, there was a sense that adults and/or peers in their school were no longer concerned about them or their learning. Consequently, psychological distress such as anxiety and depression increased.

Strategies could include utilising student mentoring programmes, some already established. Another strategy may involve student teachers who may have their practicums on hold due to Covid-19, who might assist in an online mentoring capacity, with reciprocal benefits. Old technology should not be disregarded, so where there was poor Internet connectivity, the curriculum could be taught via television.

Nearchou et al. (2020) conducted a proper systematic review of the effects of the pandemic on the mental health of young people ≤18 years old, but again did not summarise outcomes. Eight databases were searched and 700 hits led to 12 included studies ($n = 12{,}262$). All studies were of low or moderate methodological quality. All studies were quantitative, seven from China, two from Italy, one from Poland, one from Turkey and one from the United States. Most ($n = 10$) involved online questionnaires. Various forms of sampling were used, sample size varying from 17 to 8072. Only seven studies focused explicitly on children and adolescents (age range 3–18 years). Narrative reporting was used.

There was ample evidence from previous epidemics that individuals who recovered from acute viral illnesses may experience significant mental distress and go on to experience psychiatric problems. For example, the immediate aftermath of the SARS epidemic saw the emergence of various psychiatric comorbidities, with the most common presenting problems involving increased levels of anxiety, depression and features of post-traumatic stress reactions. There was also evidence that these problems could be long lasting for up to a third of people. Nor is it only those who are infected by a virus that experience negative psychosocial consequences. Evidence from Canada suggested that healthcare workers experienced intense emotional reactions during the SARS outbreak,

including fear of contagion, feelings of stigmatisation, boredom, loneliness, anger, anxiety and uncertainty.

Covid-19 had an impact on youth mental health and was particularly associated with depression and anxiety in adolescent cohorts. School closure impacted significantly on children and adolescents aged between 5 and 18. Infants and toddlers tended to be more affected by the stress that Covid-19 placed on their parents. Parental stress may be associated with child behaviour problems. Challenges created by school closure included the loss of structure and support, the loss of routine and the loss of social connection. When children were out of school, they were less physically active, spent more time on screens, had more irregular sleep patterns, and less favourable diets, tending to result in weight gain and loss of cardio-respiratory fitness. These effects were exacerbated when children were confined to their homes with limited outdoor activities and no interactions with same-aged friends.

Specifically, approximately 22% of participants reported fear for the health of relatives, almost 40% of primary school children reported being concerned about health and life threats posed by Covid-19, and 62% reported being moderately/quite worried about being infected with the virus. Older adolescents were more likely to report higher depression and anxiety symptoms than their younger counterparts. Adolescents who engaged in social distancing to protect themselves from getting sick and to avoid social judgement reported greater anxiety symptoms, but those who engaged in social distancing due to peer pressure reported greater depressive symptoms. Older adolescents tended to use social media as well as electronic media communication more frequently than younger adolescents, which may have resulted in being exposed to Covid-19 information more frequently and intensely. Frequent social media exposure was associated with increased likelihood of depression and anxiety.

Implications for policy

Darmody et al. (2020) offer a narrative review of selected studies published in 2020 (mainly research reports and grey literature), focusing on emergency remote learning and outlining implications for policy at length. However, the authors have interesting things to say about many topics, including family relationships. On the positive side, they noted that while the pandemic had an adverse impact on many individuals, nearly half (46%) of adults reported increased positive family time. A high proportion of those currently living with partners (81%) or children (83%) felt generally positive about spending more time together, and 40 per cent of the parents anticipated a closer relationship with their children once lockdown was lifted. However, despite largely positive parent-child relationships, some parents experienced stress and might suffer from depression more generally, levels varying from 14% for professionals to 40% per cent of the never-employed. By the age of 17, almost a quarter of young people had mothers who had depressive symptoms. A decline in economic circumstances was associated with an increase in maternal depression. Parents who experienced psychological distress tended to become harsher in their parenting style, showing less warmth to their children. This more distant interaction, in turn, was linked to children feeling less happy and more anxious, and displaying poorer conduct. Additionally, not being able to have face-to-face contact with friends was the single hardest impact of social restrictions on children and teenagers. While younger children turned to parents during the pandemic, teenagers considered their friends an important source of support.

Regarding education at home or home-schooling, parents with higher levels of educational attainment were more likely to teach their children directly or actively support their learning. Three parents in five reported that their children's level of motivation to learn had deteriorated. While older children were more likely to prefer learning in school, younger children preferred the home environment. Parents with both higher and lower levels of education

reported spending 1–2 h/day home-schooling their children. These figures are substantially less than a typical school day.

Turning to communication with school, most parents (80%) reported receiving text messages or emails from their child's school, and more than a half were receiving resources directly from the teacher. However, only about a quarter of children were accessing virtual classes and only one third were communicating with their teachers through educational apps. The abrupt transition to remote learning affected some activities more than others, with group work and practical work (generally seen as more engaging by students) suffering more. This was especially evident in state schools and in areas with low broadband coverage and/or lower incomes.

Regarding differences by socio-economic disadvantage, students in private schools were much more likely to have daily online classes and access to online videoconferencing with teachers. Even in state schools, higher-income parents were more likely than low-income parents to say their children received active help from the school. Children in higher social class families were significantly more likely to have access to a laptop or computer at home. Whereas 71% of children from professional families had access to these devices, only 44% of those whose parents were never employed did so. In the most disadvantaged schools, 15% of teachers reported that more than a third of their students would not have adequate access to an electronic device for learning from home, compared to only 2% in the most affluent schools.

Turning to student response to online/blended learning, 50% of teachers in private schools reported they were receiving more than three-quarters of work back, compared with 27% in the most advantaged state schools, and just 8% in the least advantaged state schools. Twenty-four per cent of teachers reported that fewer than one in four children in their class were returning work they had been set. Teachers in the most disadvantaged schools were also more than twice as likely as those in advantaged schools to report that work their students were submitting was of much lower quality than before (15% vs. 6%).

Regarding psychological well-being, in Italy and Spain the vast majority of parents (86%) reported changes in the behaviour of their children, especially difficulty concentrating (77%), boredom (52%), irritability (39%), restlessness (39%), nervousness (38%), feelings of loneliness (31%), uneasiness (30%) and worries (30%). In China younger children (age 3–6) were more likely to display clinginess and articulate fears that family members would become ill. Older age groups (6–18 years) were more likely to display inattention, with irritability common across all age groups. In Sardinia a study of parents indicated similar problems, with over half of children (aged 4–10) showing increased irritability and around a fifth displaying mood swings and having sleep problems.

This review highlights the importance of policy intervention. In addition to enhanced opportunities to engage with small group or one-to-one tutoring, given the level of unmet need in child and adolescent mental health services, access to specialist psychological and therapeutic supports through schools should be enhanced. Summer camps or programmes and after-school provision could provide important opportunities for young people to re-engage. However, low-income families would need financial support to participate in these. This highlighted the importance of community-based provision, such as local youth clubs. The transition to life after school would be very different. Higher education institutions appeared likely to use blended learning increasingly, highlighting the importance of addressing the digital divide and providing academic and social support to students to enhance their well-being and prevent dropout. Young people entering the labour market faced particular challenges and needed additional supports to avoid long-term damage to their prospects, such as early intervention to provide career guidance and promote retention, as well as targeted (re)training to respond to potential new areas of employment growth.

Original contribution of this paper

This study has a number of novel features which discriminate it from other literature in the area of online and blended learning. First, it synthesises research literature from before the pandemic as well as during the early stages of the pandemic. It does not specifically target emergency remote learning, so it speaks mainly to the situation before and after the pandemic rather than during it. Second, it includes a great many studies in schools, far more than previous reviews. Third, it specifically attributes outcomes to sectors of education: kindergarten, primary/elementary, secondary/high, with a large number of studies in the first two categories. Fourth, it explores effectiveness in relation to gender. Fifth, it comments on whether children perceived as 'low ability' in class do better or worse in the online/blended environment. Sixth, it explores the impact of self-efficacy in learning outcomes. Seventh, it comments on parallel digital technologies (educational games, CSCL and CAI), which are not widely used outside of school. Finally, it summarises outcomes from the studies, in terms of whether the intervention was better, worse or the same in relation to regular instruction.

Aims

During the coronavirus pandemic, many schools turned to online and blended learning in various forms to provide education for children who were not allowed in school. As the pandemic receded, it is probable that some schools and teachers returned gratefully to the kind of teaching they used pre-pandemic, but many others learned lessons from their digital experiences during the pandemic to improve the quality of education for their pupils. This systematic analysis aims to investigate the research literature on the effects of online or blended learning from schools to see whether it has effects which are better, worse or the same as regular face-to-face classroom learning, which will be relevant as schools return to 'normal' functioning.

Definitions

Online learning

Online learning is a style of education in which students learn complete programmes of work via electronic and online media only, so that they can completely control the time, pace, and place of their learning (*Oxford English Dictionary*). In other words, all learning happens out of school. However, in this review many studies reported what they called online learning, when in fact they had the pupils in school using the school internet system to access some form of web-based program under the supervision of the teacher. This is not what we consider to be online learning, and such projects have been categorised as computer-assisted instruction (CAI).

Blended learning

Blended learning is a style of education in which students learn via electronic and online media as well as regular face-to-face teaching in the classroom, so that they can, in part, control the time, pace and place of their learning (*Oxford English Dictionary*). In other words, part of the learning happens out of schools and part of it happens in schools. One example of this is the widespread notion of 'flipped learning' (Bond, 2020b). Typically, information-giving

happens out of school and interactive elements such as teacher and peer discussion happen in school. The in- and out-of-school aspects are usually closely synchronised—for example, mornings at home and afternoons in school. However, interactive sessions between teacher and student(s) or between student and peers can be also hosted through video conferencing, web chat, message boards or other means. In this review, online or blended learning may have taken place for an entire programme of learning, or it may only have taken place for the relatively brief project which is reported.

In this report we also consider computerised educational games and computer-supported collaborative learning (CSCL), currently mainly used in school but clearly with the potential to be used in locations outside of school.

Educational games

Computerised educational games (EG) are designed to help users learn about certain subjects, expand concepts, reinforce development, understand a historical event or culture, or assist them in learning a skill as they play, but do so within the context of a game which has rules and objectives (e.g., to 'win'). They can be 'serious' games designed especially for their educational value, or commercial games, which nonetheless have educational value and are often better produced. Games involve interactive play that teaches goals, rules, structure, adaptation and problem solving, all often represented in a story. They provide feedback and enable learning by enhancing enjoyment, engagement, motivation, creativity, social interaction and emotion, including ego gratification. Games may be for single players, for two or several players, or be part of a massively arranged system for multiple cooperative teams or adversaries, as in Massive Multiplayer Online Games (MMOGs) (Manesis, 2020).

Computer-supported collaborative learning

Computer-supported collaborative learning concerns how collaborative learning supported by technology can enhance peer interaction and work in groups, including facilitating the distribution of knowledge and expertise among community members (Lipponen, 2002). The participants use Internet-enabled software tools to support social and collaborative learning at a distance from one another and from their instructor. However, many CSCL projects have taken place within schools rather than outside them.

Computer-assisted instruction

We did not include computer-assisted instruction (CAI) keywords in our literature search. However, we found that many studies which purported to be of online or blended learning actually turned out to be school-based, where pupils had interacted with digital materials or used web-based materials in school while under the direct supervision of the teacher. Hence, we categorised these as CAI. Clearly, schools are using many forms of technology on site, which could be exported out of school, but it seems this had not been occurring to any considerable extent pre-pandemic. Consequently, this definition of CAI is given: CAI uses a combination of text, graphics, sound and/or video via a computer to present a programme of instructional material, have the student interact with it, and monitor the learning that takes place. CAI programs can be simply drill and practice, or they can be much more complex, such as involving simulations (Root et al., 2018).

Regular or traditional instruction

Many of the studies reviewed here compared digital technology to regular instruction, so it is worth giving a definition of that also. Regular or traditional or conventional classroom instruction can be defined as a structured programme of face-to-face predominantly teacher-centred instruction, including the teacher giving information and teacher-led discussion. Instructional materials include textbooks, lectures and individual written assignments. All students in the classroom usually receive a single, unified curriculum. Particularly in secondary school, subjects are often individual and independent, whereas in primary school they may be integrated (UNESCO-UNEVOC, 2022). We accept that regular instruction is very likely to vary, not least across countries, and that this comparison is with a highly heterogeneous group of practices and is highly arbitrary. Furthermore, not all papers compared the experimental condition to regular instruction.

Given these definitions, we developed the following research questions:

RQ1: Does online learning from schools have positive effects compared to full-time regular instruction in school?
RQ2: Does blended learning from schools have positive effects compared to full-time regular instruction in school?
RQ3: Do educational games have positive effects compared to full-time regular instruction in school?
RQ4: Does computer-supported collaborative learning have positive effects compared to full-time regular instruction in school?

Structure of this paper

Many studies on online and blended learning have been conducted in higher education, but this review of digital learning from schools since 2000 encompassed 1355 studies, so clearly despite the lower volume there is also a good deal of research on this. The review also investigates EG and CSCL, which are mostly used in school but have the potential to be used outside school. This is also true of CAI, which was not searched for but nonetheless featured heavily among the hits. Of course, EG, CSCL and CAI also have the potential *not* to be used in blended learning.

Definitions of blended and online learning (and synonyms) and EG, CSCL and CAI are given. Previous literature largely focusing on very recent empirical reviews and on underpinning theory is discussed. Most of the relevant empirical work is of course summarised later in the paper. Research questions lead to Method, with keywords, databases and exclusion criteria outlined. The coding framework is explicated and inter-rater reliability reported. Results are then given in summary form. The paper ends with a discussion (including limitations, links to previous literature and a critical analysis) and a conclusion (which gives the answers to the research questions). Appendix S1 lists the full references for the 1355 items in the review. Appendix S2 offers more accessible and briefer teaching points drawn from the research evidence.

LITERATURE REVIEW

Most of the empirical literature is summarised in the systematic review that follows, so we will not mention it here.

Studies contributing references to this review

The Education Endowment Foundation (EEF) (2020b) produced a rapid research review of online and blended learning from schools. Although valuable, this study only reviewed reviews and meta-analyses (and then not all of them), whereas we reviewed reviews and individual studies. We extracted the reviews and added them to our study. However, the present review categorises studies in the same way as the EEF report, that is, as Online, Blended, Games, CSCL, or CAI. In the US, Studies of Distance Learning (2020) gave a list of studies relevant to online and blended learning across all sectors of education, which partially stemmed from searching ERIC with regard to blended and online learning and partially from crowd-sourced suggestions for relevant studies. We extracted the studies concerned with schools and added them to our study. Evidence for Learning Australia (2021) (https://www.evidenceforlearning.org.au) have been analysing single studies from the EEF Rapid Evidence Assessment. We also do this, but add many other findings.

Other and more recent reviews

The peer reviewers for this paper made a number of helpful suggestions, including listing seven reviews which the search terms and databases had failed to identify. We discuss these below. They also commented on the difficulty of merging quantitative and qualitative studies in the same paper, which we have done here. Major and Savin-Baden (2011) offered a summary of ways of aggregating qualitative studies, but it is only of qualitative studies, not of qualitative and quantitative studies combined.

Of these seven reviews, only one was entirely focused on schools (Pérez-Sanagustín et al., 2017), but it did not focus on online and blended learning from schools, rather on all ICT use in schools. Likewise, Tamim et al. (2011) reviewed reviews but only on the impact of technology, not of online or blended learning, and mixed schools and post-secondary establishments without discrimination. Schindler et al. (2017) similarly focused on the impact of technology, not of online and blended learning, investigating the effects on student engagement. There was no discrimination between educational sectors at all, but the paper seemed likely to be all, or almost all, about higher education.

Turning to studies which were of online and blended learning, Cheng et al. (2018) investigated the effects of the flipped classroom, but only 16 out of 55 studies were in schools. Nonetheless, effect sizes for K–12 ($g = 0.22$) and undergraduate students ($g = 0.21$) were similar. Arts and humanities showed the largest effect sizes ($g = 0.63$), but this was only in five studies. Lundin et al. (2018) also investigated flipped learning, but used only one database and focused on highly cited papers. Publications addressing higher education contexts made up 25 of the 31 publications, compared to only four publications addressing K–12 schools. They did, however, also attempt to investigate relevant theory, complaining about its 'notable absence'.

Van Alten et al. (2019) likewise focused on the flipped classroom, but only in secondary ($n = 11$ studies) and post-secondary education ($n = 104$). This was despite commenting that the flipped classroom might work differently for students of different age groups—because of differences in self-regulated learning abilities which tended to increase at higher ages. This assumption is contradicted by other evidence. Another paper was that of Philipsen et al. (2019), who offered a systematic review of qualitative data from 15 articles on teacher professional development via online/blended learning, but of course this was of teachers in higher education or beyond. Only one database was searched, search terms were not given and quantitative research was excluded.

Theoretical literature

Behaviourist strategies can be used to teach what (facts), cognitive strategies can be used to teach how (processes and principles), and constructivist strategies can be used to teach why (higher-level thinking that promotes personal meaning, situated and contextual learning) (Cooper, 1993), but components of each can be found in modern instructional design models (Janicki & Liegle, 2001). For example, Garrison (2000) and Garrison et al. (2001) blended the cognitive, social and instructional presences in their 'Community of Inquiry Framework'.

In the West, learning has generally moved away from uni-directional teacher-student instruction to student discovery and construction of knowledge, but this is not necessarily the case in other countries, such as those with a Confucian heritage (Lam, 2020). Constructivist theorists claim that learners interpret information and the world according to their personal reality; learn by observation, processing and interpretation, and then transform the information into personal knowledge (Wilson, 1997). A feature of constructivism can be situated learning, which sees learning as contextual, allowing immediate application (Hung et al., 2004). Multi-contextual learning is used to ensure generalisation.

A separate track in the theory argues that multi-media input, as is easily arranged through computers, is important. Paivio's (1971) 'Dual Coding Theory' recognised two modes for processing information: through images and language. Learners who utilised both modes of processing simultaneously facilitated enhanced comprehension. Similarly, in Krashen's 'Comprehensible Input Theory' (Krashen, 1981), learners acquired skills when given sufficient exposure to them in multiple meaningful multimedia contexts. Again similarly, Mayer et al.'s (2014) 'Cognitive Multimedia Learning Theory' or 'Dual Channelling Theory' suggested that when learners engaged with both auditory and visual materials, they processed information through both senses, resulting in a stronger and more meaningful impact on the learning process. However, multi-media learning facilitates but does not ensure 'Active Learning' (Prince, 2004), which requires learners to be engaged in meaningful activities and tasks which encourage them to apply and transfer knowledge more effectively, possibly also heightening meta-cognitive and self-regulation skills.

A third track in the theory is whether learning should be in isolation or embedded in peer group activities. Relevant here is Vygotsky's 'Sociocultural Theory' (1962), emphasising that learning is a social process and social interaction is closely linked to cognitive development. The 'Zone of Proximal Development' is the space between areas of task difficulty which are too hard for the learner to succeed in unaided, but not too hard for them to succeed in with help from a peer. Long's (1981) interactional theory suggested that development would be enhanced when learners participated in negotiated meaningful interaction, especially between peers. Associated is Fisher's theory of information grounds (Fisher et al., 2004; Fisher & Naumer, 2006), which promotes the desirability of socially embedded peer-supported learning environments.

A fourth track in the theory considers the role of stress, affect and self-efficacy (self-confidence) on motivation in learning. Krashen's (1982) 'Affective Filter Hypothesis' suggests that as stress increases for the learner, effective learning decreases. However, this appears to be a simplification of the Yerkes-Dodson U-shaped curve (Yerkes & Dodson, 1908), which suggested that very low anxiety can be as problematic as high anxiety. Associated with this is Bandura's (1994) work on self-efficacy, concerned with people's beliefs in their ability to influence events that affect their lives. Learners engage in tasks they feel confident in, and reject or avoid those in which they don't. Self-efficacy is the belief that one can make a difference by one's actions. Without it, people have little incentive to undertake activities or to persevere in the face of difficulties. Of course, it is also possible to be *too* self-confident, and this is particularly relevant in regard to gender differences (Busch, 1995). Connected to this is Deci and Ryan's (2008) 'Self-Determination Theory', which argues that a person with

autonomous motivation is characterised by intrinsic motivation, high volition, self-initiated actions, self-regulation and psychological freedom in learning, whereas a person with controlled motivation always feels that their thoughts, feelings and/or behaviours are controlled by outside pressures.

Of course, all of these may apply to different degrees, but all equally apply to face-to-face learning. A fifth track of theory concerns the degree to which computers make a difference. Clark (1983) stated that studies showed that students gained significant learning benefits from computers as opposed to conventional instruction, but suggested that the reason is not the medium of instruction, but the instructional strategies in the learning materials. On the other hand, Kozma (2001) argues that computers are particularly relevant when it comes to simulations, and here the medium does influence the learning.

A more radical theory, 'Connectivism', has been proposed by Siemens (2004), who sees it as the integration of principles from chaos, network, complexity and self-organisation theories. Owing to the information explosion, learning is not under the control of the teacher, nor indeed wholly under that of the learner. Changing environments, innovations, changes in the discipline and in related disciplines all suggest that learners have to unlearn what they have learned in the past, learn how to learn and evaluate new information. What must be learned is continually changing. Given this, content knowledge is less important than thinking ability. 'Learning … can reside outside of ourselves (within an organization or database), is focused on connecting specialized information sets, and the connections that enable us to learn more are more important than our current state of knowing' (Siemens, 2004, p. 5).

Siemens (2004) proposes eight features of Connectionism: (1) Learning and knowledge rests in diversity of opinions; (2) Learning is a process of connecting specialised nodes or information sources; (3) Learning may reside in non-human appliances; (4) Capacity to know more is more critical than what is currently known; (5) Nurturing and maintaining connections is needed to facilitate continual learning; (6) Ability to see connections between fields, ideas and concepts is a core skill; (7) Currency (accurate, up-to-date knowledge) is the focus of all connectivist learning activities; (8) Decision making is itself a learning process—choosing what to learn is seen through the lens of a shifting reality. While there is a right answer now, it may be wrong tomorrow due to alterations in the information climate.

There are more recent theoretical papers (Picciano, 2017), but they add little to the discourse. Of these five areas of theory, the behaviourist/cognitive/constructivist dimension adds little new to processes in the digital domain, as indeed is true of the role of peer interaction (although this might be easier face-to-face than digitally). The effect of multi-media likewise operates in both domains, although it may be easier to engineer in a digital context. The role of stress and self-efficacy is likely to be variable over time. As students encounter a new digital environment, they are likely to be stressed and have low self-efficacy, but later when accommodated to it they may be less stressed and more motivated than in a regular instructional environment (Weinstein & Selman, 2014). The only theory which is applicable only to digital environments is the connectivism of Siemens (2004).

METHOD

Alexander (2020) offered a helpful analysis for education professionals of systematic review and meta-analysis principles (which were principally designed for medical professionals). She gave methodological guidance for framing a systematic review (e.g., situating the review theoretically, establishing what is already known and still needs to be known, articulating an answerable question), procedural steps and challenges (e.g., setting one's search parameters, specifying search terms and time frames, appropriately delimiting the search), consolidating and summarising challenges (e.g., recording the basic details, charting other appropriate characteristics,

beginning to consider other ways of grouping the data that are potentially relevant), and interpreting and communicating findings (e.g., recognising outcomes that are meaningful and relevant, finding ways to capture significant patterns and trends, clearly communicating the outcomes of the systematic review). We follow these steps in this paper.

Keywords

We used the following keywords in our search:

- school (including primary and elementary, middle, secondary and high, although these were not entered as search terms) AND
- blended learning OR online learning OR e-learning OR elearning OR virtual learning OR distance learning OR remote learning OR digital teaching OR flipped learning OR flipped classroom OR computer supported collaborative learning OR computer supported cooperative work OR online cooperative work OR online collaborative learning OR educational games online OR educational video game OR serious games.

We initially added a third sector (AND effect OR impact OR outcome), but found that this seriously restricted the number of hits in different databases as these words did not appear frequently in titles, abstracts or keywords, and so discarded it.

Databases and exclusion criteria

We searched the following research databases: ERIC, JSTOR, Scopus, Web of Knowledge, Google Scholar, EEF Rapid Evidence Assessment Reviews, EEF Rapid Evidence Assessment Single Studies, and Studies of Distance Learning. The inclusion/exclusion criteria are given in Table 1. Although the technology itself changed constantly, methods for engaging teachers with new technology might not have changed as rapidly, so we felt we should search back at least as far as 2000.

The PRISMA chart (Figure 1) shows the quantity of papers from different databases—there were very different numbers from different databases. Also, the number of duplications were relatively small, indicating each database was accessing a largely different range of items. Although the EEF Rapid Review listed many reviews, systematic analyses and meta-analyses of online and blended learning, games, CSCL and CAI, we found many more. Abstracts which appeared to be relevant from the title or abstract were identified—1540 of them. The full paper was obtained for each of these items and 185 did not conform to our selection criteria and so were excluded, leaving 1355 papers. The vast majority of these were peer-reviewed journal papers, although there were a few conference papers and even fewer doctoral theses. There were no papers authored by commercial entities who might have had a stake in the product.

The coding framework

The 1355 full papers were then coded according to what data they provided. The coding framework was not guided by any particular theoretical framework, but developed inductively from our preliminary scan of the full text of the included papers, including all categories that seemed to be most common. Papers which were reviews or meta-analyses were coded as such. Papers were then coded according to whether they were in the Online, Blended,

TABLE 1 Inclusion/exclusion criteria

Inclusion criteria	Exclusion criteria
Papers dated 2000 or after	Papers dated before 2000
Learning in school	Learning other than in school
Papers in English	Papers not in English
Papers including evidence (whether qualitative or qualitative)	Papers not including evidence
Peer reviewed journal papers	Books
Chapters with an abstract	Chapters which did not have an abstract
Doctoral theses	All other items without an abstract
Conference papers	Research reports unless from university or government
	Duplicates

Games, CSCL or CAI categories (irrespective of what they claimed to be) and studies which overlapped areas were multiply coded. Studies were then coded according to what sector of education they related to: Early Years or Kindergarten, Primary or Elementary, Middle, Secondary or High, Primary and Secondary together, and unclear (and again studies which overlapped areas were multiply coded). Early Years or Kindergarten was included as a category as some readers might not expect online or blended learning to be applicable to such young children.

Studies were then coded according to subject area: Reading, English, Writing, Maths, Science, STEM (Science, Technology, Engineering and Mathematics), Thinking, English as a Foreign Language (EFL), Humanities (including history, geography, social studies, economics), Health (including physical education, nutrition, exercise, aggressive behaviour), and Arts/Music (including composing, sculpting). These categories evolved from early attempts at coding and while the aggregate categories encompassed considerable diversity, they proved manageable.

Studies were then coded according to gender if that was mentioned, then according to sub-group: Socio-economically Disadvantaged, of High or Low Ability (as perceived by the school, which turned out to be almost all of low ability), whether participants had any Special Educational Needs or Disability (and the type of that disability), were of any Ethnic Minority within their host country, were mainly in a Rural setting (since online and blended learning are particularly relevant for those far from a school), were Second Language Learners (although most of the findings were related to those studying English as a Foreign Language), were Hospitalised or Sick at Home children, Gifted children, or those in a Non-English-Speaking Educational Context. In this latter case we wished to see how many different countries apart from English-speaking countries participants came from, in order to consider whether their host educational culture might be different from Western Anglophone culture.

Then Outcomes were considered: was the intervention better, worse or the same? In many cases the comparison was to regular instruction, but sometimes it was in comparison to another intervention, which might also have been computer-led. The comparator group was usually referred to in studies as a control or comparison or alternative treatment group. Some studies used a baseline design to compare student functioning with the functioning of the same students in a previous period. We did not include any indicator of the *degree* to which a study was Better/Worse, as many studies were qualitative and this proved impossible to quantify. Codings of separate results in the same direction were treated as one coding. Where a study reported more than one outcome where the outcomes contradicted each other, both contradictory outcomes were coded (this accounts for some findings appearing to have more codings than the number of papers).

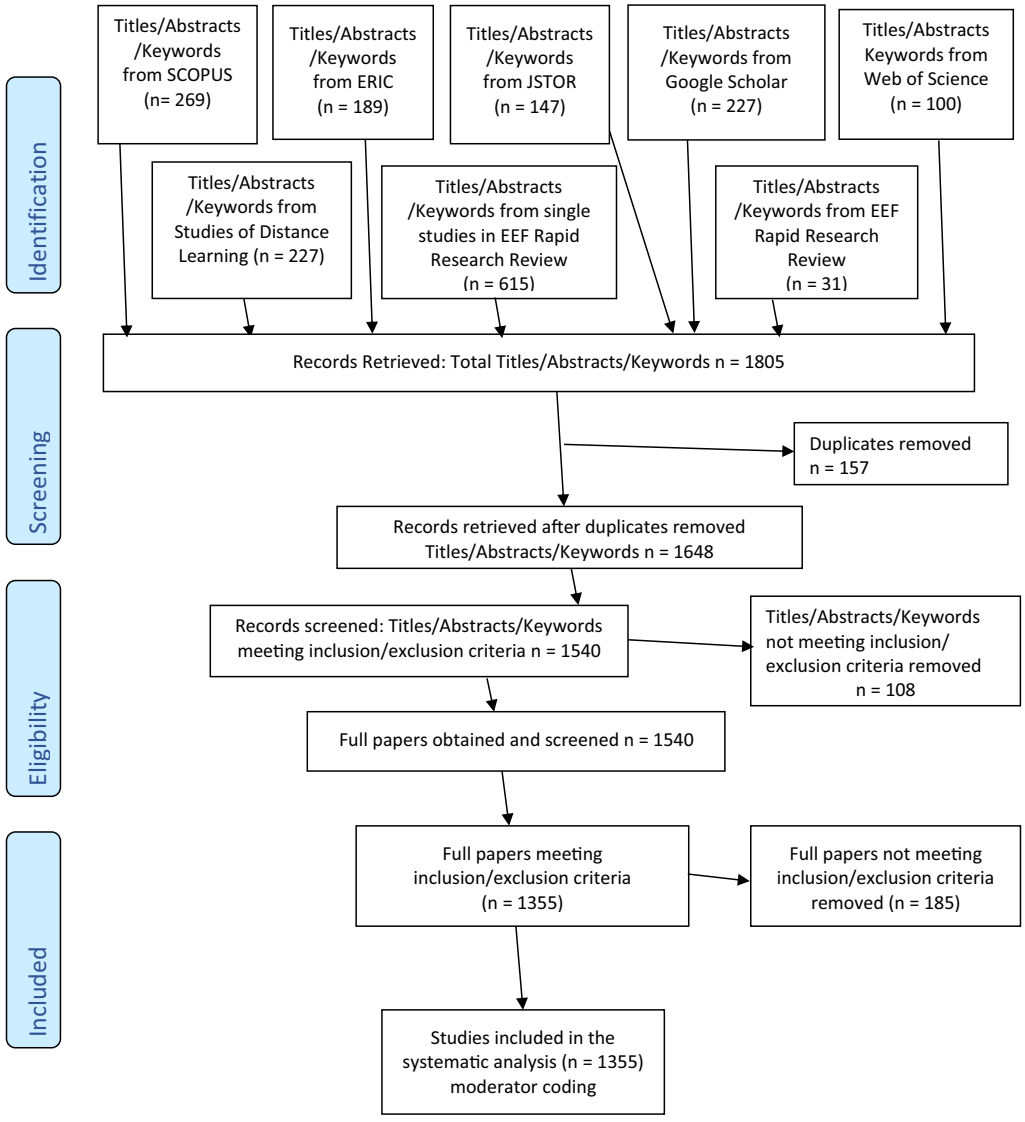

FIGURE 1 PRISMA flow diagram

Then Socio-Emotional Functioning was considered—was there evidence that the implementation of online or blended learning had improved social functioning, worsened it, or had it stayed the same? Similarly, Psychological Well-being was considered, although relatively few studies mentioned this. Finally, teaching points were also extracted, but these were not coded.

Coding definitions

Many studies which superficially appeared to be of online and blended learning did not in fact conform to our definition of these. Such studies might have had pupils using web-based learning or digital packages, but these tended to be only in school, using the school's own

internet system, rather than at home. Such studies were coded CAI rather than Online or Blended. Although we deliberately excluded CAI from our search terms, many CAI methods have *potential* for use in blended learning, and where we discovered such items, we included them and coded them as CAI. Likewise, many of the Games and CSCL we reported were used in a classroom environment, but clearly had potential to be used in a blended learning environment.

One option we did not pursue was to code online/blended learning as either synchronous or asynchronous. It has been argued that web-based learning in school is more likely to be synchronous, whereas online/blended learning outside of school is more likely to be asynchronous, offering a higher level of flexibility concerning mode, place, pace and path (Kanematsu et al., 2009). In the event, few of the abstracts mentioned this variable and some studies used both synchronous and asynchronous, and so it was not coded.

Additionally, the concept of Non-English-Speaking Educational Context (NESEC) indicated countries where English was not the native language and where the educational context might be very different from what would be expected in the West. Ethnic Minority (EM) was used for an ethnic minority within a country, rather than between countries.

Full papers and inter-rater reliability

The 1540 full papers were divided between two senior researchers very familiar with this research area, who found a further 185 papers not to meet the inclusion criteria, leaving 1355. Many of those excluded contained no outcome data related to student outcomes ($n = 90$), many were only in higher education ($n = 64$), a few were beyond the scope of the study ($n = 14$), some were so obscure as to prove unobtainable despite our best efforts ($n = 12$) and in three cases the full paper was in a language other than English (although an abstract was given in English). Inter-rater reliability of coding was assessed for a random sample of 100 items, yielding a percentage of agreement between coders of 95.6%. We noted that of the 185 papers excluded, largely on the grounds that they gave no data on student outcomes, 17 of them nonetheless felt able to give recommendations regarding how to proceed.

Data analysis

For each coding category, the number of studies that indicated the digital intervention was Better were summed, with the number that were the Same and the number that were Worse. Some studies were Unclear, and this category was also summed. Then the distribution of responses among these options was compared to the response options that would be expected 'normally', that is, a flat distribution. The resulting chi-squared statistic and probability was quoted. In a few cases the actual response options contained a zero, and in these cases chi-squared could not be used. Fisher's Exact Probability test might have been used here, but in these cases the overall numbers were very low and it was felt this would not be reliable.

Quality of studies

Rating the quality of studies was highly desirable, as otherwise we would be adding strong and weak studies together and giving them the same weight (although our first impression of the results was that the proportion of positive studies was so high—see Results

below—that both strong and weak studies must be contributing to this). We tried using the GRADE framework of Guyatt et al. (2011), who proposed rating quality of evidence as High, Moderate, Low or Very Low. Studies were rated lower if the risk of bias, inconsistency, indirectness and imprecision were serious or very serious and publication bias was likely or very likely, and higher if there was a large effect, a graded response to dose, and if all plausible confounds would reduce a demonstrated effect. The authors acknowledge that application of GRADE to ill-defined recommendations that one may call 'motherhood statements' or 'good practice recommendations' would prove problematic. Indeed, the studies in this paper were so various and so variously described that obtaining a satisfactory level of inter-rater reliability proved impossible (the best overall was Cronbach alpha 0.57).

We searched for other methods of rating quality of studies which might avoid this concern, for example in Zeng et al.'s (2015) meta-analysis of methodological quality assessment tools, but we found that although the methods with fewer stipulations regarding the categories were more likely to give higher inter-rater reliability, this was largely because the stipulations themselves were so few and so loosely defined, so the higher inter-rater reliability was artefactual. However, we then observed that within the Guyatt et al. (2011) framework, the studies in the highest category (High) actually had a higher and satisfactory Cronbach alpha (0.74) when compared to the residue of the studies. This enabled us to compare the outcomes for the high-quality studies with the outcomes for the other studies, to investigate whether the high-quality studies yielded a disproportionately high or low result compared to the other studies. Obviously, the fear was that the higher quality studies would yield a disproportionately low result compared to the other studies. The outcome of this investigation is reported in the results.

RESULTS

Analysis by intervention and outcome

Following the Guyatt et al. (2011) framework, of 1848 mentions (some papers were coded for more than one intervention—figures represent mentions, not papers) coding divided these into High Quality 425 (23%), Moderate Quality 592 (32%), Low Quality 517 (28%) and Very Low Quality 314 (17%). Table 2 shows the distribution of High versus Other categories across the Better, Same, Worse, Unclear outcome categories. Comparing frequencies using the chi-squared test, the result was not significant (χ^2 [df 3, $n = 1848$] = 0.89, $p = 0.83$). Thus, the high-quality studies showed the same pattern of outcomes as the other studies.

Analysis of data by intervention and outcomes is shown in Table 3. Categorising the interventions in the 1355 papers, we found: Online 134 studies (7%), Blended 232 (13%), CSCL 129 (7%), Games 488 (26%), and CAI 865 (47%) (a few studies encompassed more than one area of operation).

Considering the overall totals regarding outcomes, the general picture was very positive—1576 mentions in studies found digital technology better than regular instruction (85%), whereas 146 (8%) found it the same and only 46 (3%) found it worse (see Table 2). Compared to 'normal' expectations, the chi-square statistic (χ^2) was 1181.30, $p < 0.00001$. This is clearly a major finding, suggesting that digital technology is much superior to regular instruction despite the caveats, at least in research papers.

Among the interventions, CAI appears as the most effective in terms of the number of papers reporting it Better than regular instruction ($n = 784$, 91%, $\chi^2 = 778.21$, $p < 0.00001$). Next came Blended Learning ($n = 192$ Better, 13%, $\chi^2 = 159.92$, $p < 0.00001$), not far ahead of Games ($n = 397$, 81%, $\chi^2 = 319.23$, $p < 0.00001$) and CSCL ($n = 104$ Better, also 81%, $\chi^2 = 83.07$, $p < 0.00001$). A fair way behind comes Online Learning ($n = 99$, 74%, chi-squared = 67.45, $p < 0.00001$), although this also represents a good degree of effectiveness.

TABLE 2 Comparison of high quality and other papers

Intervention rating	Better	Same	Worse	Unclear
High quality	365	37	9	17
Moderate, low and very low categories	1211	109	37	63
Total 1848 mentions, 100%	1576 (85%)	146 (8%)	46 (3%)	80 (5%)

TABLE 3 Summary data for interventions

Intervention	Better	Same	Worse	Unclear
Online ($n = 134$, 7%)	99 (74%)	17 (13%)	11 (8%)	7 (5%)
Blended ($n = 232$, 13%)	192 (83%)	21 (9%)	6 (3%)	13 (5%)
CSCL ($n = 129$, 7%)	104 (81%)	9 (7%)	3 (2%)	13 (10%)
Games ($n = 488$, 26%)	397 (81%)	44 (9%)	14 (3%)	33 (7%)
CAI ($n = 865$, 47%)	784 (91%)	55 (6%)	12 (1.5%)	14 (1.5%)
Overall total ($n = 1848$, 100%)	1576 (85%)	146 (8%)	46 (3%)	80 (5%)

Note: Some papers were coded for more than one intervention. Figures represent mentions, not papers.

Blended Learning clearly did better than Online Learning, both in terms of being better but also when adding better and same results together (which is relevant for these alternatives to regular instruction). Regarding CAI, Games and CSCL, which tend to be used in schools but less so outside of schools, the Better figure is the most important, since no teacher would be interested in using a method which only gave the same results as existing methods. Here CAI is clearly the best, with Games and CSCL coming equal second. Blended Learning gave better results than either of these latter two.

Reviews and meta-analyses

Among the 1355 papers were a large number of reviews and meta-analyses which included schools—144 in total, divided into Online 19, Blended 20, Games 56, CSCL 14 and CAI 35 (see Table 4). Unfortunately for our purpose, many of these reviews (81%) covered primary, secondary and higher education (and even beyond) all together. Where this was the case, the majority of the studies usually came from higher education. It was rare for there to be any discrimination in the findings about what sector of education studies came from.

In Online Learning, 13 of 19 reviews found Online Learning better that regular instruction ($n = 13$ Better, 68%, $\chi^2 = 8.89$, $p < 0.03$), in comparison to Blended Learning ($n = 19$ Better, 95%, $\chi^2 = 17.98$, $p < 0.0004$). CSCL was $n = 13$ Better, 93%, $\chi^2 = 10.16$, $p < 0.02$; Games was $n = 49$ Better, 88%, $\chi^2 = 43.97$, $p < 0.00001$; and CAI was $n = 29$ Better, 83%, $\chi^2 = 22.93$, $p < 0.00004$. Thus, Blended Learning reviews were the most positive of all, although Games and CAI had strong chi-squared results.

Effect Sizes (ESs) are a quantitative measure of the magnitude of an effect, ranging from 0 to 1 or more, and the larger the ES the stronger the relationship between two variables. Negative ESs indicate a negative or worsening effect. By contrast, statistical significance varies with the size of the sample. In Online Learning only seven reviews gave ESs, ranging from 0.03 to 0.38 with a mean of 0.16, and none focused on a specific sector of education. In Blended Learning, only six reviews gave ESs, ranging from 0.20 to 0.98 with a mean of 0.48, and only two focused on a specific sector of education. In Games only 11 reviews gave ESs, ranging from 0.13 to 1.13 with a mean of 0.48, and only six focused on a specific

TABLE 4 Summary data for reviews and meta-analyses

Intervention	Better	Same	Worse	Unclear
Online (n = 19, 13%)	13 (68%)	5 (26)	0	1 (5%)
Blended (n = 20, 14%)	19 (95%)	1 (5%)	0	0
CSCL (n = 14, 10%)	13 (93%)	0	0	1 (7%)
Games (n = 56, 39%)	49 (88%)	2 (4%)	0	5 (8%)
CAI (n = 35, 24%)	29 (83%)	3 (8.5)	0	3 (8.5%)
Overall Total (n = 144, 100%)	129 (90%)	11 (5%)	0	10 (5%)

Note: Some papers were coded for more than one school sector. Figures represent mentions, not papers.

sector of education. In CSCL only seven reviews gave ESs, ranging from 0.20 to 0.95 with a mean of 0.49, and only two focused on a specific sector of education. In CAI only 11 reviews gave ESs, and these ranged from 0.13 to 2.5 with a mean of 0.81, and only six focused on a specific sector of education.

It might appear that from reviews CAI is the most effective intervention, while Blended Learning, Games and CSCL are of moderate effectiveness, followed by Online Learning which seems the least effective (although still somewhat effective). However, we urge extreme caution in interpreting these figures, as relatively few reviews focused solely on schools. In any event, the ESs for each intervention are extremely heterogeneous. These results are not greatly dissimilar to those for all papers separately, except that when considering the individual papers Blended Learning seems to do rather better than in the reviews.

Analysis by subject and sector

Analysis by subject and sector is shown in Table 5. The subjects were: Reading, Maths, Science, STEM, Thinking, EFL, Humanities, English, Health, Art Music, and Writing. The sectors were: Early Years Kindergarten, Primary, Middle, Secondary and Primary +High schools. We created a matrix which mapped the interactions between these factors as well as outcomes. The resulting matrix was extremely complex and here we merely summarise it.

In Table 5, the general picture for sectors was again very positive—82% of these interventions did better than regular instruction, while 10% were the same and only 3% were worse. The Primary/Elementary sector (n = 468 Better, 87%, χ^2 = 3438.53, $p < 0.00001$) and the Early Years/Kindergarten sector (n = 39 Better, 87%, χ^2 = 34.81, $p < 0.00001$) did best, the latter particularly startling for those who might not have expected digital technology to be widely used with very young children. Secondary/High came next (n = 370 Better, 80%, χ^2 = 298.75, $p < 0.00001$). Primary and Secondary came next (n = 165 Better, 77%, χ^2 = 120.88, $p < 0.00001$), below both Primary and Secondary separately, perhaps raising questions about the wisdom of trying to include participants from both sectors simultaneously. Curiously, last of all came Middle Schools (n = 93 Better, 73%, χ^2 = 63.49, $p < 0.00001$). The Primary/Elementary sector had the largest number of studies.

Turning to subjects, Science (21% of the total) and Mathematics (henceforth Maths) (20%) were the most popular subjects, followed by Thinking (10%), Reading (9%), English as a Foreign Language (EFL) (7%), Humanities (5%), Health (5%), STEM (4%), Writing (4%), English (2%), Arts/Music (1%) and Foreign Language (1%). Some papers (mainly reviews and meta-analyses but also reports covering a large geographical area) covered All Subjects (n = 166, 11%). In terms of effectiveness, all subjects were above 80% Better status except

TABLE 5 Summary data for school sectors

School sector	Better	Same	Worse	Unclear
Early Years/Kindergarten ($n = 45$, 3%)	39 (87%)	2 (4%)	1 (2%)	3 (7%)
Primary/ Elementary ($n = 536$, 39%)	468 (87%)	47 (9%)	8 (1.5%)	13 (2.5%)
Middle ($n = 127$, 9%)	93 (73%)	20 (16%)	6 (5%)	8 (6%)
Secondary/High ($n = 460$, 33%)	370 (80%)	55 (12%)	18 (4%)	17 (4%)
Primary +Secondary ($n = 214$, 16%)	165 (77%)	20 (9%)	8 (4%)	21 (10%)
Overall total ($n = 1382$, 100%)	1135 (82%)	144 (10%)	41 (3%)	62 (5%)

Note: Some papers were coded for more than one school sector. Figures represent mentions, not papers.

English (77%). There were some fairly small differences between subjects: EFL interventions were the most effective (90%), followed by Writing and STEM (89%), then Thinking (88%), Arts/Music (87%), Humanities (86%), Health and Science (84%), Reading and Maths (82%), Foreign Languages (81%) and English (77%). Thus, the most popular subjects were by no means those where the greatest effectiveness was evident.

Other moderator variables

Many other moderator variables were present in such modest numbers that it was not felt helpful to further subdivide the studies into each of the five intervention categories. Consequently, effects are generally indicated for all digital interventions together.

Gender was coded. Females were better in 27 cases (39%), Males and Females equal in 36 cases (52%), and Males better in only 6 (9%). Comparing with 'normal' expectations, overall girls did better than boys at digital learning ($\chi^2 = 13.15$, $p < 0.0014$), although many studies had girls and boys doing equally well.

Special sub-groups were coded according to whether they were: Disadvantaged, of HiLo Ability (almost all of 'low ability'), had Special Educational Needs or Disabilities, were from an Ethnic Minority, were from Rural areas (where remote learning might be more needed), were Second Language Learners (almost all EFL studies), were Hospitalised or Sick at Home children, or were Gifted, and whether they came from a Non-English-Speaking Educational Context. Studies coded as Disadvantaged numbered 46 (3%), HiLo 82 (6%), a variety of Disability 112 (8%), Ethnic Minority status 30 (2%), Rural 34 (3%), Second Language Learners 76 (6%), Hospitalised or Sick at Home 7 (1%), Gifted 10 (1%) and Non-English-Speaking Educational Context 329 (24%).

The Disadvantaged subgroup numbered only 46, and in many cases such children performed better than regular instruction ($n = 37$, 79%), with only six Same and two Worse codings ($\chi^2 = 23.11$, $p < 0.00001$). This is a slightly worse outcome than for all students (85%), but nonetheless suggests that Disadvantaged children can make almost as large gains as other children, when usually they would be expected to perform much more poorly. Nonetheless, it also indicates that digital learning widens the gap between the advantaged and the disadvantaged. We expected to encounter some discussion of the presumed difficulties for socioeconomically disadvantaged children of accessing computers and the Internet at home, but there was very little of this, other than some comment on such children possibly having less support from parents. Some schools had made iPads or laptop computers available to Disadvantaged children to take home, with reported positive results. A few studies had tried using mobile phones rather than computers and the Internet. There was also comment on the loss of educational capability during the summer break for disadvantaged children.

The code HiLo almost exclusively related to 'low ability' children (as they might be perceived by the school based on performance in regular class-based subjects) ($n = 82$), and in the vast majority of cases these children performed better than in regular instruction ($n = 76$, 92%), with only three Same and two Worse codings ($\chi^2 = 64.06$, $p < 0.00001$). This is a better outcome than for all students combined (85%), suggesting that apparently 'low ability' children operating in the digital space can prove surprisingly competent when given the opportunity and are not disadvantaged by digital learning.

The Special Educational Need or Disability code ($n = 112$) covered children with very various special educational needs (SEN) (see Table 6), including: All unspecified SEN 24 (21%), Learning Difficulties 30 (27%), Autistic Spectrum Disorder 14 (13%), Attention Deficit and Hyperactivity Disorder 10 (9%), Visual Impairment 5 (4%), Emotional and Behavioural Disorder 6 (5%), Deaf and Hard of Hearing 5 (4%), Dyslexia/Specific Learning Difficulties 5 (4%), Down Syndrome 3 (3%) and Writing Difficulty 3 (3%). Overall, 89 studies (80%) found Better outcomes, not much lower than the average for all students ($\chi^2 = 71.76$, $p < 0.00001$). However, this still means that the gap between Special Needs students and the rest of the student population was continuing to widen. This aggregate figure is worth disentangling into types of disability, which gives more complex results.

On the positive side, Deaf and Hard of Hearing, Down Syndrome and Writing Difficulty students had 100% Better ratings, albeit with very modest numbers of studies (3–5) which did not permit the calculation of χ^2. Attention Deficit and Hyperactivity Disorder (ADHD) students had a 90% Better rating (out of 14 studies, no χ^2), followed closely by Learning Difficulty (87% of 30 papers—the highest number of papers, $\chi^2 = 21.06$, $p < 0.0001$) and Autistic Spectrum Disorder (ASD) (86%; 14 studies, $\chi^2 = 9.38$, $p < 0.025$) students. All of these were above the average for all students.

On the less positive side, Unspecified Disabilities (16 out of 24 studies, 67%, $\chi^2 = 9.52$, $p < 0.03$) and Emotional and Behavioural Difficulties (four out of six studies, 67%, no χ^2) both only had a modest Better rating—but it was still a positive Better rating. Visual Impairment was 60% (five studies, although three were Better and two were Unclear, no χ^2). Dyslexia/Specific Learning Difficulty was least effective (50%), with two Betters, one Same and one Unclear (no χ^2). However, of these 112 SEN studies, only one had a Worse outcome (1%).

TABLE 6 Summary data for special educational need or disability sectors

Impact rank	SEN/disability	Better	Same	Worse	Unclear
7	Unspecified ($n = 24$, 21%)	16 (67%)	3 (12.5%)	1 (4%)	4 (16.5%)
5	Learning Difficulty ($n = 30$, 27%)	26 (87%)	3 (10%)	0	1 (3%)
6	Autistic Spectrum Disorder ($n = 14$, 13%)	12 (86%)	1 (7%)	0	1 (7%)
4	ADHD ($n = 10$, 9%)	9 (90%)	0	0	1 (10%)
9	Visual Impairment ($n = 5$, 4%)	3 (60%)	0	0	2 (40%)
7	EBD ($n = 6$, 5%)	4 67%	2 (33%)	0	0
1	Deaf & HoH ($n = 5$, 4%)	5 (100%)	0	0	0
10	Dyslexia/SLD ($n = 4$, 4%)	2 (50%)	1 (25%)	0	1 (25%)
1	Down Syndrome ($n = 3$, 3%)	3 (100%)	0	0	0
1	Writing Difficulty ($n = 3$, 3%)	3 (100%)	0	0	0
	Miscellaneous ($n = 8$, 7%)	6 (75%)	1 (12.5%)		1 (12.5%)
	Overall Total ($n = 112$, 100%)	89 (80%)	11 (9.5%)	1 (1%)	11 (19.5%)

Note: Some papers were coded for more than one school sector. Figures represent mentions, not papers.

Ethnic Minority Status ($n = 30$) was often seen in US studies as Afro-American (7 studies) or Hispanic (8 studies). However, East Africans also featured (2 studies), as did Native Americans and Maori (1 study each). Ten studies spoke of all ethnicities without specifying. With Ethnic Minorities, 25 (78%) of studies reported Better performance, five (16%) Same and two (6%) Worse ($\chi^2 = 13.91$, $p < 0.0010$). This was similar to the performance of Disadvantaged students, indicating that EM children can make almost as large gains as other children, when some Ethnic Minorities might be expected to perform more poorly. Nonetheless, the gap between Ethnic Minority and other pupils continued to widen.

Rural schools ($n = 34$) were often seen as problematic (du Plessis, 2014), not just because of a lack of hardware and Internet connectivity, but also because in some countries rural school income could be less and the teachers could be less qualified. However, 29 studies (85%) found that digital innovations were better than regular instruction in rural schools, while two were the Same (6%) and two were Worse (6%) ($\chi^2 = 20.56$, $p < 0.00003$). This was similar to the performance of all students taken together (85%), so there was no evidence that rural schools were under-performing.

We hoped the coding for Second Language Learners ($n = 76$) would help us discover students who were not speaking their native language in class. However, the bulk of this category was taken up by English as a Foreign Language learners. Nonetheless, the mostly EFL students showed high levels of performance: 69 (90%) proved Better than regular instruction, while 7 (9%) were the Same and only one (1%) Worse ($\chi^2 = 53.68$, $p < 0.00001$). Thus, these mostly EFL students were performing at a higher level than was average for all students.

A few studies ($n = 7$) focused on children who were Hospitalised or Sick at Home. Some studies focused on children with chronic conditions (e.g., leukaemia) who were hospitalised for a long time, while other studies considered short-term conditions. Six of these (86%) found the children were Better after the intervention, and only one was the Same (14%) (no χ^2). Thus, performance was virtually the same as for all children taken together.

The Gifted coding ($n = 10$) showed eight studies Better (80%) and two the Same (20%) (no χ^2). Thus, the performance for Gifted was a little below the average for all students. Two studies compared gifted and non-gifted students and expressed surprise when the non-gifted did as well as the gifted. This perhaps suggests that 'giftedness' is traditionally defined by the school and pupils with digital competencies may achieve unexpected high performance in that area.

Non-English-Speaking Educational Context (NESEC) was coded whenever the study had taken place in a country which did not have English as its first language (although the study was of course written in the English language). This was intended to show whether different cultural factors might be operating that could influence the success or otherwise of the digital technology. Table 7 shows the countries so identified (in order of frequency of studies and then alphabetically), with the number of studies originating in each. There were 59 countries, 329 papers, 24% of the total number. Clearly, interest in this area occurs all over the world. It seems very likely that there are cultural features that need taking into account when implementing any digital technology (Idris et al., 2016). The size and population of a country (and to an extent its level of economic development) seems to bear no relationship to its research productivity in this area. Interestingly, when we examined the effectiveness of the studies that had taken place in a country which did not have English as its first language, we found that they had higher effectiveness than other studies (Better ratings 293 out of 329, 89%). Only four studies (1%) found Worse outcomes, while Same ($n = 16$) and Unclear ($n = 16$) were 5% each ($\chi^2 = 278.36$, $p < 0.00001$).

Table 8 reports changes in Socio-Emotional Functioning (SEF) as a result of engagement with digital technology. Overall, improvement was reported by 221 studies (87%), whereas only 9 (4%) reported a deterioration. A further small number (19, 7%) found no difference

and six (2%) were unclear ($\chi^2 = 199.94$, $p < 0.00001$). However, a large number of these 221 positive studies were reporting an improvement in self-efficacy, rather than social functioning per se. This suggests that Socio-Emotional Functioning, and in particular self-efficacy, makes an important contribution to effectiveness.

Psychological Well-Being (PWB) was a rarely coded option, only 14 studies (79%) noting an improvement in this area. However, only two studies stayed the same (14%) and one study was coded as Unclear (7%) ($\chi^2 = 9.40$, $p < 0.02$). Thus, PWB does not seem to make as large a contribution to outcomes as SEF, although the numbers are so small that caution is needed in interpretation.

DISCUSSION

Summary

This systematic review included 1355 studies, divided into: Online 134 studies (7%), Blended 232 (13%), CSCL 129 (7%), Games 488 (26%), and CAI 865 (47%) (a few studies were coded as of more than one type). Overall, digital technology was found more effective than regular instruction in 85% of studies, while 8% found it the same and only 3% found it worse. Blended Learning was considerably more effective than Online Learning (83% of studies more effective than regular instruction compared to 74%). Introducing other conditions which generally do not currently take place out of school, CAI was the most effective (91%), a little ahead of Games and CSCL (both 81%).

Reviews and meta-analyses numbered 144 in total: Online 19, Blended 20, Games 56, CSCL 14 and CAI 35. Unfortunately, 81% of these covered primary, secondary and higher education (and even beyond) together, and made no attempt to distinguish between them. Nonetheless, Blended Learning had 95% of review studies Better than regular instruction, CSCL 93%, Games 88%, CAI 83% and Online 72%. Mean Effect Sizes (ESs) (such as were available) suggested CAI was the most effective, followed by Blended Learning, Games and CSCL, followed by Online Learning. While this was in keeping with the overall finding for all separate studies, great caution is needed in interpreting these figures.

Regarding school sector, Primary and Early Years/Kindergarten both were 87% Better than regular instruction, while Secondary/High came next with 80%. Primary + Secondary came next, and last of all Middle Schools with 73%. This latter still suggests three quarters of studies were more effective than regular instruction. Turning to subjects, Science and Maths were the most popular subjects (around 20% of the total each), then in order of popularity Thinking, Reading, English as a Foreign Language (EFL), Humanities, Health, STEM, Writing, English, Arts/Music and Foreign Languages (from 10% to 1%). EFL interventions were the most effective (90%), then Writing and STEM (89%), Thinking (88%), Arts/Music (87%), Humanities (86%), Health and Science (84%), Reading and Maths (82%), Foreign Languages (81%) and English (77%). Thus, the most popular subjects were by no means those where the greatest effectiveness was evident. Regarding gender, Females were Better in 27 cases (39%) and Males Better in only 6 (9%), but Males and Females were equal in 36 cases (52%). Thus, overall, Females did better.

Studies coded as Disadvantaged numbered 46 (3%), HiLo 82 (6%), a variety of Disability 112 (8%), Ethnic Minority status 30 (2%), Rural 34 (3%), Second Language Learners 76 (6%), Hospitalised or Sick at Home 7 (1%), Gifted 10 (1%) and Non-English-Speaking Educational Context 329 (24%). The Disadvantaged subgroup ($n = 46$) had 37 (79%) of children performing Better than regular instruction. This is a slightly worse outcome than for all students combined, but nonetheless suggests that Disadvantaged children can make almost as large gains as other children, when usually they would be expected to perform much more poorly.

TABLE 7 Non-English-speaking countries originating included papers

Rank	Country	Number of included papers
1	Taiwan	67
2	Indonesia	32
3	Turkey	29
4	China	19
5	Hong Kong	14
6	Spain	12
7	Malaysia	12
8	India	11
9	South Korea	10
10	Iran	9
11	Netherlands	8
12	Thailand	8
13	Greece	5
14	Singapore	5
15	Chile	5
16	Israel	4
17	Mexico	4
18	Brazil	4
19	Cyprus	3
20	Czechia	3
21	Nigeria	3
22	Columbia	3
23	France	3
24	Italy	3
25	Philippines	3
26	Saudi Arabia	3
27	Sweden	3
28	Denmark	2
29	Ethiopia	2
30	Finland	2
31	Germany	2
32	Japan	2
33	Kenya	2
34	Kuwait	2
35	Norway	2
36	Oman	2
37	Poland	2
38	Sudan	2
39	Austria	1
40	Belgium	1

TABLE 7 (Continued)

Rank	Country	Number of included papers
41	Costa Rica	1
42	Ecuador	1
43	Ghana	1
44	Hungary	1
45	Iraq	1
46	Jordan	1
47	Malawi	1
48	Morocco	1
49	Nepal	1
50	Palestine	1
51	Peru	1
52	Portugal	1
53	Slovakia	1
54	Slovenia	1
55	South Africa	1
56	Sri Lanka	1
57	Switzerland	1
58	Ukraine	1
59	United Arab Emirates	1

TABLE 8 Summary data for socio-emotional functioning (SEF)

SEF factor	Better	Same	Worse	Unclear
SEF+ ($n = 228$, 90%)	207 (90%)	15 (7%)	2 (1%)	4 (2%)
SEF ($n = 8$, 3%)	6 (75%)	1 (12.5%)	0	1 (12.5%)
SEF− ($n = 19$, 7%)	8 (42%)	3 (16%)	7 (37%)	1 (5%)
Overall ($n = 255$, 100%)	221 (87%)	19 (7%)	9 (4%)	6(2%)

Note: Some papers were coded for more than one school sector. Figures represent mentions, not papers.

HiLo studies were mostly of children perceived as of low ability, and in the vast majority of cases HiLo children performed Better than regular instruction ($n = 76$, 92%), suggesting the digital space allowed them to flourish.

Turning to studies including children with Special Educational Needs (SEN) or Disability, unspecified SEN were 18 (16%), Learning Difficulties 34 (30%), Autistic Spectrum Disorder (ASD) 14 (13%), Attention Deficit and Hyperactivity Disorder (ADHD) 10 (9%), Visual Impairment 7 (6%), Emotional and Behavioural Disorder (EBD) 6 (5%), Deaf and Hard of Hearing 5 (4%), Dyslexia/Specific Learning Difficulties 5 (4%), Down Syndrome 3 (3%) and Writing Difficulty 3 (3%), with other conditions less than 3%. Overall, 89 studies (80%) found Better outcomes, which is not much lower than the average for all students, but this aggregate figure can be disentangled into types of disability. On the positive side, Deaf and Hard of Hearing, Down Syndrome and Writing Difficulty students had 100% Better ratings, albeit with very modest numbers of studies (3–5). ADHD students had a 90% Better rating

(10 studies), followed closely by Learning Difficulty (87%) (with the highest number of papers—30) and ASD (86%; 14 studies) students. All of these were above the average for all students.

Regarding Ethnic Minorities (EM), 25 (78%) of studies reported Better performance, five (16%) the Same and two (6%) Worse. This was similar to the performance of Disadvantaged students, indicating that EM children can make almost as large gains as other children, when usually some minorities would be expected to perform more poorly. Regarding Rural schools, 29 (85%) found that digital technology was Better than regular instruction, thus appearing to perform as well as the average for all schools. The Second Language Learner (SLL) code mainly encompassed EFL learners, who showed high levels of performance, 69 (90%) Better than regular instruction.

Regarding Hospitalised or Sick at Home studies, six of these (86%) found the children were Better after the intervention, and only one was the Same (14%). Thus, performance was virtually the same as for all students taken together. Regarding Gifted studies ($n = 10$), eight studies were Better (80%) and two the Same (20%). Thus, the performance for Gifted was a little below the average for all students. Regarding Non-English-Speaking Educational Context (NESEC), these studies had higher effectiveness than other studies (Better ratings for 293 out of 329, 89%).

Socio-Emotional Functioning (SEF) as a result of engagement with digital technology (mostly focused on self-efficacy) was reported by 225 studies, and 91% (207) of these were Better than regular instruction. Regarding Psychological Well-Being (PWB) 14 studies out of 17 (79%) noted an improvement. Thus, PWB does not seem to make as large a contribution to outcomes as SEF, although the numbers are so small that caution is needed in interpretation.

A large number of these studies were noted as having implications for Teaching, so we broke them down by intervention and readers will find them in Appendix S2.

The original contribution of this study was that it: synthesised pre-pandemic as well as pandemic literature but did not target emergency remote learning, included far more studies in schools than previous reviews, specifically attributed outcomes to sectors of education, explored effectiveness in relation to gender, commented on whether children perceived as 'low ability' in class did better or worse in the online/blended environment, explored the impact of self-efficacy in learning outcomes, commented on parallel digital technologies (Educational Games, CSCL and CAI) which were not widely used outside of school, and reported outcomes in terms of whether the intervention was better, worse or the same in relation to regular instruction.

Limitations and strengths

We think that the broad terms of the search, the large number and variety of databases used, and the inclusion of both quantitative and qualitative studies are positive features of this paper. As the PRISMA chart shows, different databases tended to yield very different studies, and so we had a wide range of studies. However, the issue of publication bias must be considered. Authors tend to submit and editors tend to publish papers that report positive or significant results (Dwan et al., 2013), so the published papers here might not be a representative sample of all research that was done, let alone all practice in the field. They may give a somewhat unrealistically positive picture of effectiveness. We included doctoral theses and conference papers, which were not peer reviewed in the normal way and more often gave weaker results. However, we excluded books, chapters without an abstract and any forms of 'grey' literature.

We did not search for any studies dated prior to 2000, although we accept that there are examples of interesting and substantial work which were published before that. We did not insert any search terms specifically designed to extricate CAI interventions, but nonetheless CAI was the largest category of interventions. Additionally, Subjects were grouped to conform to a relatively small number of categories, which sometimes included great variety, but was necessary on the grounds of manageability.

We accept that the categorisation of studies into those which yielded a Better, Same, Worse or Unclear effect than regular instruction (or other comparison group) is crude (although the criterion of two-thirds of results needing to be Better to allow categorisation as Better somewhat ameliorates this). Some readers of a positivist persuasion might wish for some indication of the degree to which interventions were 'Better' or 'Worse', supported by tests of statistical significance, but statistical significance is largely a function of sample size and in any event some papers were purely qualitative. Nonetheless, studies containing multiple contradictory findings were multiply coded, and the inter-rater reliability was very high, so at least there was consistency between coders.

We intended to investigate the correlation between study sample size and mean Effect Size (ES) where available, to see whether small studies were more likely to produce small ESs, that is, that publication bias was not militating against small studies. However, of the Effect Sizes available, 81% came from studies involving higher education students as well as school students, and it proved impossible to conduct this analysis. For the same reason, we could not employ measures such as Orwin's (1983) failsafe N to investigate how many missing studies would be needed to bring the ESs to the $p = 0.05$ level.

There is also a question regarding interpretation of results. For Online and Blended Learning, which are intended as substitutes for regular classroom instruction, a 'Same' outcome is positive (although not as positive as a 'Better' outcome). However, this is not true of Games, CAI or CSCL, which so far are usually an addition to the normal classroom curriculum. In these cases, 'Same' would mean the outcome was just the same as normal classroom instruction, but the teacher would have incurred expense in purchasing the technology and time in learning how to use and manage it, so 'Same' would not be a positive outcome.

Finally, we note that research studies are typically moderately well organised, so they may not be typical of what has recently happened in schools—a requirement to deliver support to pupils via online or blended learning with little notice and time to prepare. Particularly during remote emergency learning, implementation might have been of uncertain quality in some schools and from some teachers, which might make these practices less related to the research outcomes (Karma et al., 2021).

Relationship to previous literature

The literature discussed in the Introduction about emergency remote learning is very interesting but not entirely pertinent to our search, which is more about online/blended learning before and after the pandemic. In the literature review we discussed the sources which led to this review (e.g., the EEF [2020b] rapid research review, which included many fewer reviews than the present study), and note that they did not comment on the relative effectiveness of online and blended learning, although both were found effective. Then we discussed the additional reviews missed by our search, but these were much the same as the papers already reviewed. Then we discussed theoretical perspectives (as has been called for many times but never done), noting that most pedagogical theories applied equally to face-to-face and online/blended learning (e.g., behaviourism, cognitivism, constructivism, situated learning, active learning and peer interaction). Self-efficacy and self-determination theory were

motivational theories relevant to learning, but again they related to both face-to-face and digital learning.

The reader may wonder: were we not able to code for these pedagogical elements, in order to explicate not only the nature of the digital technology and its effects, but also the pedagogical underpinnings of the intervention? In fact, we tried and were not successful. So many papers failed to report on all the pedagogical elements and others described them in so many different ways that the inter-rater reliability of attempted coding was so low as to preclude the inclusion of this aspect. It seemed that many papers were authored by digital enthusiasts who were less enthusiastic about pedagogy.

Although constructivist pedagogical strategies are probably more fashionable in the English-language literature at the moment (Hung et al., 2004; Wilson, 1997), they tend to be more time-consuming, so teachers may be unlikely to use them for all aspects of teaching. In any case, behavioural strategies may be more efficient for knowledge transfer and cognitive strategies more efficient for learning processes and principles, which is why Janicki and Liegle (2001) identified all these components in everyday teaching. We would expect all to feature in digital learning, although different countries would doubtless have different balances between these aspects.

Multi-media input may be easier through computers, but is perfectly possible in regular instruction, so the theorists advocating both visual and auditory processing (Krashen, 1981; Mayer et al., 2014; Paivio, 1971; Prince, 2004) are not actually saying anything specifically about digital technology.

One might assume that online and blended learning would have adverse social effects, since separating children from their friends as well as opportunities for direct interaction with the teacher and all other students seems intuitively likely to be damaging. However, when we consider the role of peer interaction and social embeddedness (Fisher et al., 2004; Fisher & Naumer, 2006; Garrison et al., 2001; Garrison, 2000; Long, 1981; Vygotsky, 1962), although these are seen as important in some digital technology studies, they are also found in face-to-face classrooms. Nonetheless, those studies which addressed social and emotional learning and psychological well-being tended to report a positive effect.

Turning to the impact of self-efficacy or self-belief (Bandura, 1994; Busch, 1995; Deci & Ryan, 2008; Krashen, 1982), again we find these are relevant to face-to-face as well as digital learning, and in both cases may be somewhat context-specific, varying over time as the student accommodates to new and challenging tasks. However, we have relevant empirical results here. As noted above, SEF studies (mostly focused on self-efficacy) showed high effectiveness compared to being in school full-time, 91% Better than regular instruction. Thus, we confirmed the heightened importance of Self-efficacy in a digital environment.

Overall, however, it is unsurprising that Clark (1983) argues that the instructional strategy is important, not the medium of instruction, although Kozma (2001) argues that computers may be essential in simulations. Nonetheless, we have found some evidence that social and emotional learning and self-efficacy are even more important in digital learning.

The Connectivism of Siemens (2004) appears the only theory solely addressing digital technology, emphasising constant unlearning and relearning as ‘knowledge’ changes, with ‘knowledge’ coming from many sources, not just school, making thinking skills much more important than content knowledge. Connectivism emphasises flexible learning, especially asynchronous learning, with the learner much more independent, undertaking learning tasks in any order, adding new materials not prescribed by the school, and so forth. However, this description of a more autonomous and self-regulated learner may not apply to all learners—some might remain unmotivated and be more prone to procrastination. Thus, while the other principles are pedagogical principles which apply equally to face-to-face and digital learning (and remain important), Connectivism yields new insights into the future digital world. Nonetheless, given the general lack of theory in the papers reviewed, we were not

able to identify any individual paper as providing proof of Connectivism, nor directly relate any of the papers in this review to it.

Further critical analysis

While we have found that online and blended learning are effective, and often more effective than regular instruction, both of them beg the question of degree of parental or other carer supervision at home. In countries where extended families are not common, and especially where both parents need to work and therefore be absent from the house, the child-minding function of school becomes much more significant. Obviously with young children there are legal issues to consider, but even with older children the question of how on-task they remain when in the house on their own is another issue. When online or blended work done at home is assessed quickly, the school can see which students are failing to keep on task and perform, and can consider requiring that they attend school in order to complete their online or blended tasks in an environment where there is a degree of supervision. This might be particularly relevant for disadvantaged and/or procrastinating pupils, some of whom might have difficulty accessing computers and the internet at home. Of course, the digital task presented might not be well matched to the student's capability, which could also be alleviated if the student were in school. Once the student is performing more satisfactorily, they might be allowed to resume studies at home with appropriate digital support. Thus, the student can see that working at home is a facility that can be withdrawn if performance does not keep up to standard.

Ideally, online and blended learning should be available on a variety of devices to improve access, including smart phones and games consoles as well as desktop/laptop computers and tablets requiring an internet connection. However, some applications do not run on all platforms. One of the advantages of games is that they often do not require an internet connection, but for example a spreadsheet task presents particular problems when viewed on a smartphone, and the device needs to suit the software being deployed. However, multiple device platforms should be considered.

While we had anticipated a high number of studies in CAI, we were surprised by the very large number of studies on Games. This is clearly an area of great productivity. However, schools might be using Games in class, but using them out of class is another step which is less common. There is a time allocation tension—simple Games might be discarded by teachers as being insufficiently educational, while complex Games might be discarded if teachers cannot find the time to learn them, even if just in order to assess their suitability. This is also true of other digital applications.

Worryingly, some studies of purely virtual schools in the USA (schools teaching only online which children can opt to ‘attend’ rather than traditional schools) have reported that achievement in such schools was below that in traditional schools (Ahn & McEachin, 2017; Miron & Gulosino, 2016; Miron et al., 2018; Mislevy et al., 2020; Poelmans et al., 2018). This could be related to the fact that disadvantaged Caucasian students were over-represented in such schools. This contraindication needs to be further investigated and corrected if possible.

We believe that digital technology needs to enshrine sound pedagogical principles if it is to succeed long term, otherwise short-term studies are only likely to be measuring a novelty effect, even in relation to control groups. Teachers are best placed to ensure that this happens, but this will require self-confidence or self-efficacy—important for teachers as well as for students, and for both hopefully increasing over time.

Implications for future practice, policy and research

Practice

We hope that teachers will be further motivated to access the Teaching Points online (Appendix S2), and/or be guided to these by other professionals. We also hope that teachers' experience of online and blended learning acquired during the lockdowns and school closures will be used productively as schools move forward. Even though marshalled at high speed without the benefit of much planning time, these innovations do indicate a new way of working for schools. Post-pandemic, while wholly online learning is probably not relevant for most pupils except those in remote areas, blended learning certainly offers promise. A system of accessing learning at home during the morning with activities and discussion relating to that learning at school in the afternoon is certainly one schools might wish to experiment with. This could also help to increase engagement and performance in homework activities and support young people on study leave preparing for exams. We hope that our view that the pedagogy inherent in digital technology (with which teachers should be more familiar) is probably more important than the technology itself will be reassuring to teachers.

Teachers might also be reassured if future emergencies require similar measures. Of course, there may be partial lockdowns which require only a limited number of children to be in school, and blended learning would be ideal in that instance. Equally, online learning may be good for some sub-groups of students but not others, and this needs to be made clear also. Post-pandemic, skills developed in delivering online and blended learning should be sustained rather than lost.

Policy

Local and national government needs to develop local and national policies for online and particularly blended learning from schools. This does not mean that local and national government should be delivering continuing professional development (CPD) themselves (especially compulsory CPD), as they might tend to be rather slow and ponderous in what is a rapidly changing world. Nor should they attempt to fully quality control and certify CPD offerings, for the same reason. There are already many separate platforms offering online material, some requiring payment and some not, but some of these are only accessible to teachers. National government could usefully create a dashboard which gives access to all these sites, giving a brief description (free of marketing exaggeration) of each and suitability for age ranges. Indeed, the dashboard could easily add simple learning analytics, including the strategic clustering of learner ages and types and discrimination of novice and experienced users, so that users could see which applications were the most popular in relation to their own situation (Matcha et al., 2020). Those sites which are only accessible to teachers should seriously consider making themselves freely available to parents, to widen student choice and also address the needs of permanent home schoolers.

Governmental thinking needs to be better informed by the research base. A national policy is desirable which is actually implemented in practice. Certainly, a great deal more teacher professional development on digital technology is needed, which could be partly online and partly face-to-face. Our results suggest that more CPD is particularly needed in how to deliver CSCL, Games and CAI outside of the school environment, given that there is already expertise on using them within school. In terms of sectors, the Primary sector seems to need this CPD least, the Secondary sector needing it somewhat more and Middle schools needing it most. CPD addressing more than one sector does not seem particularly effective. Science and mathematics are the most popular areas for digital interventions, but

effectiveness does not follow popularity, so science and mathematics need more CPD work to raise their effectiveness.

There is some evidence that females are better than males in digital interventions, and both genders should be made aware of this in CPD to dispel any previous notions to the contrary. There was evidence that disadvantaged children and ethnic minority pupils could make as nearly as large gains as advantaged pupils, but this needs to be improved so that disadvantaged pupils make at least the same gains as advantaged pupils, otherwise the gap will widen. CPD should also focus on ways of making devices and the internet available to pupils from disadvantaged homes, since merely supplying these in the short term will not be sufficient. CPD should heighten awareness that pupils traditionally perceived as 'low-ability' tend to do better with digital interventions once they have access to devices and the internet.

Regarding Special Needs pupils, Deaf and Hard of Hearing, Down Syndrome, Writing Difficulty, Attention Deficit and Hyperactivity Disorder (ADHD), Learning Difficulty and Autistic Spectrum Disorder (ASD) pupils did well, but more studies are needed to confirm this. CPD should focus more on the fields which performed below average, that is, Emotional and Behavioural Difficulties, Visual Impairment, Dyslexia/Specific Learning Difficulty and Gifted (although again more studies are needed to confirm this). There is evidence that self-efficacy is an important feature of successful digital interventions, so CPD should have raising the self-efficacy of pupils (and teachers) as an important objective.

It may be that teacher unions will be concerned by the suggested shift in teacher practice, but there is no suggestion that fewer teachers will be needed; if anything, more (but better trained) teachers will be needed.

Research

This analysis of 1355 full papers proved to be time-consuming, but we would urge that future attempts take the trouble to encompass all five areas, since they are obviously interlinked. Our definitions and coding system have proved reliable and future research may seek to adopt them. Equally obviously, this systematic review will need updating as time goes by. We hope that future research will acknowledge the difficulties we have encountered in undertaking this review and seek to find ways of resolving them. The heterogeneity of effect sizes and their lack of attribution to specific school sectors is an example of such a difficulty. Some system of analysis of study quality is very necessary, but such quality analyses are very various and the choice must be both detailed and informative, but also practically feasible and reliable. Researchers must guard against the Hawthorne effect and longer-term follow-up of effectiveness needs to feature in many more studies, together with more evidence on implementation integrity.

Further research is needed into the balance between national and local government delivered CPD (which might be compulsory) and that stemming from individual teachers or groups of teachers from one or many schools investigating a range of public, freeware and private CPD opportunities and making their own decisions about what to pursue and implement in their classrooms. The first option might have coherence and leadership, while the second might be more up-to-date and relevant to classroom practice. Our results suggest that more research on how to deliver CSCL, games and CAI outside of the school environment is also sorely needed. Why middle schools perform so poorly is in need of investigation, as is why interventions which attempt to bridge sectors tend to do poorly. Science and mathematics are popular areas but less effective, and research should explore why—perhaps science and mathematics interventions tend to have weaker underlying pedagogy.

The gender disparity merits further investigation, and this should be coupled with investigation of self-efficacy, since the two seem likely to be connected. Disadvantaged and

ethnic minority pupils do better than one might expect, but these labels are not explanatory—it would be helpful to detail what features of them are causally linked to performance in digital interventions. Research should also investigate the longevity of devices and internet connections made available to disadvantaged pupils, with associated exploration of the full range of multiple uses to which the devices are put when made available. Research might also investigate the difference in perceptions of traditionally 'low-ability' pupils (as defined by the school), their successful performance on digital interventions, and the implications that might have for pupil self-efficacy and adjusting school perceptions. Research might also explore changing patterns of teacher self-efficacy and resistances to deploying digital interventions—there tends to be a good deal of focus on what motivated teachers can do, but there also needs to be investigation of less-motivated and/or conservative teachers, who form a significant part of the workforce. Finally, in the special needs area, emotional and behavioural difficulties, visual impairment, dyslexia/specific learning difficulty and gifted studies were all weaker, and research should seek to investigate why this is and how it may be remedied. Future research work should also consider an analysis of types of pedagogy underlying the different types of digital technology. We urge authors of individual studies to describe their intervention in sufficient detail as to make it replicable.

CONCLUSIONS

Overall, 85% of studies showed mostly positive effects and only 3% found it worse, but of course simple vote counting counts of studies can be misleading (although our procedure was a little more complex than simple vote counting). Nonetheless, this difference is so large that digital technology does appear to have positive effects. The review studies largely supported the overall analysis, which adds weight to this finding. Blended learning was more effective than online learning, although both were effective but to different degrees, and online learning may be the only option for students remote from school and those unable to attend school owing to pandemics or other causes. CAI was more effective than either, but much of CAI was in school, and it cannot be assumed that it would be as effective if delivered to homes. The effectiveness of games and CSCL fell somewhere between blended and online learning, but again much of this was in school, and it cannot be assumed that it would be as effective if delivered elsewhere.

In relation to the research questions:

RQ1: Does online learning from schools have positive effects compared to full-time regular instruction in school? The answer is 'yes'.
RQ2: Does blended learning from schools have positive effects compared to full-time regular instruction in school? The answer is 'yes', but blended learning seems to have stronger effects than online learning.
RQ3: Do educational games have positive effects compared to full-time regular instruction in school? The answer is 'yes'; more effective than online learning but less effective than blended learning, but much of this was in schools rather than elsewhere.
RQ4: Does computer-supported collaborative learning have positive effects compared to full-time regular instruction in school? The answer is 'yes'; more effective than online learning but less effective than blended learning, and less effective than games, but much of this was in schools rather than elsewhere.

Primary and early years/kindergarten showed the most positive effects, followed by secondary/high schools. Middle schools did not seem as effective. A wide range of subjects shows positive results, not only reading, maths and science/STEM, but also English,

writing, English as foreign language, critical thinking, humanities, art and music, and health. Popularity of subject and effectiveness were not the same. Effects were different by gender; girls generally did better than boys. Studies of 'low ability' students and second language learners (mainly EFL) showed high levels of performance, greater than all studies combined. Studies of rural students and hospitalised or sick at home students showed equal effectiveness to all studies combined. Studies of disadvantaged, gifted or SEN or disabled students showed effects slightly lower than average for all studies. Within SEN or disability, studies of deaf and hard of hearing, Down Syndrome, writing difficulty, ADHD and ASD students showed greater effectiveness than was average for all students. Studies of ethnic minority students gave similar results. Students in Non-English-Speaking Educational Contexts (NESEC) showed high effectiveness (well above the average for all students), even though the 59 countries they lived in might have had cultural issues in relation to digital learning. Socio-emotional functioning studies (mostly focused on self-efficacy) showed high effectiveness. Psychological well-being studies were much fewer but 79% noted an improvement.

Beyond these questions, the issue of the relationship between types of pedagogy and types of digital technology arose. Although we found it impossible to analyse this relationship empirically, we noted that many pedagogical features of face-to-face learning could also be found in digital technology. Digital technology's main advantage may be the possibility for enhanced task flexibility and learner autonomy, encouraging greater self-regulation and the following of the student's own patterns of motivation. However, despite the finding that blended and online learning are generally more effective that face-to-face instruction, this may not be an advantage for all. There is a great deal of information on teaching points for effective online/blended learning which can be drawn from evidence-based studies, and these are summarised in Appendix S2.

ACKNOWLEDGEMENT

The report which was the mainspring for the development of this paper (which includes design principles not in the present paper) can be downloaded from: https://discovery.dundee.ac.uk/en/publications/the-effectiveness-of-online-and-blended-learning-from-schoos-a-sy.

ETHICS STATEMENT

This research did not access any raw data. The reviews on which it was based aggregated studies which had already received ethical approval. Consequently no additional ethical approval was necessary.

CONFLICT OF INTEREST

The results of this study do not create a conflict of interest for the authors.

DATA AVAILABILITY STATEMENT

The systematically analysed papers are available from the corresponding author on request.

ORCID

Keith J. Topping https://orcid.org/0000-0002-0589-6796

REFERENCES

Ahn, J., & McEachin, A. (2017). Student enrollment patterns and achievement in Ohio's online charter schools. *Educational Researcher*, *46*(1), 44–57.

Alexander, P. A. (2020). Methodological guidance paper: The art and science of quality systematic reviews. *Review of Educational Research*, *90*(1), 6–23. https://doi.org/10.3102/0034654319854352
Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (vol. *4*, pp. 71–81). Academic Press.
Barbour, M. K., Labonte, R., Kelly, K., Hodges, C., Moore, S., Lockee, B., & Trust, T. (2020). *Understanding pandemic pedagogy: Differences between emergency remote, remote, and online teaching*. Canadian eLearning Network. https://doi.org/10.13140/RG.2.2.31848.70401
Berson, I. R., Luo, W. W., & Yang, W. P. (2022). Special issue: Narrowing the digital divide in early childhood: Technological advances and curriculum reforms. *Early Education and Development*, *33*(1), 183–185. https://doi.org/10.1080/10409289.2022.1989740
Bond, M. (2020a). Schools and emergency remote education during the COVID-19 pandemic: A living rapid systematic review. *Asian Journal of Distance Education*, *15*(2), 191–247.
Bond, M. (2020b). Facilitating student engagement through the flipped learning approach in K-12: A systematic review. *Computers & Education*, *151*, 1–36. https://doi.org/10.1016/j.compedu.2020.103819
Bond, M., Bergdahl, N., Mendizabal-Espinosa, R., Kneale, D., Bolan, F., Hull, P., & Ramadani, F. (2021). *Global emergency remote education in secondary schools during the COVID-19 pandemic: A systematic review*. EPPI Centre, University College London. https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3847
Busch, T. (1995). Gender differences in self-efficacy and attitudes toward computers. *Journal of Educational Computing Research*, *12*(2), 147–158. https://doi.org/10.2190/H7E1-XMM7-GU9B-3HWR
Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2018). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Etr&D – Educational Technology Research and Development*, *44*(3), 176. https://doi.org/10.1007/s11423-018-9633-7
Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, *53*(4), 445–459. https://doi.org/10.3102/00346543053004445
Cooper, P. A. (1993). Paradigm shifts in designing instruction: From behaviorism to cognitivism to constructivism. *Educational Technology*, *33*(5), 12–19.
Crompton, H., Burke, D., Jordan, K., & Wilson, S. W. G. (2021). Learning with technology during emergencies: A systematic review of K-12 education. *British Journal of Educational Technology*, *52*(4), 1554–1575. https://doi.org/10.1111/bjet.13114
Darmody, M., Smyth, E., & Russell, H. (2020). *The implications of the COVID-19 pandemic for policy in relation to children and young people: A research review*. Economic and Social Research Institute. https://doi.org/10.26504/sustat94
Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology/Psychologie Canadienne*, *49*(3), 182–185. https://doi.org/10.1037/a0012801
Drane, C. F., Vernon, L., & O'Shea, S. (2020). Vulnerable learners in the age of COVID-19: A scoping review. *Australian Educational Researcher*, *48*(4), 585–604. https://doi.org/10.1007/s13384-020-00409-5
du Plessis, P. (2014). Problems and complexities in rural schools: Challenges of education and social development. *Mediterranean Journal of Social Sciences*, *5*(20), 1109.
Dwan, K., Gamble, C., Williamson, P. R., Kirkham, J. J., & for the Reporting Bias Group (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias: An updated review. *PLoS One*, *8*(7), e66844. https://doi.org/10.1371/journal.pone.0066844
Education Endowment Foundation (2020a). *Rapid evidence assessment: Impact of school closures on the attainment gap*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/school-closures-rapid-evidence-assessment
Educational Endowment Foundation (2020b). *Remote learning: Rapid evidence assessment*. Education Endowment Foundation. https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/remote-learning-for-pupils
Feenberg, A. (2019). The internet as network, world, co-construction, and mode of governance. *The Information Society*, *35*(4), 229–243. https://doi.org/10.1080/01972243.2019.1617211
Fisher, K. E., Durrance, J. C., & Hinton, M. B. (2004). Information grounds and the use of need-based services by immigrants in Queens, New York: A context-based, outcome evaluation approach. *Journal of the American Society for Information Science and Technology*, *55*(8), 754–766. https://doi.org/10.1002/asi.20019
Fisher, K. E., & Naumer, C. M. (2006). Information grounds: Theoretical basis and empirical findings on information flow in social settings. In A. Spink, & C. Cole (Eds.), *New directions in human information behavior* (pp. 93–111). Springer.
Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23. https://doi.org/10.1080/08923640109527071
Garrison, R. (2000). Theoretical challenges for distance education in the 21st century: A shift from structural to transactional issues. *International Review of Research in Open and Distance Learning*, *1*(1), 1–17. http://www.icaap.org/iuicode?149.1.1.2

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., & Schünemann, H. J. (2011). GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, *64*(4), 383–394.

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). *The difference between emergency remote teaching and online learning*. EDUCAUSE. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Hung, D., Looi, C. K., & Koh, T. S. (2004). Situated cognition and communities of practice: First-person 'lived experiences' vs. third-person perspectives. *Educational Technology & Society*, *7*(4), 193–200.

Idris, M. Z., Mustaffa, N. B., & Yusoff, S. O. S. (2016). Preservation of intangible cultural heritage using advance digital technology: Issues and challenges. *Harmonia: Journal of Arts Research and Education*, *16*(1), 1–13. https://doi.org/10.15294/harmonia.v16i1.6353

Janicki, T., & Liegle, J. O. (2001). Development and evaluation of a framework for creating web-based learning modules: A pedagogical and systems approach. *Journal of Asynchronous Learning Networks*, *5*(1), 58–84. http://www.sloan-c.org/publications/jaln/v5n1/pdf/v5n1_janicki.pdf

Kanematsu, H., Fukumura, Y., Ogawa, N., Okuda, A., Taguchi, R., Nagai, H., & Barry, D. M. (2009). Problem based learning in Metaverse as a digitized synchronous type learning. In *Proceedings of ICEE – iCEER-2009, Seoul, South Korea, August 23–28*.

Karma, I. G. M., Darma, I. K., & Santiana, I. M. A. (2021). Blended learning as an educational innovation and solution during the covid-19 pandemic. *International Research Journal of Engineering, IT & Scientific Research*, *7*(1), 1–9. https://doi.org/10.21744/irjeis.v7n1.1176

Kozma, R. B. (2001). Counterpoint theory of 'learning with media'. In R. E. Clark (Ed.), *Learning from media: Arguments, analysis, and evidence* (pp. 137–178). Information Age Publishing Inc.

Krashen, S. D. (1981). *Bilingual education and second language acquisition theory. Schooling and language minority students: A theoretical framework* (pp. 51–79). Evaluation, Dissemination and Assessment Center, California State University.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Prentice-Hall International.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of Covid-19 school closures on academic achievement. *Educational Researcher*, *49*(8), 549–565. https://doi.org/10.33102/0013189X20965918

Lam, C. M. (Ed.) (2020). *Philosophy for children in Confucian societies in theory and practice*. Routledge.

Lipponen, L. (2002). Exploring foundations for computer-supported collaborative learning. Computer Support for Collaborative Learning. In *Proceedings of CSCL 2002, January 7-12, Boulder, Colorado* (pp. 72–81). http://www.helsinki.fi/science/networkedlearning/texts/lipponen2002.pdf

Liu, S., Liu, Y., & Liu, Y. (2020). Somatic symptoms and concern regarding COVID-19 among Chinese college and primary school students: A cross-sectional survey. *Psychiatry Research*, *289*, 113070. https://doi.org/10.1016/j.psychres.2020.113070

Long, M. H. (1981). Input, interaction, and second-language acquisition. *Annals of the New York Academy of Sciences*, *379*, 259–278. https://doi.org/10.1111/j.1749-6632.1981.tb42014.x

Lundin, M., Bergviken Rensfeldt, A., Hillman, T., Lantz-Andersson, A., & Peterson, L. (2018). Higher education dominance and siloed knowledge: A systematic review of flipped classroom research. *International Journal of Educational Technology in Higher Education*, *15*(1), 1. https://doi.org/10.1186/s41239-018-0101-6

Maity, S., Sahu, T. N., & Sen, N. (2020). Panoramic view of digital education in COVID-19: A new explored avenue. *Review of Education*, *9*(2), 405–423. https://doi.org/10.1002/rev3.3250

Major, C. H., & Savin-Baden, M. (2011). Integration of qualitative evidence: Towards construction of academic knowledge in social science and professional fields. *Qualitative Research*, *11*(6), 645–663. https://doi.org/10.1177/1468794111413367

Manesis, D. (2020). Digital games in primary education. In I. Deliyannis (Ed.), *Game design and intelligent interaction*. IntechOpen.

Matcha, W., Uzir, N. A., Gasevic, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies*, *13*(2), 226–245. https://doi.org/10.1109/TLT.2019.2916802

Mayer, R. E., Lee, H. J., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*, *28*(5), 653–660.

Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when and how*. Routledge.

Miron, G., & Gulosino, C. (2016). *Virtual schools report 2016: Directory and performance review*. National Education Policy Center. ERIC Number: ED574701.

Miron, G., Shank, C., & Davidson, C. (2018). *Full-time virtual and blended schools: Enrollment, student characteristics, and performance*. National Education Policy Centre. ERIC Number: ED591990.

Mislevy, J., Schmidt, R., Puma, M., Ezekoye, A., & Saucedo, D. (2020). *Comparing the achievement of students in Virtual Virginia and face-to-face courses*. Regional Educational Laboratory Appalachia. https://eric.ed.gov/?id=ED602892

Nearchou, F., Flinn, C., Niland, R., Subramaniam, S. S., & Hennessy, E. (2020). Exploring the impact of COVID-19 on mental health outcomes in children and adolescents: A systematic review. *International Journal of Environmental Research & Public Health*, *17*(22), 8479. https://doi.org/10.3390/ijerph17228479

Organisation for Economic Co-operation and Development. (2020). *Education and COVID-19: Focusing on the long-term impact of school closures*. OECD. https://read.oecd-ilibrary.org/view/?ref=135_135187-1piyg9kc7w&title=Education-and-COVID-19-Focusingon-the-long-term-impact-of-school-closures&_ga=2.62634668.1411069064.1627572200-1309661371162757220 0

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*(2), 157–159. https://doi.org/10.2307/1164923

Paivio, A. (1971). *Imagery and cognitive processes*. Holt, Rinehart & Winston.

Pérez-Sanagustín, M., Nussbaum, M., Hilliger, I., Alario-Hoyos, C., Heller, R. S., Twining, P., & Tsai, C.-C. (2017). Research on ICT in K-12 schools – A review of experimental and survey-based studies in computers & education 2011 to 2015. *Computers & Education*, *104*, A1–A15. https://doi.org/10.1016/j.compedu.2016.09.006

Philipsen, B., Tondeur, J., Pareja Roblin, N., Vanslambrouck, S., & Zhu, C. (2019). Improving teacher professional development for online and blended learning: A systematic meta-aggregative review. *Etr&D – Educational Technology Research and Development*, *67*(5), 1145–1174. https://doi.org/10.1007/s11423-019-09645-8

Picciano, A. G. (2017). Theories and frameworks for online education: Seeking an integrated model. *Online Learning*, *21*(3), 166–190. https://doi.org/10.24059/olj.v21i3.1225

Poelmans, S., Goeman, K., & Wautelet, Y. (2018). Net benefits of face-to-face versus online instruction at school: A repetitive factorial experiment in an ecological setting. In *Paper presented at the International Association for Development of the Information Society (IADIS) International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA) (15th, Budapest, Hungary, Oct 21-23, 2018)*. https://eric.ed.gov/?id=ED600592

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Root, J. R., Stevenson, B., & Davis, L. L. (2018). Computer-assisted instruction to teach academic skills. In F. Volkmar (Ed.), *Encyclopedia of autism spectrum disorders*. Springer.

Schindler, L. A., Burkholder, G. J., Morad, O. A., & Marsh, C. (2017). Computer-based technology and student engagement: A critical review of the literature. *International Journal of Educational Technology in Higher Education*, *14*(1), 253. https://doi.org/10.1186/s41239-017-0063-0

Short, C. R., Graham, C. R., Holmes, T., Oviatt, L., & Bateman, H. (2021). Preparing teachers to teach in K-12 blended environments: A systematic mapping review of research trends, impact, and themes. *TechTrends: for Leaders, Education & Training*, *65*(6), 993–1009. https://doi.org/10.1007/s11528-021-00626-4

Siemens, G. (2004). *Connectivism: A learning theory for the digital age*. ElearnSpace.Org. http://www.elearnspace.org/Articles/connectivism.htm

Studies of Distance Learning (2020). *Washington, DC: What works clearinghouse, institute of education sciences*. U.S. Department of Education. https://ies.ed.gov/ncee/wwc/distancelearningstudy

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, *81*(1), 4–28. https://doi.org/10.3102/0034654310393361

UNESCO (2020). *COVID-19 educational disruption and response*. UNESCO. https://en.unesco.org/covid19/educationresponse

UNESCO-UNEVOC International Centre (2022). *TVETipedia glossary*. UNESCO-UNEVOC. https://unevoc.unesco.org/home/TVETipedia+Glossary/filt=all/id=706#:~:text=Traditional%20instruction%20%E2%80%93%20a%20structured%20education,age%2C%20and%20possibly%20also%20ability

van Alten, D. C., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, *28*, 100281. https://doi.org/10.1016/j.edurev.2019.05.003

Vincent-Lancrin, S., C. C. Romaní, & F. Reimers (Eds.) (2022). *How learning continued during the Covid-19 pandemic: Global lessons from initiatives to support learners and teachers*. OECD. https://doi.org/10.1787/bbeca162-en

Viner, R. M., Russell, S. J., Croker, H., Packer, J., Ward, J., Stansfield, C., Mytton, O., Bonell, C., & Booy, R. (2020). School closure and management practices during coronavirus outbreaks including COVID-19: A rapid systematic review. *The Lancet Child & Adolescent Health*, *4*(5), 397–404. https://doi.org/10.1016/S2352-4642(20)30095-X

Vygotsky, L. S. (1962). *Thought and language* (translated by A. Kozulin). MIT Press.

Weinstein, E. C., & Selman, R. L. (2014). Digital stress: Adolescents' personal accounts. *New Media and Society*, *18*(3), 391–409. https://doi.org/10.1177/1461444814543989

Wilson, B. G. (1997). Reflections on constructivism and instructional design. In C. R. Dills, & A. J. Romiszowski (Eds.), *Instructional development paradigms* (pp. 63–80). Educational Technology Publications.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482. https://doi.org/10.1002/cne.920180503

Zeng, X. T., Zhang, Y. G., Kwong, J. S. W., Zhang, C., Li, S., Sun, F., Niu, Y. M., & Du, L. (2015). The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: A systematic review. *Journal of Evidence-Based Medicine*, *8*, 2–10. https://doi.org/10.1111/jebm.12141

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Topping, K. J., Douglas, W., Robertson, D., & Ferguson, N. (2022). Effectiveness of online and blended learning from schools: A systematic review. *Review of Education*, *10*, e3353. https://doi.org/10.1002/rev3.3353