Mirazchiyski and Gershteyn
Large-scale Assessments in Education (2024) 12:33
https://doi.org/10.1186/s40536-024-00224-9

Large-scale Assessments in Education

RESEARCH Open Access

Technology factors related to the differences in paper and online reading scores in PIRLS 2016

Plamen V. Mirazchiyski[1*] and Vadim Gershteyn[2]

*Correspondence: plamen.mirazchiyski@pei.si

[1] Educational Research Institute, Ljubljana, Slovenia
[2] Humanities, Alma Mater Europea University, Maribor, Slovenia

Abstract

The Progress in International Reading Literacy Study (PIRLS) was conducted in paper and online reading modes in 2016 using the same samples of students in a number of countries. Differences in reading literacy scores were found in several European countries. In some countries, the differences favored the electronic reading mode. Yet in others, the paper reading mode was favored. As the electronic reading mode differs substantially in the cognitive demands compared to the paper mode, it can be expected that the differences between the two modes are related to the variables related to technology: availability and access, general use, use for educational purposes in class or out-of-school, and self-efficacy with technology. This study investigates the Information and Communication Technology (ICT) factors related to the differences in paper and online reading in six European countries participating in both modes in PIRLS 2016. This study uses linear regression models as the application of multilevel modeling is not suitable because of the low between-school variances across countries. The results from this study show limited support for the relative effect that the student individual, school, and classroom ICT variables have on the differences between paper and electronic reading. Access to technology is related to mode differences only in Italy, and the use of computer devices in and out of school is related to the mode differences in Italy and Portugal. Student self-efficacy is related to the mode differences in Portugal and Slovenia. School resources show significant effects in Denmark (computers to students ratio) and Italy (instruction affected by digital resource shortages). None of the classroom variables showed any significant relationship in any of the countries. In addition, socio-economic status (which is proxied by the variable on home resources for learning) is a significant predictor in half of the countries. In addition to these findings, the general technological context within countries is discussed as part of an evaluation of the difference in reading in the two test delivery modes. The general uptake of technology in different social and economic aspects, as measured by the Digital Economy and Society Index (DESI), follows the differences between the two reading modes.

Keywords: Online reading literacy, Paper reading literacy, Reading mode comparison, PIRLS, Country comparison, Technology factors

Introduction

Because of the prevalence of technology in education and in general, it is assumed in the literature that reading on devices (or e-reading/online reading) will be a part of the current and future modality of learning (Delgado et al., 2018). Yet, there has been insufficient evaluation of student reading outcomes between e-reading and paper reading modalities (i.e., books and printed materials). As Delgado et al. (2018) put it, "Although there are clear advantages of digital-based assessment and learning, including reduced costs and increased individualization, research indicates that there may be disadvantages as well" (Delgado et al., 2018, p. 3). In particular, one reason that there is insufficient evidence comparing e-reading and paper reading modalities is that paper reading and e-reading contain different elements and potentially employ different skills. For instance, digital texts may include features such as hyperlinks, animations, or adaptive texts that are qualitatively different from paper reading and may "confound and hide media effects on learning processes" (Delgado et al., 2018, p. 9). Potential avenues for analysis in education evaluation include head-to-head comparisons of e-reading and paper reading, as well as longitudinal analyses based on the prevalence of technology in students' lives and this technology's effect on both e-reading and paper reading outcomes.

Summary of differences between e-reading and paper reading

In this study, the authors argue that, though the evidence for the effectiveness of e-reading in producing the same or better outcomes than paper reading modalities is limited, e-reading ought to be adapted to in pedagogical settings on account of its ubiquity in the classroom. That is to say, because students already use e-reading, we ought to make the best of this situation even if the evidence for the parity between e-reading and paper reading is low. Despite this inevitability of e-reading use in modern pedagogy, e-reading's detractors are quite vocal, and possibly for justified reasons, though these reasons may diminish when e-reading technology improves. Delgado et al. allege that "mere experience with digital technology does not improve students' comprehension skills, but instead has a detrimental effect," showing there is not an a priori benefit of e-reading (Delgado et al., 2018, p. 11). Moreover, "digital environments may not always be best suited to fostering deep comprehension and learning" (Delgado et al., 2018, p. 26). These criticisms are attributed to the fact that students use computers or other e-reading delivery systems for myriad purposes besides reading, which suggests there is a diminishing return for computer use for e-learning purposes. Two meta-analyses performed by Delgado et al. (2018) revealed that reading comprehension diminished when using digital texts compared to printed texts. Rosén and Gustafsson (2016) hypothesized that "increased computer use at home has a negative effect on reading achievement and that this can be explained by displacement theories" (Rosen & Gustafsson, 2016, p. 2). Displacement theories hold that, because students use digital technology for purposes besides reading, digital technology is insufficient as a reading-only platform when other means of entertainment are readily available on the same device. Research has shown that multitasking has a negative effect on learning achievement. Delgado et al. (2018) uncovered that there is an advantage for paper reading between participants and within participants compared to electronic reading, but the ubiquity of e-learning makes a suggestion to adopt paper reading problematic, as previously mentioned.

Summary of factors related to the differences

Computer use at home is a proxy for high socio-economic status (SES) and/or parental educational attainment, especially in the early 2000s where many of these studies were conducted. In one study, it was shown that "low-income students who received a computer achieved better results on several educational outcome variables than the minority low income students who did not" but this effect is not generalizable to the whole population (Rosen & Gustafsson, 2016, p. 4). Data from PIRLS 2001 and TIMMS 1999 revealed "a positive correlation between computer access at home and student performance in reading, mathematics, and science" (Rosen & Gustafsson, 2016, p. 4). In particular, it was found that computer use by nine-year olds was associated with higher reading and mathematics scores (Rosen & Gustafsson, 2016). Overall, it was concluded in a study that the presence of technology in the home improved e-reading scores relative to paper reading in PRILS 2001, though Rosen and Gustafsson (2016) warn: "increased availability of high-speed internet access [is] associated with less frequent self-reported computer use for homework, in addition to significantly lower test scores in mathematics and reading" (Rosen & Gustafsson, 2016, p. 5).

Detailed differences between e-reading and paper reading modalities that explain the differential in outcomes

Differences in delivery

According to Clinton, cost "is a driving force in the development of electronic texts" (Clinton, 2019, p. 2). It is worth noting that, compared to e-books, paper books are a physical product that one can own and feel in their hands, contributing to ergonomics in terms of holding the book which incurs some benefits to reading. Consumers complain that, even if ebooks are cheaper, the cost is not "worth it" because one can own a book physically, which is not equivalent to the "ownership" of a file one reads on a computer or Kindle (Clinton, 2019). As mentioned, e-books do provide consumers with the benefit of storing many books in one place, though this comes with the equivalent downside of not being able to store physical books, which consumers appreciate for an aesthetic pleasure. Moreover, "The experience of reading from screens is frequently described as less pleasant and less engaging than that of reading from paper" (Clinton, 2019, p. 2). Concerns related to e-reading include eye strain from reading text on screens and the time it takes without any perceived benefit compared to paper reading modalities (Clinton, 2019). In short, e-reading may suffer from screen inferiority, wherein "readers have weaker performance and metacognitive awareness of their performance, on assessments based on reading from screens compared to paper" (Clinton, 2019, p. 2).

Differences in types of understanding

The conditions that lead to understanding, and what type of understanding that is, may differ based on the modality of e-reading or paper reading. According to Clinton (2019), what this means is that "it is possible that differences between media in performance would be noted for inferential, but not literal understanding measures" (Clinton, 2019, p. 3). Literal understanding of a text involves accounting for facts and information. Meanwhile, inferential understanding of a text involves making conclusions based on prior knowledge. Overall, it is unclear whether reading times are affected by the modality of reading, though readers appear to experience wandering thoughts and lack of concentration with e-reading (Clinton, 2019). Wandering thoughts and a lack of concentration have been shown to reduce comprehension and increase reading time. In particular, "Readers report that it is more difficult to focus when reading from screens compared to paper" while "mind wandering while reading has been found to be negatively associated with reading performance" (Clinton, 2019, pp. 30–31). It is important to note that self-assessment of performance, which is an indicator of performance itself, is consistently higher with e-reading than paper reading, even if outcomes do not bear out this positive self-assessment. In other words, e-reading has lower calibration accuracy than paper reading. Finally, it was discovered that e-reading is generally better for narrative rather than expository texts (Clinton, 2019). It appears that, if effort is required to understand a piece of writing, this effort is most efficiently returned in the paper modality of reading.

Park and Lee (2021) concur with Clinton (2019) that "literal reading comprehension level increased significantly in the tablet reading group compared to the print reading and textbook-based groups. However, reading printed books was more effective in enhancing inferential reading comprehension compared to reading e-books on tablet" (Park & Lee, 2021, p. 52). It was discovered that, in a head-to-head comparison between e-reading and paper reading, "only the group reading printed books had a significant increase in grammatical knowledge over the 11 weeks" (Park & Lee, 2021, p. 52). In short, tablets are a potential benefit to students developing literal reading comprehension, though possibly not for inferential reading. The lack of an ability to perform inferential reading on e-books compared to paper reading can be attributed to the Shallowing Hypothesis, which holds that "frequent use of digital and social media (e.g., short messages, tweets, and social networking service posts) allows quick interactions, immediate feedback, easy portability, and consistent connection to the Internet, leading to shallow cognitive processing and decreased reflective thought" (Park & Lee, 2021, p. 54). Furthermore, it was discovered that, despite the presence of technology in the home for students in Norway, Norwegian English as a Second Language (ESL) students still performed better when tested on e-reading (Park & Lee, 2021).

Effects of age and generation

The newness of the e-reading modality may mean the effects of age or generation are not accounted for. In other words, it is possible that "children may be more accustomed to reading from screens and readers who are adults likely learned how to read from paper" (Clinton, 2019, p. 5). Since metacognition, which involves thinking about performance, and the capacity for accurate self-assessment improves with age, it is also possible that disparities currently seen between e-reading and paper reading modalities will decrease longitudinally (Clinton, 2019). Overall, since there is better performance for paper reading compared to e-reading even without a decrease in time, paper reading modalities "yield better performance on assessments than reading from screens" overall (Clinton, 2019, p. 30). Finally, it is possible that calibration accuracy, or the sense of understanding a text, is simply lower for e-reading texts possibly due to the lack of ergonomic handling of a physical book, leading to poorer performance. Clinton uncovered that, "logically, one's previous experience with reading from screens could lead to more comfort with the medium and a preference for screens over paper, but this has not necessarily been shown in previous findings" (Clinton, 2019, p. 31).

Differences in mental representation of the text

English as a Second Language (ESL) classrooms are a good case study for studying e-reading versus paper reading modalities as a heavy intake of reading is required for mastery of a language. Park and Lee point out that, despite the economic benefits of e-books to consumers, e-book "features hamper the ability of readers to construct and maintain mental representations of the text [and] add to the cognitive burden and visual fatigue, and prevent deep understanding of the text" (Park & Lee, 2021, p. 39). Overall, no reading and comprehension differences were observed between e-reading and paper reading modalities in Park and Lee's study, though it was found that those with lower reading abilities had greater difficulty understanding expository texts on e-reading compared to paper reading. As stated, reading is crucial for ESL because exposure to English texts is a reliable indicator for increasing English skills. In particular, "extensive reading forces the readers to repeatedly receive language input and understand its context, thereby increasing the familiarity with sentence structures in real situations" (Park & Lee, 2021, p. 41). Moreover, "learners develop superficial and fast reading habits while using digital devices for entertainment, such as surfing the web, chatting with friends in real time, and exchanging messages on social network services" (Park & Lee, 2021, p. 40). It is possible that the relationship students have to technology (that is, what they use it for, such as reading, gaming, or talking to friends, and what use patterns of it they have already habituated) will frame their capacity to understand information that is delivered by this technology. At the same time, it is possible that no technological innovation will allow for the same mental framing of text incurred in paper reading.

Effects of distraction

A potential reason that students routinely perform better with paper reading protocols than e-reading can be the level of distraction inherent to e-reading. Liu states that "The same digital code that expresses words and numbers can, if the parameters of expression are adjusted, generate sounds and images" (Liu, 2005, p. 701). Like there is a differential between literal and inferential reading, there is also one between searching and consuming information on both modalities. In short, e-reading is better for searching whereas consuming information is best done through book-form (Liu, 2005). In his study, Liu uncovered that "undergraduate students who read online text find the text more difficult to understand, less interesting, and the authors less credible than those who read the printed version" (Liu, 2005, p. 702). Moreover, Liu points out that there has been a shift in how people read (emphasis in original):

> Around the year 1750, there was a dramatic change in the way people read documents. Before this time, people were reading *intensively*. They had only a few books to read and they read them over and over again. By the early 1800s, however, people started to read things *extensively*. (Liu, 2005, p. 705).

Now that people are reading extensively, rather than intensively, an incentive exists to skim works since there are many other options for both informative and entertaining reading. One user of e-reading reports: "I skim much more html pages than I do with printed materials" (Liu, 2005, p. 705). Another e-reader recounts: "I find that my patience with reading long documents is decreasing. I want to skip ahead to the end of long articles" (Liu, 2005, p. 8). Moreover, the change in the way people read includes a tendency towards "picture reading." With this approach, readers are "looking for illustrations to explain charts and pictures. Any document with texts only will bore many savvy IT users" (Liu, 2005, p. 8).

Limitations of comparative evaluations of e-reading vs. paper reading modalities in the literature

Among authors in the education evaluation literature, there is a concern that e-reading materials are qualitatively different from printed materials, and studies are inadequately teasing out these differences in their comparative analyses. For instance, Mangen et al. contend that "compared with the amount of research on hypertext reading, the number of studies specifically addressing the potential differences between sequential and continuous reading of linear, narrative and nonnarrative texts in print and on screen is small" (Mangen et al., 2013, p. 62). Alexander and Singer (2017) likewise lament that foundational concepts are not adequately described in the literature. In particular, "only five articles (13.89%) included a definition of digital reading in any form. Within this small subset, two definitions were explicit and the remaining three were implicit" (Alexander & Singer, 2017, p. 11). Despite these inconsistencies, it was determined by Mangen et al. (2013) that e-reading is inferior to paper reading due to the observation that reading performance is weaker even if reading time among e-reading and paper reading is the same. Finally, "LCD computer screens like the ones used in this study are known to cause visual fatigue due to their emitting light" though this has been addressed in recent technological innovations in e-reading (Mangen et al., 2013, p. 66).

Phenomenology and evaluation of e-reading vs. paper reading

E-learning is the future of pedagogy despite the limitations identified for e-reading. Alexander and Singer (2017) concur that "the ubiquity of reading digitally has already answered that question [of whether digital reading belongs in our society]. In fact, as time and technology progress, the convenience of reading digitally fortifies its stake" (Margolin et al., 2013, p. 25). In an astute point, Alexander and Singer (2017) add that "for those invested in understanding and promoting student learning, therefore, there is little gained from setting up a false dichotomy between reading and digital reading" and "one medium will not and should not be regarded as routinely better for comprehension" (Alexander & Singer, 2017, p. 29). Zhang and Kudva agree in stating that "some researchers have concluded that new media simply complement old media, citing the effects of television on radio and the VCR on movie theaters as examples" (Zhang & Kudva, 2014, p. 1696). Moreover, as mentioned, e-books are not seen as so cost-efficient when one considers that it is not replacing a physical product of a book. For this reason, printed books and e-books can complement each other. It is unfortunate that the literature of e-reading and paper learning modalities ignores readers who read in both formats. Zhang and Kudva conclude that "the most frequent readers are those who read both print books and e-books, signifying that those who like to read will read books in any medium" (Zhang & Kudva, 2014, p. 1705).

If the shift to digital modalities of learning is going on as expected, educators must be adaptable to the new reality. Mangen et al. describe e-learning as a case of "human-technology interaction" (Mangen et al., 2019, p. 1). These authors quote Baron et al. (2017) in asserting that, phenomenologically, smell, sight, and touch are relevant and important functions to reading. These authors also qualify the findings by Delgado et al. (2018) in showing that the advantage for paper reading was stronger in time-constrained reading than in self-paced reading (Mangen et al., 2019). What this means is that, for leisurely or self-directed reading, compared to reading for a test, the disadvantages to e-reading may be mitigated due to context.

It has been reported by digital readers that "they feel it difficult to have a clear representation on the entirety of the text and to localize a given part of information within the text" (Mangen et al., 2019, p. 3). The intangibility of text on an e-reading platform and a lack of fixed cues about progress "contribute to a loss of orientation with respect to readers' assessment of the temporal relations between events in the text" (Mangen et al., 2019, p. 8). Finally, when reading certain texts, it is helpful to flip back to earlier parts to remind ourselves of how different facets of the text are connected. While this is possible on an e-reading platform, it is not nearly as natural or intuitive.

Some of the uptake of digital modalities of learning is driven by economics. Baron et al. (2017) point out that "Within both lower and higher education, adoption of eBooks has been significantly driven by economic considerations, since digital books are typically less costly than print equivalents, at least when purchased new" (Baron et al., 2017, p. 2). For instance, university students in the United States report that cost is their primary consideration in choosing between e-books and print books. For print books, students remarked that it was easier to underline and make marginal notes compared to e-books (Baron et al., 2017). However, e-books carried advantages as well. In particular, students "missed the ease of searching that is available with digital texts" compared to printed texts, and they viewed paper consumption as bad for the environment (Baron et al., 2017, p. 6). It was also determined that it is convenient to store all books in one place in an e-reading platform (Baron et al., 2017).

Students revealed their preferences to be paper books, though there were some nuances in this evaluation. In particular, when reading long texts for school, 86.4% of responders preferred paper books. What students "liked most" about paper books included an emotional and aesthetic response to the book (Baron et al., 2017). Meanwhile, "regarding cost, 86.9% of participants said that if cost were the same for digital and print materials for schoolwork, they would choose print, with 80.9% opting for print for pleasure reading, assuming cost parity" (Baron et al., 2017, p. 21). A total of 91.8% of respondents indicated that they concentrate best when reading paper books (Baron et al., 2017).

Outcomes in e-reading vs. paper reading and mitigating factors

Performance in reading is related to intrinsic motivation for both paper and digital reading modalities. Michael Becker et al. (2010) show that a motivation to read is a prerequisite from deriving meaning from print and improving reading comprehension. Becker et al. (2010) contend that Grade 4 intrinsic motivation to read may positively predict Grade 6 literacy and is mediated by Grade 4 reading amount. It has been argued that intrinsically motivated students “invest more time and effort to fully understand texts. As a result, they tend to achieve deeper levels of text comprehension” (Becker et al., 2010, p. 781). Becker et al. (2010) describe that students fail to progress in reading because they do not experience progress and competence, which builds intrinsic motivation. Ultimately, “Student reading motivated by the wish to please parents or teachers does not promote achievement gains over time” (Becker et al., 2010, p. 782).

Reading comprehension is assisted by prior knowledge which is gained through sustained reading. Chen defines prior knowledge “as accurate and inaccurate ideas about the target contents and discipline and the discourse conventions of the reading material” (Chen, 2017, p. 1). Moreover, “the processes of searching for and retrieving information from digital texts involve socially and culturally derived inferences about a range of text and situational cues,” contributing to reading comprehension (Chen, 2017, p. 4). What else contributes to reading comprehension is parental educational attainment, which can be seen as a proxy for SES because the two are strongly correlated (Duncan et al., 2002). Higher educational attainment typically leads to better job placement and therefore higher income (Duncan et al., 2002). It was discovered that parents who are more involved in their children’s reading efforts and provide reading-related resources had children with greater reading comprehension and intrinsic motivation to read (Chen, 2017). Moreover, it was discovered that parental involvement affected children’s resource literacy. Reading online involves being discriminating about sources in terms of epistemological and metacognitive beliefs (Chen, 2017). Ultimately, “the processes of searching for and retrieving information from digital texts involve socially and culturally derived inferences about a range of text and situational cues,” so it is recommended that policy-makers incentivize parents to be involved in their children’s reading efforts (Chen, 2017, p. 17). This is especially important because, in the e-reading modality, a reader uses explicit and embedded hyperlinks, non-sequential page structures, and “global content representation devices” (Rasmusson & Aberg-Bengtsson, 2014). Finally, e-reading texts also include pictures, sounds, and videos, in addition to words, which potentially require different skill sets to analyze (Rasmusson & Aberg-Bengtsson, 2014).

Contextual and background characteristics as mitigating factors

SES of children’s families is related to the mode of reading as well. Students from lower-SES families have greater difficulty in comprehending digital books compared to print books (Furenes et al., 2021). The possible reason for this is that lower-SES students tend to use electronic devices more for game-related activities and, thus, may focus more on the interactive features of the electronic texts and less on their actual content. In addition, when electronic reading activities appear in group context (i.e., school classes), the group sessions are more difficult to reconcile due to the interactive text enhancements (Furenes et al., 2021). In studies that include students from lower SES, paper readers outperform digital ones; however, studies with medium or high SES found no differences (Furenes et al., 2021). Besides the SES, the location of the school (large city, small town or rural area) can be related to outcomes from digital learning merely because of the access to technology (at home and at school), experience with technology, and level of skills involved. Results from different countries comparing e-learning with paper learning modalities are mixed in terms of establishing the effectiveness of one modality over another. In some countries, students from urban areas have more skills and experience, better access to technologies at home and/or at schools, and more frequent use of technology reported. In other countries the results favor rural students (Tran et al., 2020). In addition to this, and depending on the country, urban schools may have local access to more resources than the rural ones, can provide a better environment for learning due to staffing conditions, and contain students from more advantaged families. In other countries, different circumstances apply, like urban schools being surrounded with high poverty, low community support, and considerable crime and violence (Hooper et al., 2015).

Use of technology as a mitigating factor

Besides the basic demographic characteristics, use of technology itself can be related to the reading outcomes. Experience with technology can be a moderator of digital text comprehension. It can be assumed that, with enough experience in using technologies, the difficulties in comprehending electronic text will disappear. If this assumption holds, then, every next generation will be more surrounded by digital devices at an earlier age. As a result, newer generations will have equivalent or even better levels in comprehending electronic text compared to the paper ones (Delgado et al., 2018). Regardless of whether the children have been born and grown into a digital environment, it can be assumed that the general use of computers is associated with reading achievement, and previous research has found such an association: students who used a computer at home have shown higher scores in mathematics and reading (see Lee et al., 2009, for example).

Instructor competence as a mitigating factor

Instructor competence is another factor that influences children's reading comprehension and achievement. Gil-Flores et al. (2014) found that teachers with 3+years of instruction experience have greater success in instigating reading comprehension among students compared to teachers that have only been instructing for a year, though this effect diminishes after five years of instructing. It was also discovered that teachers have an accurate and reliable ability to evaluate reading comprehension ability among their students (Gil-Flores et al., 2014). As the incorporation of digital resources in reading instruction becomes a growing aspect, using technology in the classroom and presence of teachers who are familiar and comfortable with it also becomes significant. Teacher use of technology in class depends on their attitudes, beliefs and comfort, but also on the availability of training materials (Hooper et al., 2015). Access to technology is an important factor to e-reading and paper reading differentials: software, hardware, and subscription to educational websites, in particular, play a role in determining reading outcomes. As Hooper et al. state, "with the importance of online reading for informational purposes, student access to computers, the Internet, and support for their online educational research are increasingly important to expanding literacy competencies" (Hooper et al., 2015, p. 41).

This study

The Progress in International Reading Literacy Study (PIRLS) is an international large-scale assessment of reading literacy conducted by the International Association for the Evaluation of Educational Achievement (IEA), an independent international cooperative of national research institutions and government agencies. PIRLS is conducted in regular five-year cycles starting from 2001, providing trends of fourth-grade students for 20 years. A total of 50 countries and 11 benchmarking entities participated in PIRLS 2016. Results uncovered that there are more fourth-grade students reading at high proficiency than there were fifteen years ago when PIRLS 2001 took place (Mullis et al., 2017a, 2017b). In addition to evaluating changes in literacy, PIRLS 2016 was the first cycle to examine the effects of an e-learning modality on PIRLS outcomes. As the Internet is becoming a central part of obtaining additional information for students and they start to rely on the Internet as a source of information, the importance of the online/e-reading disparity becomes increasingly important because of the required skills and competencies related to it (Mullis & Martin, 2015). ePIRLS is a valuable extension of PIRLS due to the different mode of reading that allows for comparisons between paper and electronic modes by the design of the study, as “the reading comprehension skills and strategies assessed in ePIRLS will parallel those assessed in PIRLS” (Mullis et al., 2015a, 2015b, p. 24). Further, “as an extension of PIRLS, ePIRLS results can be considered in the context of the PIRLS results, including comparative achievement on PIRLS and in relation to the PIRLS context questionnaire data” (Mullis et al., 2017a, 2017b, p. 3). ePIRLS provides a simulated Internet environment with a browser providing search engine, texts to navigate through, the ability to select and process information to respond to complex reading tasks. These reading tasks are focused on only one component of reading literacy: reading for information purposes (Mullis & Martin, 2015). Seven European countries (Denmark, Ireland, Italy, Norway, Portugal, Slovenia, and Sweden) that participated in PIRLS 2016 took the ePIRLS (the electronic reading component of PIRLS) along with the paper instruments (PIRLS). The sampled students had two testing sessions. In one of these sessions, they took the paper test, and in the other one they took the electronic one. The paper-reading test and e-reading test had common items which allowed for concurrent calibration using Item Response Theory (IRT) in putting the data from both assessments on a common scale and allowing for comparison in reading for information purposes between the two modes (Foy & Yin, 2017) (for more details on the scaling methodology and “plausible values” [PVs], see the “Methodology” section). This, in turn, allowed for making comparisons between the student reading achievement on paper and electronically.

This paper investigates the differences in reading achievement of grade 4 students using data from PIRLS and ePIRLS and the degree of association of the technology factors (background and ICT characteristics, home and school resources, and computer use variables for different purposes) have with these differences across countries. In this paper “mode” is understood as the way of delivering the test content. “Mode differences” are understood as differences in the measured latent construct – reading literacy – in PIRLS and ePIRLS 2016. The reading literacy is measured on a number of items and is provided as composite test scores in PIRLS and ePIRLS (see the “Methods” section). “Mode effect” is understood as the driver of the differences in reading literacy due to the delivery mode of the content to students (paper vs. electronic). It is important to note that PIRLS 2016 is not the first international large-scale assessment that tested students using both paper and pencil and computer devices for testing. The Programme for International Student Assessment (PISA) added electronic testing mode to the paper and pencil one in 2012. However, a country could take either one of the modes, but not both (OECD, 2014). Hence, PIRLS 2016 represents the first unique opportunity to test the differences in performance between paper and pencil and computer-based test delivery system.

It was found that there were differences between e-reading and paper reading achievement in a number of countries. Though, the differences did not trend the same way in different countries or were immediately explicable. In some countries, the e-reading scores were higher, and in others the opposite difference was observed. As Fig. 1 shows, in Slovenia, Italy, and Portugal, the e-reading scores are significantly lower compared to the paper-reading scores. On the other hand, e-reading scores are significantly higher in Sweden, Norway, and Denmark. While the e-reading scores are higher than the paper-reading in Ireland, the difference between the two modes is not statistically significant. As such, this study is an attempt to find if the following factors associated with the differences in two reading modes across European countries from the student, home, teacher, and school level could have had an effect on outcomes: variables related to availability of digital technology and use of digital technology for teaching and learning in reading instruction.

One of the possible factors affecting the differences in achievement between the two modes is the method of administration itself. The literature review so far has shown that reading on screen and paper results in different processes and cognitive demands. As Fishbein et al. (2018) note, "It is acknowledged that changing from paper-and-pencil to the new PC- and tablet-based administration could have substantial and unpredictable effects on student performance" (Fishbein et al., 2018, p. 2). A study on the mode effect in the Trends in International Mathematics and Science Study (TIMSS) 2019 was conducted by the International Study Center (ISC) in Boston College, as TIMSS 2019 was the first IEA assessment to use both modes. The study found that the results on item level were generally unaffected and the standard deviations and standard errors of the scores were similar across the modes. The cross-mode correlations of the scores were 0.95 which reflects similar distribution shapes and student rankings (Fishbein et al., 2018). There is no such comprehensive study for all countries participating in PIRLS

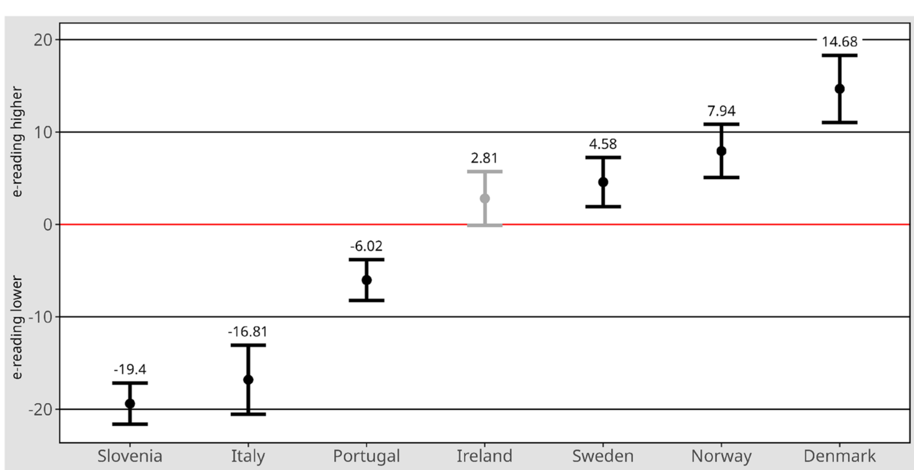

Fig. 1 Differences in paper and e-reading in European countries participating in PIRLS and ePIRLS 2016

2016 yet. However, Støle et al. (2020) found no mode effect in Norwegian data from PIRLS 2016. In addition, if there is a mode effect that stems from the instruments, it is unclear how and why the effect has different directions across the seven European countries in this study. Only European countries are included in this study, as they have more similar contexts in general, but also in terms of educational policies compared to the rest of the countries taking PIRLS and ePIRLS in 2016 (Abu Dhabi [UAE], Canada, Dubai [UAE], Georgia, Israel, United Arab Emirates, and United States).

The interest in the mode differences stems from the fact that the "Internet reading increasingly is becoming one of the central ways students are acquiring information" and "reading curricula around the world are beginning to emphasize the importance of developing online reading skills and competencies such as reading for information" (Mullis and Martin, 2015, p. 5). As reading is shifting more and more toward using computer devices, this has a lot in common with the general understanding of the "digital divide" with its three levels: (1) access to/possession of technology; (2) use of technology; and (3) ability (as outcome) (see Hohlfeld et al., 2008). Reading literacy, as outcome in ePIRLS 2016, is at the highest level in this study, frequency of use is at the lower level and access is at the lowest. The research question, hypotheses and the selection of variables (see the "Methods" section and Table A.1 in the Appendix) in this study stems from the theoretical model of digital divide briefly outlined above.

Given the literature review above and the lack of evidence for a mode effect in PIRLS 2016, the research question of this study is as follows:

Which background and ICT characteristics, home and school resources, and computer use variables for different purposes exert a relative effect on the different achievement levels in reading by mode of administration across the European countries participating in PIRLS 2016 where differences are found?

Our hypotheses are as follows:

1. Variables related to availability of digital technology at home and at school have a relative effect on the differences reading achievement by modes of reading;
2. The use of digital technology for teaching and learning in reading instruction has a relative effect on the differences reading achievement by modes of reading; and
3. Student familiarity with computer technology has an effect on the differences in reading achievement by modes of reading.

The hypotheses are tested on the population in each European country participating in PIRLS and ePIRLS 2016.

Methodology

Data and measures

The data for this study stems from the PIRLS (paper reading mode) and ePIRLS 2016 (electronic reading mode) International Data Base (or shortly IDB, IEA & TIMSS & PIRLS ISC at BC, 2018) from European countries taking the test in both modes with the same sample of students. In PIRLS 2016, the countries could choose to participate in ePIRLS as well and 14 entities chose to do so under the condition they take both the paper PIRLS and ePIRLS with the same sample of students. That is, the same students take both the paper and electronic PIRLS assessments. However, not all sampled students in each country took part in both of the assessments. In general, the overlap between students taking PIRLS and ePIRLS in European countries was quite high (see Foy & Yin, 2017 for detailed statistics). The only exception is Ireland where, due to a different sampling design implementation stemming from the availability of computers to conduct the ePIRLS sessions, the overlap was just 53.7% (Foy & Yin, 2017). Besides this, the difference between the paper and electronic mode is small and not statistically significant. This is why Ireland had to be removed from this study. The average overlap of the PIRLS and ePIRLS samples is 86.6% which is above the threshold of 85% for sampling participation for highest quality of the samples in PIRLS (see Laroche & Foy, 2017) with a minimum in Denmark (71.4%) and a maximum in Portugal (98.2%). With the exception of Denmark, all other countries had ePIRLS sampling participation above 85%. Only students that participated in both modes of administration were retained in the final ePIRLS samples (Foy & Yin, 2017).

PIRLS has a complex sampling and assessment design. Samples were drawn on a country level where each country had its own sampling frame of schools teaching students in the target population (grade 4). The sampling is multistage, cluster with probability proportional to the size (PPS) of the sampling units (i.e. schools). In the first stage, the schools are sampled with PPS where the number of students in the target population is the measure of the size of the school. At least 150 schools had to be sampled in each country to ensure the precision of the samples. In the second stage, one or two intact classrooms of students are sampled at random within each school. The minimum required sample of students across the sampled school was 4000. School and student samples are representative for the populations of schools and students in each country. The teachers sampled in PIRLS, on the other hand, do not constitute a representative sample of the teachers within a country. These are only the teachers teaching the sampled classes. As a result, teacher data in PIRLS and ePIRLS 2016 cannot be analyzed on their own, but only with student data (Foy, 2018; LaRoche & Foy, 2017; LaRoche et al., 2017). The complex assessment design of PIRLS 2016 is multiple-matrix sampling (MMS) wherein the items are grouped in 10 blocks for literary and informational experience. The blocks are rotated into 15 booklets, so that there is one block in every next booklet which overlaps with one block from the previous booklet to ensure linking across booklets. The last two blocks for informational and literary experience and literary experience are in a separate booklet not linked to any other booklet. ePIRLS assessment design uses the same approach as the paper PIRLS. In ePIRLS, four tasks are rotated across 12 task combinations with every two consecutive task combinations linked through a common task. Each block in PIRLS and task combination in ePIRLS consists of stimulus material (printed reading passage or simulated web pages respectively) (Martin et al., 2015). PIRLS and ePIRLS use items on informational reading, a total of 85 items in PIRLS, and 91 items in ePIRLS. PIRLS and ePIRLS were scaled concurrently to put the student achievement from both assessments on a common scale with PIRLS item parameters being fixed to its own calibration. Placing the the ePIRLS achievement scores on the PIRLS reporting scale was done using PIRLS linear transformation constants. The concurrent calibration of the two datasets used the same conditioning variables from student, teacher and school questionnaires (for more details Foy & Yin, 2017). The IRT parameters from the concurrent calibration were used along with conditioning background variables derived from student, school principal and teacher questionnaires' data to estimate student proficiency in reading for informational purposes in the two different modes of administration (paper and electronic). The generation of scores involves the so-called "plausible values" (PVs) methodology which, in essence, is an imputation technique. Five student scores, also called "plausible values" after the methodology used to derive them, were generated for each student and represent the students' reading scores in overall reading and in reading in different domains, both in paper and electronic modes (Foy, 2018). For more information on what PVs are, see von Davier et al. (2009); for detailed description of the scaling methodology and the generation of student reading achievement scores as PVs, see Foy and Yin (2017). These sets of PVs are part of PIRLS (paper reading) and ePIRLS (e-reading) are stored as variables in the IDB files (IEA & TIMSS & PIRLS ISC at BC, 2018). Across the countries used in this study, the PIRLS test the Cronbach alpha reliability ranges from 0.8 to 0.88 and for ePIRLS from 0.89 to 0.80 (see Foy et al., 2017). These reliabilities are more than satisfactory for the purpose of this paper.

The consequences of the complex sampling and assessment designs is that the usual analysis techniques would provide biased population estimates. Thus, each analysis involving PVs needs to be repeated with each PV and the final results have to be aggregated. All estimates need weights to be applied. When computing the standard errors, the analyses have to be performed with each PV, the full weight, and replication weights due to the unequal sampling probabilities. The estimates from each PV with the full weight and each replicate are aggregated to account for both the sampling and measurement variance and to compute the final standard errors. For more details, see Foy and LaRoche (2017).

The outcome measure in this study is the difference between paper reading and e-reading scores. As the scores in paper reading and e-reading are presented as two sets of five PVs each, the difference is computed as from each e-reading PV the corresponding paper reading PV is subtracted. As the five PVs for e-reading for information purposes are named from ASEREA01 to ASERA05 and the five PVs for paper reading for information purposes are named from ASRINF01 to ASRINF05, the difference is computed as shown in Eq. 1. That is, the differences are computed as ASEREA01 − ASRINF01, ASEREA02 − ASRINF02, and so on. Computed this way, the differences in e-reading and paper reading scores represent a new set of PVs. As these five differences originate from the original PVs and represent the differences in imputed measures, all analyses using these news scores are further performed under the same assumptions and follow the computational routines pertinent to computations with PVs as presented by Foy and LaRoche (2017) (also see above for a short explanation). That is, any analysis involving the differences between e-reading and paper reading for information purposes is performed five times (once with each difference), the results are summarized and the standard errors are computed using both the sampling and imputation variance components. None of the analyses in this paper are performed or reported on individual level, as PIRLS and ePIRLS are not designed for this, but only for reporting on population level in each country.

This study uses variables from the ePIRLS 2016 student, school, and teacher questionnaires. As informed by the literature review, a number of variables is included in the analyses. These are presented in Table A.1 in the Appendix. The table shows the respondents providing the information (students, school principals or teachers), the variable names and the actual question (in case questionnaire questions were used), and the variable names of the constructed scales. Table A.1 also indicates whether a variable is categorical (providing the original response categories) or continuous (providing the measurement properties—center point and standard deviation). As for some of the categorical variables the original values were in descending order (i.e. higher values indicate lesser amounts), the last column in the table indicates if the variable, where appropriate, was reverse-coded before using it in any analysis.

$$DIFF_{INF} = ASEREA_i - ASRINF_i \tag{1}$$

where

$DIFF_{INF}$- differences between e-reading and paper reading for information purposes scores.

ASEREA—vector of PVs in e-reading for information purposes.

ASRINF—vector of PVs in paper reading for information purposes.

i—number of a PV in a set

There are two types of scales used in this study – some are readily available in the ePIRLS datasets (ASBGHRL, ASBGDDH, ASBGSEC and ACBGDRS). The Student Home Resources for Learning (ASBGHRL) was used as a proxy for SES is available from the PIRLS 2016 IDB (IEA & TIMSS & PIRLS ISC at BC, 2018) and is created by the PIRLS international study center using the variables on the number of books at home, number of study supports (internet connection and own room), number of children's books at home, and the highest educational and occupational levels of either parent. The Digital Devices in the Home (ASBGDDH) scale is available in the IDB (IEA & TIMSS & PIRLS ISC at BC, 2018) and was created using data on the availability of internet connection, number of digital devices, and availability of digital devices for reading. The Student Self-Efficacy for Computer Use scale (ASBGSEC) was created using IRT with data on student opinion about how well they feel about their own proficiency with computers, how well they can type, and how easy it is for them to find information on the internet. The Instruction Affected by Digital Resource Shortages (ACBGDRS) scale was created from the school principal questionnaire data on technologically competent staff, audio-visual resources for instruction, computer technology for teaching and learning, and computers and software specifically for reading instruction. More information on how these scales were constructed by the international study center for the IDB (IEA & TIMSS & PIRLS ISC at BC, 2018) can be found in the PIRLS 2016 technical documentation (Martin et al., 2017).

To answer the research questions and test the hypotheses, additional scales were constructed for the purpose of this study using the ePIRLS student and teacher data (see the column "Source"). The Combined Frequency of Using Computer Devices in and Out of School (FREQDEVACT) is an IRT scale constructed for the purpose of this study. The variables used to construct this scale are as follows: (1) time spent per day on finding and reading information for school work; (2) preparing reports and presentations for school work; (3) daily time spend at a computer to play games; (4) daily time spent at a computer to watch videos; (5) daily time spent at a computer chatting; and (6) daily time spent at a computer surfing internet. The Use of Computers in Classroom Reading Activities (FREQCLASSUSE) scale was constructed for this study using questions from the teacher questionnaire on how often teachers do the following activities in reading classes with their students: (1) ask them to read digital texts; (2) teach them strategies for reading digital texts; (3) teach them to be critical when reading on internet; (4) ask them to look up information; (5) ask them to research particular topic or problem; and (6) ask them to write stories or other texts. These two scales were created following the same methodology as the international study center used for creating all other scales in PIRLS and ePIRLS 2016 (Martin et al., 2017).

The student FREQDEVACT and the teacher FREQCLASSUSE scales were created using the IRT partial credit model (PCM) (Masters & Wright, 1997). The scaling procedure under the PCM model involves estimation of the item location and the item step parameter. It then uses these parameters to estimate individual (student or the teacher) scores, respectively, on the latent trait (Masters & Wright, 1997). To construct these two scales, the mirt R package (Chalmers et al., 2023) was used through PCM, utilizing the expectation–maximization (EM) estimation algorithm. The construction of the FREQDEVACT student scale used the student senate weight and the construction of the FREQCLASSUSE teacher scale used the teacher senate weight to ensure equal contribution of all countries to the location and step parameters regardless of their population sizes. As the student and teacher individual scores for these two scales were obtained as regular standardized scores ($N[0, 1]$) from the IRT modeling, their metric was altered to have the same metric as the scales produced by the PIRLS international study center (N[10, 2]) (see Martin et al., 2017). The item parameters for these two constructed scales can be found in Table A.2 and Table A.3 in the Appendix. The Cronbach alpha reliability for the FREQDEVACT equals 0.71, and for the FREQCLASSUSE it is 0.84. These reliability coefficients fall within the range of acceptable values for Cronbach alpha chosen by the international study center for the PIRLS and ePIRLS contextual scales (see Martin et al., 2017). Further, the student Home Resources for Learning scale was aggregated on the school level and used in the analyses of school IT factors as well. The ratio of computers to students in the schools (COMSTRAT) was computed by dividing the number of computers (including tablets) the school has for use by fourth-grade students by the total number of target grade students enrolled in the school in the year of testing (2016).

Analysis methods

The analyses pertinent to PIRLS data would utilize the power of multilevel modeling (MLM) techniques (see Finch & Bolin, 2017, for example), as the students are nested in classes and schools in a strict hierarchy. However, an initial analysis of the variance in the difference between paper and e-reading showed that the between-school variance is very low. In all countries the between-school variance is below 6%. In Denmark, Norway, and Slovenia, it is even below 3%. The actual intraclass correlation coefficients (ICCs) were computed using

Table 1 Intraclass correlations and design effects

Countries	ICC	DEFF (school)	DEFF (teacher)
Denmark	0.013	1.20	1.22
Italy	0.056	2.18	1.95
Norway	0.014	1.26	1.23
Portugal	0.034	1.64	1.57
Slovenia	0.028	1.55	1.46
Sweden	0.051	2.04	1.93

Mplus version 8.9 (see Muthén & Muthén, 2017) and are presented in Table 1. The estimated ICCs for the NULL models using student and school data, and student and teacher data, are identical and, thus, there is only one column presenting the ICCs. As Nezlek (2008) points out, the low ICC alone is not enough to justify why not using MLM, although Goldstein (2011) states that "When a variance partition coefficient [i.e. intraclass correlation] is small, we can expect reasonably good agreement between the multilevel estimates and the simpler OLS ones" (Goldstein, 2011, p. 27). As Maas and Hox (2005) contend, however, it is not so much the ICC, but the design effect that is the more important statistic in this case, as it indicates how much the standard errors in a complex sample are underestimated when compared to a simple random sample. A design effect of 2.0 is considered as small (Maas & Hox, 2005). Furthermore, "If the design effect is smaller than 2, using single-level analysis on multilevel data does not seem to lead to overly misleading results" (Hox & Maas, 2001, p. 165). Also, "A design effect larger than two indicates the necessity of a multilevel analysis, and a design effect around two indicates the sufficiency of a single-level analysis" (Wang & Qiu, 2019, p. 40). In this study, the design effect was computed using the formula provided by Maas and Hox (2005) and Wang and Qiu (2019):

$$DEFF = 1 + \left(\overline{N} - 1\right) \times ICC \tag{2}$$

where.

$DEFF$—design effect.

$\overline{N}$—average cluster size.

ICC—intraclass correlation coefficient.

The design effects for the student and school data and student and teacher data are presented in Table 1, alongside the ICCs.

In most cases across the six countries it appeared to be lower than 2.0 for models using student and school data. The only exceptions were Italy (2.18) and Sweden (2.04). However, in both cases this design effect is just over 2.0, i.e. still very small. For models using student and teacher data, the design effects are below 2.0 in all countries. Thus, there is no reason using MLM in this study over single-level OLS regression, as it will not be any more beneficial, and using OLS regression will not introduce bias to the estimates. Thus, the analyses for all countries' data are performed using OLS regression to ensure consistency and comparability of the results.

The low percentages of between-school variances does not make the use of MLM feasible, as it will not reveal differences between clusters (classes or schools). Thus, this paper uses ordinary least square regression (OLS) models wherein the difference between paper and electronic reading are used as a dependent variable and the variables from Table A.1 in the Appendix are used as independent variables. As per the research question and the hypotheses of this study, there are two sets of models. The first set has two models and uses the school and student variables as independent. The second one uses only teacher variables as independent.

The first model with student and school data adds student variables only: (1) amount of time spent using computer daily (ASBE01); (2) amount of time spent finding and reading information daily (ASBE02); (3) use of computer for reading schoolwork (ASBG10A), use of computer for preparing schoolwork (ASBG10B); (4) the combined frequency of using computer devices in activities in and out of school scale (FREQDEVACT); (5) the computer self-efficacy in computer use scale (ASBGSEC); and (6) the access to various digital devices at home scale (ASBGDDH).

The second model with student and school data adds the student home resources for learning scale (ASBGHRL) and the school variables related to the ICT school characteristics: (1) access to digital books (ACBG10); (2) instruction affected by digital resources scale (ACBGDRSD); (3) computers to students ratio at school (COMSRAT); (4) school location (ACBG05B); and (5) student home resources for learning aggregated at school level (HRLAGGR). The purpose of fitting two different regression models (one with student variables only and one adding the school variables) is to determine to what extent the relative effect of student personal characteristics and behaviors related to general purpose use of technology and use of technology for study purposes change after controlling for school variables related to technology. Equations 3 and 4 are formal representations of Model 1 and Model 2 using student and school data.

$$\begin{aligned} Y_i = &\beta_{0i} + \beta_{1i}ASBE01 + \beta_{2i}ASBE02 + \beta_{3i}ASBG10A \\ &+ \beta_{4i}ASBG10B + \beta_{5i}FREQDEVACT \\ &+ \beta_{6i}ASBGSEC + \beta_{7i}ASBGDDH + \epsilon_i \end{aligned} \tag{3}$$

Table 3 Differences in paper and online reading in PIRLS 2016 against DESI 2016 index

Countries	**Average difference in achievement (e-reading and paper reading)**	**(SE)**		**DESI 2016**
Denmark	14.68	(1.85)	▲	0.68
Sweden	4.58	(1.36)	▲	0.67
Norway	7.94	(1.47)	▲	0.65
Portugal	− 6.02	(1.13)	▼	0.53
Slovenia	− 19.4	(1.14)	▼	0.49
Italy	− 16.81	(1.90)	▼	0.40

▲—e-reading significantly higher.

▼—e-reading significantly lower.

$$\begin{aligned} Y_i = &\beta_{0i} + \beta_{1i}ASBE01 + \beta_{2i}ASBE02 + \beta_{3i}ASBG10A \\ &+ \beta_{4i}ASBG10B + \beta_{5i}FREQDEVACT \\ &+ \beta_{6i}ASBGSEC + \beta_{7i}ASBGDDH \\ &+ \beta_{8i}ACBG10 + \beta_{9i}ACBGDRSD \\ &+ \beta_{10i}COMSRAT + \beta_{11i}ACBG05B + \beta_{12i}HRLAGGR + \epsilon_i \end{aligned} \tag{4}$$

where.

Y_i—dependent variable (ith difference from the electronic and paper reading PVs).

ϵ_i—error term.

There is just one model using teacher and student data which uses the use of computers in classroom reading activities scale (FREQCLASSUSE) constructed for the purpose of this study (see the previous subsection) and the total number of years students' teachers have been teaching (ATBG01). The formal presentation of the model can be found in Eq. 5.

$$Y_i = \beta_{0i} + \beta_{1i}FREQCLASSUSE + \beta_{2i}ATBG01 + \epsilon_i \tag{5}$$

The variable names in Eqs. 3, 4, 5 can be matched to their full meaning in Table A.1 in the Appendix. All of the analyses have been performed using the appropriate weights and their replicates. For the models using student and school data, the student total weights were used, as when merging the student and school datasets the school characteristics become property of the students and, thus, student weights have to be used, as explained in the PIRLS 2016 User Guide (Foy, 2018). For the model involving teacher data, teacher total weights were used, as required (see Foy, 2018). As the differences between the two modes in reading for information purposes is computed as the differences in PVs in both modes (paper and e-reading), these are five sets of differences in the scores in reading for informational purposes between the modes and, thus, still need to be treated as imputed variables when computing statistics and their standard errors. To handle these analytical challenges, all computations are performed using the R package RALSA (Mirazchiyski, 2021), which is designed to analyze data from large-scale assessments and can compute correct estimates and their standard errors taking into account the complex sampling and assessment design issues. The computations for these models use standardized coefficients for the dependent and independent variables as these are more suitable for interpretation considering the varying differences in paper and e-reading across countries, as well as their different signs (paper reading higher or e-reading higher).

Results

All results from this study are reported and interpreted by country. The results from Model 1 and Model 2, along with the diagnostic model statistics are presented in Table A.4 through Table A.7 in the Appendix. The results from Model 1 using student data (Table A.4 in the Appendix) show that the student computer use and home digital resources demonstrate effects in a limited number of countries. The combined frequency of using computer devices in activities in and out of school has an effect in Italy and Portugal. The student self-efficacy in computer use has an effect in Portugal and Slovenia. The access to digital devices at home has an effect only in Italy. All coefficients for these variables in these countries are positive, i.e., these variables are related to better performance in e-reading. It is worth noting, however, that although significant, these coefficients are very small, 0.10 or lower. None of the aforementioned variables or any other in the model has any significant effect in other countries in this study.

The results from Model 2 wherein school digital resources (availability of electronic books, digital resource shortages, and computers to student ratio), school location, individual student home resources for learning, and home resources for learning are aggregated at the school level are presented in Table A.6 in the Appendix. In Italy, the combined frequency of computer use in and out of school became insignificant after controlling for the newly introduced control variables, but the coefficient for access to digital devices at home remained significant. In Portugal, the combined frequency in using computer devices in and out of school and student self-efficacy in computer use remained significant, although the size of the coefficients for these two variables slightly decreased. That is, the control variables did not change the relationship between these two variables and the dependent one (difference between paper and electronic reading). At the same time, the coefficient for access to digital devices in the home became significant in Portugal. In Slovenia, the coefficient for student self-efficacy in computer use became insignificant. Concerning the newly introduced control variables, the individual home resources for learning have significant coefficients in Italy, Slovenia and Norway where the coefficients are negative, i.e., the individual home resources for learning have a negative effect on the e-reading in these three countries. Computer to students ratio has significant positive effect only in Denmark (i.e., higher ratio of computers to students tends to affect e-reading positively). The coefficient for the instruction being affected by shortage of digital resources is positive and significant only in Italy. As the lower values of this scale mean that instruction is more affected and higher values mean that the instruction is less affected by digital resource shortages, the positive coefficient indicates that students studying in schools where there are less shortages tend to have higher scores in e-reading and vice-versa.

The teacher model using the use of computers for classroom reading activities scale and the number of years teaching did not reveal any statistically significant results in any of the countries. That is, none of these two classroom variables is related to the difference between the paper and e-reading results.

In addition to these models, the associations between home resources for learning and school location, on the one hand, and the differences between paper and e-reading on the other were tested. These tests were done using linear regression models wherein these two variables were added as contrast coded predictors to test the differences between groups (male–female, size of the communities where students study). None of these two analyses revealed significant differences between groups of different students based on these two variables.

Table 2 Countries correlation coefficients between the Home Resources for Learning scale and the differences between paper and e-reading

Countries	Correlation	(SE)	p
Denmark	− 0.09	(0.03)	0.002
Italy	− 0.04	(0.02)	0.055
Norway	− 0.08	(0.02)	< 0.001
Portugal	− 0.02	(0.02)	0.446
Slovenia	− 0.13	(0.03)	< 0.001
Sweden	− 0.02	(0.02)	0.327

Finally, the correlation between the individual home resources for learning and the difference between paper and e-reading was tested as well. The correlation coefficients are presented in Table 2. The negative sign shows that the students with more home resources for learning tend to have higher scores on the e-reading test. Significant relationships were found in Denmark, Norway, and Slovenia, although these relationships are very weak—the strongest is in Slovenia, − 0.13 ($p < 0.001$). Additionally to Model 2 with student and school data, a model wherein the individual and the aggregated home educational resources for learning were interacted and there was no interaction found.

Summary and discussion

This article investigated (1) the effect of availability of student technology; (2) student use of digital technology for school and out-of-school purposes; and (3) teacher characteristics and use of technology in the classroom in relation to the differences in paper and e-reading performance in European countries participating in PIRLS 2016, taking both modes of reading (e-reading and paper reading). The results show support, although not in all countries, to the hypotheses stated earlier in the article for just a few countries.

As for the first hypothesis, the availability of technology at home is related to the differences in e-learning in just one country (Italy) where the coefficient remains significant after controlling for all other variables. Access to technology at school is related to the differences in the two modes of reading only in two countries: Denmark (computers to students ratio) and Italy (instruction affected due to digital resource shortages).

Similarly, for the second hypothesis, using digital technology in teaching and learning is related to the differences between the two modes in just two countries (Denmark and Portugal) where the combined frequency of using computer devices in and out of school has a significant effect. However, after controlling for ICT-related variables on school level and individual and aggregated home resources for learning, the result remains significant only in one of them (Portugal). The use of computers in classroom reading activities did not reveal any statistically significant differences in any of the countries in this study, nor did the number of years of being a teacher.

As for the third hypothesis, student familiarity with technology had a significant effect in two countries (Portugal and Slovenia), but when controlling for school ICT-related variables and individual and aggregated home resources for learning, the relationship remains significant in only one country (Portugal).

In addition to the findings from these models, this study finds no significant differences between male and female students in terms of differences between the two reading modes. Similarly, there were no differences between the students studying in communities with different sizes. The home resources for learning at individual level in the regression Model 2 did show significant effects in Norway and Slovenia. Testing the relationship without any other variables in the model (i.e. correlation) also added Denmark to this group.

These results bring insights to ICT-related variables and their relationship in the differences between the two reading modes, but to a limited extent. At a first glance, the results suggest that in countries located in the south of Europe (Italy, Portugal and Slovenia) there are more prominent effects compared to the countries located in the north (Denmark, Norway and Sweden). However, the variables related to these differences are rather different across these countries. In addition, there are too few countries in Europe that took the test in both reading modes to provide support with their data for such a definitive claim. That is, there needs to be a larger number of countries to obtain a deeper insight.

Besides the data and findings on ICT-related variables collected in PIRLS 2016, it is important to discuss the differences in the two reading modes (paper and electronic) against the general context of use of technologies across the countries. The Digital Economy and Society Index (DESI) (European Commission, n.d. 2023) collects a large amount of data and provides a summary on a large number of ICT-related indicators. The index aggregates information on four major strands: (1) human capital (internet user skills, and advanced skills and development); (2) connectivity (fixed broadband take-up, fixed broadband coverage, and mobile broadband); (3) integration of digital technology, digital technologies for business, and e-commerce); and (4) digital public services (e-government) (European Commission, n.d. 2023). The information on the differences in paper and online reading in PIRLS 2016 from Fig. 1 was cross-referenced to DESI from 2016. The summary is presented in Table 3.

As Table 3 shows, the countries in this study with higher DESI have differences favoring online reading while in those countries with lower DESI, the differences favor paper reading. Moreover, the larger the differences become in either direction (paper or online being higher), the smaller or bigger the values of DESI become. That is, there is a clear indication that the general digital development of a country can be related with the outcomes in online and paper reading. However, this cannot be regarded as the sole evidence that the overall digital development of a society favors online reading literacy for informational purposes. Given this, an additional important question can be asked: Is online reading completely independent from computer and information literacy (CIL)? Reading in an online environment would require some basic skills to navigate through the content and process the information. On the other hand, it is unclear if CIL can be developed without reading literacy but also the other way around – that CIL is a prerequisite for successful online reading. In the case of online reading skills for retrieving information is an important prerequisite, so is CIL. The differences between paper and electronic reading in different aspects were made quite clear in the introduction of this paper. Online texts are multimodal and distributed across different sources, which often requires creating one's own reading paths using navigation strategies. Readers also need to maintain focus on the task while navigating through the interconnected sources. This fact poses greater comprehension demands on the reader, as searching for information in such a complex environment requires self-regulatory processes (Mullis et al., 2015a, 2015b). The above is quite close to the concept of information literacy wherein "identifying information needs, searching for and locating information, and evaluating the quality of information… include the ways in which the collected information can be transformed and used to communicate ideas" (Fraillon et al., 2013, p. 16). That is, online reading literacy for information purposes may be intersected with information literacy and computer literacy. The results in Table 3 may also be related with the general framework of the "digital divide" which has three distinct levels (see Hohlfeld et al., 2008 for example):

1. Equitable access to hardware (lowest);
2. Frequency of use of technology by teachers and students in class (medium); and
3. Student's ability to use ICT for their own empowerment (highest).

Concerning the last level, early adopters in technology have the advantage of having continuously more experience than the later adopters (Hohlfeld et al., 2008; M.-C. Kim & Kim, 2001). The different rates at which technology permeated different countries may explain the results in Table 3.

As the results from this study also show, in some countries (Denmark, Norway, and Slovenia), traditional SES differences (proxied by the home resources for learning) also exerts their influence on distinctions among early and later adopters. Thus, additional studies on the differences between online and paper reading are warranted. Future studies shall focus more on the relationship between the ICT and CIL variables and the mode of reading. Unfortunately, PIRLS does not provide any scores on CIL, which is understandable as it has different focus and goals. Having a study where the CIL is measured would be beneficial to explain the differences in reading modes.

The main implications from this study are as follows:

1. Although the access to/possession of technology is considered as having become universal with the spread of devices, software and services, the availability in the classrooms, but also in the out-of-school locations, is still important. This is the case in both less and more developed countries, especially the availability in schools. Hence, there is still need to continuously equip classrooms and provide maintenance to the devices and infrastructure, also given that technology is continuously developing at rapid rates.
2. The use of technology in classroom instruction and in out-of-school activities has an "effect" on online reading literacy. The use in out-of-school activities, however, may not be necessarily related to learning activities, but for leisure, for example. Further, the relationship between reading literacy and computer use is affected by the home resources for learning and school ICT variables. Thus, more frequent use of com-

puter devices in instruction can help improving online reading by compensating for the home differences in resources for learning.

3. Although use of electronic devices have permeated all spheres of life, student familiarity with technology remains as an important prerequisite for online reading ability. The mitigation effect of school ICT variables (i.e. the effect of familiarity disappears after controlling for school ICT) is a clear sign that through its ICT equipment, infrastructure and usage schools can compensate for different levels of familiarity with technology.
4. The overall spread of technology across different countries' economies and societies shows a clear pattern in favoring paper (countries in South Europe) vs. electronic (countries in North Europe) reading. That is, differences in terms of human capital related to technology, connectivity, integration of digital technology, and availability and use of digital public services across the countries still exist and these gaps may be related, even indirectly, use of technology in instruction and, subsequently, achievement in different subjects.
5. Related to the previous, reading literacy when using technology may not be an isolated construct, but related to the more general computer and information literacy. Strengthening the latter one could help improving the former.

Limitations

This study does not come without limitations. Some of the limitations are methodological, other are conceptual.

The first limitation is related to the differences of the scores between PIRLS and ePIRLS. Although the data is cross-sectional, uses IRT with concurrent calibration of PIRLS and ePIRLS items, and the achievement scores are multi-item composites (see the "Data and measures", there still can be an issue related to the reliability of the differences between PIRLS and ePIRLS scores. This, however, is not possible to estimate precisely.

Second, there is limitation related to the sampling within countries for the different modes of test delivery. Although each country had to use the same sample in both modes (paper and electronic), countries could not ensure that all sampled students took both tests. In each country some students were absent in the ePIRLS session. Although there were such cases in each country, the median overlap in both testing sessions was rather high—86.6% which is still above than the acceptable threshold of 85% for sampling participation in PIRLS 2016 without using replacement schools (highest category of participation to ensure quality) (LaRoche & Foy, 2017) and ensures representativeness of the samples. In almost all countries the participation in both PIRLS and ePIRLS was above 85%. In some countries these are close to 100%, for example in Portugal the overlap is 98.2%. The effect of this overlap cannot be known, but shall be considered when interpreting the results.

Third, there can be other ICT-related variables that mediate the mode differences, as computer anxiety, efficiency in ICT usage, prior use of ICT, efficiency in using ICT, etc.

Testing the mediation effect of these variables would be an important contribution to the topic. However, PIRLS and ePIRLS 2016 do not collect data on these variables.

Fourth, the test-taking behavior can be an important component in research on the mode effect. For example, the literature review provided information on the mind wandering (for example, see Mangen et al., 2019; and Delgado et al., 2018). Extending the research would require access to process data from ePIRLS 2016. Unfortunately, such data was not collected in ePIRLS 2016.

Fifth, we could not control for the order in which PIRLS and ePIRLS tests took place. The only information available is that “PIRLS, including the student questionnaire, was always administered before ePIRLS” (Johansone, 2017, p. 6.11) and that ePIRLS assessment was conducted “typically on the next day” (Mullis et al., 2017a, 2017b). As this information is insufficient, the database lacks information on the order and exact timing, and there could still be variation in the actual implementation across countries, the potential effect the order of the testing sessions has on the results cannot be estimated.

Lastly, in this article the differences in PV reading scores were computed by subtracting each paper reading PV from its corresponding electronic reading PV (ASEREA01—ASRINF01, ASEREA02—ASRINF02, and so on, see Eq. 1) which results in five score differences, as many as the number of PVs. From a methodological point of view, however, the ideal estimation of the differences between electronic and paper reading scores using the PVs and their use in the analyses would be to compute the differences between all possible combinations of electronic reading PVs. This would result in 25 possible score differences instead of just five (one for each pair of PVs’ differences) and the current specialized software for analyzing PIRLS data can handle five PVs, as this is per PIRLS’ design. Further, given that the PVs themselves are random draws from the conditional distribution of the scale proficiencies given the students’ item responses, the student’s background variables, and model parameters for the items (see Foy & Yin, 2017), it is unlikely that the existing differences between PVs in a set will be big. In Norway, for example, the largest differences in PV averages of paper reading PVs is just 1.72 score points, for the electronic reading is 1.48. Given the distribution of the PVs (center point of 500 and standard deviation of 100), these are really small differences, i.e. it is probably unlikely to find very different results using the two approaches. While the authors of this paper consider the question of the differences in these two estimation procedures (differences between pairs of PVs vs. all possible combinations), this goes beyond the scope of the paper.

Appendix

See Tables 4, 5, 6, 7, 8, 9, 10

Table 4 Variables from different respondent types used in the analyses

Respondent	Variable name	Question/scale	Source	Type	Categories/ Measurement properties	Reversed
Student	ASBE01	About how much time do you spend using a computer each day?	Instrument question	Categorical	1–Less than 30 min; 2–30 min up to 1 h; 3–From 1 h up to 2 h; 4–2 h or more	No
Student	ASBE02	About how much time do you spend each day finding and reading information on the internet?	Instrument question	Categorical	1–Less than 30 min; 2–30 min up to 1 h; 3–From 1 h up to 2 h; 4–2 h or more	No
Student	ASBG10A	Using computer for schoolwork—Finding and reading information	Instrument question	Categorical	1–No time; 2–30 min or less; 3–More than 30 min	No
Student	ASBG10B	Using computer for schoolwork—Preparing reports and presentations	Instrument question	Categorical	1–No time; 2–30 min or less; 3–More than 30 min	No
Student	FREQDEVACT	Combined frequency of using computer devices in activities in and out of school	Constructed scale	Continuous	M = 10; SD = 2	–
Student	ASBGHRL	Home Resources for Learning scale	PIRLS scale	Continuous	M = 10; SD = 2	–
Student	ASBGDDH	Digital Devices in the Home scale	PIRLS scale	Continuous	M = 10; SD = 2	–
Student	ASBGSEC	Students Self-Efficacy Computer use	PIRLS scale	Continuous	M = 10; SD = 2	–
School	ACBG05B	School location	Instrument question	Categorical	1—Urban–Densely populated; 2—Suburban–On fringe or outskirts of urban area; 3—Medium size city or large town; 4—Small town or village; 5—Remote rural	Yes
School	ACBG10	Does the school provide access to digital books?	Instrument question	Categorical	1—Yes; 2—No	Yes
School	ACBGDRS	Instruction Affected by Digital Shortages scale	PIRLS scale	Continuous	M = 10; SD = 2	–

Table 4 (continued)

Respondent	Variable name	Question/scale	Source	Type	Categories/ Measurement properties	Reversed
School	HRLAGGR	Home Resources for Learning scale aggregated at school level	PIRLS scale	Continuous	M = 10; SD = 2	–
School	COMSTRAT	Ratio of computers to students	Constructed scale	Continuous	M = 10; SD = 2	–
Teacher	FREQCLASSUSE	Use of Computers in Classroom Reading Activities scale	Constructed scale	Continuous	M = 10; SD = 2	–
Teacher	ATBG01	By the end of this school year, how many years will you have been teaching altogether?	Instrument question	Continuous	M = 21.09; SD = 11.23	–

Table 5 IRT item parameters for the FREQDEVACT scale

Items	Location	Step 1	Step 2	Step 3	Step 4	Infit
ASBG11A	− 0.05444	− 1.59362	− 0.02645	0.91756	0.48475	0.78
ASBG11B	0.14537	− 1.37005	0.12585	0.73692	1.08878	0.77
ASBG11C	0.80303	− 0.14238	1.13558	1.38245	0.83647	0.88
ASBG11D	0.51168	− 0.82986	0.85889	1.28509	0.73260	0.77

Table 6 IRT item parameters for the FREQCLASSUSE scale

Items	Location	Step 1	Step 2	Step 3	Infit
ATBR14CA	0.82911	− 1.52839	0.65751	3.35820	0.77
ATBR14CB	1.93041	− 0.63898	2.03525	4.39496	0.81
ATBR14CC	0.34716	− 2.47889	0.88731	2.63306	0.91
ATBR14CD	− 0.44007	− 4.05435	0.01459	2.71955	0.68
ATBR14CE	0.40317	− 2.70399	0.69191	3.22160	0.74
ATBR14CF	0.13650	− 3.07250	0.24234	3.23968	1.17

Table 7 Results from Model 1 (student data)—student computer use and home digital resources

Variables	Countries	Coefficients (SE)	p	Countries	Coefficients (SE)	p
Time spent using computer daily	Denmark	0.04 (0.03)	0.210	Portugal	− 0.01 (0.03)	0.768
Time spent using computer to finding and reading information daily	Denmark	− 0.01 (0.03)	0.669	Portugal	− 0.02 (0.02)	0.486
Using computer or tablet for schoolwork—reading	Denmark	− 0.02 (0.03)	0.508	Portugal	0.00 (0.03)	0.967
Using computer or tablet for schoolwork—preparing	Denmark	− 0.01 (0.03)	0.708	Portugal	0.04 (0.03)	0.221
Combined frequency of using computer devices in activities in and out of school	Denmark	0.04 (0.04)	0.272	Portugal	0.10 (0.03)	0.002
Student self-efficacy in computer use	Denmark	− -0.01 (0.04)	0.779	Portugal	0.04 (0.02)	0.049
Access to digital devices in the home	Denmark	− 0.03 (0.03)	0.282	Portugal	0.05 (0.03)	0.080
Time spent using computer daily	Italy	0.03 (0.03)	0.303	Slovenia	0.02 (0.04)	0.675
Time spent using computer to finding and reading information daily	Italy	− 0.03 (0.02)	0.193	Slovenia	0.04 (0.03)	0.112
Using computer or tablet for schoolwork—reading	Italy	0.03 (0.03)	0.333	Slovenia	0.01 (0.02)	0.745
Using computer or tablet for schoolwork—preparing	Italy	− 0.02 (0.03)	0.460	Slovenia	− 0.01 (0.02)	0.618
Combined frequency of using computer devices in activities in and out of school	Italy	0.06 (0.03)	0.049	Slovenia	0.03 (0.05)	0.497
Student self-efficacy in computer use	Italy	0.04 (0.03)	0.190	Slovenia	0.06 (0.03)	0.040
Access to digital devices in the home	Italy	0.06 (0.02)	0.011	Slovenia	− 0.02 (0.03)	0.436
Time spent using computer daily	Norway	0.01 (0.03)	0.735	Sweden	0.00 (0.03)	0.958
Time spent using computer to finding and reading information daily	Norway	− 0.01 (0.02)	0.649	Sweden	0.02 (0.03)	0.617
Using computer or tablet for schoolwork—reading	Norway	0.03 (0.03)	0.250	Sweden	0.02 (0.03)	0.448
Using computer or tablet for schoolwork—preparing	Norway	− 0.01 (0.03)	0.635	Sweden	0.05 (0.04)	0.162
Combined frequency of using computer devices in activities in and out of school	Norway	− 0.03 (0.04)	0.510	Sweden	0.05 (0.03)	0.137
Student self-efficacy in computer use	Norway	0.02 (0.04)	0.634	Sweden	0.02 (0.03)	0.392
Access to digital devices in the home	Norway	0.02 (0.04)	0.515	Sweden	− 0.01 (0.04)	0.735

Table 8 Model statistics from Model 1 (student data)

Countries	Statistic	Estimate	(SE)	p
Denmark	R-Squared	0.01	(0.005)	–
	Adjusted R-Squared	0.00	(0.005)	–
	F-Statistic	2.49	(1.503)	0.097
	DF	2141.00	(130.173)	–
Italy	R-Squared	0.02	(0.006)	–
	Adjusted R-Squared	0.01	(0.006)	–
	F-Statistic	7.37	(2.651)	0.085
	DF	3139.00	(198.329)	–
Norway	R-Squared	0.00	(0.004)	–
	Adjusted R-Squared	0.00	(0.004)	–
	F-Statistic	2.20	(2.009)	0.274
	DF	3104.00	(203.310)	–
Portugal	R-Squared	0.02	(0.006)	–
	Adjusted R-Squared	0.02	(0.006)	–
	F-Statistic	10.95	(3.476)	0.002
	DF	4148.00	(264.868)	–
Slovenia	R-Squared	0.01	(0.006)	–
	Adjusted R-Squared	0.01	(0.006)	–
	F-Statistic	6.82	(3.354)	0.042
	DF	3795.00	(248.684)	–
Sweden	R-Squared	0.01	(0.007)	–
	Adjusted R-Squared	0.01	(0.007)	–
	F-Statistic	4.22	(2.669)	0.114
	DF	2818.00	(186.926)	–

Table 9 Results from Model 2 (student and school data)—student use, home and school digital resources, home resources for learning, and school location

Variables	Countries	Coefficients (SE)	p	Countries	Coefficients (SE)	p
Time spent using computer daily	Denmark	0.02 (0.04)	0.532	Portugal	0.00 (0.03)	0.861
Time spent using computer to finding and reading information daily	Denmark	− 0.01 (0.04)	0.718	Portugal	− 0.01 (0.02)	0.653
Using computer or tablet for schoolwork—reading	Denmark	− 0.02 (0.03)	0.604	Portugal	0.01 (0.03)	0.821
Using computer or tablet for schoolwork—preparing	Denmark	− 0.01 (0.03)	0.801	Portugal	0.04 (0.03)	0.251
Combined frequency of using computer devices in activities in and out of school	Denmark	0.04 (0.05)	0.426	Portugal	0.08 (0.03)	0.004
Student self-efficacy in computer use	Denmark	0.00 (0.04)	0.949	Portugal	0.05 (0.02)	0.014
Access to digital devices in the home	Denmark	0.00 (0.04)	0.941	Portugal	0.06 (0.03)	0.022
Home resources for learning	Denmark	− 0.08 (0.05)	0.090	Portugal	− 0.05 (0.03)	0.117
Access to digital books	Denmark	0.05 (0.04)	0.217	Portugal	− 0.02 (0.02)	0.412
Instruction affected by digital resource shortage	Denmark	− 0.02 (0.03)	0.603	Portugal	0.04 (0.03)	0.117
Computers to students ratio	Denmark	0.07 (0.03)	0.018	Portugal	− 0.01 (0.02)	0.759

Table 9 (continued)

Variables	Countries	Coefficients (SE)	p	Countries	Coefficients (SE)	p
School location	Denmark	0.02 (0.04)	0.645	Portugal	0.03 (0.03)	0.265
Home resources for learning (school aggr.)	Denmark	− 0.02 (0.04)	0.590	Portugal	− 0.01 (0.04)	0.786
Time spent using computer daily	Italy	0.02 (0.03)	0.604	Slovenia	0.01 (0.04)	0.833
Time spent using computer to finding and reading information daily	Italy	− 0.03 (0.02)	0.242	Slovenia	0.04 (0.03)	0.187
Using computer or tablet for schoolwork—reading	Italy	0.00 (0.03)	0.979	Slovenia	0.01 (0.03)	0.725
Using computer or tablet for schoolwork—preparing	Italy	− 0.01 (0.03)	0.667	Slovenia	− 0.01 (0.03)	0.683
Combined frequency of using computer devices in activities in and out of school	Italy	0.07 (0.04)	0.065	Slovenia	0.02 (0.05)	0.641
Student self-efficacy in computer use	Italy	0.04 (0.03)	0.180	Slovenia	0.05 (0.03)	0.098
Access to digital devices in the home	Italy	0.08 (0.03)	0.002	Slovenia	0.01 (0.03)	0.704
Home resources for learning	Italy	− 0.06 (0.02)	0.020	Slovenia	− 0.12 (0.03)	< 0.001
Access to digital books	Italy	0.00 (0.03)	0.939	Slovenia	0.00 (0.03)	0.939
Instruction affected by digital resource shortage	Italy	0.08 (0.04)	0.040	Slovenia	− 0.01 (0.03)	0.644
Computers to students ratio	Italy	− 0.01 (0.03)	0.808	Slovenia	0.00 (0.03)	0.883
School location	Italy	0.00 (0.04)	0.965	Slovenia	0.00 (0.03)	0.893
Home resources for learning (school aggr.)	Italy	− 0.02 (0.03)	0.608	Slovenia	− 0.01 (0.03)	0.833
Time spent using computer daily	Norway	0.00 (0.03)	0.947	Sweden	0.01 (0.03)	0.846
Time spent using computer to finding and reading information daily	Norway	− 0.02 (0.02)	0.511	Sweden	0.00 (0.04)	0.952
Using computer or tablet for schoolwork—reading	Norway	0.02 (0.03)	0.361	Sweden	0.03 (0.03)	0.250
Using computer or tablet for schoolwork—preparing	Norway	− 0.01 (0.03)	0.816	Sweden	0.06 (0.04)	0.121
Combined frequency of using computer devices in activities in and out of school	Norway	− 0.03 (0.04)	0.438	Sweden	0.03 (0.03)	0.371
Student self-efficacy in computer use	Norway	0.01 (0.04)	0.790	Sweden	0.03 (0.03)	0.293
Access to digital devices in the home	Norway	0.05 (0.04)	0.190	Sweden	0.00 (0.04)	0.920
Home resources for learning	Norway	− 0.09 (0.03)	0.001	Sweden	0.01 (0.03)	0.666
Access to digital books	Norway	0.02 (0.03)	0.417	Sweden	0.00 (0.04)	0.975
Instruction affected by digital resource shortage	Norway	− 0.03 (0.03)	0.371	Sweden	0.01 (0.05)	0.829
Computers to students ratio	Norway	− 0.01 (0.04)	0.859	Sweden	0.01 (0.02)	0.683
School location	Norway	0.00 (0.03)	0.892	Sweden	− 0.01 (0.03)	0.715
Home resources for learning (school aggr.)	Norway	− 0.02 (0.03)	0.488	Sweden	− 0.08 (0.04)	0.092

Table 10 Model statistics from Model 2 (student and school data)

Countries	Statistic	Estimate	(SE)	p
Denmark	R-Squared	0.03	(0.011)	–
	Adjusted R-Squared	0.02	(0.011)	–
	F-Statistic	3.88	(1.610)	0.016
	DF	1748.00	(118.008)	–
Italy	R-Squared	0.03	(0.009)	–
	Adjusted R-Squared	0.02	(0.009)	–
	F-Statistic	6.00	(2.026)	0.003
	DF	2796.00	(185.413)	–
Norway	R-Squared	0.02	(0.006)	–
	Adjusted R-Squared	0.01	(0.006)	–
	F-Statistic	3.82	(1.494)	0.011
	DF	2980.00	(198.436)	–
Portugal	R-Squared	0.02	(0.007)	–
	Adjusted R-Squared	0.02	(0.007)	–
	F-Statistic	6.88	(2.221)	0.002
	DF	4032.00	(260.467)	–
Slovenia	R-Squared	0.03	(0.009)	–
	Adjusted R-Squared	0.02	(0.009)	–
	F-Statistic	7.48	(2.549)	0.003
	DF	3491.00	(237.296)	–
Sweden	R-Squared	0.02	(0.012)	–
	Adjusted R-Squared	0.01	(0.012)	–
	F-Statistic	3.82	(2.423)	0.115
	DF	2605.00	(180.147)	–

Acknowledgements
The authors would like to express their acknowledgments to the Slovenian Research Agency for funding this research.

Author contributions
PVM prepared the conception of the paper, prepared the datasets, analyzed the data, described and interpreted the results, prepared the discussion and conclusions. VG prepared the introduction, background of the study, finalized the conclusion and discussion.

Funding
The authors acknowledge the financial support from the Slovenian Research Agency (research core funding No. P5-0106 [Educational Research], and project J5-2553 "New domains of inequality: The digital divide in Slovenia").

Availability of data and materials
The data used for the analyses in this article are publicly available and can be found on IEA's data repository (https://www.iea.nl/data-tools/repository).

Declarations

Ethics approval and consent to participate
In this manuscript the officially published PIRLS 2016 data sets were used for the analyses. These data sets were downloaded as public use files from the IEA's website (https://www.iea.nl/data-tools/repository). Therefore, neither consent to participate or consent for publication nor ethics approval were required for these analyses.

Consent for publication
The authors provide their consent to publish this manuscript upon publication in the Springer open journal Large-scale Assessments in Education.

Competing interests
The authors declare that they have no competing interests.

Received: 12 October 2023 Accepted: 30 September 2024
Published online: 09 October 2024

References

Alexander, P., & Singer, L. (2017). Reading on paper and digitally: what the past decades of empirical research reveal. *Review of Educational Research, 20*(10), 1–35. https://doi.org/10.3102/0034654317722961

Baron, N., Calixte, R., & Havewala, M. (2017). The persistence of print among university students. *Telematics and Informatics, 34*(5), 590–604.

Becker, M., McEvanly, N., & Kortenbuck, M. (2010). Intrinsic and extrinsic reading motivation as predictors of reading literacy: a longitudinal study. *Journal of Educational Psychology, 102*(4), 773–785.

Chalmers, P., Robitzsch, A., Zoltak, M., Kim, K. H., Flak, C. F., Meade, A., Schneider, L., King, D., Liu, C. W., & Oguzhan, O. (2023). *mirt: Multidimensional Item Response Theory [Computer software]*. Retrieved from https://CRAN.R-project.org/package=mirt.

Chen, S. F. (2017). Modeling the influences of upper-elementary school students' digital reading literacy, socioeconomic factors, and self-regulated learning strategies. *Research in Science and Technological Education, 35*(3), 330–348.

Clinton, V. (2019). Reading from paper compared to screens: a systematic review and meta-analysis. *Journal of Research in Reading, 00*(00), 1–38. https://doi.org/10.1111/1467-9817.12269

Delgado, P., Vargas, C., Ackerman, R., & Salmeron, L. (2018). Don't throw away your printed books: a meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review, 25*, 23–38. https://doi.org/10.1016/j.edurev.2018.09.003

Duncan, G. J., Daly, M. C., McDonough, P., & Williams, D. R. (2002). Optimal indicators of socioeconomic status for health research. *American Journal of Public Health, 92*(7), 1151–1157. https://doi.org/10.2105/AJPH.92.7.1151

European Commission. (n.d.). *The Digital Economy and Society Index (DESI) | Shaping Europe's digital future*. Retrieved June 26, 2023, from https://digital-strategy.ec.europa.eu/en/policies/desi

Finch, W. H., & Bolin, J. H. (2017). *Multilevel modeling using Mplus*. CRC Press.

Fishbein, B., Martin, M. O., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 item equivalence study: examining mode effects for computer-based assessment and implications for measuring trends. *Large-Scale Assessments in Education, 6*(1), 11. https://doi.org/10.1186/s40536-018-0064-z

Foy, P. (2018). *PIRLS 2016 user guide for the international database*. TIMSS & PIRLS Study Center.

Foy, P., & LaRoche, S. (2017). Estimating standard errors in the PIRLS 2016 results. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016*. Lynch School of Education, Boston College.

Foy, P., Martin, M. O., Mullis, I. V. S., & Yin, L. (2017). Reviewing the PIRLS 2016 achievement item statistics. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016*. TIMSS & PIRLS International Study Center.

Foy, P., & Yin, L. (2017). Scaling the PIRLS 2016 achievement data. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016*. TIMSS & PIRLS International Study Center.

Fraillon, J., Schulz, W., & Ainley, J. (2013). *International Computer and Information Literacy Study: Assessment Framework*. International Association for the Evaluation of Educational Achievement.

Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: a meta-analysis. *Review of Educational Research, 91*(4), 483–517. https://doi.org/10.3102/0034654321998074

Gil-Flores, J., Johansson, S., Rosen, M., & Myrberg, E. (2014). The role of online reader experience in explaining students' performance in digital reading. *School Effectiveness and School Improvement, 25*(3), 394–407.

Goldstein, H. (2011). *Multilevel statistical models*. Wiley.

Hohlfeld, T., Ritzhaupt, A., Barron, A., & Kemker, K. (2008). Examining the digital divide in K-12 public schools: Four-year trends for supporting ICT literacy in Florida. *Computers & Education, 51*(4), 1648–1663. https://doi.org/10.1016/j.compedu.2008.04.002

Hooper, M., Mullis, I. V. S., & Martin, M. O. (2015). PIRLS 2016 context questionnaire framework. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2016 assessment framework*. TIMSS & PIRLS International Study Center & IEA.

Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(2), 157–174. https://doi.org/10.1207/S15328007SEM0802_1

IEA & TIMSS and PIRLS ISC at BC. (2018). *PIRLS 2016 International Database: Progress in International Reading Literacy Study* (Version 1). IEA. https://doi.org/10.58150/PIRLS_2016_DATA

Johansone, I. (2017). Survey operations procedures in PIRLS, 2016. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016*. Lynch School of Education, Boston College.

Kim, M.-C., & Kim, J.-K. (2001). Digital divide: conceptual discussions and prospect. In W. Kim, T.-W. Ling, Y.-J. Lee, & S.-S. Park (Eds.), *The human society and the internet internet-related socio-economic issues* (pp. 78–91). Springer-Verlag.

LaRoche, S., & Foy, P. (2017). Sample implementation in PIRLS, 2016. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016*. Lynch School of Education, Boston College.

LaRoche, S., Joncas, M., & Foy, P. (2017). Sample design in PIRLS 2016. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016*. Lynch School of Education, Boston College.

Lee, S. M., Brescia, W., & Kissinger, D. (2009). Computer use and academic development in secondary schools. *Computers in the Schools, 26*(3), 224–235. https://doi.org/10.1080/07380560903095204

Liu, Z. (2005). Reading behavior in the digital environment: changes in reading behavior over the past ten years. *Journal of Documentation, 61*(6), 700–712. https://doi.org/10.1108/00220410510632040

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86–92. https://doi.org/10.1027/1614-1881.1.3.86
Mangen, A., Olivier, G., & Velay, J. L. (2019). Comparing comprehension of a long text read in print book and on kindle: where in the text and when in the story? *Frontiers in Psychology, 10*, 1–11.
Mangen, A., Walgermo, B., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: effects on reading comprehension. *International Journal of Educational Research, 58*, 61–68. https://doi.org/10.1016/j.ijer.2012.12.002
Margolin, S., Driscoll, C., Toland, M., & Kegler, J. (2013). E-readers, computer screens, or paper: does reading comprehension change across media platforms? *Applied Cognitive Psychology, 27*, 512–519.
Martin, M. O., Mullis, I. V. S., & Foy, P. (2015). Assessment design for PIRLS, PIRLS literacy, and ePIRLS in 2016. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2016 assessment framework*. Lynch School of Education, Boston College.
Martin, M. O., Mullis, I. V. S., Hooper, M., Yin, L., Foy, P., Fishbein, B., & Liu, J. (2017). Creating and interpreting the PIRLS 2016 context questionnaire scales. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016.* Lynch School of Education, Boston College.
Masters, G. N., & Wright, B. D. (1997). The partial credit model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 101–122). Springer.
Mirazchiyski, P. V. (2021). RALSA: The R analyzer for large-scale assessments. *Large-Scale Assessments in Education*, **9**(1) , 1–24. https://doi.org/10.1186/s40536-021-00114-4
Mullis, I. V. S., & Martin, M. O. (Eds.). (2015a). *PIRLS 2016 assessment framework*. TIMSS & PIRLS International Study Center.
Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017a). *ePIRLS 2016 international results in online informational reading*. TIMSS & PIRLS International Study Center & IEA.
Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. IEA TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College
Mullis, I. V. S., Martin, M. O., & Sainsbury, M. (2015b). PIRLS 2016 reading framework. In I. V. S. Mullis & M. O. Martin (Eds.), *PIRLS 2016 assessment framework*. TIMSS & PIRLS International Study Center.
Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide*. Muthén & Muthén.
Nezlek, J. B. (2008). An introduction to multilevel modeling for social and personality psychology: multilevel analyses. *Social and Personality Psychology Compass, 2*(2), 842–860. https://doi.org/10.1111/j.1751-9004.2007.00059.x
OECD. (2014). *PISA 2012 technical report*. OECD.
Park, J., & Lee, J. (2021). Effects of E-Books and printed books on EFL learners' reading comprehension and grammatical knowledge. *English Teaching, 76*(3), 35–61. https://doi.org/10.1585/engtea.76.3.202109.35
Rasmusson, M., & Aberg-Bengtsson, L. (2014). Does performance in digital reading relate to computer game playing? A study of factor structure and gender patterns in 15-year-olds' reading literacy performance. *Scandinavian Journal of Educational Research, 59*(6), 691–709.
Rosen, M., & Gustafsson, J. E. (2016). Is computer availability at home causally related to reading achievement in grade 4? A longitudinal difference in differences approach to IEA data from 1991 to 2006. *Large-Scale Assessments in Education, 4*(5), 1–19. https://doi.org/10.1186/s40536-016-0020-8
Støle, H., Mangen, A., & Schwippert, K. (2020). Assessing children's reading comprehension on paper and screen: a mode-effect study. *Computers & Education, 151*, 1–13. https://doi.org/10.1016/j.compedu.2020.103861
Tran, T., Ho, M. T., Pham, T. H., Nguyen, M. H., Nguyen, K. L. P., Vuong, T. T., Nguyen, T. H., Nguyen, T. D., Nguyen, T.-L., Khuc, Q., La, V. P., & Vuong, Q. H. (2020). How digital natives learn and thrive in the digital age: evidence from an emerging economy. *Sustainability, 12*(9), 1–24. https://doi.org/10.3390/su12093819
von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series, 2*(1), 9–36.
Wang, W. C., & Qiu, X. L. (2019). Multilevel modeling of cognitive diagnostic assessment: the multilevel DINA example. *Applied Psychological Measurement, 43*(1), 34–50. https://doi.org/10.1177/0146621618765713
Zhang, Y., & Kudva, S. (2014). E-books versus print books: readers' choices and preferences across contexts. *Journal of the Association for Information Science and Technology, 8*, 1695–1706.

Publisher's Note